Metabolism of host-derived carbohydrates by *Bifidobacterium breve* UCC2003

Ollscoil na hÉireann, Corcaigh

THE NATIONAL UNIVERSITY OF IRELAND, CORK

A Thesis presented to the National University of Ireland for the

Degree of Doctor of Philosophy by

Muireann Egan, B.Sc

Department of Microbiology

University College Cork

Supervisor: Prof. Douwe van Sinderen

April 2015

TABLE OF CONTENTS

Declaration

I hereby declare that the research presented in this thesis is my own work and effort, and that it has not been submitted for any other degree, either at University College Cork or elsewhere. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

This work was completed under the guidance of Prof. Douwe van Sinderen at the Alimentary Pharmabiotic Centre & School of Microbiology, Biosciences Institute, University College Cork.

Signature:……

Date:……

LIST OF FIGURES

LIST OF TABLES

This thesis is dedicated to my family,

Go raibh míle maith agaibh.

LIST OF PUBLICATIONS

Chapter II: **Egan, M., O'Connell Motherway, M., Ventura, M. & van Sinderen, D. (2014).** Metabolism of sialic acid by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **80**, 4414-4426.

Chapter III: **Egan, M., O'Connell Motherway, M. & van Sinderen, D.** A GntR-type transcriptional repressor controls sialic acid utilisation in *Bifidobacterium breve* UCC2003. *FEMS Microbiol Lett* **362,** 1-9

Chapter IV: **Egan, M., O' Connell Motherway, M., Kilcoyne, M., Kane, M., Joshi, L., Ventura, M. & van Sinderen, D. (2014).** Cross-feeding by *Bifidobacterium breve* UCC2003 during co-cultivation with *Bifidobacterium bifidum* PRL2010 in a mucin-based medium**.** *BMC Microbiol* **14,** 282**.**

ABBREVIATIONS

ABC-type transporters: ATP-Binding Cassette Transporters

AdoMet: *S*-adenosyl-L-methionine

ADP: Adenosine Diphosphate

Amp: Ampicillin

anSME: anaerobic Sulfatase Maturation Enzyme

APC: Alimentary Pharmabiotic Centre

ATP: Adenosine Triphosphate

BLAST: Basic Local Alignment Search Tool

CCR: Carbon Catabolite Repression

cDNA: complementary DNA

CFS: Cell Free Supernatant

CFU: Colony Forming Units

CGH: Comparative Genome Hybridisation

Cm: Chloramphenicol

COG: Cluster of Orthologous Groups of proteins

DNA: Deoxyribonucleic Acid

DTT: Dithiothreitol

EDTA: Ethylenediaminetetraacetic Acid

Em: Erythromycin

EMSA: Electrophoretic Mobility Shift Assay

FGE: Formylglycine Generating Enzyme

FGly: Formylglycine; 3-oxoalanine

F6PPK: Fructose-6-Phosphate Phosphoketolase

GalE: UDP-glucose-4-epimerase

GalNAc: *N*-acetylgalactosamine

GalT: Galactose-1-phosphate uridylyltransferase

GEO: Gene Expression Omnibus

GH: Glycosyl Hydrolase

GIT: Gastrointestinal tract

GlcNAc: *N*-acetylglucosamine

GNB: Galacto-*N*-biose

HMO: Human milk oligosaccharides

HPAEC-PAD: High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection

HPTLC: High Performance Thin Layer Chromatography

HTH: Helix-Turn-Helix

IPTG: Isopropyl-β-D-thiogalactopyranoside

JCM: Japanese Collection of Microorganisms

Kan: Kanamycin

kDA: kilo Daltons

LAB: Lactic Acid Bacteria

LB: Luria Bertani broth

LMG: Belgian Coordinated Collection of Microorganisms

LNB: Lacto-*N*-biose

LnbP: Lacto-*N*-biose phosphorylase

LNDFH: Lacto-*N*-difucohexaose

LNDH: Lacto-*N*- difucosylpentaose

LNH: Lacto-*N*-hexaose

LNT: Lacto-*N*-tetraose

LNnT: Lacto-*N*-neotetraose

MRS: de Man, Rogosa and Sharpe medium

mMRS: modified de Man, Rogosa and Sharpe medium

mRNA: messenger RNA

MW: Molecular Weight

NCBI: National Centre for Biotechnology Information

NCFB: National Collection of Food Bacteria

NCIMB: National Collection of Industrial and Marine Bacteria

NCTC: National Collection of Type Cultures

NIZO: Nizo Food Research

OD: Optical Density

ORF: Open Reading Frame

PCR: Polymerase Chain Reaction

PEP-PTS: Phosphoenolpyruvate-dependent Phosphotransferase System

Poly[d(I-C)]: Poly- deoxyinosinic-deoxycytidylic acid

PRL: Culture collection of probiogenomics, University of Parma

RBS: Ribosome Binding Site

RCA: Reinforced Clostridial Agar

RCM: Reinforced Clostridial Medium

RNA: Ribonucleic Acid

ROK: Repressor Open reading frame Kinase

RP-HPLC: Reverse Phase High Performance Liquid Chromatography

RT-PCR: Reverse Transcription-PCR

SCFA: Short Chain Fatty Acids

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Tet: Tetracycline

Tris-HCl: Tris(hydroxymethyl)aminomethane- Hydrochoride

UCC: University College Cork Culture Collection

wt/vol: weight/volume

X-gal: 5-bromo-4-chloro-3-indolyl-D-galactopyranoside

ABSTRACT

Bifidobacteria are Gram positive, anaerobic, typically Y-shaped bacteria which are naturally found in the digestive tract of certain mammals, birds and insects. *Bifidobacterium breve* strains are numerically prevalent among the gut microbiota of many healthy breast-fed infants. The prototypical *B. breve* strain UCC2003 has previously been shown to utilise numerous carbohydrates of plant origin. Various aspects of host-derived carbohydrate metabolism occurring in this bacterium will be described in this thesis.

Chapter II describes *B. breve* UCC2003 utilisation of sialic acid, a nine-carbon monosaccharide, which is found in human milk oligosaccharides (HMOs) and the mucin glycoprotein. *B. breve* UCC2003 was also shown to cross-feed on sialic acid released from 3' sialyllactose, a prominent HMO, by the extracellular sialidase activity of *Bifidobacterium bifidum* PRL2010.

Chapter III reports on the transcriptional regulation of sialic acid metabolism in *B. breve* UCC2003 by a transcriptional repressor encoded by the *nanR* gene. NanR belongs to the GntR-family of transcriptional regulators and represents the first bifidobacterial member of this family to be characterised.

Chapter IV investigates *B. breve* UCC2003 utilisation of mucin. *B. breve* UCC2003 was shown to be incapable of degrading mucin; however when grown in co-culture with *B. bifidum* PRL2010 it exhibits enhanced growth and survival properties. A number of methods were used to investigate and identify the mucin components supporting this enhanced growth/viability phenotype.

Chapter V describes the characterisation of two sulfatase-encoding gene clusters from *B. breve* UCC2003. The transcriptional regulation of both sulfatase-encoding gene clusters was also investigated.

The work presented in this thesis represents new information on the metabolism of host-derived carbohydrates in bifidobacteria, thus increasing our understanding of how these gut commensals are able to colonise and persist in the gastrointestinal tract.

Chapter I

General Introduction

1.1 Introduction

The gastrointestinal tract (GIT) of mammals is the natural habitat for a diverse community of facultative and obligate anaerobic bacteria, collectively known as the gut microbiota. The human GIT is home to an estimated 10^{14} microorganisms, 10^{11}-10^{12} of which reside in the large intestine (Ley *et al.*, 2006). The overall gene set carried by the gut microbiota, also called the gut microbiome, of which 99 % is deemed to be of bacterial origin, is approximately 150 times larger than that of the human genome (Gill *et al.*, 2006; Qin *et al.*, 2010).

Immediately following birth, the essentially sterile GIT becomes colonised by microbes, followed by an increase in microbial diversity throughout early development (Koenig *et al.*, 2011; Palmer *et al.*, 2007; Vaishampayan *et al.*, 2010). A number of microbiota and metagenomic studies have provided an in depth analysis of the microbial diversity of a typical infant gut (Avershina *et al.*, 2013; Koenig *et al.*, 2011; Palmer *et al.*, 2007; Turroni *et al.*, 2012b; Vaishampayan *et al.*, 2010; Yatsunenko *et al.*, 2012). In one such study, which monitored the diversity of the gut microbiota through the first year of life, it was found that while members of the genera *Escherichia* and *Bacteroides* were dominant after one month, they were quickly supplanted by bacteria belonging to the genus *Bifidobacterium*, which was then clearly dominant for up to eleven months (Vaishampayan *et al.*, 2010). These data were substantiated by later studies, for example showing that members of the *Bifidobacteriaceae* family dominated the gut microbiota of infants between one and five months in age (Turroni *et al.*, 2012b), or that (in a study of 87 infants and based on 16S rRNA sequencing) bifidobacteria constitute approximately 60 % of the total bacterial load at four months of age (Avershina *et al.*, 2013). Another comprehensive metagenomic study characterised bacterial species in faecal samples from infants and

adults from the United States, the Amazonas of Venezuela and rural Malawi. Perhaps unsurprisingly, it was found that there were significant differences in the phylogenetic diversity between the three countries, and while bifidobacteria were again dominant in the three groups, their proportional representation diminished during the first year of life of the investigated infants (Yatsunenko *et al.*, 2012).

A number of factors have been shown to influence the establishment and consequent development of the gut microbiota in infants. The mode of delivery is considered an important factor which provides the initial microbial components to shape the infant intestinal microbiota: studies have shown that infants born by Caesarean section have lower numbers of bifidobacteria compared with vaginally delivered infants (Biasucci *et al.*, 2010; Huurre *et al.*, 2008). Also the method of feeding, i.e. breast-fed versus formula-fed, influences the development and diversity of the infant microbiota (Fanaro *et al.*, 2003; Roger *et al.*, 2010). Furthermore, studies have described the influence of antibiotic treatment on the development of the infant gut microbiota, with the administration of antibiotics being associated with a reduction in phylogenetic diversity (Antonopoulos *et al.*, 2009; Fouhy *et al.*, 2012; Koenig *et al.*, 2011; Palmer *et al.*, 2007). Within the first three years of human life the composition of the associated bacterial community in the gut evolves towards an adult-like microbiota, with one study noting that genes predicted to encode enzymes for plant-derived fibre metabolism, indicative of an adult diet, are present in the microbiome prior to the introduction of solid foods (Koenig *et al.*, 2011; Yatsunenko *et al.*, 2012).

Analysis of the GIT microbiota in healthy adults has shown that the bacterial composition varies throughout the intestinal tract. In the upper part of the small intestine, native conditions such as the acidic pH (the pH of the duodenum is 6.0,

rising to 7.4 in the terminal ileum), rapid luminal flow and presence of bile salts are responsible for keeping the cell biomass relatively low (Fallingborg, 1999; Walter & Ley, 2011). The diversity of microbial populations in the small intestine is also relatively low, with the duodenum and jejunum dominated by acid-tolerant bacteria such as *Streptococcus* and lactobacilli, while members of the genera *Bacteroides* and *Clostridia* dominate the ileum (Wang *et al.*, 2005). Bifidobacteria have also been identified in the ileum of children (aged between ten and fourteen years) and adults (Reuter, 2001).

The native conditions of the large intestine are more conducive to bacterial populations with higher densities, due to a near neutral pH, lower bile levels (the terminal ileum is the main site of bile acid reabsorption (Dawson *et al.*, 2009)) and a slower rate of peristalsis compared to the small intestine. Unlike the small intestine, the epithelial layer of the large intestine lacks Peyer's patches, the purpose of which is to sample bacteria and direct immune responses, thereby resulting in an increased tolerance for bacteria in the large intestine (Heel *et al.*, 1997; Walter & Ley, 2011). Similar to the small intestine, members of the genera *Bacteroides* and *Clostridium* are abundantly present, in addition to members of the phyla *Actinobacteria, Verrucomicrobia, Proteobacteria* and *Fusobacteria* (Eckburg *et al*., 2005; Turnbaugh *et al*., 2008). Large scale metagenomic sequencing projects have demonstrated the variability in the composition of the microbiota among individuals, as even identical twins share less than 50 % of their species level bacterial taxa (Turnbaugh *et al.*, 2010). Lifestyle factors, such as a shift in diet from a high fat/low fibre to a low fat/high fibre diet, can cause significant changes in the microbiota within a day (Turnbaugh *et al.*, 2009; Wu *et al.*, 2011). It was also found that there is increased inter-personal variability in older people (in a study of 178 people aged

between 64 and 102) compared to younger adults (mean age of 36), and that diet is the main factor that influences the composition of the gut microbiota (Claesson *et al.*, 2012). This study found that the gut microbiota of older people in long term care was significantly less diverse compared to their counterparts living in the community and this correlated with the moderate to high fat/low fibre diet typical of the individuals in long term care (Claesson *et al.*, 2012).

Given the sheer size and diversity of the gut microbiome, it is no surprise that the gut microbiota has been implicated in various health and disease states of the host. The (development of the) gut microbiota is associated with maturation and activity of the immune system. Germ-free animals tend to show defects in the development of gut-associated lymphoid tissues, as well as having smaller Peyer's patches and mesenteric lymph nodes and they are also more susceptible to infection by certain bacteria, viruses and parasites (reviewed in Round & Mazmanian, 2009). The commensal microbiota can induce repair of damaged intestinal epithelium (Pull *et al.*, 2005), while in mice the gut microbiota has been shown to protect against Type I diabetes, an autoimmune disease (Wen *et al.*, 2008). Microorganisms also impact on the structure of the protective mucus layer of the GIT, a number of studies have found variations in the biochemical composition and thickness of the mucus layer in germ-free animals as compared to their conventionally raised counterparts (Meslin *et al.*, 1999; Petersson *et al.*, 2011; Sharma *et al.*, 1995; Szentkuti *et al.*, 1990). The genomic content of the gut microbiota, and particularly the microbiome of the colon, is enriched with genes involved in the metabolism of carbohydrates, the fermentation products of which are short chain fatty acids (SCFA) that account for approximately 10 % of the calories extracted from a Western-style diet per day, while also playing a role in the regulation of inflammation and protection against colonic carcinogenesis

(Gill *et al.*, 2006; McNeil, 1984; Velázquez *et al.*, 1996; Vinolo *et al.*, 2011). Members of the gut microbiota, in particular representatives of the lactic acid bacteria (LAB) and bifidobacteria, have also been shown to synthesise certain B vitamins which cannot be produced by the host (LeBlanc *et al.*, 2011; LeBlanc *et al.*, 2013; Pompei *et al.*, 2007).

Changes in the composition and diversity of the gut microbiota have been implicated in a number of diseases, including Crohn's disease, ulcerative colitis, irritable bowel disease and colorectal cancer (de Vos & de Vos, 2012), as well as increased frailty in the elderly (Claesson *et al.*, 2012). A number of large scale metagenomic studies have focused on finding correlations between the composition of the microbiota and specific disease states such as obesity (Turnbaugh *et al.*, 2008), irritable bowel disease (Frank *et al.*, 2007; Greenblum *et al.*, 2012) and diabetes (Karlsson *et al.*, 2013; Larsen *et al.*, 2010). The purpose of these studies is to identify particular microbiota-related changes which can then be used as potential biomarkers for impending or fully manifest diseases, while also leading to a greater understanding of how the composition of the gut microbiota relates to its function in the gut (Bäckhed *et al.*, 2005).

1.2 General features of bifidobacteria

Bifidobacteria are Gram positive, saccharolytic, typically Y-shaped bacteria which are members of the *Bifidobacteriaceae* family and the *Actinobacteria* phylum. They were first isolated in 1900 by Henri Tissier, when he worked in the Pasteur Institute, from the faeces of a breast-fed infant. Tissier recommended that such bifidobacteria be administered to infants suffering from diarrhoea, believing that they would

displace the microbes that were causing this gastrointestinal malfunction (Tissier, 1900). Today various members of the *Bifidobacterium* genus, as well as strains belonging to the lactobacilli, are commercially exploited as so-called probiotic bacteria, defined as live microorganisms that, when administered in adequate amounts, confer a health benefit on the host (Hill *et al.*, 2014; Stanton *et al.*, 2005).

As of 2014, the *Bifidobacterium* genus consists of 48 species, which fall within six phylogenetic clusters, namely (i) *Bifidobacterium pseudolongum*, (ii) *Bifidobacterium pullorum*, (iii) *Bifidobacterium longum*, (iv) *Bifidobacterium adolescentis*, (v) *Bifidobacterium asteroides* and (vi) *Bifidobacterium boum* (Lugli *et al.*, 2014; Ventura *et al.*, 2006). The vast majority of bifidobacterial species have been isolated from the GIT of mammals, birds or insects, as well as from sewage and fermented milk (Ventura *et al.*, 2007; Ventura *et al.*, 2014). The first bifidobacterial genome to be sequenced was that of *B. longum* subsp. *longum* NCC2705 (Schell *et al.*, 2002) and as of December 2014, forty complete bifidobacterial genome sequences are available in the database of the National Centre of Biotechnology Information (located at the folllowing web site: http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi) (Table 1.1), while a recent study investigated another forty two draft genome sequences (Milani *et al.*, 2014). Nine species are represented by complete genome sequences, as well as recently published pangenome analyses of the *Bifidobacterium breve* and *Bifidobacterium animalis* subsp. *lactis* species, along with a core genome analysis of the *B. adolescentis* phylogenetic group, which includes the *B. adolescentis*, *Bifidobacterium catenulatum* and *Bifidobacterium pseudocatenulatum* species (Bottacini *et al.*, 2014; Duranti *et al.*, 2012; Milani *et al.*, 2013). Bifidobacterial genomes typically have a GC content that varies between 58 and 61 %, and range in

size between 1.73 and 3.25 Mb, with the genome of *Bifidobacterium biavatii* the largest sequenced to date (Milani *et al.*, 2014; Ventura *et al.*, 2014). This rapidly increasing collection of genomic data has allowed for comprehensive investigations into the phenotypic adaptations which enable bifidobacteria to colonise the GIT.

Table 1.1: List of complete sequenced bifidobacterial genomes.

Species	Genome Size (MB)	GC content %	Genes	Proteins	Source	Genbank No.	Reference
Bifidobacterium adolescentis ATCC 15703	2.1	59.20	1702	1632	Human GIT	NC_008618	Unpublished
Bifidobacterium adolescentis 22L	2.2	59.30	1812	1725	Human milk	CP007443	(Duranti *et al.*, 2014)
Bifidobacterium animalis subsp. *lactis* AD011	1.93	60.50	1603	1527	Infant faeces	NC_011835	(Kim *et al.*, 2009)
Bifidobacterium animalis subsp. *lactis* BI-04	1.94	60.50	1631	1567	Human faeces	NC_012841	(Barrangou *et al.*, 2009)
Bifidobacterium animalis subsp. *lactis* DSM 10140	1.94	60.50	1628	1565	Fermented milk	NC_012815	(Barrangou *et al.*, 2009)
Bifidobacterium animalis subsp. *lactis* BB-12	1.84	60.50	1706	1642	Human GIT	NC_017214	(Garrigues *et al.*, 2010)
Bifidobacterium animalis subsp. *lactis* V9	1.94	60.50	1636	1572	Human faeces	NC_017217	(Sun *et al.*, 2010)
Bifidobacterium animalis subsp. *lactis* CNCM I-2494	1.94	60.50	1724	1660	Human faeces	NC_017215	(Chervaux *et al.*, 2011)
Bifidobacterium animalis subsp. *lactis* BLC1	1.94	60.50	1607	1518	Human GIT	NC_017216	(Bottacini *et al.*, 2011)
Bifidobacterium animalis subsp. *lactis* ATCC 25527	1.93	60.50	1601	1538	Human GIT	NC_017834	(Loquasto *et al.*, 2013)
Bifidobacterium animalis subsp. *lactis* B420	1.94	60.50	1625	1561	Human GIT	NC_017866	(Stahl & Barrangou, 2012)
Bifidobacterium animalis subsp. *lactis* BI-07	1.94	60.50	1661	1597	Human faeces	NC_017867	(Stahl & Barrangou, 2012)
Bifidobacterium animalis subsp. *lactis* Bl12	1.94	60.50	1607	1518	Human GIT	NC_021593	(Milani *et al.*, 2013)
Bifidobacterium animalis subsp. *lactis* ATCC 27673	1.96	60.50	1620	1558	Human GIT	NC_022523	(Loquasto *et al.*, 2013)
Bifidobacterium animalis RH	1.93	60.50	1603	1539	Human faeces	CP007755	(Liu *et al.*, 2014)
Bifidobacterium asteroides PRL2011	2.17	60.10	1731	1658	Insect hindgut	NC_018720	(Bottacini *et al.*, 2012)
Bifidobacterium bifidum PRL2010	2.21	62.70	1791	1706	Infant faeces	NC_014638	(Turroni *et al.*, 2010)
Bifidobacterium bifidum S17	2.19	62.80	1845	1783	Infant faeces	NC_014616	(Zhurina *et al.*, 2011)
Bifidobacterium bifidum BGN4	2.22	62.60	1902	1834	Infant faeces	NC_017991	(Yu *et al.*, 2012)
Bifidobacterium breve ACS-071-V-Sch8b	2.33	58.70	2011	1826	Human GIT	NC_017218	Unpublished
Bifidobacterium breve UCC2003	2.42	58.70	1985	1854	Infant faeces	NC_020517	(O'Connell Motherway *et al.*, 2011b)
Bifidobacterium breve 12L	2.25	58.90	1847	1765	Human milk	CP006711	(Bottacini *et al.*, 2014)
Bifidobacterium breve JCM 7017	2.29	58.70	1873	1770	Infant faeces	CP006712	(Bottacini *et al.*, 2014)

Bifidobacterium breve JCM 7019	2.36	58.60	1999	1915	Human faeces	CP006713	(Bottacini *et al.*, 2014)
Bifidobacterium breve NCFB 2258	2.32	58.70	1873	1770	Infant faeces	CP006714	(Bottacini *et al.*, 2014)
Bifidobacterium breve 689b	2.33	58.70	1914	1821	Infant faeces	CP006715	(Bottacini *et al.*, 2014)
Bifidobacterium breve S27	2.3	58.70	1896	1748	Infant faeces	CP006716	(Bottacini *et al.*, 2014)
Bifidobacterium coryneforme LMG 18911	1.76	60.50	1422	1364	Honeybee gut	CP007287	(Milani *et al.*, 2014)
Bifidobacterium dentium Bd1	2.64	58.50	2195	2127	Dental caries	NC_013714	(Ventura *et al.*, 2009)
Bifidobacterium longum NCC2705	2.26	60.11	1799	1728	Infant faeces	NC_004307	(Schell *et al.*, 2002)
Bifidobacterium longum DJO10A	2.39	60.12	2073	2001	Human GIT	NC_010816	(Lee *et al.*, 2008)
Bifidobacterium longum subsp. *infantis* ATCC 15697	2.83	59.90	2588	2416	Infant GIT	NC_011593	(Sela *et al.*, 2008)
Bifidobacterium longum subsp. *longum* JDM301	2.48	59.80	2035	1958	Human faeces	NC_014169	(Wei *et al.*, 2010)
Bifidobacterium longum subsp. *longum* BBMN68	2.27	59.90	1876	1804	Human faeces	NC_014656	(Hao *et al.*, 2011)
Bifidobacterium longum subsp. *longum* JCM1217	2.39	60.30	2009	1924	Human faeces	NC_015067	(Fukuda *et al.*, 2011)
Bifidobacterium longum subsp. *infantis* 157F	2.41	60.11	2070	1999	Infant faeces	NC_015052	(Fukuda *et al.*, 2011)
Bifidobacterium longum subsp. *longum* KACC 91563	2.4	59.81	2050	1985	Infant faeces	NC_017221	(Ham *et al.*, 2011)
Bifidobacterium longum BXY01	2.48	59.80	1997	1901	Human GIT	CP008885	Unpublished
Bifidobacterium longum subsp. *longum* GT15	2.34	60.00	1989	1893	Human GIT	CP006741	Unpublished
Bifidobacterium thermophilum RBL67	2.29	60.10	1904	1845	Infant faeces	NC_020456	(Jans *et al.*, 2013)

1.3 Carbohydrate metabolism in bifidobacteria

The carbohydrate-containing fraction present in the human diet can be separated at a basic level into two groups, based on their digestibility by host-encoded enzymes in the GIT. The first group consists of simple sugars, which are absorbed in the small intestine either directly or following hydrolysis by host-encoded enzymes. The second group consists of complex carbohydrates, often plant cell wall polysaccharides, such as cellulose, pectin, (arabino)xylan, arabinan and resistant starch, that are resistant to digestion by host-encoded glycosyl hydrolases and absorption in the small intestine (Bond *et al.*, 1980; Englyst *et al.*, 1992; Lattimer & Haub, 2010). It has been estimated that between 10 and 60 g of dietary carbohydrate reaches the colon on a daily basis (Cummings & Macfarlane, 1991), providing a consistent nutrient supply to the gut microbiota. In addition, the mucus layer, which can extend up to 150 µm from the epithelial surface (Johansson *et al.*, 2011), provides another source of carbohydrate in the form of the mucin glycoprotein (Salyers *et al.*, 1977a). Other host-derived carbohydrates in the large intestine include heparan sulfate, which is present in low quantities in the mucus layer of the large intestine (Oshiro *et al.*, 2001) and chondroitin sulfate which is also present in the colonic mucosa (Eliakim *et al.*, 1986). In total, it is estimated that mucin and related muco-polysaccharides such as chondroitin sulfate provide 3-5 g of carbohydrate for fermentation by the intestinal microbiota per day (Stephen *et al.*, 1983).

Bifidobacteria are saccharolytic organisms whose ability to colonise and survive in the large intestine is largely dependent on the ability to metabolise certain diet- and/or host-derived carbohydrates as mentioned above. The majority of the gut microbiota is assumed to employ the glycolytic pathway when metabolising

carbohydrates (Macfarlane & Macfarlane, 2003). In contrast, the characteristic central metabolic pathway of bifidobacteria is the fructose-6-phosphate phosphoketolase pathway, also known as the bifid shunt (de Vries & Stouthamer, 1967; Scardovi & Trovatelli, 1965) (Fig. 1.1). The key enzyme of this pathway is fructose-6-phosphate phosphoketolase (F6PPK), the presence of which is considered a reliable indicator of the presence of bifidobacteria, although it is also produced by *Gardnerella vaginalis*, which belongs, like bifidobacteria, to the *Bifidobacteriaceae* family (Gao & Gupta, 2012; Gavini *et al.*, 1996; Palframan *et al.*, 2003). From 1 mole of fermented glucose, the fructose-6-phosphate phosphoketolase pathway yields 2.5 ATP molecules, thus providing more energy than the carbohydrate fermentation pathways present in other gut bacteria, such as lactic acid bacteria (Palframan *et al.*, 2003). However, it should be noted that this does not take into account the energy cost of internalising the carbohydrate substrate. In general, bifidobacteria internalise sugars through ABC-type transport systems, which require ATP (Davidson & Chen, 2004), as opposed to the cost-neutral phosphoenolpyruvate-phosphotransferase systems (PEP-PTS; see below) (Postma *et al.*, 1993). The fructose-6-phosphate phosphoketolase pathway produces SCFAs which are perceived to be beneficial to the host (Palframan *et al.*, 2003). The bifid shunt-mediated fermentation of 1 mole of hexose sugar theoretically yields 1.5 moles of acetate and 1 mole of lactate, while pentose sugars result in a 1:1 ratio of acetate to lactate, although these ratios can vary depending on the carbon source, growth phase of the cells and the environmental pH (Macfarlane & Macfarlane, 2003; Palframan *et al.*, 2003).

When the first bifidobacterial genome, that of *B. longum* subsp. *longum* NCC2705, was sequenced in 2002 it was noted that more than 8.5 % of the predicted proteins

were assigned to the cluster of orthologous groups (COG) in the carbohydrate transport-metabolism category (Schell *et al.*, 2002), a characteristic which was later found to be shared among the genomes of many other bifidobacterial species (Lugli *et al.*, 2014; Milani *et al.*, 2014; Ventura *et al.*, 2007).

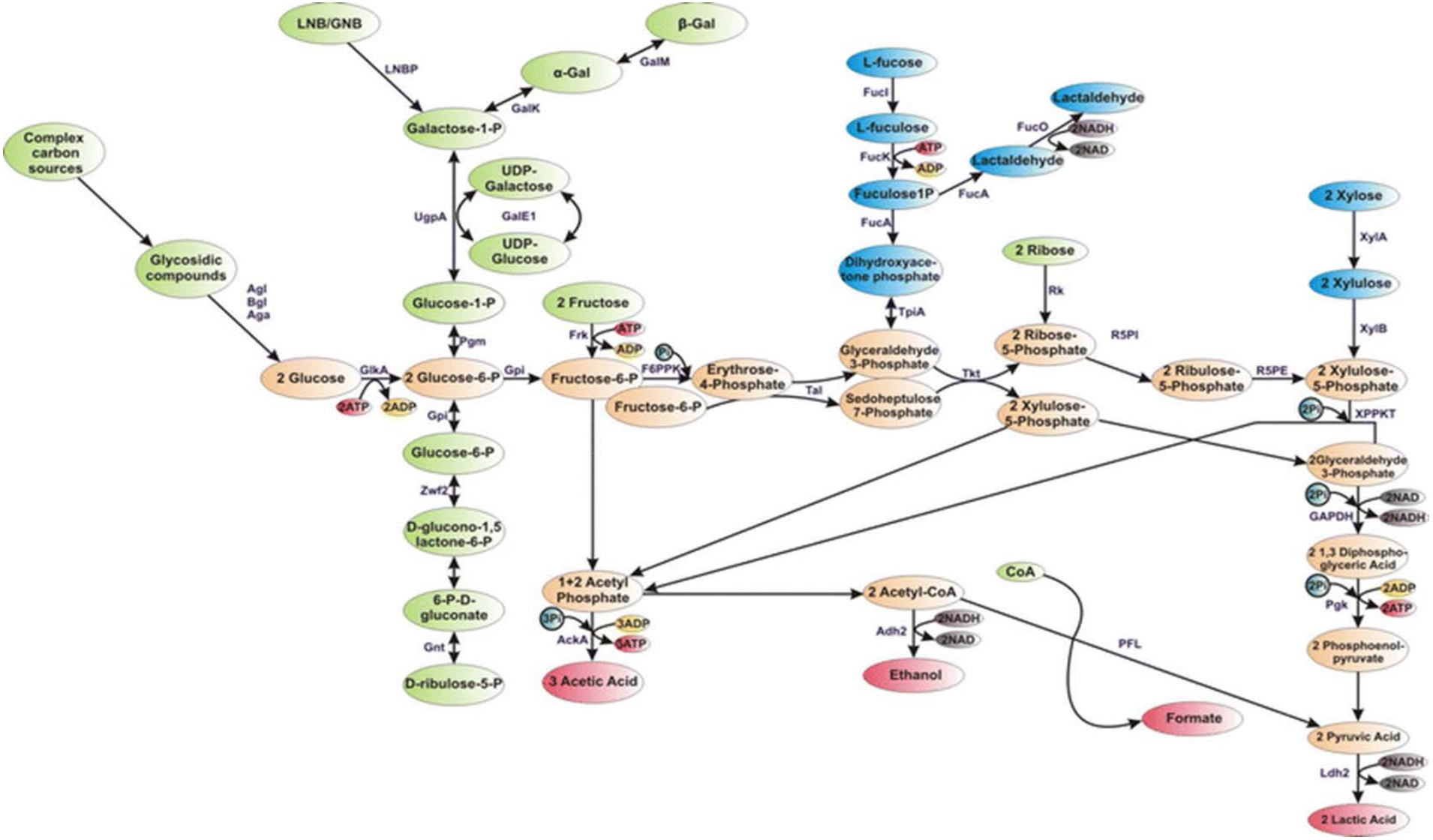

LNB/GNB
LNBP
β-Gal
GalM
α-Gal
GalK
Galactose-1-P
UDP-Galactose
UgpA
GalE1
UDP-Glucose
Complex carbon sources
Glycosidic compounds
AgI
BgI
Aga
Glucose-1-P
Pgm
2 Glucose
GlkA
2ATP
2ADP
2 Glucose-6-P
Gpi
Glucose-6-P
Zwf2
D-glucono-1,5 lactone-6-P
6-P-D-gluconate
Gnt
D-ribulose-5-P
2 Fructose
Frk
ATP
ADP
Fructose-6-P
Pi
F6PPK
Erythrose-4-Phosphate
Tal
Fructose-6-P
Sedoheptulose 7-Phosphate
Glyceraldehyde 3-Phosphate
TpiA
Dihydroxyace-tone phosphate
L-fucose
FucI
L-fuculose
FucK
ATP
ADP
Fuculose1P
FucA
FucA
Lactaldehyde
FucO
2NADH
2NAD
Lactaldehyde
2 Ribose
Rk
2 Ribose-5-Phosphate
R5PI
Tkt
2 Ribulose-5-Phosphate
R5PE
2 Xylulose-5-Phosphate
2 Xylulose-5-Phosphate
2 Xylose
XylA
2 Xylulose
XylB
2Pi
XPPKT
2Glyceraldehyde 3-Phosphate
2Pi
2NAD
2NADH
GAPDH
2 1,3 Diphospho-glyceric Acid
2Pi
2ADP
2ATP
Pgk
2 Phosphoenol-pyruvate
1+2 Acetyl Phosphate
3Pi
3ADP
3ATP
AckA
3 Acetic Acid
2 Acetyl-CoA
Adh2
2NADH
2NAD
Ethanol
CoA
PFL
Formate
2 Pyruvic Acid
Ldh2
2NADH
2NAD
2 Lactic Acid

Figure 1.1: Schematic representation of carbohydrate degradation through “bifid shunt” in bifidobacteria. Abbreviations: AckA, acetate kinase; Adh2, aldehyde-alcohol dehydrogenase 2; Aga, α-galactosidase; Agl, α-glucosidase; Bgl, β-glucosidase; GalE1, UDP-glucose 4-epimerase; GalK, galactokinase; GalM, galactose mutarotase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase C; GlkA, glucokinase; Gnt, 6-phosphogluconate dehydrogenase; Gpi, glucose 6-phosphate isomerase; Frk, fruktokinase; F6PPK, fructose-6-phosphoketolase; FucI, L-fucose isomerase; FucK, L-fuculose kinase; FucA, L-fuculose-1P aldolase; FucO, lactaldehyde reductase; Ldh2, lactate dehydrogenase; LNBP, lacto-*N*-biose phosphorylase;Pgk, phosphoglyceric kinase; Pgm, phosphoglucomutase; Pfl, formate acetyltransferase; Rk, ribokinase; R5PI, ribose-5-phosphate isomerase; R5PE, ribulose-5-phosphate epimerase; Tal, transaldolase; Tkt, transketolase; TpiA, triosephosphate isomerase;UgpA, UTP-glucose-1-phosphate uridylyltransferase; XPPKT, xylulose-5-phosphate/fructose-6-phosphate phosphoketolase; XylA, xylose isomerase;XylB, xylulose kinase; Zwf2, glucose-6-phosphate 1-dehydrogenase; Pi, phosphate (figure taken from a previous review (Pokusaeva *et al.,* 2011a)).

1.4 Carbohydrate uptake by bifidobacteria

Carbohydrate uptake by bifidobacteria is achieved through a combination of ATP-binding cassette (ABC) transporters, proton symporters and proton motive force-driven permeases (Pokusaeva *et al.*, 2011a; Schell *et al.*, 2002; Ventura *et al.*, 2009). PEP-PTS (phosphoenolpyruvate-phosphotransferase systems) systems are less frequently identified in bifidobacterial genomes compared to other uptake systems, with *Bifidobacterium bifidum* PRL2010 being an exception. This strain is predicted to encode just two ABC-type transport systems, four PEP-PTS systems and four secondary transporters (Turroni *et al.*, 2012a). PEP-PTS systems have been associated with the transport of glucose and fructose in *B. breve* NCFB 2257 (Degnan & Macfarlane, 1993). Four putative PEP-PTS systems were identified on the *B. breve* UCC2003 genome (Maze *et al.*, 2007), however, carbohydrate uptake in this strain typically involves ABC-type transport systems (Egan *et al.*, 2014a; O'Connell *et al.*, 2013; O'Connell Motherway *et al.*, 2013; Pokusaeva *et al.*, 2011a). An in-depth study of the transport systems of *B. longum* subsp. *longum* NCC2705 identified 13 ABC-type transport systems, predicted to be involved in the transport of di-, tri- and higher order oligosaccharides, three permeases of the major facilitator superfamily predicted to internalise glucose, lactose and sucrose, and just one PEP-PTS for the transport of glucose (Parche *et al.*, 2007). The predominance of ABC transporters on bifidobacterial genomes is perhaps a reflection of the availability of more complex carbohydrates rather than monosaccharides in the large intestine, as mentioned above. ABC transporters, in particular the CUT1 subfamily, transport a variety of oligosaccharides (Schneider, 2001), whereas the PEP-PTS system is typically limited to monosaccharides or certain disaccharides such as sucrose (Deutscher *et al.*, 2006; Lengeler *et al.*, 1982).

1.5 Glycosyl hydrolases in bifidobacteria

As mentioned above, the carbohydrates present in the large intestine are believed to be predominantly di-, tri-, or higher order oligosaccharides, thus their metabolic processing requires them to be degraded to their constituent monosaccharides. This necessitates a specific type of enzyme, namely a glycosyl hydrolase (GH). GHs are responsible for the hydrolysis of the glycosidic bond between monosaccharides and are found in all three kingdoms (*Archaebacteria*, *Eubacteria* and *Eukaryota*) (Cantarel *et al.*, 2009; Henrissat, 1991). In 1991, Bernard Henrissat devised a method of classifying GHs based on their amino acid sequence similarity and, as of December 2014, 133 different GH families have been recognised and are listed at www.cazy.org (Henrissat, 1991; Lombard *et al.*, 2014). According to the current information on www.cazy.org (December 2014), bifidobacterial genomes are predicted to encode between 25 and 87 GHs, which in turn represent between 12 and 28 GH families per genome (Table 1.2). The currently available information on this website lists the genome of *Bifidobacterium dentium* Bd1 as encoding the highest number of predicted GHs among bifidobacteria (87) (Ventura *et al.*, 2009). When this genome was sequenced it was suggested that the relatively high number of predicted GHs is a reflection of the typical niche of this bacterium, i.e. the oral cavity, where the bacteria are exposed to the full contents of the ingested foods (Ventura *et al.*, 2009). The number of predicted GHs on the genomes of *B. animalis* subsp. *lactis* DSM 10140 and Bl-04 are among the lowest (35 and 38, respectively) and this correlates with the relatively small size of the typical *B. animalis* subsp. *lactis* genome as compared to other bifidobacteria (Barrangou *et al.*, 2009).

Table 1.2: Glycosyl hydrolase (GH)-encoding genes on bifidobacterial genomes.

Strain Name	GH families	Number of GHs
Bifidobacterium adolescentis 22L	22	61
Bifidobacterium adolescentis ATCC 15703	22	58
Bifidobacterium animalis RH	16	39
Bifidobacterium animalis subsp. *animalis* ATCC25527	13	34
Bifidobacterium animalis subsp. *lactis* AD011	16	38
Bifidobacterium animalis subsp. *lactis* ATCC 27673	17	35
Bifidobacterium animalis subsp. *lactis* B420	16	38
Bifidobacterium animalis subsp. *lactis* BB12	16	42
Bifidobacterium animalis subsp. *lactis* Bi-07	16	40
Bifidobacterium animalis subsp. *lactis* Bl-04	16	38
Bifidobacterium animalis subsp. *lactis* Bl-12	15	36
Bifidobacterium animalis subsp. *lactis* BLC1	16	38
Bifidobacterium animalis subsp. *lactis* CNCM I-2494	16	40
Bifidobacterium animalis subsp. *lactis* DSM 10140	16	35
Bifidobacterium animalis subsp. *lactis* V9	16	38
Bifidobacterium asteroides PRL2011	22	48
Bifidobacterium bifidum BGN4	22	41
Bifidobacterium bifidum PRL2010	22	43
Bifidobacterium bifidum S17	22	39
Bifidobacterium breve 12L	22	47
Bifidobacterium breve 689b	27	55
Bifidobacterium breve ACS-071-V-Sch8b	22	49
Bifidobacterium breve JCM 7017	24	51
Bifidobacterium breve JCM 7019	26	54
Bifidobacterium breve NCFB 2258	25	51
Bifidobacterium breve S27	24	50
Bifidobacterium breve UCC2003	26	55
Bifidobacterium coryneforme LMG18911	14	26
Bifidobacterium dentium Bd1	25	87
Bifidobacterium indicum LMG 11587	13	25
Bifidobacterium longum 105-A	23	70
Bifidobacterium longum BXY01	28	65
Bifidobacterium longum DJO10A	25	67
Bifidobacterium longum NCC2705	24	56
Bifidobacterium longum subsp. *infantis* 157F	28	61
Bifidobacterium longum subsp. *infantis* ATCC15697	22	46
Bifidobacterium longum subsp. *infantis* JCM1222	23	49
Bifidobacterium longum subsp. *longum* BBMN68	24	60
Bifidobacterium longum subsp. *longum* F8	23	54
Bifidobacterium longum subsp. *longum* GT15	25	59
Bifidobacterium longum subsp. *longum* JCM1217	23	59
Bifidobacterium longum subsp. *longum* JDM301	28	64
Bifidobacterium longum subsp. *longum* KACC91563	22	55
Bifidobacterium pseudocatenulatum D2CA	20	60
Bifidobacterium thermophilum RBL67	12	31

Interest in GHs encoded by bifidobacteria has grown, which is not surprising when it is considered that GH activity of a bifidobacterial species or strain is likely to correspond to the ability to utilise certain prebiotics. A prebiotic is defined as "a selectively fermented ingredient that allows specific changes, both in the composition and/or activity in the gastrointestinal microflora that confers benefits upon host well-being and health" (Gibson *et al.*, 2004). In order to be considered a prebiotic, an ingredient must be scientifically proven to (i) resist gastric acidity, hydrolysis by mammalian enzymes and gastrointestinal absorption, (ii) be fermented by the intestinal microbiota, and (iii) stimulate the growth and/or activity of intestinal bacteria associated with health and well-being (Gibson *et al.*, 2004). As mentioned above, a significant proportion of the carbohydrate ingested by the host reaches the colon undigested, making them candidates for prebiotic status. Such prebiotic candidates include inulin, fructo-oligosaccharides (FOS), (trans)galacto-oligosaccharides (TOS/GOS) and lactulose (Gibson *et al.*, 2004; Macfarlane *et al.*, 2006). Inulin, which belongs to a class of dietary fibre known as fructans, has long been associated with promoting bifidobacterial growth, using both *in vivo* and *in vitro* studies (reviewed in Gibson *et al.*, 2004). TOS is a mixture of oligosaccharides derived from lactose by enzymatic transglycosylation (Crittenden & Playne, 1996). In a clinical trial, TOS has been shown to increase bifidobacterial numbers in patients with irritable bowel syndrome (IBS) and alleviate IBS-associated symptoms (Silk *et al.*, 2009). Lactulose is a disaccharide formed by the isomerisation of lactose. In a clinical trial using 16 healthy volunteers, it was found that 10 g of lactulose per day could significantly increase faecal bifidobacterial counts (Bouhnik *et al.*, 2004). Other potential prebiotics include xylo-oligosaccharides (XOS), fructo-

oligosaccharides (FOS) and isomalto-oligosaccharides (Gibson *et al.*, 2004; Macfarlane *et al.*, 2006).

There have been previous review articles dedicated to GHs characterised from bifidobacteria and their relation to prebiotic metabolism (Pokusaeva *et al.*, 2011a; van den Broek & Voragen, 2008; van den Broek *et al.*, 2008) and since then additional GHs specific for potential prebiotics have been characterised. For example, *B. breve* UCC2003 encodes an extracellular β1-4 endogalactanase (GalA) that hydrolyses galactan to produce galacto-oligosaccharides, which are then internalised and further metabolised to galactose by a β-galactosidase (GalG) (O'Connell Motherway *et al.*, 2011a). Four β-galactosidases were characterised from *B. longum* subsp. *infantis* ATCC15697, which release monosaccharides from GOS (Garrido *et al.*, 2013), while an exo-β1-3-galactanase from *B. longum* subsp. *longum* JCM1217 was active on β1-3 linked GOS as well as arabinogalactan (Fujita *et al.*, 2014). A novel α-glucosidase, MelD, as well as two α-galactosidases, RafD and MelE, from *B. breve* UCC2003, were shown to exhibit activity towards plant-derived raffinose and melibiose (O'Connell *et al.*, 2013). A β1-4 xylanase capable of the complete hydrolysis of XOS was characterised from *B. adolescentis* DSMZ 18350 (Amaretti *et al.*, 2013). Finally, two arabinofuranosidases, AbfA and AbfB, were characterised from *B. adolescentis* LMG10502 which were shown to hydrolyse the arabinose substitutions of arabinoxylan (Lagaert *et al.*, 2010).

1.6 Metabolism of host derived carbohydrates by bifidobacteria

1.6.1 Mucin metabolism by bifidobacteria

The GIT is coated with a continuous layer of mucus, the main component of which is mucin glycoprotein. Given its ubiquity, and the fact that carbohydrate constitutes approximately 80 % of the total mucin mass (Larsson *et al.*, 2009), it is surprising that only a small proportion of the culturable intestinal microbiota is believed to encode enzymes required for (partial) mucin degradation into free sugars, including members of the *Bifidobacterium, Bacteroides* and *Ruminococcus* genera, and the more recently characterised gut commensal *Akkermansia muciniphila* (Crost *et al.*, 2013; Derrien *et al.*, 2004; Hoskins *et al.*, 1985; Salyers *et al.*, 1977b; Sonnenburg *et al.*, 2005; Turroni *et al.*, 2010). Mucins are both highly complex and diverse. The original investigations on the (bio)chemical composition of human colonic mucin described twenty-one discrete oligosaccharide structures (Podolsky, 1985). It is now presumed that MUC2, the prominent secretory mucin in the colon, contains more than 100 structurally different O-linked glycans, with four distinct core structures, which can be elongated by the addition of galactose, *N*-acetylgalactosamine (GalNAc) and *N*-acetylglucosamine (GlcNAc), and substituted with sialic acid, fucose or sulfate residues in terminal or branched positions (Capon *et al.*, 2001; Larsson *et al.*, 2009; Podolsky, 1985; Robbe *et al.*, 2004) (illustrated in Fig. 1.2).

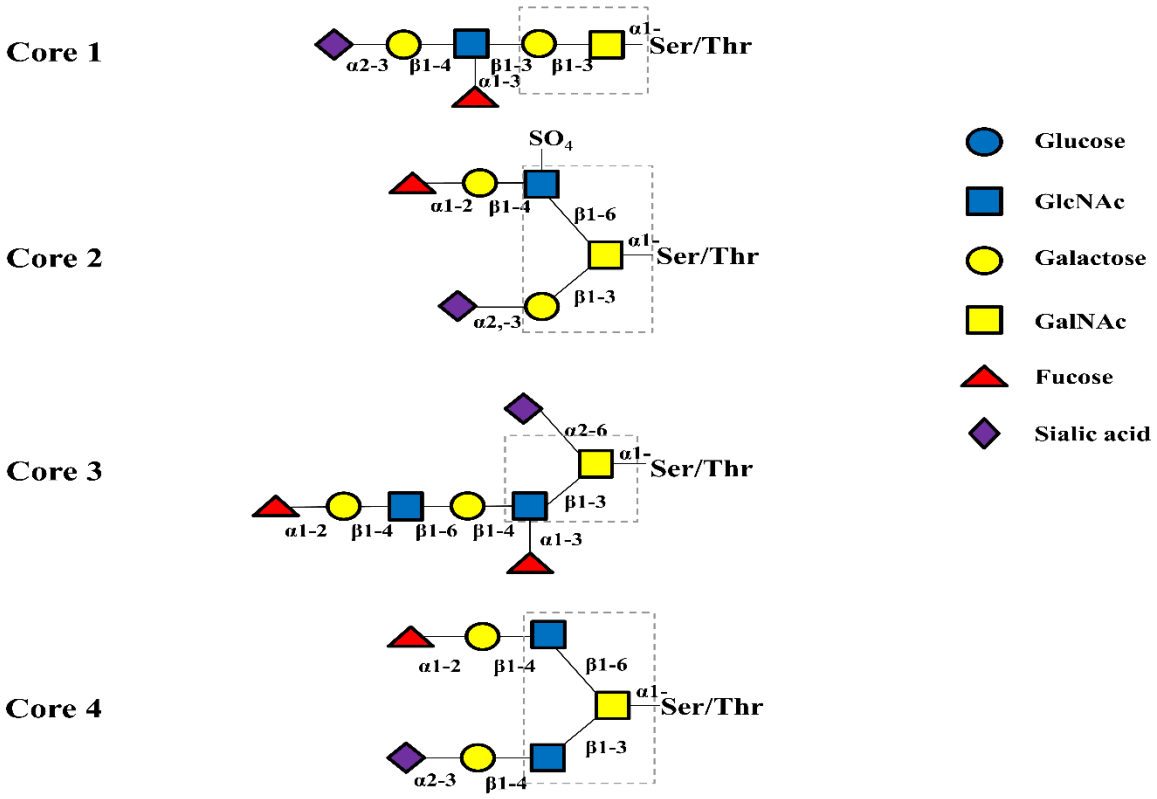


Figure 1.2: Representative examples of complex O-GalNAc glycans from human mucin (Varki *et al.*, 2009). The core structures, 1-4, are indicated with a grey box. Each core structure can be elongated by the addition of galactose, GalNAc and GlcNAc. The oligosaccharide chains can be substituted with sialic acid, fucose or sulfate residues in. Ser, serine; Thr, threonine. The monosaccharide key is shown on the right.

Mucin-degrading bifidobacteria, first described in 1985, were shown to constitutively express extracellular enzymes capable of degrading oligosaccharide side chains of gut mucins (Hoskins *et al.*, 1985). The degradation of mucin occurs sequentially with the removal of component monosaccharides, thus requiring a multitude of enzymes with various glycosidic specificities (Corfield *et al.*, 2002) (illustrated in Fig. 1.3). Cell surface-anchored GHs were later characterised, including two α-L-fucosidases, AfcA and AfcB, from *B. bifidum* JCM1254 (Ashida *et al.*, 2009; Katayama *et al.*, 2004) and an endo-α-*N*-acetylgalactosaminidase, EngBF, from *B. longum* subsp. *longum* JCM1217, which hydrolyses the linkage between GalNAc of the core 1 disaccharide and the serine or threonine residue of the proteinaceous backbone (Fujita *et al.*, 2005). Additionally, an extracellular, membrane bound, exo-α-sialidase, SiaBb, that releases sialic acid from porcine gastric mucin was characterised from *B. bifidum* JCM1254 (Kiyohara *et al.*, 2011). Recently, a cell-membrane anchored α-*N*-acetylglucosaminidase, designated AgnB, which is specific to the GlcNAc-α1-4-Gal structure found in gastroduodenal and colonic mucin, was characterised from *B. bifidum* JCM1254 (Shimada *et al.*, 2014). Interestingly, the only intracellular mucin-degrading enzyme characterised from *B. bifidum* is an α-*N*-acetylgalactosaminidase (designated NagBb) which hydrolyses the α1-linkage between GalNAc and the serine or threonine residue of mucin core 3 structures. It was hypothesised by the authors that extended mucin core 3 structures are extracellularly hydrolysed by sialidase (SiaBb), β-galactosidase (BbgIII) and β-*N*-acetylhexosaminidase (BbhI) activity, resulting in GalNac-α1-Ser/Thr (also known as the Tn antigen), which is assimilated through unknown transporters and intracellularly hydrolysed by NagBb (Kiyohara *et al.*, 2011; Kiyohara *et al.*, 2012; Miwa *et al.*, 2010) (Fig. 1.3).

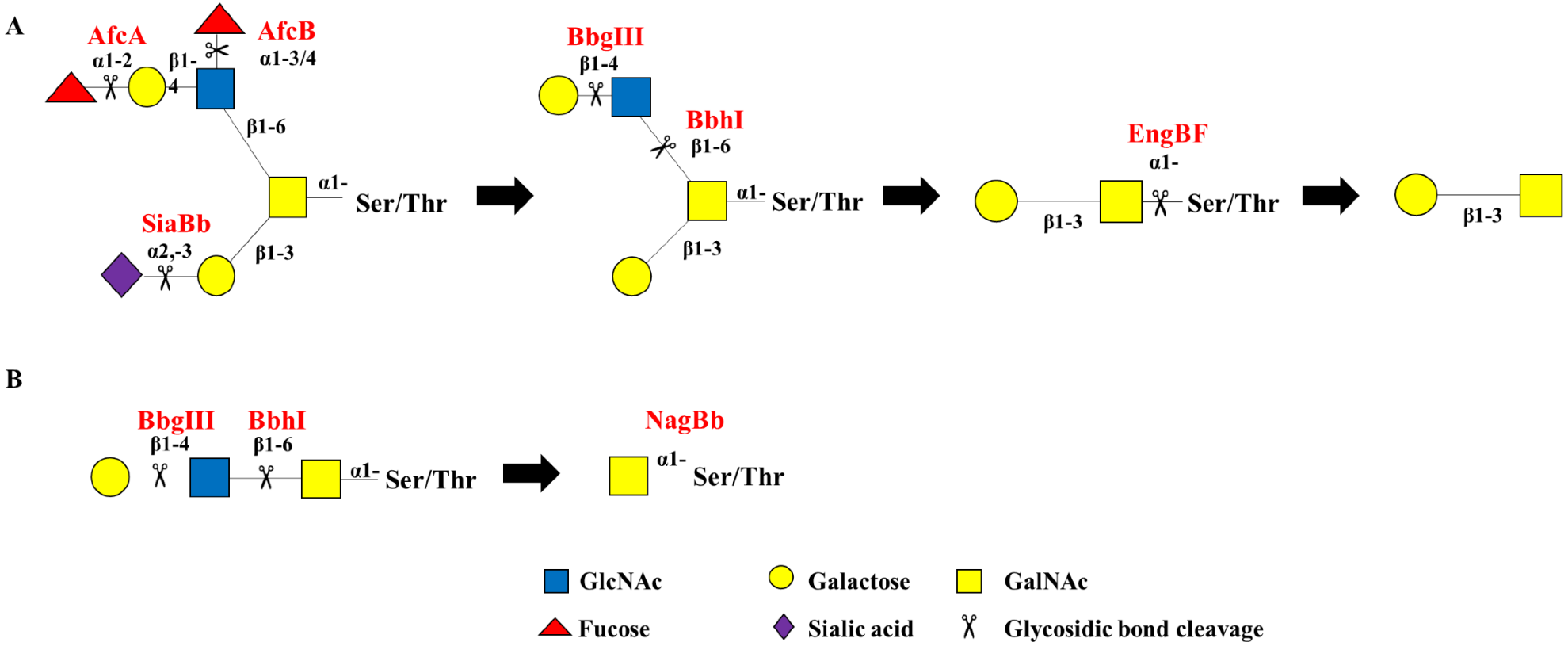

A
AfcA
α1-2
β1-4
AfcB
α1-3/4
β1-6
α1-
Ser/Thr
SiaBb
α2,-3
β1-3
BbgIII
β1-4
BbhI
β1-6
α1-
Ser/Thr
β1-3
EngBF
α1-
β1-3
Ser/Thr
β1-3
B
BbgIII
β1-4
BbhI
β1-6
α1-
Ser/Thr
NagBb
α1-
Ser/Thr
GlcNAc
Galactose
GalNAc
Fucose
Sialic acid
Glycosidic bond cleavage

Figure 1.3: Schematic representation of mucin degradation by *B. bifidum* JCM1254 glycosyl hydrolases. (A) In extended core 1 and core 2 structures, terminal sialic acid or fucose residues are removed by sialidase (SiaBb) or fucosidase (AfcA or AfcB) activity, respectively. The oligosaccharide chain is sequentially degraded by β-galactosidase (BbgIII) and β-*N*-acetylhexosaminidase (BbhI) activity. The resulting GNB is cleaved from the Ser/Thr amino acids by an endo-α-*N*-acetylgalactosaminidase (EngBF) and assimilated into the cell through a GNB/LNB-specific solute binding protein of an ABC transport system. (B) Core 3 mucin structures are similarly degraded, followed by the hydrolysis of the GalNAc-Ser/Thr linkage by the NagBb enzyme. The characterised glycosyl hydrolases are indicated in red. The monosaccharide key is shown at the bottom of the figure.

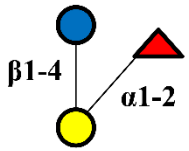
When the genome of *B. bifidum* PRL2010 was sequenced in 2010, it was observed that 60 % of the identified GHs are linked to mucin degradation (Turroni *et al.*, 2010). This included two putative exo-α-sialidases, two putative α-L-fucosidases and a predicted endo-α-*N*-acetylgalactosaminidase, all of which are presumed to be extracellular based on to the presence of a putative signal peptide (Turroni *et al.*, 2010). Other possible mucin-degrading enzymes encoded by the genome of *B. bifidum* PRL2010 include four *N*-acetyl-β-hexosaminidases (two of which are predicted to be extracellular) and four β-galactosidases (one of which is presumed to be extracellular) (Turroni *et al.*, 2010). Comparative genome hybridisation analysis revealed that most of the genes encoding the aforementioned enzymes are conserved within the examined members of the *B. bifidum* species and of a number of bifidobacterial strains tested, only members of this species were capable of growth in media containing porcine gastric mucin as the sole carbon source (Turroni *et al.*, 2010). That such a large number of predicted GHs are extracellular seems significant, especially when it is considered that despite the presence of genes encoding a predicted sialidase and fucosidase on its genome, sialic acid and fucose only support poor/no growth of *B. bifidum* PRL2010 (Turroni *et al.*, 2012a). In a recent *in vitro* study it was shown that fucose and galactose accumulate in the medium during *B. bifidum* PRL2010 growth on porcine gastric mucin (Egan *et al.*, 2014b). Indeed, it has been suggested that the function of these enzymes is to provide the strain access to specific mucin components, for example galacto-*N*-biose (Galβ1-3GalNAc; GNB) and galacto-*N*-tetraose (Galβ1-3-GalNAc-β1-3Galβ1-4GlcNAc; GNT) (Turroni *et al.*, 2010; Turroni *et al.*, 2014). The metabolic cluster required for the utilisation of GNB, as previously characterised in *B. longum* subsp.

longum JCM1217, is conserved among *B. bifidum* strains (Kitaoka *et al.*, 2005; Nishimoto & Kitaoka, 2007; Turroni *et al.*, 2010).

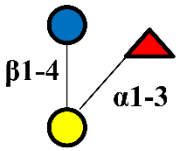
1.6.2 HMO metabolism by bifidobacteria

Mucin-derived oligosaccharides and human milk oligosaccharides (HMOs) share a number of primary monosaccharide components including galactose, GlcNAc, glucose, sialic acid and fucose (Bode & Jantscher-Krenn, 2012; Urashima *et al.*, 2012). There are at least 200 known HMO structures, yet they are not believed to be metabolised by the infant and hence reach the colon undigested (Ninonuevo *et al.*, 2006). The type 1 structure found among HMOs is lacto-*N*-biose (Galβ1-3GlcNAc; LNB), which requires the same metabolic pathway as GNB from mucin (Kitaoka *et al.*, 2005; Nishimoto & Kitaoka, 2007), and the type II structure is *N*-acetyllactosamine (Galβ1-4GlcNAc). HMOs containing the type I structure are more abundant than those containing the type II structure (Urashima *et al.*, 2012). Both LNB and *N*-acetyllactosamine are linked to a lactose residue at their reducing end by a β1-3 linkage to form lacto-*N*-tetraose (Galβ1-3GlcNAcβ1-3Galβ1-4Glc; LNT) and lacto-*N*-neotetraose (Galβ1-4GlcNAcβ1-3Galβ1-4Glc; LnNT), respectively (Bode & Jantscher-Krenn, 2012). The oligosaccharide chains can be elongated with up to a dozen repeats of LNB or *N*-acetyllactosamine, with further modification of the backbone through the addition of fucose and sialic acid to the non-reducing end or to an internal GlcNAc (Bode & Jantscher-Krenn, 2012). In addition, a lactose residue can be fucosylated through an α1-2/3 linkage or sialylated through an α2-3/6 linkage, to produce fucosyllactose (2’ or 3’ fucosyllactose) or sialyllactose (3’ or 6’ sialyllactose), respectively (illustrated in Fig. 1.4). Approximately 77 % of HMOs are fucosylated and 16 % are sialylated (Ninonuevo *et al.*, 2006). A small number of HMO structures contain both sialic acid and fucose, such as 3’sialyl-3-fucosyllactose (Bode & Jantscher-Krenn, 2012) and monofucosyl-monosialyllacto-*N*-hexaose (Ninonuevo *et al.*, 2006). In quantitative terms, three of the four most abundant

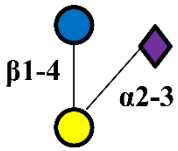
HMOs are fucosylated, namely 2'-fucosyllactose (0.45 to 3.93 g L^{-1}), lacto-*N*-fucopentaose (0.67 to 2.08 g L^{-1}) and lacto-*N*-difucohexaose (0.5 to 1.87 g L^{-1}) (Asakuma *et al.*, 2007; Chaturvedi *et al.*, 2001; Coppa *et al.*, 1993; Kunz *et al.*, 2000; Thurl *et al.*, 2010) (Fig. 1.4).



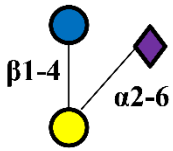
2'- fucosyllactose



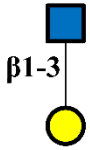
3'- fucosyllactose



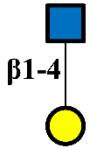
3'- sialyllactose



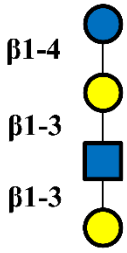
6'- sialyllactose



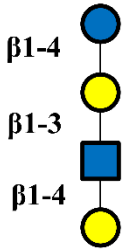
Lacto-*N*-biose



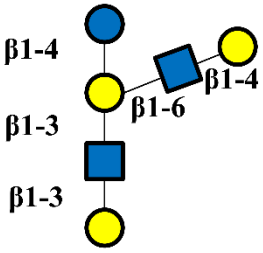
N-acetyllactosamine



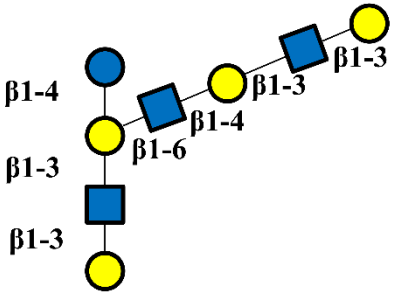
Lacto-*N*-tetraose



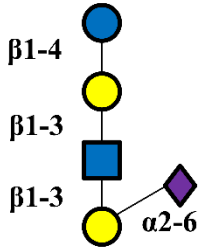
Lacto-*N*-neotetraose



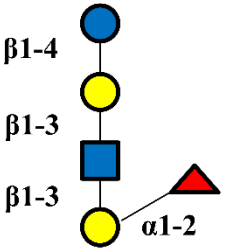
Lacto-*N*-hexaose



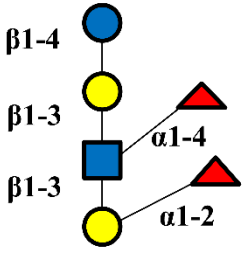
Iso-lacto-*N*-octaose



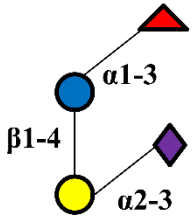
Sialyl-lacto-*N*-tetraose a



Lacto-*N*-fucopentaose I



Lacto-*N*-difucohexaose I



3' sialyl-3-fucosyllactose

Figure 1.4: Structures of selected HMO mentioned in this review. The addition of fucose (α1-2 or α1-3 linkage) or sialic acid (α2-3 or α2-6) linkage to lactose results in fucosyllactose or sialyllactose, respectively. Alternatively, lactose can be elongated with LNB (Type I structures) or *N*-acetyllactosamine (Type II structures). Further additions in a β1-3 linkage extends the chain (*para* HMO) or a β1-6 linkage introduces chain branching (*iso* HMO). The monosaccharide key is shown at the bottom of the figure.

HMOs have long been associated with promoting bifidobacterial growth (György *et al.*, 1954) and in recent years the metabolic pathways by which certain species of bifidobacteria utilise HMOs have come under intense scrutiny. Much of this research has focused on *B. bifidum* and *B. longum* subsp. *infantis*, and it seems that these two species are the most adapted to HMO utilisation. Given the similar oligosaccharide structures and monosaccharide components of mucin and HMOs, it is not surprising that *B. bifidum* PRL2010 was shown to achieve good growth on HMOs (Turroni *et al.*, 2010). A later study showed that *B. bifidum* JCM1254 also achieves high levels of growth on HMOs, while providing interesting analysis of the HMO components utilised by this strain (Asakuma *et al.*, 2011). Similar to *B. bifidum* PRL2010 degradation of mucin, it was found that hydrolysis of HMO components occurs extracellularly, as evident from a decrease in LNT and 2'-fucosyllactose and a concurrent increase in LNB, fucose and lactose in the media in the early logarithmic growth phase, indicating extracellular hydrolysis of these oligosaccharides (Asakuma *et al.*, 2011). *B. bifidum* JCM1254 encodes an extracellular α1-2 fucosidase, which is believed to be required for growth on 2'-fucosyllactose (Ashida *et al.*, 2009; Katayama *et al.*, 2004). A predicted extracellular lacto-*N*-biosidase which liberates LNB from LNT was also characterised from this strain (Wada *et al.*, 2008) (Fig. 1.5B). By mid-exponential phase, 2'-fucosyllactose, LNB and LNT were absent, while only trace amounts of lactose remained. After 24 h of growth, all detectable oligosaccharides had been consumed, with only fucose and galactose remaining in the growth medium (Asakuma *et al.*, 2011). The results of this study were in accordance with previous findings, which showed that *B. bifidum* utilises high molecular weight HMOs, and that fucose accumulates in the growth medium (Ward *et al.*, 2007). In a very recent publication, *B. bifidum* PRL2010 was shown to

extracellularly hydrolyse 3’ sialyllactose, of which it only utilises lactose, with sialic acid remaining in the medium (Egan *et al.*, 2014a).

In the aforementioned study (Ward *et al.*, 2007), five bifidobacterial strains (representing four bifidobacterial species) namely *B. longum* subsp. *longum* ATCC15707, *B. longum* subsp. *infantis* ATCC15697, *B. breve* ATCC27539, *B. bifidum* ATCC29521 and *B. adolescentis* ATCC15703, were tested for their ability to utilise HMOs. *B. longum* subsp. *infantis* ATCC15697 achieved three times the cell density compared to that of the other strains during growth on HMOs as a sole carbon source. Interestingly, *B. longum* subsp. *infantis* ATCC15697 and *B. breve* ATCC27539 were shown to utilise the five monosaccharide components of HMOs, namely glucose, galactose, GlcNAc, fucose and sialic acid, whereas *B. bifidum* ATCC29521, *B. adolescentis* ATCC15703 and *B. longum* subsp. *longum* ATCC15707 were shown to only utilise glucose and galactose (Ward *et al.*, 2007). Another study showed that *B. longum* subsp. *infantis* ATCC15697 consumed a significantly larger proportion of available HMOs, compared to *B. longum* subsp. *longum* DJO10A and *B. breve* ATCC15700 (63.9 %, 35.2 % and 24.4 %, respectively) (LoCascio *et al.*, 2007). These authors furthermore showed that crude cell extract of the *B. longum* subsp. *infantis* ATCC15697 strain grown on HMOs contains sialidase and fucosidase activity, which was absent from the crude cell extract of the other tested bifidobacterial strains, indicative of the presence of intracellular fucosidase and sialidase enzymes in *B. longum* subsp. *infantis* ATCC15697 (LoCascio *et al.*, 2007). When the genome of *B. longum* subsp. *infantis* ATCC15697 was first sequenced, it was noted that it contains a 43 kb gene cluster, comprised of 30 genes, dedicated to the metabolism of HMO (Sela *et al.*, 2008). This cluster is conserved among *B. longum* subsp. *infantis* strains, yet interestingly, is

absent in the genomes of *B. longum* subsp. *longum* strains, which consequently cannot utilise HMOs as a sole carbon source (LoCascio *et al.*, 2010; Schell *et al.*, 2002; Sela *et al.*, 2008). Instead, *B. longum* subsp. *longum* strains possess gene clusters required for the metabolism of the plant-derived pentose sugars, xylose and arabinose, genes which are absent or partially present as remnants in *B. longum* subsp. *infantis* (Schell *et al.*, 2002; Sela & Mills, 2010).

Some of the GHs involved in HMO metabolism in *B. longum* subsp. *infantis* ATCC15697 include an α1-2 fucosidase, an α1-3/4 fucosidase, two β-galactosidases, three *N*-acetylhexosaminidases and an α2-3/6 sialidase, all of which lack a signal sequence at the N-terminal and are therefore presumed to be intracellular (Sela *et al.*, 2008; Sela *et al.*, 2011; Sela *et al.*, 2012; Yoshida *et al.*, 2012). These findings suggest that *B. longum* subsp. *infantis* ATCC15697 internalises particular intact HMOs, which are then degraded sequentially by intracellular GHs (Sela & Mills, 2010) (Fig. 1.5). For this purpose, *B. longum* subsp. *infantis* ATCC15697 encodes twenty genes predicted to be Family 1 solute binding proteins (associated with fourteen predicted ABC-type transport systems) for the transport of oligosaccharides, twice as many as other bifidobacterial genomes (Garrido *et al.*, 2011; Sela *et al.*, 2008; Tam & Saier, 1993). Of these, the solute binding proteins corresponding to locus tags Blon_0883 and Blon_2177 recognise the type I glycans LNB and GNB, while that specified by Blon_2177 also recognises the abundant HMO structures LNT, lacto-*N*-hexaose (LNH) and lacto-*N*-octaose (Garrido *et al.*, 2011; Kogelberg *et al.*, 2004). The predicted solute binding proteins specified by Blon_2344 and Blon_2347, which are located within the previously mentioned HMO cluster on the genome of *B. longum* subsp. *infantis* ATCC15697, recognise type II glycans such as LNnT and lacto-*N*-difucohexaose (LNDFH) (Garrido *et al.*, 2011; Sela *et al.*, 2008).

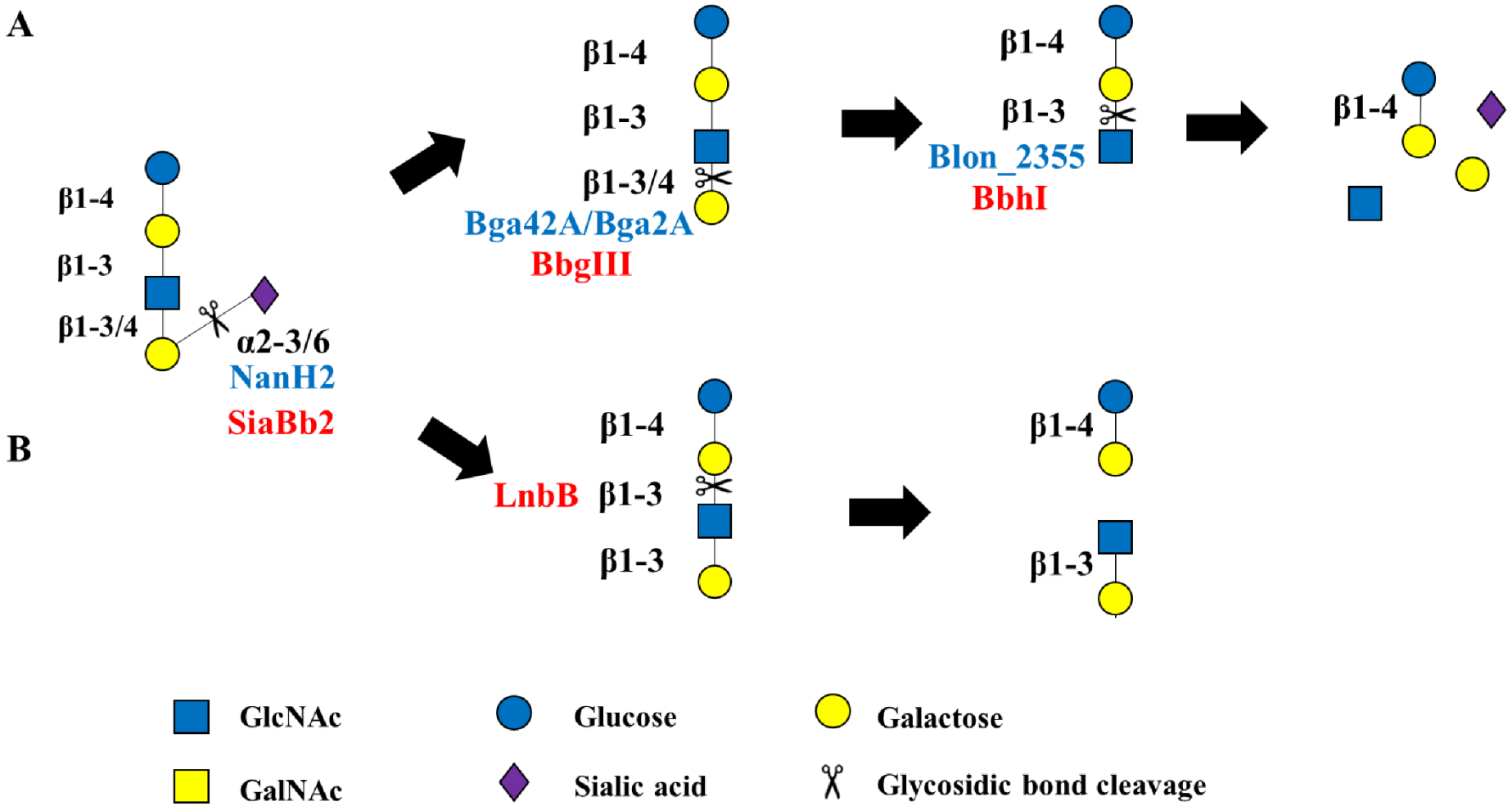

A
β1-4
β1-3
β1-3/4
α2-3/6
NanH2
SiaBb2
β1-4
β1-3
β1-3/4
Bga42A/Bga2A
BbgIII
β1-4
β1-3
Blon_2355
BbhI
β1-4
B
LnbB
β1-4
β1-3
β1-3
β1-4
β1-3
GlcNAc
GalNAc
Glucose
Sialic acid
Galactose
Glycosidic bond cleavage

Figure 1.5: Schematic representation of the degradation of sialylated Type I and Type II HMO by *B. longum* subsp. *infantis* ATCC15697 and *B. bifidum* JCM1254. (A) *B. longum* subsp. *infantis* ATCC15697 degrades sialylated-LNT and sialylated-LNnT sequentially, beginning with the removal of sialic acid by NanH2 activity, followed by the removal of Gal from the non-reducing end by the β-galactosidases, Bga42A (LNT) or Bga2A (LNnT). The resulting lacto-*N*-triose is hydrolysed to lactose and GlcNAc by *N*-acetylglucosaminidase (Blon_2355) activity. *B. bifidum* JCM1254 extracellularly degrades Type II HMO in a similar manner. (B) *B. bifidum* JCM 1254 extraceullularly cleaves Type I HMO through lacto-*N*-biosidase (LnbB) activity, producing lactose and GNB which are assimilated into the cell. Glycosyl hydrolases characterised from *B. longum* subsp. *infantis* ATCC15697 and *B. bifidum* JCM1254 are indicated in blue and red, respectively. The monosaccharide key is shown at the bottom of the figure.

Thus, this strain can assimilate a substantial number of different HMOs and degrade them intracellularly to their monosaccharide components for further metabolic processing. The limitation to this strategy is that only HMOs of a certain degree of depolymerisation (DP ≤ 8) can be consumed by the strain (LoCascio *et al.*, 2007), due to transporter specificity or the steric factors which inhibit the transport of high molecular weight substrates across the cell membrane (Sela & Mills, 2010). This is in contrast to the strategy of extracellular hydrolysis employed by *B. bifidum*, which allows it to depolymerise the high molecular weight HMOs that are inaccessible to other bifidobacterial species.

B. longum subsp. *infantis* and *B. bifidum* have therefore adopted divergent strategies in the utilisation of similar host-derived carbohydrates. A third possible strategy employed by certain bifidobacterial species is cross-feeding on the mucin and HMO components released by the extracellular activity of another (bifido)bacterial species. It has long been hypothesised that extracellular degradation of large polysaccharides such as mucin may provide nutritional support to other enteric bacteria (Hoskins *et al.*, 1985), while it was later suggested that cross-feeding on HMO may occur between bifidobacterial species, with particular reference to *B. bifidum* and *B. breve* (Ward *et al.*, 2007). Certain strains of *B. breve* can metabolise fucosylated HMOs, although the overall utilisation of HMOs is moderate compared to *B. longum* subsp. *infantis* ATCC15697 (Ruiz-Moyano *et al.*, 2013). However, as mentioned above, it was previously shown that *B. breve* ATCC15700 can utilise all monosaccharide components of HMO, as well as LNT and LNnT (LoCascio *et al.*, 2007; LoCascio *et al.*, 2009; Ward *et al.*, 2007). The metabolic pathway by which *B. breve* UCC2003 utilises sialic acid has been described (Egan *et al.*, 2014a). It is possible that *B. breve* relies on its ability to utilise a broad range of mono- and oligosaccharides

(Pokusaeva *et al.*, 2011a) to scavenge the mucin and HMO components released by the extracellular mucin and HMO-degrading activities of *B. bifidum* and perhaps other (bifido)bacteria. Such cross-feeding between bifidobacterial species was shown for α2-3 linked sialyllactose: *B. breve* UCC2003 was shown to metabolise sialic acid released by the exo-sialidase activity of *B. bifidum* PRL2010 (Egan *et al.*, 2014a). Similarly, *B. breve* UCC2003 was shown to exhibit growth and sustain viability in medium containing mucin, yet only in the presence of *B. bifidum* PRL2010 (Egan *et al.*, 2014b). Therefore, cross-feeding represents a third strategy by which certain bifidobacteria can utilise complex, host-derived carbohydrates (illustrated in Fig. 1.6).

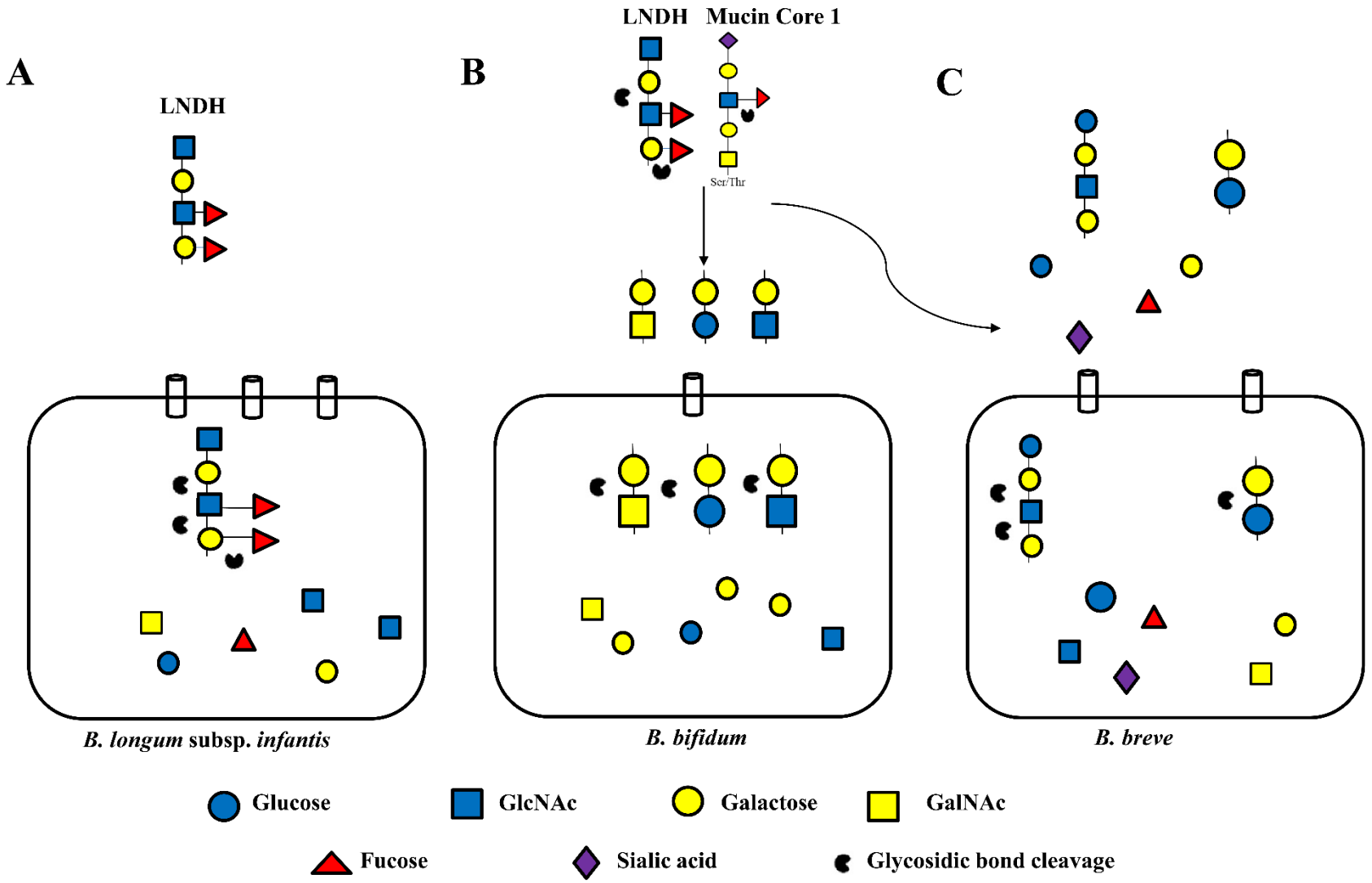

A
LNDH
B
LNDH
Mucin Core 1
Ser/Thr
C
B. longum subsp. infantis
B. bifidum
B. breve
Glucose
GlcNAc
Galactose
GalNAc
Fucose
Sialic acid
Glycosidic bond cleavage

Figure 1.5: Schematic representation of the strategies employed by three bifidobacterial species to utilise host derived glycans using lacto-*N*-difucosylhexaose (LNDH) and extended mucin core 1 structures as examples. (A) *B. longum* subsp. *infantis* internalises intact HMO through a considerable number of ABC transporters, followed by intracellular hydrolysis of the HMO to its constituent parts. (B) *B. bifidum* hydrolyses large HMO and mucin oligosaccharides using a number of extracellular GHs. The resulting disaccharides such as GNB and LNB are internalised, but many monosaccharides are not. (C) *B. breve* internalises the monosaccharides released by *B. bifidum* (and other intestinal mucin-degrading bacteria). *B. breve* may also internalise a small number of HMOs such as LNT and LNnT. The monosaccharide key is shown at the bottom of the figure. GNB, galacto-*N*-biose; LNB, lacto-*N*-biose.

1.6.3 Metabolism of N-linked glycoproteins

Proteins are the fourth most abundant component of human milk, behind lactose, lipids and HMO (Table 1.3), and it is estimated that 70 % of the most abundant proteins in milk are glycosylated (Froehlich *et al.*, 2010). Like mucin, glycosylation can be O-linked to a serine or threonine residue, or *N*-linked to an asparagine residue via a GlcNAc monomer (Froehlich *et al.*, 2010; Picariello *et al.*, 2008). All *N*-linked glycans contain the core pentasaccharide $Man_3GlcNAc_2$, which can be further modified to produce one of three structures. In high mannose structures, the core pentasaccharide is extended by another two to six mannose residues. Complex structures contain no additional mannose residues but may have up to five antennae consisting of GlcNAc, galactose and fucose, and are often decorated with sialic acid residues. Hybrid glycans comprise of a combination of high mannose and complex branches (Morelle & Michalski, 2007). Bacterial metabolism of *N*-glycans has primarily been associated with pathogens such as *Enterococcus faecalis*, which utilises the high mannose-type *N*-glycan RNaseB (Roberts *et al.*, 2000), and *Capnocytophaga canimorsus*, which utilises the complex-type *N*-glycans fetuin and IgG (Renzi *et al.*, 2011). The ability to utilise *N*-glycans is associated with the presence of an extracellular enzyme, endo-β-*N*-acetylglucosaminidase, which cleaves the pentasaccharide core between the GlcNAc residues, releasing the oligosaccharide chain (Roberts *et al.*, 2001; Varki *et al.*, 2009).

Recent research has shown that some bifidobacterial species also display extracellular endo-β-*N*-acetylglucosaminidase activity (Garrido *et al.*, 2012). *B. longum* subsp. *infantis* ATCC15697 was shown to degrade the high mannose-containing glycoprotein RNaseB, yet interestingly, other *B. longum* subsp. *infantis* strains were shown to exhibit very little hydrolytic activity towards this substrate,

suggesting that this is a strain-specific characteristic (Garrido *et al.*, 2012). Of the nine *B. longum* subsp. *longum* strains tested, only four demonstrated endo-β-*N*-acetylglucosaminidase activity, an observation that is consistent with this subspecies' preference for plant-derived polysaccharides (Schell *et al.*, 2002), while members of the *B. bifidum* species display no activity towards this substrate (Garrido *et al.*, 2012). In contrast, the *B. breve* strains JCM 7019 and KA179 were shown to display high activity towards this substrate, completely deglycosylating RNaseB and consistent with this it was shown that heavily *N*-glycosylated cell wall proteins from *Saccharomyces cerevisiae* can support growth of *B. breve* strains SC139 and KA179 (Garrido *et al.*, 2012). This presents an interesting scenario where *B. bifidum* and *B. breve* may collaborate to access *N*-glycans, whereby the endo-β-*N*-acetylglucosaminidase activity of *B. breve* liberates the oligosaccharide chain, which is then further degraded by the sialidase and fucosidase activities of *B. bifidum* to remove sialic acid and fucose residues, thus introducing a mutually beneficial partnership between these two species of bifidobacteria. Characterisation of two endo-β-*N*-acetylglucosaminidases, EndoBl-1 and EndoBl-2, from *B. longum* subsp. *infantis* ATCC15697 and *B. longum* subsp. *infantis* SC142, respectively, revealed activity against both high mannose and complex glycans, while Endo-BL1 was also active on lactoferrin from human milk, another example of this strain's adaptation to the infant gut (Garrido *et al.*, 2012).

Table 1.3: Abundance of the four primary components of human milk.

Milk Component	Quantity (g L^{-1})	References
Lactose	43-70	(Fusch *et al.*, 2011; Mitoulas *et al.*, 2002; Silvestre *et al.*, 2014; Thurl *et al.*, 2010)
Lipids	36-43	(Mitoulas *et al.*, 2002; Silvestre *et al.*, 2014)
HMOs	5-23	(Coppa *et al.*, 1993; Kunz *et al.*, 2000; Thurl *et al.*, 2010)
Protein	6-12	(Mitoulas *et al.*, 2002; Silvestre *et al.*, 2014)

1.6.4 Sulfatase activity in bacteria

Host-derived glycoproteins such as mucin, as well as proteoglycans such as chondroitin sulfate and heparan sulfate which are found in the colonic mucosa and human milk (Eliakim *et al.*, 1986; Newburg *et al.*, 1995; Oshiro *et al.*, 2001), are often highly sulfated (Fig. 1.7). Three sulfated oligosaccharide structures have also been isolated from human milk, although they are present only in minute quantities (>100 μg ml^{-1}) (Guérardel *et al.*, 1999). As described above, the *O*-linked glycans of human colonic mucin have four distinct core structures bound to a serine or threonine residue, namely Galβ1-3GalNAc (Core 1, GNB), GlcNAcβ1-6(Galβ1-3)GalNAc (Core 2), GlcNAcβ1-3GalNAc (Core 3) and GlcNAcβ1-6(GlcNAcβ1-3)GalNAc (Core 4) (Fig. 1.2). In human colonic mucin, these structures can carry sulfate residues, a potential purpose of which is to protect the mucin against bacterial glycosidases (Brockhausen, 2003). However, glycosulfatase activity has been identified in enteric bacteria. *Bacteroides thetaiotaomicron* and *Bacteroides ovatus* were initially predicted to produce sulfatases due to their ability to utilise the highly sulfated sugars, chondroitin sulfate and heparan (Salyers *et al.*, 1977b). Subsequently, a glycosulfatase purified from *Prevotella* strain RS2 was shown to remove sulfate groups from mucus glycoproteins (Roberton *et al.*, 1993).

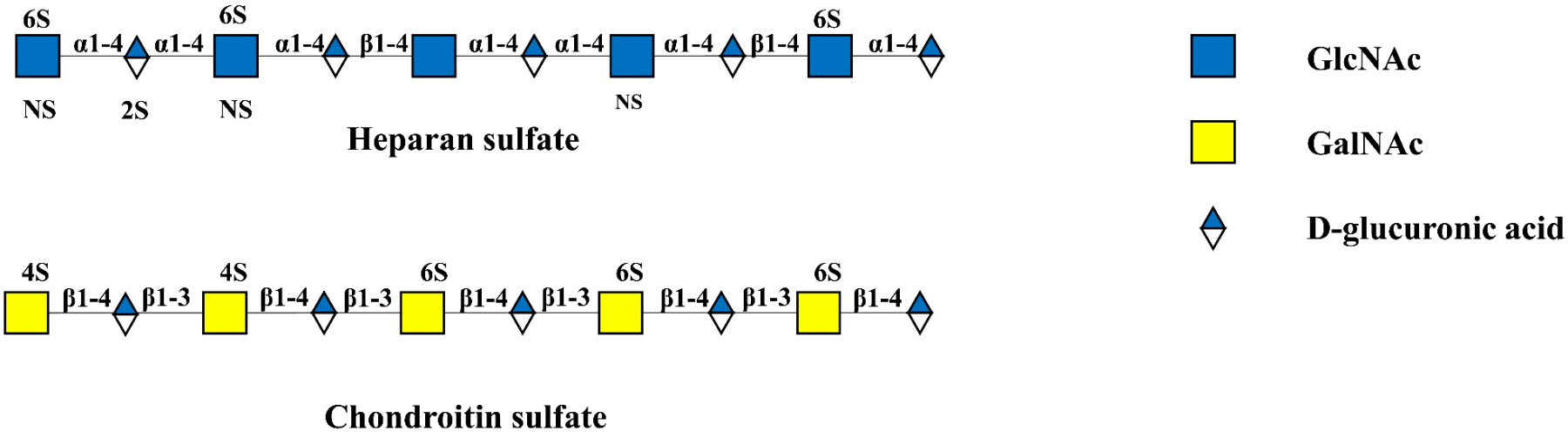


Figure 1.7: Representative examples of the structures of heparan sulfate and chondroitin sulfate (modified from Varki *et al.*, 2009). Heparan sulfate is a linear polysaccharide consisting of repeating disaccharide units of D-glucosamine and β-D-glucuronic acid. The D-glucosamine monomer can be *N*-acetylated or *N*-sulfated (represented by NS), both of which can subsequently 6-O-sulfated. The glucuronic acid monomer can be 2-O-sulfated. Chondroitin sulfate is a linear polysaccharide consisting of repeating disaccharide units of D-glucuronic acid and GalNAc. The GalNAc monomers can be 4 or 6-O-sulfated. The monosaccharide key is shown on the right.

Sulfatases constitute a family of hydrolases that are highly conserved between prokaryotes and eukaryotes, and whose substrate specificity can range from glycoproteins and glycolipids to hydroxysteroids (von Figura *et al.*, 1998). In both prokaryotes and eukaryotes, sulfatases are unique in that they require a 3-oxoalanine (typically called Cα-formylglycine or FGly) residue at the active site (Bond *et al.*, 1997; Lukatela *et al.*, 1998). The absence of this modification results in severely decreased sulfatase activity (Schmidt *et al.*, 1995). Eukaryotic sulfatases carry a conserved cysteine (Cys) residue, which is post-translationally modified to FGly in the endoplasmic reticulum (Dierks *et al.*, 1997; Schmidt *et al.*, 1995). Prokaryotic sulfatases carry either a Cys or a serine (Ser) residue which requires post-translational conversion to FGly in the cytosol in order to activate the enzyme (Beil *et al.*, 1995; Marquordt *et al.*, 2003; Miech *et al.*, 1998; Szameit *et al.*, 1999).

Initially, modification of Ser-type or Cys-type sulfatases in bacteria was thought to follow one of two pathways. In *Klebsiella pneumoniae* DSM 681, the conversion of the Ser_{72} residue to FGly was found to be catalysed by an iron-sulfur protein encoded by the *atsB* gene (Szameit *et al.*, 1999). The *atsB* gene is a homolog of the *chuR* gene from *Ba. thetaiotaomicron* BT4001 which was previously implicated in the metabolism of chondroitin sulfate and heparan sulfate in this species, although at the time it was predicted to encode a regulatory protein (Cheng & Salyers, 1995). AtsB is a member of the *S*-adenosyl-L-methionine (AdoMet)-dependent family of radical enzymes, whose activity was believed to be limited to the maturation of Ser-type sulfatases (Fang *et al.*, 2004). In humans and certain prokaryotes such as *Mycobacterium tuberculosis* H37Rv, modification of Cys-type sulfatases was shown to require a formylglycine-generating enzyme (FGE), an enzyme which is strictly dependent on oxygen (Carlson *et al.*, 2008; Cosma *et al.*, 2003; Dierks *et al.*, 2003;

Roeser *et al.*, 2006). However, it was notable that anaerobic bacteria encoding Cys-type sulfatases do not possess FGE-encoding genes, but, rather, *astB*-related genes (Berteau *et al.*, 2006). Since then, a prokaryotic maturase, which is active on Cys-type sulfatases under anaerobic conditions, has been characterised from *Clostridium perfringens* ATCC13124, a discovery which explained the absence of FGE-related genes in anaerobic bacteria (Berteau *et al.*, 2006). The authors of this study proposed that AtsB and the characterised maturase from *C. perfrigens* ATCC13124 belong to a new family named anSMEs (anaerobic sulfatase maturation enzymes), being separate from the oxygen-dependent FGE family (Berteau *et al.*, 2006). Subsequent research has shown that the anSME from *C. perfringens* ATCC13124, as well as the ChuR protein from *Ba. thetaiotaomicron* VPI-5481, were active on both Ser-type and Cys-type sulfatases (Benjdia *et al.*, 2008).

Metagenomic sequence analysis has revealed that genes that encode homologues of anSME enzymes are enriched in the gut microbiomes of humans compared to non-gut microbial communities (Kurokawa *et al.*, 2007), suggesting an important role in bacterial colonisation of the gut. *Ba. thetaiotaomicron* type strain VPI-5481 possesses 28 predicted sulfatase-encoding genes, but only a single gene for a predicted anSME (Benjdia *et al.*, 2011). Nevertheless, this anSME was shown to have significant importance in this strain's colonisation of the gut, as an isogenic derivative of this strain (designated ΔanSME) carrying a deletion in the anSME-encoding gene displayed reduced fitness *in vivo* even in the presence of dietary glycans (Benjdia *et al.*, 2011). As expected, growth of the ΔanSME strain was also impaired on sulfated carbohydrates, such as mucin, chondroitin sulphate and heparin, thus demonstrating the importance of sulfatase activity to carbohydrate metabolism in the gut (Benjdia *et al.*, 2011).

1.7 Carbon catabolite repression in bacteria

Carbon catabolite repression (CCR) is a regulatory process whereby, when presented with a number of carbon sources, bacteria will preferentially utilise the substrate that most effectively yields (the highest amount of) energy, while inhibiting the expression or activity of proteins involved in the uptake and catabolism of other substrates (Gorke & Stülke, 2008; Stülke & Hillen, 1999). The two main forms of CCR are (i) the inhibition of expression of genes involved in the uptake or utilisation of secondary substrates (Cohn & Horibata, 1959), and (ii) the inhibition of proteins involved in the uptake of secondary substrates, also known as inducer exclusion (Dills *et al.*, 1980; Gorke & Stulke, 2008). CCR is defined as "a regulatory phenomenon by which the expression of functions for the use of secondary carbon sources and the activities of the corresponding enzymes are reduced in the presence of a preferred carbon source" (Gorke & Stulke, 2008).

The methods by which CCR operates in the Gram negative bacterium *Escherichia coli* and the low-GC, Gram positive bacterium *Bacillus subtilis* have been studied extensively and have been shown to be hugely dependent on the PEP-PTS transport system (Postma *et al.*, 1993; Saier Jr & Ramseier, 1996). In *E. coli*, the key factor for CCR is the phosphorylation state of the IIA component of the glucose specific-PTS, $EIIA^{Glc}$. In times of high flux through glycolysis, the ratio of PEP (the high energy phosphate donor) to pyruvate in the cell is low, resulting in dephosphorylated $EIIA^{Glc}$. Concurrently, the availability of a PTS sugar, such as glucose, results in a phosphate group being transferred from $EIIA^{Glc}$ to the sugar. Therefore, during times of growth on a PTS sugar, $EIIA^{Glc}$ is dephosphorylated. In this form, the protein binds and inactivates enzymes and transporters involved in the metabolism of secondary metabolites, such as the LacY permease which transports lactose into the

cell, in an example of inducer exclusion as defined above. In the absence of a PTS sugar and when the PEP to pyruvate ratio is high, $EIIA^{Glc}$ is phosphorylated, and in this form $EIIA^{Glc}$ binds and activates adenylate cyclase, which leads to cyclic AMP (cAMP) synthesis. This triggers the formation of the cAMP-CRP complex, which in turn binds and activates the promoters of many catabolic genes and operons (Bettenbrock *et al.*, 2007; Feucht & Saier, 1980; Harwood *et al.*, 1976; Hogema *et al.*, 1998).

In *B. subtilis* the key player of CCR is the phosphorylation state of the histidine protein (HPr) of the PTS. Phosphorylated HPr, HPr-(Ser-P), binds to carbon catabolite protein A (CcpA) to promote its binding to *cre* sites on the DNA, in turn leading to the repression of transcription of other catabolic gene clusters, although in certain cases binding of CcpA to the *cre* site may also lead to transcriptional activation. This was observed for the *ackA* and *pta* genes of *B. subtilis* which are required for the excretion of acetate (Gorke & Stulke, 2008; Grundy *et al.*, 1993; Presecan-Siedel *et al.*, 1999). The phosphorylation of HPr is catalysed by the HprK protein. In times of high glycolytic activity, the levels of fructose-1,6-biphosphate in the cell are high, inducing HPrK kinase activity. However, when nutrients are limited, inorganic phosphate (Pi) accumulates, inducing HPrK phosphorylase activity, resulting in dephosphorylated HPr which cannot interact with CcpA, thus transcription of other catabolic gene clusters can proceed (Deutscher *et al.*, 1995; Jones *et al.*, 1997; Reizer *et al.*, 1998; Saier Jr *et al.*, 1996).

1.8 CCR-type phenomena and LacI type repression in bifidobacteria

CCR systems comparable to those described in *E. coli* and *B. subtilis* have so far not been identified in bifidobacteria, although phenomena reminiscent of CCR have been described in a number of bifidobacterial species. In *B. breve* UCC2003, expression of an operon involved in the metabolism of FOS was shown to be induced in the presence of sucrose and Actilight (a commercial source of short chain FOS), yet repressed when the strain was grown in a medium containing a mixture of glucose and sucrose, or fructose and sucrose, indicating the preferred utilisation of glucose or fructose (Ryan *et al.*, 2005). Similarly, when this strain was grown in a medium containing ribose or a combination of ribose and glucose, genes encoding the predicted fructose-specific PEP-PTS, as well as other carbohydrate-related ABC transport systems, were down-regulated (Pokusaeva *et al.*, 2010). The *B. longum* subsp. *longum* strain NCC2705 was shown to preferentially utilise lactose over glucose, a phenomenon previously only observed in *Streptococcus thermophilus* CNRZ302 and another *B. longum* strain, SH2 (Kim *et al.*, 2003; Parche *et al.*, 2006; van den Bogaard *et al.*, 2000). The preferential utilisation of lactose over glucose, was achieved by transcriptional down-regulation of the *glcP* gene, a glucose-specific permease of the major facilitator superfamily, in the presence of lactose (Parche *et al.*, 2006). In *B. animalis* subsp. *lactis*, sucrose metabolising activity was induced in the presence of sucrose, raffinose or oligofructose, but repressed in the presence of glucose (Trindade *et al.*, 2003). Another CCR-related metabolic phenomenon was observed when *B. longum* subsp. *infantis* ATCC15697 was grown on a complex mixture of HMOs. It was observed that of the five fucosidase-encoding genes identified on the *B. longum* subsp. *infantis* ATCC15697 genome, one was up-

regulated, two were down-regulated and two showed no change in transcription, compared to when the strain was grown in lactose (Sela *et al.*, 2012).

Transcriptional repressors are DNA binding proteins which physically interact with an operator sequence in the vicinity of a regulated promoter, thereby preventing the binding of RNA polymerase and transcription initiation. Perhaps the best characterised transcriptional repressor in bacteria is the LacI protein of *E. coli*, which represses transcription of the *lac* operon, required for the uptake and utilisation of lactose (Gilbert & Müller-Hill, 1966; Jacob & Monod, 1961). LacI monomers are divided into three distinct domains. The N-terminus contains a helix-turn-helix DNA binding domain and a short hinge region, both of which are involved in specific binding to an operator sequence upstream of the *lac* operon, thereby preventing transcription. The core domain, or effector binding domain, binds a sugar ligand, while at the C-terminal, the multi-merisation domain is essential for the formation of the LacI dimer (the native form of LacI) or the LacI tetramer (Lewis *et al.*, 1996). In the case of the *lac* operon, the effector molecule is allolactose, produced from lactose by β-galactosidase activity. Binding of allolactose to the core domain of the LacI repressor results in a conformational change that prevents binding of the protein to the operator sequence, thus permitting transcription of the *lac* operon (Jacob & Monod, 1961). Members of the LacI/GalR family of transcriptional regulators all function in the same manner as the LacI protein from *E. coli*, yet have specific effectors and DNA binding properties (Swint-Kruse & Matthews, 2009; Weickert & Adhya, 1992).

In a recent study on the comparative genomics of LacI-family transcription factors (LacI-TFs), the largest average number of such transcription factors per genome was found in lineages of the *Actinobacteria* phylum, including the *Bifidobacteriaceae*

family (Ravcheev *et al.*, 2014). Bifidobacterial genomes have been shown to encode a particularly high number of repressor proteins, a characteristic believed to allow the bacteria to adapt faster to fluctuations in carbohydrate sources (Schell *et al.*, 2002). For example, the genome of *B. longum* subsp. *longum* NCC2705 encodes 22 predicted LacI-type transcriptional repressors, all of which contain a sugar-binding motif, suggesting their involvement in the regulation of carbohydrate metabolism (Schell *et al.*, 2002). Just one carbohydrate-responsive transcriptional activator has so far been characterised in a bifidobacterial species: the ROK (Repressor open reading frame kinase)-type regulator RafR, which was shown to activate transcription of the raffinose utilisation cluster in *B. breve* UCC2003 (O'Connell *et al.*, 2014).

Six predicted LacI-type regulators encoded by *B. breve* UCC2003 have so far been characterised, including $LacI_{fos}$, which regulates transcription of the *fos* operon (Ryan *et al.*, 2005), GalR, which controls transcription of a galactan utilisation cluster (O'Connell Motherway *et al.*, 2011a), CldR which controls cellodextrin utilisation (Pokusaeva *et al.*, 2011b) and RbsR, which controls transcription of the *rbsACBDK* cluster for ribose utilisation (Pokusaeva *et al.*, 2010). Recently, two LacI-type regulators, designated MelR1 and MelR2, were shown to control transcription of the melezitose utilisation cluster in *B. breve* UCC2003 (O'Connell *et al.*, 2014). The characterised LacI-family transcriptional regulators from *B. breve* UCC2003 and their predicted or experimentally determined operator sequences are presented in Table 1.4. These repressors demonstrate the typical characteristics of LacI-TFs as reviewed recently (Ravcheev *et al.*, 2014). The latter review found that the majority of LacI-TFs (69 % of those studied) regulate one or two gene operons, and that 90 % of the investigated LacI-TFs are local regulators, as they control a

single metabolic pathway, usually involved in the catabolism of carbohydrates. This is in agreement with the characterised LacI-family repressors from *B. breve* UCC2003. In terms of the DNA binding motifs of the operators, the identified DNA binding sequences of LacI-family repressors from *B. breve* UCC2003 share the characteristic CG pair in the centre of the binding sites of other LacI-TFs, with the exception of the predicted binding site of $LacI_{fos}$ (Ryan *et al.,* 2005). Approximately 20 % of operons regulated by LacI-TFs contain more than one upstream binding site, similar to what was observed in the *gal* and *cld* operons of *B. breve* UCC2003 (O'Connell Motherway *et al.*, 2011a; Pokusaeva *et al.*, 2011b). Finally, 75 % of the binding motifs of LacI-TFs are located within 30 and 140 bp upstream of the closest regulated gene. This is in agreement with the LacI-binding sites identified in *B. breve* UCC2003, where only in one case the binding site is present at a distance of more than 200 bp from the start of regulated gene (O'Connell Motherway *et al.*, 2011a).

Table 1.4: Characterised LacI-family transcriptional regulators of *B. breve* UCC2003.

Protein Name	Locus Tag	Binding site	Gene cluster regulated (No. of genes)	No. of binding sites	Reference
CldR	Bbr_0105	xTxxTGGAAACGxTxCCAxxAA	*cldEFGC* (4)	2	(Pokusaeva *et al.*, 2011b)
AtsR1	Bbr_0351	xxGTxTxACGTTAxxxTx	Bbr_0349-0355 (7)	2	Chapter V of this thesis
GalR	Bbr_0421	xxxxTGxTACACxGxTxxAxCA	*galCDEGRA* (6)	4	(O'Connell Motherway *et al.*, 2011a)
$LacI_{fos}$	Bbr_1321	GGCCTCCGTTCAGCACGGAGGCC	*fosABC* (3)	1	(Ryan *et al.*, 2005)
RbsR	Bbr_1420	TGATTAAACGTTTAAATCA	*rbsACBDK* (5)	1	(Pokusaeva *et al.*, 2010)
MelR1	Bbr_1863	TGCATAAGCGCTTAGCAA	*melABCDE* (5)	1	(O'Connell *et al.*, 2014)
MelR2	Bbr_1864	TGCGTAATCGATATCGCA	Bbr_1861-1862 (2)	1	(O'Connell *et al.*, 2014)

1.9 Conclusion

As the number of complete bifidobacterial genome sequences continues to grow, so too does our understanding of the species and strain-specific attributes responsible for the metabolic diversity described in this review. An understanding of bifidobacterial HMO metabolism, as well as that of the structurally related oligosaccharides in mucin, will provide insights into how bifidobacterial species, such as *B. breve* and *B. bifidum*, can be effective and abundant colonisers of the infant gut, yet can also be present in significant numbers in the adult microbiota (Turroni *et al.*, 2012b), despite the dramatic shift in carbohydrate sources that occurs during weaning. Knowledge concerning the factors that contribute to bifidobacterial utilisation of mucin is also of increasing importance as questions are raised on the potential harm caused by degradation of the protective mucus layer and the effect of the Western-style diet and its associated microbiota on mucin degradation (Mahowald *et al.*, 2009; Marcobal *et al.*, 2013). Furthermore, an understanding of the characteristic metabolic traits of the various species of bifidobacteria, and their effects on other members of the gut microbiota is of crucial importance in the development of the next generation of probiotics and prebiotics.

1.10 Thesis outline

Chapter I (above) is a general introduction of this thesis which covers carbohydrate metabolism in bifidobacteria with particular emphasis on host-derived carbohydrates.

Chapter II describes the metabolism of sialic acid by *B. breve* UCC2003.

Chapter III reports on the regulation of sialic acid metabolism in *B. breve* UCC2003 by the transcriptional repressor, NanR.

Chapter IV describes the co-cultivation of *B. breve* UCC2003 and *B. bifidum* PRL2010 on a mucin-based medium.

Chapter V reports on the identification two sulfatase gene clusters in *B. breve* UCC2003, their transcriptional regulation and their role in the metabolism of sulfated monosaccharides.

Chapter VI contains a general discussion and the overall conclusions of the work performed as part of this thesis.

1.11 References

Amaretti, A., Bernardi, T., Leonardi, A., Raimondi, S., Zanoni, S. & Rossi, M. (2013). Fermentation of xylo-oligosaccharides by *Bifidobacterium adolescentis* DSMZ 18350: kinetics, metabolism, and β-xylosidase activities. *Appl Microbiol Biotechnol* **97**, 3109-3117.

Antonopoulos, D. A., Huse, S. M., Morrison, H. G., Schmidt, T. M., Sogin, M. L. & Young, V. B. (2009). Reproducible community dynamics of the gastrointestinal microbiota following antibiotic perturbation. *Infect Immun* **77**, 2367-2375.

Asakuma, S., Urashima, T., Akahori, M., Obayashi, H., Nakamura, T., Kimura, K., Watanabe, Y., Arai, I. & Sanai, Y. (2007). Variation of major neutral oligosaccharides levels in human colostrum. *Eur J Clin Nutr* **62**, 488-494.

Asakuma, S., Hatakeyama, E., Urashima, T., Yoshida, E., Katayama, T., Yamamoto, K., Kumagai, H., Ashida, H., Hirose, J. & other authors (2011). Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. *J Biol Chem* **286**, 34583-34592.

Ashida, H., Miyake, A., Kiyohara, M., Wada, J., Yoshida, E., Kumagai, H., Katayama, T. & Yamamoto, K. (2009). Two distinct α-L-fucosidases from *Bifidobacterium bifidum* are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. *Glycobiology* **19**, 1010-1017.

Avershina, E., Storrø, O., Øien, T., Johnsen, R., Wilson, R., Egeland, T. & Rudi, K. (2013). Bifidobacterial succession and correlation networks in a large unselected cohort of mothers and their children. *Appl Environ Microbiol* **79**, 497-507.

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920.

Barrangou, R., Briczinski, E. P., Traeger, L. L., Loquasto, J. R., Richards, M., Horvath, P., Coûté-Monvoisin, A. C., Leyer, G., Rendulic, S. & other authors (2009). Comparison of the complete genome sequences of *Bifidobacterium animalis* subsp. *lactis* DSM 10140 and Bl-04. *J Bacteriol* **191**, 4144-4151.

Beil, S., Kehrli, H., James, P., Staudenmann, W., Cook, A. M., Leisinger, T. & Kertesz, M. A. (1995). Purification and characterization of the arylsulfatase synthesized by *Pseudomonas aeruginosa* PAO during growth in sulfate-free medium and cloning of the arylsulfatase gene (*atsA*). *Eur J Biochem* **229**, 385-394.

Benjdia, A., Subramanian, S., Leprince, J., Vaudry, H., Johnson, M. K. & Berteau, O. (2008). Anaerobic sulfatase-maturating enzymes, first dual substrate radical S-adenosylmethionine enzymes. *J Biol Chem* **283**, 17815-17826.

Benjdia, A., Martens, E. C., Gordon, J. I. & Berteau, O. (2011). Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont, *Bacteroides thetaiotaomicron*. *J Biol Chem* **286**, 25973-25982.

Berteau, O., Guillot, A., Benjdia, A. & Rabot, S. (2006). A new type of bacterial sulfatase reveals a novel maturation pathway in prokaryotes. *J Biol Chem* **281**, 22464-22470.

Bettenbrock, K., Sauter, T., Jahreis, K., Kremling, A., Lengeler, J. W. & Gilles, E. D. (2007). Correlation between growth rates, EIIACrr phosphorylation, and intracellular cyclic AMP levels in *Escherichia coli* K-12. *J Bacteriol* **189**, 6891-6900.

Biasucci, G., Rubini, M., Riboni, S., Morelli, L., Bessi, E. & Retetangos, C. (2010). Mode of delivery affects the bacterial community in the newborn gut. *Early Hum Dev* **86**, 13-15.

Bode, L. & Jantscher-Krenn, E. (2012). Structure-function relationships of human milk oligosaccharides. *Adv Nutr* **3**, 383S-391S.

Bond, C. S., Clements, P. R., Ashby, S. J., Collyer, C. A., Harrop, S. J., Hopwood, J. J. & Guss, J. M. (1997). Structure of a human lysosomal sulfatase. *Structure* **5**, 277-289.

Bond, J. H., Currier, B. E., Buchwald, H. & Levitt, M. D. (1980). Colonic conservation of malabsorbed carbohydrate. *Gastroenterology* **78**, 444-447.

Bottacini, F., Dal Bello, F., Turroni, F., Milani, C., Duranti, S., Foroni, E., Viappiani, A., Strati, F., Mora, D. & other authors (2011). Complete genome

sequence of *Bifidobacterium animalis* subsp. *lactis* BLC1. *J Bacteriol* **193**, 6387-6388.

Bottacini, F., Milani, C., Turroni, F., Sánchez, B., Foroni, E., Duranti, S., Serafini, F., Viappiani, A., Strati, F. & other authors (2012). *Bifidobacterium asteroides* PRL2011 genome analysis reveals clues for colonization of the insect gut. *PloS one* **7**, e44229.

Bottacini, F., O'Connell Motherway, M., Kuczynski, J., OConnell, K., Serafini, F., Duranti, S., Milani, C., Turroni, F., Lugli, G. & other authors (2014). Comparative genomics of the *Bifidobacterium breve* taxon. *BMC Genomics* **15**, 170.

Bouhnik, Y., Attar, A., Joly, F. A., Riottot, M., Dyard, F. & Flourie, B. (2004). Lactulose ingestion increases faecal bifidobacterial counts: A randomised double-blind study in healthy humans. *Eur J Clin Nutr* **58**, 462-466.

Brockhausen, I. (2003). Sulphotransferases acting on mucin-type oligosaccharides. *Biochem Soc Trans* **31**, 318-325.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37**, D233-D238.

Capon, C., Maes, E., Michalski, J., Leffler, H. & Kim, Y. S. (2001). Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. *Biochem J* **358**, 657-664.

Carlson, B. L., Ballister, E. R., Skordalakes, E., King, D. S., Breidenbach, M. A., Gilmore, S. A., Berger, J. M. & Bertozzi, C. R. (2008). Function and structure of a prokaryotic formylglycine-generating enzyme. *J Biol Chem* **283**, 20117-20125.

Chaturvedi, P., Warren, C. D., Altaye, M., Morrow, A. L., Ruiz-Palacios, G., Pickering, L. K. & Newburg, D. S. (2001). Fucosylated human milk oligosaccharides vary between individuals and over the course of lactation. *Glycobiology* **11**, 365-372.

Cheng, Q. & Salyers, A. A. (1995). Use of suppressor analysis to find genes involved in the colonization deficiency of a *Bacteroides thetaiotaomicron* mutant unable to grow on the host-derived mucopolysaccharides chondroitin sulfate and heparin. *Appl Environ Microbiol* **61**, 734-740.

Chervaux, C., Grimaldi, C., Bolotin, A., Quinquis, B., Legrain-Raspaud, S., van Hylckama Vlieg, J. E. T., Denariaz, G. & Smokvina, T. (2011). Genome sequence of the probiotic strain *Bifidobacterium animalis* subsp. *lactis* CNCM I-2494. *J Bacteriol* **193**, 5560-5561.

Claesson, M. J., Jeffery, I. B., Conde, S., Power, S. E., O'Connor, E. M., Cusack, S., Harris, H. M. B., Coakley, M., Lakshminarayanan, B. & other authors (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* **488**, 178-184.

Cohn, M. & Horibata, K. (1959). Inhibition by glucose of the induced synthesis of the beta-galactoside-enzyme system of *Escherichia coli*. Analysis of maintenance. *J Bacteriol* **78**, 601-612.

Coppa, G. V., Gabrielli, O., Pierani, P., Catassi, C., Carlucci, A. & Giorgi, P. L. (1993). Changes in carbohydrate composition in human milk over 4 months of lactation. *Pediatrics* **91**, 637-641.

Corfield, A. P., Wagner, S. A., Clamp, J., Kriaris, M. & Hoskins, L. (1992). Mucin degradation in the human colon: production of sialidase, sialate *O*-acetylesterase, *N*-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria. *Infect Immun* **60**, 3971-3978

Cosma, M. P., Pepe, S., Annunziata, I., Newbold, R. F., Grompe, M., Parenti, G. & Ballabio, A. (2003). The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases. *Cell* **113**, 445-456.

Crittenden, R. G. & Playne, M. J. (1996). Production, properties and applications of food-grade oligosaccharides. *Trends Food Sci Tech* **7**, 353-361.

Crost, E. H., Tailford, L. E., Le Gall, G., Fons, M., Henrissat, B. & Juge, N. (2013). Utilisation of mucin glycans by the human gut symbiont *Ruminococcus gnavus* is strain-dependent. *PLoS ONE* **8**, e76341.

Cummings, J. & Macfarlane, G. (1991). The control and consequences of bacterial fermentation in the human colon. *J Appl Bacteriol* **70**, 443-459.

Davidson, A. L. & Chen, J. (2004). ATP-binding cassette transporters in bacteria. *Ann Rev Biochem* **73**, 241-268.

Dawson, P. A., Lan, T. & Rao, A. (2009). Bile acid transporters. *J Lipid Res* **50**, 2340-2357.

de Vos, W. M. & de Vos, E. A. J. (2012). Role of the intestinal microbiome in health and disease: from correlation to causation. *Nutr Rev* **70**, S45-S56.

de Vries, W. & Stouthamer, A. (1967). Pathway of glucose fermentation in relation to the taxonomy of bifidobacteria. *J Bacteriol* **93**, 574-576.

Degnan, B. A. & Macfarlane, G. T. (1993). Transport and metabolism of glucose and arabinose in *Bifidobacterium breve*. *Arch Microbiol* **160**, 144-151.

Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* **54**, 1469-1476.

Deutscher, J., Küster, E., Bergstedt, U., Charrier, V. & Hillen, W. (1995). Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in Gram-positive bacteria. *Mol Microbiol* **15**, 1049-1053.

Deutscher, J., Francke, C. & Postma, P. W. (2006). How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* **70**, 939-1031.

Dierks, T., Schmidt, B. & von Figura, K. (1997). Conversion of cysteine to formylglycine: A protein modification in the endoplasmic reticulum. *Proc Natl Acad Sci USA* **94**, 11963-11968.

Dierks, T., Schmidt, B., Borissenko, L. V., Peng, J., Preusser, A., Mariappan, M. & von Figura, K. (2003). Multiple sulfatase deficiency is caused by mutations in the gene encoding the human Cα-formylglycine generating enzyme. *Cell* **113**, 435-444.

Dills, S. S., Apperson, A., Schmidt, M. R. & Saier Jr, M. H. (1980). Carbohydrate transport in bacteria. *Microbiol Rev* **44**, 385.

Duranti, S., Turroni, F., Milani, C., Foroni, E., Bottacini, F., Dal Bello, F., Ferrarini, A., Delledonne, M., van Sinderen, D. & other authors (2012). Exploration of genomic diversity and core genome of the *Bifidobacterium adolescentis* phylogenetic group by means of a polyphasic approach. *Appl Environ Microbiol* **79**, 336-346.

Duranti, S., Turroni, F., Lugli, G. A., Milani, C., Viappiani, A., Mangifesta, M., Gioiosa, L., Palanza, P., van Sinderen, D. & other authors (2014). Genomic characterization and transcriptional studies of the starch-utilizing strain *Bifidobacterium adolescentis* 22L. *Appl Environ Microbiol* **80**, 6080-6090.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638.

Egan, M., O'Connell Motherway, M., Ventura, M. & van Sinderen, D. (2014a). Metabolism of sialic acid by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **80**, 4414-4426.

Egan, M., O' Connell Motherway, M., Kilcoyne, M., Kane, M., Joshi, L., Ventura, M. & van Sinderen, D. (2014b). Cross-feeding by *Bifidobacterium breve* UCC2003 during co-cultivation with *Bifidobacterium bifidum* PRL2010 in a mucin-based medium. *BMC Microbiol* **14**, 282.

Eliakim, R., Gilead, L., Ligumsky, M., Okon, E., Rachmilewitz, D. & Razin, E. (1986). Histamine and chondroitin sulfate E proteoglycan released by cultured human colonic mucosa: indication for possible presence of E mast cells. *Proc Natl Acad Sci USA* **83**, 461-464.

Englyst, H. N., Kingman, S. & Cummings, J. (1992). Classification and measurement of nutritionally important starch fractions. *Eur J Clin Nutr* **46**, S33-50.

Fallingborg, J. (1999). Intraluminal pH of the human gastrointestinal tract. *Dan Med Bull* **46**, 183-196.

Fanaro, S., Chierici, R., Guerrini, P. & Vigi, V. (2003). Intestinal microflora in early infancy: composition and development. *Acta paediatrica* **92**, 48-55.

Fang, Q., Peng, J. & Dierks, T. (2004). Post-translational formylglycine modification of bacterial sulfatases by the radical S-adenosylmethionine protein AtsB. *J Biol Chem* **279**, 14570-14578.

Feucht, B. U. & Saier, M. (1980). Fine control of adenylate cyclase by the phosphoenolpyruvate: sugar phosphotransferase systems in *Escherichia coli* and *Salmonella typhimurium*. *J Bacteriol* **141**, 603-610.

Fouhy, F., Guinane, C. M., Hussey, S., Wall, R., Ryan, C. A., Dempsey, E. M., Murphy, B., Ross, R. P., Fitzgerald, G. F. & other authors (2012). High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin. *Antimicrob Agents Chemother* **56**, 5811-5820.

Frank, D. N., St. Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N. & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* **104**, 13780-13785.

Froehlich, J. W., Dodds, E. D., Barboza, M., McJimpsey, E. L., Seipert, R. R., Francis, J., An, H. J., Freeman, S., German, J. B. & other authors (2010). Glycoprotein expression in human milk during lactation. *J Agric Food Chem* **58**, 6440-6448.

Fujita, K., Oura, F., Nagamine, N., Katayama, T., Hiratake, J., Sakata, K., Kumagai, H. & Yamamoto, K. (2005). Identification and molecular cloning of a novel glycoside hydrolase family of core 1 type O-glycan-specific endo-α-*N*-acetylgalactosaminidase from *Bifidobacterium longum*. *J Biol Chem* **280**, 37415-37422.

Fujita, K., Sakaguchi, T., Sakamoto, A., Shimokawa, M. & Kitahara, K. (2014). *Bifidobacterium longum* subsp. *longum* exo-β-1,3-galactanase is an enzyme for the degradation of type II arabinogalactan. *Appl Environ Microbiol* **80**, 4577-4584.

Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., Tobe, T., Clarke, J. M., Topping, D. L. & other authors (2011). Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543-547.

Fusch, G., Choi, A., Rochow, N. & Fusch, C. (2011). Quantification of lactose content in human and cow's milk using UPLC–tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **879**, 3759-3762.

Gao, B. & Gupta, R. S. (2012). Phylogenetic framework and molecular signatures for the main clades of the phylum *Actinobacteria*. *Microbiol Mol Biol Rev* **76**, 66-112.

Garrido, D., Kim, J. H., German, J. B., Raybould, H. E. & Mills, D. A. (2011). Oligosaccharide binding proteins from *Bifidobacterium longum* subsp. *infantis* reveal a preference for host glycans. *PLoS ONE* **6**, e17315.

Garrido, D., Nwosu, C., Ruiz-Moyano, S., Aldredge, D., German, J. B., Lebrilla, C. B. & Mills, D. A. (2012). Endo-β-*N*-acetylglucosaminidases from infant gut-associated bifidobacteria release complex *N*-glycans from human milk glycoproteins. *Mol Cell Proteomics* **11**, 775-785.

Garrido, D., Ruiz-Moyano, S., Jimenez-Espinoza, R., Eom, H. J., Block, D. E. & Mills, D. A. (2013). Utilization of galactooligosaccharides by *Bifidobacterium longum* subsp. *infantis* isolates. *Food Microbiol* **33**, 262-270.

Garrigues, C., Johansen, E. & Pedersen, M. B. (2010). Complete genome sequence of *Bifidobacterium animalis* subsp. *lactis* BB-12, a widely consumed probiotic strain. *J Bacteriol* **192**, 2467-2468.

Gavini, F., Van Esbroeck, M., Touzel, J. P., Fourment, A. & Goossens, H. (1996). Detection of fructose-6-phosphate phosphoketolase (F6PPK), a key enzyme of the bifid-shunt, in *Gardnerella vaginalis*. *Anaerobe* **2**, 191-193.

Gibson, G., Probert, H., Van Loo, J., Rastall, R. & Roberfroid, M. (2004). Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutr Res Rev* **17**, 259-275.

Gilbert, W. & Müller-Hill, B. (1966). Isolation of the lac repressor. *Proc Natl Acad Sci USA* **56**, 1891.

Gill, S., Pop, M., DeBoy, R., Eckburg, P., Turnbaugh, P., Samuel, B., Gordon, J., Relman, D., Fraser-Liggett, C. & other authors (2006). Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355-1359.

Gorke, B. & Stulke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Micro* **6**, 613-624.

Greenblum, S., Turnbaugh, P. J. & Borenstein, E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc Natl Acad Sci USA* **109**, 594-599.

Grundy, F. J., Waters, D. A., Allen, S. H. & Henkin, T. M. (1993). Regulation of the *Bacillus subtilis* acetate kinase gene by CcpA. *J Bacteriol* **175**, 7348-7355.

Guérardel, Y., Morelle, W., Plancke, Y., Lemoine, J. & Strecker, G. (1999). Structural analysis of three sulfated oligosaccharides isolated from human milk. *Carbohydr Res* **320**, 230-238.

György, P., Norris, R. F. & Rose, C. S. (1954). Bifidus factor. A variant of *Lactobacillus bifidus* requiring a special growth factor. *Arch Biochem Biophys* **48**, 193-201.

Ham, J. S., Lee, T., Byun, M. J., Lee, K. T., Kim, M. K., Han, G. S., Jeong, S. G., Oh, M. H., Kim, D. H. & other authors (2011). Complete genome sequence of *Bifidobacterium longum* subsp. *longum* KACC 91563. *J Bacteriol* **193**, 5044.

Hao, Y., Huang, D., Guo, H., Xiao, M., An, H., Zhao, L., Zuo, F., Zhang, B., Hu, S. & other authors (2011). Complete genome sequence of *Bifidobacterium longum* subsp. *longum* BBMN68, a new strain from a healthy Chinese centenarian. *J Bacteriol* **193**, 787-788.

Harwood, J. P., Gazdar, C., Prasad, C., Peterkofsky, A., Curtis, S. & Epstein, W. (1976). Involvement of the glucose enzymes II of the sugar phosphotransferase

system in the regulation of adenylate cyclase by glucose in *Escherichia coli*. *J Biol Chem* **251**, 2462-2468.

Heel, K. A., McCauley, R. D., Papadimitriou, J. M. & Hall, J. C. (1997). Review: Peyer's patches. *J Gastroenterol Hepatol* **12**, 122-136.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **280 (Pt 2)**, 309-316.

Hill, C., Guarner, F., Reid, G., Gibson, G. R., Merenstein, D. J., Pot, B., Morelli, L., Canani, R. B., Flint, H. J. & other authors (2014). Expert consensus document: The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat Rev Gastroenterol Hepatol* **11**, 506-514.

Hogema, B. M., Arents, J. C., Bader, R., Eijkemans, K., Yoshida, H., Takahashi, H., Aiba, H. & Postma, P. W. (1998). Inducer exclusion in *Escherichia coli* by non-PTS substrates: the role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIAGlc. *Mol Microbiol* **30**, 487-498.

Hoskins, L. C., Agustines, M., McKee, W. B., Boulding, E. T., Kriaris, M. & Niedermeyer, G. (1985). Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *J Clin Invest* **75**, 944-953.

Huurre, A., Kalliomäki, M., Rautava, S., Rinne, M., Salminen, S. & Isolauri, E. (2008). Mode of delivery - effects on gut microbiota and humoral immunity. *Neonatology* **93**, 236-240.

Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**, 318-356.

Jans, C., Lacroix, C., Follador, R. & Stevens, M. J. A. (2013). Complete genome sequence of the probiotic *Bifidobacterium thermophilum* strain RBL67. *Genome Announc* **1**.

Johansson, M. E. V., Larsson, J. M. H. & Hansson, G. C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host–microbial interactions. *Proc Natl Acad Sci USA* **108**, 4659-4665.

Jones, B. E., Dossonnet, V., Küster, E., Hillen, W., Deutscher, J. & Klevit, R. E. (1997). Binding of the catabolite repressor protein CcpA to its DNA target is regulated by phosphorylation of its corepressor HPr. *J Biol Chem* **272**, 26530-26535.

Karlsson, F. H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C. J., Fagerberg, B., Nielsen, J. & Backhed, F. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99-103.

Katayama, T., Sakuma, A., Kimura, T., Makimura, Y., Hiratake, J., Sakata, K., Yamanoi, T., Kumagai, H. & Yamamoto, K. (2004). Molecular cloning and

characterization of *Bifidobacterium bifidum* 1,2-α-l-fucosidase (*afcA*), a novel inverting glycosidase (glycoside hydrolase family 95). *J Bacteriol* **186**, 4885-4893.

Kim, J. F., Jeong, H., Yu, D. S., Choi, S.-H., Hur, C.-G., Park, M.-S., Yoon, S. H., Kim, D.-W., Ji, G. E. & other authors (2009). Genome sequence of the probiotic bacterium *Bifidobacterium animalis* subsp. *lactis* AD011. *J Bacteriol* **191**, 678-679.

Kim, T. B., Song, S. H., Kang, S. C. & Oh, D. K. (2003). Quantitative comparison of lactose and glucose utilization in *Bifidobacterium longum* cultures. *Biotechnol Progr* **19**, 672-675.

Kitaoka, M., Tian, J. & Nishimoto, M. (2005). Novel putative galactose operon involving lacto-*N*-biose phosphorylase in *Bifidobacterium longum*. *Appl Environ Microbiol* **71**, 3158-3162.

Kiyohara, M., Tanigawa, K., Chaiwangsri, T., Katayama, T., Ashida, H. & Yamamoto, K. (2011). An exo-α-sialidase from bifidobacteria involved in the degradation of sialyloligosaccharides in human milk and intestinal glycoconjugates. *Glycobiology* **21**, 437-447.

Kiyohara, M., Nakatomi, T., Kurihara, S., Fushinobu, S., Suzuki, H., Tanaka, T., Shoda, S.-i., Kitaoka, M., Katayama, T. & other authors (2012). α-*N*-acetylgalactosaminidase from infant-associated bifidobacteria belonging to novel glycoside hydrolase family 129 is implicated in alternative mucin degradation pathway. *J Biol Chem* **287**, 693-700.

Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., Angenent, L. T. & Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci USA* **108**, 4578-4585.

Kogelberg, H., Piskarev, V. E., Zhang, Y., Lawson, A. M. & Chai, W. (2004). Determination by electrospray mass spectrometry and 1H-NMR spectroscopy of primary structures of variously fucosylated neutral oligosaccharides based on the iso-lacto-*N*-octaose core. *Eur J Biochem* **271**, 1172-1186.

Kunz, C., Rudloff, S., Baier, W., Klein, N. & Strobel, S. (2000). Oligosaccharides in human milk: Structural, functional, and metabolic aspects. *Annu Rev Nutr* **20**, 699-722.

Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V. K. & other authors (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**, 169-181.

Lagaert, S., Pollet, A., Delcour, J. A., Lavigne, R., Courtin, C. M. & Volckaert, G. (2010). Substrate specificity of three recombinant α-l-arabinofuranosidases from *Bifidobacterium adolescentis* and their divergent action on arabinoxylan and arabinoxylan oligosaccharides. *Biochem Biophys Res Commun* **402**, 644-650.

Larsen, N., Vogensen, F. K., van den Berg, F. W. J., Nielsen, D. S., Andreasen, A. S., Pedersen, B. K., Al-Soud, W. A., Sørensen, S. J., Hansen, L. H. & other

authors (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS ONE* **5**, e9085.

Larsson, J. M., Karlsson, H., Sjovall, H. & Hansson, G. C. (2009). A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19**, 756-766.

Lattimer, J. M. & Haub, M. D. (2010). Effects of dietary fiber and its components on metabolic health. *Nutrients* **2**, 1266-1289.

LeBlanc, J. G., Laiño, J. E., del Valle, M. J., Vannini, V., van Sinderen, D., Taranto, M. P., de Valdez, G. F., de Giori, G. S. & Sesma, F. (2011). B-Group vitamin production by lactic acid bacteria – current knowledge and potential applications. *J Appl Microbiol* **111**, 1297-1309.

LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D. & Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr Opin Biotechnol* **24**, 160-168.

Lee, J. H., Karamychev, V., Kozyavkin, S., Mills, D., Pavlov, A., Pavlova, N., Polouchine, N., Richardson, P., Shakhova, V. & other authors (2008). Comparative genomic analysis of the gut bacterium *Bifidobacterium longum* reveals loci susceptible to deletion during pure culture growth. *BMC genomics* **9**, 247.

Lengeler, J. W., Mayer, R. J. & Schmid, K. (1982). Phosphoenolpyruvate-dependent phosphotransferase system enzyme III and plasmid-encoded sucrose transport in *Escherichia coli* K-12. *J Bacteriol* **151**, 468-471.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247-1254.

Ley, R. E., Peterson, D. A. & Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837-848.

Liu, L., Qin, Y., Wang, Y., Li, H., Shang, N. & Li, P. (2014). Complete genome sequence of *Bifidobacterium animalis* RH, a probiotic bacterium producing exopolysaccharides. *J Biotechnol* **189c**, 86-87.

LoCascio, R., Ninonuevo, M., Freeman, S., Sela, D., Grimm, R., Lebrilla, C., Mills, D. & German, B. (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**, 8914-8919.

LoCascio, R., Ninonuevo, M., Kronewitter, S., Freeman, S., German, B., Lebrilla, C. & Mills, D. (2009). A versatile and scalable strategy for glycoprofiling bifidobacterial consumption of human milk oligosaccharides. *Microb Biotechnol* **2**, 333-342.

LoCascio, R. G., Desai, P., Sela, D. A., Weimer, B. & Mills, D. A. (2010). Broad conservation of milk utilization genes in *Bifidobacterium longum* subsp. *infantis* as revealed by comparative genomic hybridization. *Appl Environ Microbiol* **76**, 7373-7381.

Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-D495.

Loquasto, J. R., Barrangou, R., Dudley, E. G., Stahl, B., Chen, C. & Roberts, R. F. (2013). *Bifidobacterium animalis* subsp. *lactis* ATCC 27673 is a genomically unique strain within its conserved subspecies. *Appl Environ Microbiol* **79**, 6903-6910.

Lugli, G. A., Milani, C., Turroni, F., Duranti, S., Ferrario, C., Viappiani, A., Mancabelli, L., Mangifesta, M., Taminiau, B. & other authors (2014). Investigation of the evolutionary development of the genus *Bifidobacterium* by comparative genomics. *Appl Environ Microbiol* **80**, 6383-6394.

Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., Von Figura, K. & Saenger, W. (1998). Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* **37**, 3654-3664.

Macfarlane, S. & Macfarlane, G. T. (2003). Regulation of short-chain fatty acid production. *Proc Nutr Soc* **62**, 67-72.

Macfarlane, S., Macfarlane, G. T. & Cummings, J. H. (2006). Review article: prebiotics in the gastrointestinal tract. *Aliment Pharmacol Ther* **24**, 701-714.

Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., Shah, N., Wang, C., Magrini, V. & other authors (2009). Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc Natl Acad Sci USA* **106**, 5859-5864.

Marcobal, A., Southwick, A. M., Earle, K. A. & Sonnenburg, J. L. (2013). A refined palate: Bacterial consumption of host glycans in the gut. *Glycobiology* **23**, 1038-1046.

Marquordt, C., Fang, Q., Will, E., Peng, J., von Figura, K. & Dierks, T. (2003). Posttranslational modification of serine to formylglycine in bacterial sulfatases recognition of the modification motif by the iron-sulfur protein AtsB. *J Biol Chem* **278**, 2212-2218.

Maze, A., O'Connell Motherway, M., Fitzgerald, G., Deutscher, J. & van Sinderen, D. (2007). Identification and characterization of a fructose phosphotransferase system in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **73**, 545-553.

McNeil, N. I. (1984). The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr* **39**, 338-342.

Meslin, J. C., Fontaine, N. & Andrieux, C. (1999). Variation of mucin distribution in the rat intestine, caecum and colon: effect of the bacterial flora. *Comp Biochem Physiol, Part A: Mol Integr Physiol* **123**, 235-239.

Miech, C., Dierks, T., Selmer, T., von Figura, K. & Schmidt, B. (1998). Arylsulfatase from *Klebsiella pneumoniae* carries a formylglycine generated from a serine. *J Biol Chem* **273**, 4835-4837.

Milani, C., Duranti, S., Lugli, G. A., Bottacini, F., Strati, F., Arioli, S., Foroni, E., Turroni, F., van Sinderen, D. & other authors (2013). Comparative genomics of *Bifidobacterium animalis* subsp. *lactis* reveals a strict monophyletic bifidobacterial taxon. *Appl Environ Microbiol* **79**, 4304-4315.

Milani, C., Lugli, G. A., Duranti, S., Turroni, F., Bottacini, F., Mangifesta, M., Sanchez, B., Viappiani, A., Mancabelli, L. & other authors (2014). Genome encyclopaedia of type strains of the genus *Bifidobacterium*. *Appl Environ Microbiol* **80**, 6290-6302.

Mitoulas, L. R., Kent, J. C., Cox, D. B., Owens, R. A., Sherriff, J. L. & Hartmann, P. E. (2002). Variation in fat, lactose and protein in human milk over 24h and throughout the first year of lactation. *Br J Nutr* **88**, 29-37.

Miwa, M., Horimoto, T., Kiyohara, M., Katayama, T., Kitaoka, M., Ashida, H. & Yamamoto, K. (2010). Cooperation of β-galactosidase and β-*N*-acetylhexosaminidase from bifidobacteria in assimilation of human milk oligosaccharides with type 2 structure. *Glycobiology* **20**, 1402-1409.

Morelle, W. & Michalski, J. C. (2007). Analysis of protein glycosylation by mass spectrometry. *Nat Protocols* **2**, 1585-1602.

Newburg, D., Linhardt, R., Ampofo, S. & Yolken, R. (1995). Human milk glycosaminoglycans inhibit HIV glycoprotein gp120 binding to its host cell CD4. *J Nutr* **125**, 419.

Ninonuevo, M., Park, Y., Yin, H., Zhang, J., Ward, R., Clowers, B., German, B., Freeman, S., Killeen, K. & other authors (2006). A strategy for annotating the human milk glycome. *J Agric Food Chem* **54**, 7471-7480.

Nishimoto, M. & Kitaoka, M. (2007). Identification of *N*-acetylhexosamine 1-kinase in the complete lacto-*N*-biose I/galacto-*N*-biose metabolic pathway in *Bifidobacterium longum*. *Appl Environ Microbiol* **73**, 6444-6449.

O'Connell, K. J., O'Connell Motherway, M., O'Callaghan, J., Fitzgerald, G. F., Ross, R. P., Ventura, M., Stanton, C. & van Sinderen, D. (2013). Metabolism of four α-glycosidic linkage-containing oligosaccharides by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **79**, 6280-6292.

O'Connell, K. J., O'Connell Motherway, M., Liedtke, A., Fitzgerald, G. F., Ross, R. P., Stanton, C., Zomer, A. & van Sinderen, D. (2014). Transcription of two adjacent carbohydrate utilization gene clusters in *Bifidobacterium breve* UCC2003 is controlled by LacI- and Repressor Open reading frame Kinase (ROK)-type regulators. *Appl Environ Microbiol* **80**, 3604-3614.

O'Connell Motherway, M., Fitzgerald, G. F. & van Sinderen, D. (2011a). Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **4**, 403-416.

O'Connell Motherway, M., Zomer, A., Leahy, S. C., Reunanen, J., Bottacini, F., Claesson, M. J., O'Brien, F., Flynn, K., Casey, P. G. & other authors (2011b). Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc Natl Acad Sci USA* **108**, 11217-11222.

O'Connell Motherway, M., Kinsella, M., Fitzgerald, G. F. & van Sinderen, D. (2013). Transcriptional and functional characterization of genetic elements involved in galacto-oligosaccharide utilization by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **6**, 67-79.

Oshiro, M., Ono, K., Suzuki, Y., Ota, H., Katsuyama, T. & Mori, N. (2001). Immunohistochemical localization of heparan sulfate proteoglycan in human gastrointestinal tract. *Histochem Cell Biol* **115**, 373-380.

Palframan, R. J., Gibson, G. R. & Rastall, R. A. (2003). Carbohydrate preferences of *Bifidobacterium* species isolated from the human gut. *Curr Iss Intest Microbiol* **4**, 71-75.

Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol* **5**, e177.

Parche, S., Beleut, M., Rezzonico, E., Jacobs, D., Arigoni, F., Titgemeyer, F. & Jankovic, I. (2006). Lactose-over-glucose preference in *Bifidobacterium longum* NCC2705: *glcP*, encoding a glucose transporter, is subject to lactose repression. *J Bacteriol* **188**, 1260-1265.

Parche, S., Amon, J., Jankovic, I., Rezzonico, E., Beleut, M., Barutcu, H., Schendel, I., Eddy, M. P., Burkovski, A. & other authors (2007). Sugar transport systems of *Bifidobacterium longum* NCC2705. *J Mol Microbiol Biotechnol* **12**, 9-19.

Petersson, J., Schreiber, O., Hansson, G. C., Gendler, S. J., Velcich, A., Lundberg, J. O., Roos, S., Holm, L. & Phillipson, M. (2011). Importance and regulation of the colonic mucus barrier in a mouse model of colitis. *Am J Physiol Gastrointest Liver Physiol* **300**, G327-G333.

Picariello, G., Ferranti, P., Mamone, G., Roepstorff, P. & Addeo, F. (2008). Identification of *N*-linked glycoproteins in human milk by hydrophilic interaction liquid chromatography and mass spectrometry. *Proteomics* **8**, 3833-3847.

Podolsky, D. K. (1985). Oligosaccharide structures of human colonic mucin. *J Biol Chem* **260**, 8262-8271.

Pokusaeva, K., Neves, A. R., Zomer, A., O'Connell-Motherway, M., MacSharry, J., Curley, P., Fitzgerald, G. F. & van Sinderen, D. (2010). Ribose utilization by the human commensal *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **3**, 311-323.

Pokusaeva, K., Fitzgerald, G. F. & Sinderen, D. (2011a). Carbohydrate metabolism in bifidobacteria. *Genes Nutr* **6**, 285-306.

Pokusaeva, K., O'Connell Motherway, M., Zomer, A., MacSharry, J., Fitzgerald, G. F. & van Sinderen, D. (2011b). Cellodextrin utilization by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **77**, 1681-1690.

Pompei, A., Cordisco, L., Amaretti, A., Zanoni, S., Matteuzzi, D. & Rossi, M. (2007). Folate production by bifidobacteria as a potential probiotic property. *Appl Environ Microbiol* **73**, 179-185.

Postma, P., Lengeler, J. & Jacobson, G. (1993). Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* **57**, 543.

Presecan-Siedel, E., Galinier, A., Longin, R., Deutscher, J., Danchin, A., Glaser, P. & Martin-Verstraete, I. (1999). Catabolite regulation of the *pta* gene as part of carbon flow pathways in *Bacillus subtilis*. *J Bacteriol* **181**, 6889-6897.

Pull, S. L., Doherty, J. M., Mills, J. C., Gordon, J. I. & Stappenbeck, T. S. (2005). Activated macrophages are an adaptive element of the colonic epithelial progenitor niche necessary for regenerative responses to injury. *Proc Natl Acad Sci U S A* **102**, 99-104.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F. & other authors (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65.

Ravcheev, D. A., Khoroshkin, M. S., Laikova, O. N., Tsoy, O. V., Sernova, N. V., Petrova, S. A., Rakhmaninova, A. B., Novichkov, P. S., Gelfand, M. S. & other authors (2014). Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Front Microbiol* **5**: 294

Reizer, J., Hoischen, C., Titgemeyer, F., Rivolta, C., Rabus, R., Stülke, J., Karamata, D., Saier Jr, M. H. & Hillen, W. (1998). A novel protein kinase that controls carbon catabolite repression in bacteria. *Mol Microbiol* **27**, 1157-1169.

Renzi, F., Manfredi, P., Mally, M., Moes, S., Jenö, P. & Cornelis, G. R. (2011). The *N*-glycan glycoprotein deglycosylation complex (Gpd) from *Capnocytophaga canimorsus* deglycosylates human IgG. *PLoS Pathog* **7**, e1002118.

Reuter, G. (2001). The *Lactobacillus* and *Bifidobacterium* microflora of the human intestine: composition and succession. *Curr Iss Intest Microbiol* **2**, 43-53.

Robbe, C., Capon, C., Coddeville, B. & Michalski, J. (2004). Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem J* 307-316.

Roberton, A., McKenzie, C., Sharfe, N. & Stubbs, L. (1993). A glycosulphatase that removes sulphate from mucus glycoprotein. *Biochem J* **293**, 683-689.

Roberts, G., Tarelli, E., Homer, K. A., Philpott-Howard, J. & Beighton, D. (2000). Production of an endo-β-*N*-acetylglucosaminidase activity mediates growth

of *Enterococcus faecalis* on a high-mannose-type glycoprotein. *J Bacteriol* **182**, 882-890.

Roberts, G., Homer, K. A., Tarelli, E., Philpott-Howard, J., Devriese, L. A. & Beighton, D. (2001). Distribution of endo-β-*N*-acetylglucosaminidase amongst enterococci. *J Med Microbiol* **50**, 620-626.

Roeser, D., Preusser-Kunze, A., Schmidt, B., Gasow, K., Wittmann, J. G., Dierks, T., von Figura, K. & Rudolph, M. G. (2006). A general binding mechanism for all human sulfatases by the formylglycine-generating enzyme. *Proc Natl Acad Sci USA* **103**, 81-86.

Roger, L. C., Costabile, A., Holland, D. T., Hoyles, L. & McCartney, A. L. (2010). Examination of faecal *Bifidobacterium* populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology* **156**, 3329-3341.

Round, J. L. & Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313-323.

Ruiz-Moyano, S., Totten, S. M., Garrido, D. A., Smilowitz, J. T., German, J. B., Lebrilla, C. B. & Mills, D. A. (2013). Variation in consumption of human milk oligosaccharides by infant gut-associated strains of *Bifidobacterium breve*. *Appl Environ Microbiol* **79**, 6040-6049.

Ryan, S., Fitzgerald, G. & van Sinderen, D. (2005). Transcriptional regulation and characterization of a novel β-fructofuranosidase-encoding gene from *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **71**, 3475-3482.

Saier Jr, M. H. & Ramseier, T. M. (1996). The catabolite repressor/activator (Cra) protein of enteric bacteria. *J Bacteriol* **178**, 3411.

Saier Jr, M.H., Chauvaux, S., Cook, G. M., Deutscher, J., Paulsen, I. T., Reizer, J. & Ye, J. J. (1996). Catabolite repression and inducer control in Gram-positive bacteria. *Microbiology* **142**, 217-230.

Salyers, A. A., West, S. E., Vercellotti, J. R. & Wilkins, T. D. (1977a). Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Appl Environ Microbiol* **34**, 529-533.

Salyers, A. A., Vercellotti, J. R., West, S. E. & Wilkins, T. D. (1977b). Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. *Appl Environ Microbiol* **33**, 319-322.

Scardovi, V. & Trovatelli, I. D. (1965). The fructose-6-phosphate shunt as a peculiar pattern of hexose degradation in the genus *Bifidobacterium*. *Ann Microbiol*, 19-29.

Schell, M., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M.-C., Desiere, F., Bork, P. & other authors (2002). The genome

sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci USA* **99**, 14422-14427.

Schmidt, B., Selmer, T., Ingendoh, A. & Figurat, K. v. (1995). A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* **82**, 271-278.

Schneider, E. (2001). ABC transporters catalyzing carbohydrate uptake. *Res Microbiol* **152**, 303-310.

Sela, D. & Mills, D. (2010). Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol* **18**, 298-307.

Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., Lapidus, A., Rokhsar, D. S., Lebrilla, C. B. & other authors (2008). The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci USA* **105**, 18964-18969.

Sela, D. A., Li, Y., Lerno, L., Wu, S., Marcobal, A. M., German, J. B., Chen, X., Lebrilla, C. B. & Mills, D. A. (2011). An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides. *J Biol Chem* **286**, 11909-11918.

Sela, D. A., Garrido, D., Lerno, L., Wu, S., Tan, K., Eom, H.-J., Joachimiak, A., Lebrilla, C. B. & Mills, D. A. (2012). *Bifidobacterium longum* subsp. *infantis* ATCC 15697 α-fucosidases are active on fucosylated human milk oligosaccharides. *Appl Environ Microbiol* **78**, 795-803.

Sharma, R., Schumacher, U., Ronaasen, V. & Coates, M. (1995). Rat intestinal mucosal responses to a microbial flora and different diets. *Gut* **36**, 209-214.

Shimada, Y., Watanabe, Y., Wakinaka, T., Funeno, Y., Kubota, M., Chaiwangsri, T., Kurihara, S., Yamamoto, K., Katayama, T. & other authors (2014). α-*N*-acetylglucosaminidase from *Bifidobacterium bifidum* specifically hydrolyzes α-linked *N*-acetylglucosamine at nonreducing terminus of O-glycan on gastric mucin. *Appl Microbiol Biotechnol*, 1-8.

Silk, D. B. A., Davis, A., Vulevic, J., Tzortzis, G. & Gibson, G. R. (2009). Clinical trial: the effects of a trans-galactooligosaccharide prebiotic on faecal microbiota and symptoms in irritable bowel syndrome. *Aliment Pharmacol Ther* **29**, 508-518.

Silvestre, D., Fraga, M., Gormaz, M., Torres, E. & Vento, M. (2014). Comparison of mid-infrared transmission spectroscopy with biochemical methods for the determination of macronutrients in human milk. *Matern Child Nutr* **10**, 373-382.

Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C.-H., Westover, B. P., Weatherford, J., Buhler, J. D. & Gordon, J. I. (2005). Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307**, 1955-1959.

Stahl, B. & Barrangou, R. (2012). Complete genome sequences of probiotic strains *Bifidobacterium animalis* subsp. *lactis* B420 and Bi-07. *J Bacteriol* **194**, 4131-4132.

Stanton, C., Ross, R. P., Fitzgerald, G. F. & Sinderen, D. V. (2005). Fermented functional foods based on probiotics and their biogenic metabolites. *Curr Opin Biotechnol* **16**, 198-203.

Stephen, A. M., Haddad, A. C. & Phillips, S. F. (1983). Passage of carbohydrate into the colon. Direct measurements in humans. *Gastroenterology* **85**, 589-595.

Stülke, J. & Hillen, W. (1999). Carbon catabolite repression in bacteria. *Curr Opin Microbiol* **2**, 195-201.

Sun, Z., Chen, X., Wang, J., Gao, P., Zhou, Z., Ren, Y., Sun, T., Wang, L., Meng, H. & other authors (2010). Complete genome sequence of probiotic *Bifidobacterium animalis* subsp. *lactis* strain V9. *J Bacteriol* **192**, 4080-4081.

Swint-Kruse, L. & Matthews, K. S. (2009). Allostery in the LacI/GalR family: variations on a theme. *Curr Opin Microbiol* **12**, 129-137.

Szameit, C., Miech, C., Balleininger, M., Schmidt, B., von Figura, K. & Dierks, T. (1999). The iron sulfur protein AtsB is required for posttranslational formation of formylglycine in the *Klebsiella* sulfatase. *J Biol Chem* **274**, 15375-15381.

Szentkuti, L., Riedesel, H., Enss, M. L., Gaertner, K. & von Engelhardt, W. (1990). Pre-epithelial mucus layer in the colon of conventional and germ-free rats. *Histochem J* **22**, 491-497.

Tam, R. & Saier, M. H. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Rev* **57**, 320-346.

Thurl, S., Munzert, M., Henker, J., Boehm, G., Müller-Werner, B., Jelinek, J. & Stahl, B. (2010). Variation of human milk oligosaccharides in relation to milk groups and lactational periods. *Br J Nutr* **104**, 1261-1271.

Tissier, H. (1900). *Recherchers sur la flora intestinale normale pathologique du nourisson.*, The University of Paris, Paris.

Trindade, M. I., Abratt, V. R. & Reid, S. J. (2003). Induction of sucrose utilization genes from *Bifidobacterium lactis* by sucrose and raffinose. *Appl Environ Microbiol* **69**, 24-32.

Turnbaugh, P., Hamady, M., Yatsunenko, T., Cantarel, B., Duncan, A., Ley, R., Sogin, M., Jones, W., Roe, B. & other authors (2008). A core gut microbiome in obese and lean twins. *Nature* **457**, 480-484.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R. & Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* **1**, 6ra14-16ra14.

Turnbaugh, P. J., Quince, C., Faith, J. J., McHardy, A. C., Yatsunenko, T., Niazi, F., Affourtit, J., Egholm, M., Henrissat, B. & other authors (2010).

Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc Natl Acad Sci USA* **107**, 7503-7508.

Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J.-H., Zomer, A., Sanchez, B., Bidossi, A., Ferrarini, A. & other authors (2010). Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc Natl Acad Sci USA* **107**, 19514-19519.

Turroni, F., Strati, F., Foroni, E., Serafini, F., Duranti, S., van Sinderen, D. & Ventura, M. (2012a). Analysis of predicted carbohydrate transport systems encoded by *Bifidobacterium bifidum* PRL2010. *Appl Environ Microbiol* **78**, 5002-5012.

Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., Kerr, C., Hourihane, J., Murray, D. & other authors (2012b). Diversity of bifidobacteria within the infant gut microbiota. *PLoS ONE* **7**, e36957.

Turroni, F., Duranti, S., Bottacini, F., Guglielmetti, S., Van Sinderen, D. & Ventura, M. (2014). *Bifidobacterium bifidum* as an example of a specialized human gut commensal. *Front Microbiol* **5**.

Urashima, T., Asakuma, S., Leo, F., Fukuda, K., Messer, M. & Oftedal, O. T. (2012). The predominance of type I oligosaccharides is a feature specific to human breast milk. *Adv Nutr* **3**, 473S-482S.

Vaishampayan, P. A., Kuehl, J. V., Froula, J. L., Morgan, J. L., Ochman, H. & Francino, M. P. (2010). Comparative metagenomics and population dynamics of the gut microbiota in mother and infant. *Genome Biol Evol* **2**, 53-66.

van den Bogaard, P. T. C., Kleerebezem, M., Kuipers, O. P. & de Vos, W. M. (2000). Control of lactose transport, β-galactosidase activity, and glycolysis by CcpA in *Streptococcus thermophilus:* Evidence for carbon catabolite repression by a non-phosphoenolpyruvate-dependent phosphotransferase system sugar. *J Bacteriol* **182**, 5982-5989.

van den Broek, L. & Voragen, A. (2008). Bifidobacterium glycoside hydrolases and (potential) prebiotics. *Innov Food Sci Emerg Technol* **9**, 401-407.

van den Broek, L., Hinz, S., Beldman, G., Vincken, J.-P. & Voragen, A. (2008). *Bifidobacterium* carbohydrases-their role in breakdown and synthesis of (potential) prebiotics. *Mol Nutr Food Res* **52**, 146-163.

Varki, A., Cummings, R. D., Esko, J. D., Freeze, H., Stanley, P., Bertozzi, C. R., Hart, G. W. & Etzler, M. E. (2009). *Essentials of Glycobiology*, 2nd edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Velázquez, O., Lederer, H. & Rombeau, J. (1996). Butyrate and the colonocyte. *Digest Dis Sci* **41**, 727-739.

Ventura, M., Canchaya, C., Casale, A., Dellaglio, F., Neviani, E., Fitzgerald, G. & van Sinderen, D. (2006). Analysis of bifidobacterial evolution using a multilocus approach. *Int J Syst Evol Microbiol* **56**, 2783-2792.

Ventura, M., Canchaya, C., Fitzgerald, G., Gupta, R. & Sinderen, D. (2007). Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria. *Antonie van Leeuwenhoek* **92**, 265-265.

Ventura, M., Turroni, F., Zomer, A., Foroni, E., Giubellini, V., Bottacini, F., Canchaya, C., Claesson, M. J., He, F. & other authors (2009). The *Bifidobacterium dentium* Bd1 genome sequence reflects its genetic adaptation to the human oral cavity. *PLoS Genet* **5**, e1000785.

Ventura, M., Turroni, F., Lugli, G. A. & van Sinderen, D. (2014). Bifidobacteria and humans: our special friends, from ecological to genomics perspectives. *J Sci Food Agr* **94**, 163-168.

Vinolo, M. A., Rodrigues, H. G., Nachbar, R. T. & Curi, R. (2011). Regulation of inflammation by short chain fatty acids. *Nutrients* **3**, 858-876.

von Figura, K., Schmidt, B., Selmer, T. & Dierks, T. (1998). A novel protein modification generating an aldehyde group in sulfatases: its role in catalysis and disease. *BioEssays* **20**, 505-510.

Wada, J., Ando, T., Kiyohara, M., Ashida, H., Kitaoka, M., Yamaguchi, M., Kumagai, H., Katayama, T. & Yamamoto, K. (2008). *Bifidobacterium bifidum*

lacto-*N*-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. *Appl Environ Microbiol* **74**, 3996-4004.

Walter, J. & Ley, R. (2011). The human gut microbiome: Ecology and recent evolutionary changes. *Annu Rev Microbiol* **65**, 411-429.

Wang, M., Ahrné, S., Jeppsson, B. & Molin, G. (2005). Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. *FEMS Microbiol Ecol* **54**, 219-231.

Ward, R., Ninonuevo, M., Mills, D., Lebrilla, C. & German, B. (2007). In vitro fermentability of human milk oligosaccharides by several strains of bifidobacteria. *Mol Nutr Food Res* **51**, 1398-1405.

Wei, Y.-X., Zhang, Z.-Y., Liu, C., Zhu, Y.-Z., Zhu, Y.-Q., Zheng, H., Zhao, G.-P., Wang, S. & Guo, X.-K. (2010). Complete genome sequence of *Bifidobacterium longum* JDM301. *J Bacteriol* **192**, 4076-4077.

Weickert, M. J. & Adhya, S. (1992). A family of bacterial regulators homologous to Gal and Lac repressors. *J Biol Chem* **267**, 15869-15874.

Wen, L., Ley, R. E., Volchkov, P. Y., Stranges, P. B., Avanesyan, L., Stonebraker, A. C., Hu, C., Wong, F. S., Szot, G. L. & other authors (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455**, 1109-1113.

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A. & other authors (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105-108.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N. & other authors (2012). Human gut microbiome viewed across age and geography. *Nature* **486**, 222-227.

Yoshida, E., Sakurama, H., Kiyohara, M., Nakajima, M., Kitaoka, M., Ashida, H., Hirose, J., Katayama, T., Yamamoto, K. & other authors (2012). *Bifidobacterium longum* subsp. *infantis* uses two different β-galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides. *Glycobiology* **22**, 361-368.

Yu, D. S., Jeong, H., Lee, D.-H., Kwon, S.-K., Song, J. Y., Kim, B. K., Park, M.-S., Ji, G. E., Oh, T. K. & other authors (2012). Complete genome sequence of the probiotic bacterium *Bifidobacterium bifidum* strain BGN4. *J Bacteriol* **194**, 4757-4758.

Zhurina, D., Zomer, A., Gleinser, M., Brancaccio, V. F., Auchter, M., Waidmann, M. S., Westermann, C., van Sinderen, D. & Riedel, C. U. (2011). Complete genome sequence of *Bifidobacterium bifidum* S17. *J Bacteriol* **193**, 301-302.

Chapter II

Metabolism of sialic acid by *Bifidobacterium breve* UCC2003.

DNA microarray analysis was performed by Dr. Mary O' Connell Motherway.

This chapter was published in:

Egan, M., O'Connell Motherway, M., Ventura, M. & van Sinderen, D. (2014). Metabolism of sialic acid by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **80**, 4414-4426.

2.1 ABSTRACT

Bifidobacteria constitute a specific group of commensal bacteria, which inhabit the gastrointestinal tract of humans and other mammals. *Bifidobacterium breve* UCC2003 has previously been shown to utilise several plant-derived carbohydrates that include cellodextrins, starch and galactan. In the current study, we investigated the ability of this strain to utilise the mucin- and human milk oligosaccharide (HMO)-derived carbohydrate, sialic acid. Using a combination of transcriptomic and functional genomic approaches, we identified a gene cluster dedicated to the uptake and metabolism of sialic acid. Furthermore, we demonstrate that *B. breve* UCC2003 can cross-feed on sialic acid derived from the metabolism of 3' sialyllactose, an abundant HMO, by another infant gut bifidobacterial strain, *Bifidobacterium bifidum* PRL2010.

2.2 INTRODUCTION

Bifidobacteria are Gram positive, anaerobic, typically Y-shaped bacteria which are members of the *Bifidobacteriaceae* family and the *Actinobacteria* phylum. Discovered in 1900 (Tissier, 1900), bifidobacteria are naturally found in the digestive tract of mammals and insects, while they have also been isolated from the human oral cavity and sewage (Ventura *et al.*, 2007). The large intestine is the natural habitat for a large and diverse bacterial community, which is known to significantly contribute to the well-being of humans. This positive contribution includes metabolic activities to provide energy and nutrients to the host (LeBlanc *et al.*, 2013; Vaughan *et al.*, 2005), development of the immune system (Round & Mazmanian, 2009; Ventura *et al.*, 2012) and protection against pathogenic bacteria (Servin, 2004).

Bifidobacteria are saccharolytic organisms and their survival and growth in the large intestine requires a variety of extracellular and cytoplasmic glycosyl hydrolases, which they employ to metabolise carbohydrates prevalent in this environment. The characteristic central metabolic pathway in bifidobacteria is the fructose-6-phosphate phosphoketolase pathway, also known as the bifid-shunt (Scardovi & Trovatelli, 1965). Because of the diversity of carbohydrate sources present in the gastrointestinal tract and the dependence of bifidobacteria on such carbon and energy sources, it is not surprising that an estimated 8 % of the typical bifidobacterial genome is devoted to carbohydrate metabolism (Schell *et al.*, 2002; Ventura *et al.*, 2007).

Bifidobacterium breve UCC2003 has already been shown to utilise quite a variety of plant-derived poly-, oligo-, and mono-saccharides, including cellodextrins, galactan,

starch, raffinose and melezitose (O'Connell *et al.*, 2013; O'Connell Motherway *et al.*, 2008; O'Connell Motherway *et al.*, 2011a; Pokusaeva *et al.*, 2011; Ryan *et al.*, 2006), which reflects the apparent flexibility of the carbohydrate utilisation profile of this strain. Interestingly, *B. breve* strains are numerically prevalent among the gut microbiota of healthy breast-fed infants (Turroni *et al.*, 2012b), although the ability of members of this bifidobacterial species to utilise human milk- and host-derived carbohydrates is relatively unexplored.

Sialic acids comprise a large family of nine-carbon monosaccharides called neuraminic acids, the most common of which is *N*-acetylneuraminic acid (Neu5Ac), for which the name of sialic acid is also used. Approximately 16 % of human milk oligosaccharides (HMOs) are sialylated (Ninonuevo *et al.*, 2006), with 3' sialyllactose and 6' sialyllactose being abundant sialylated HMO components (Bao *et al.*, 2007; Kunz *et al.*, 2000; Martin-Sosa *et al.*, 2003). Total sialic acid concentrations are highest in colostrum, reaching levels of 5.04 mmol L^{-1}, though this level will then decrease by almost 80 % after 3 months (Wang *et al.*, 2001). Sialic acid also features prominently at the surface-exposed end of human colonic mucin (Podolsky, 1985) and in healthy adults approximately 300 µg of sialic acid is present per mg of colonic mucin (Corfield *et al.*, 1999).

Representative *Clostridium perfringens* strains were the first bacteria shown to be capable of utilising sialic acid as the sole carbon source (Nees & Schauer, 1974; Nees *et al.*, 1976). Nevertheless, sialic acid metabolism was first explored in 1983 in *Escherichia coli*, with the identification of a transporter, designated *nanT*, and an *N*-acetylneuraminate lyase, designated *nanA* (Vimr & Troy, 1985a; Vimr & Troy, 1985b), followed by the description of the complete, five-step metabolic pathway in 1999 (Fig. 2.1) (Plumbridge & Vimr, 1999). Other gut inhabitants, including the

pathogen *Vibrio cholerae* (Almagro-Moreno & Boyd, 2009) and the commensals *Bacteroides fragilis* (Brigham *et al.*, 2009) and *Lactobacillus sakei* (Anba-Mondoloni *et al.*, 2013) have since been shown to utilise sialic acid as a carbon source.

The ability of *Bifidobacterium* species to metabolise sialic acid has not been widely studied. Genes predicted to be involved in the breakdown of sialic acid, which correspond with those outlined in Figure 2.1, were identified on the genome of *Bifidobacterium longum* subsp. *infantis* ATCC15697 and in accordance this strain was shown to be capable of utilising sialic acid (Sela *et al.*, 2008) or the HMO sialylated lacto-*N*-tetraose as a sole carbon source (Sela *et al.*, 2011). A recent study on the consumption of HMOs by strains of *B. breve* revealed that all *B. breve* strains tested can utilise sialylated lacto-*N*-tetraose (Ruiz-Moyano *et al.*, 2013). Another study examined the utilisation of the monosaccharide constituents of HMO by five bifidobacterial strains, representing four bifidobacterial species (*B. longum, B. breve, Bifidobacterium bifidum* and *Bifidobacterium adolescentis*), and found that, of the strains examined, only *B. longum* subsp. *infantis* ATCC15697 and *B. breve* ATCC15700 are capable of utilising sialic acid (Ward *et al.*, 2007). This suggests that the ability to utilise sialic acid and/or sialylated-HMOs is a species/strain-specific attribute.

The aim of this study was to investigate how *B. breve* UCC2003 utilises sialic acid, a characteristic which is likely to support particular bifidobacterial species in their ability to colonise the infant gut. Furthermore, the manner by which *B. breve* UCC2003 can cross-feed on sialic acid, when derived from the metabolism of 3' sialyllactose by *B. bifidum* PRL2010, was characterised.

2.3 MATERIALS AND METHODS

Bacterial strains, plasmids, media and culture conditions. Bacterial strains and plasmids used in this study are listed in Table 2.1. *B. breve* UCC2003 was routinely cultured in Reinforced Clostridial Medium (RCM; Oxoid Ltd., Basingstoke, Hampshire, United Kingdom). *B. bifidum* PRL2010 was cultured in modified deMan Rogosa Sharpe (mMRS) medium made from first principles (De Man *et al.*, 1960), though excluding a carbohydrate source, and supplemented with 0.05 % (wt/vol) L-cysteine HCl and 1 % (wt/vol) lactose (unless otherwise specified). Carbohydrate utilisation by bifidobacteria was examined in mMRS medium supplemented with 0.05 % (wt/vol) L-cysteine HCl and a particular carbohydrate source (0.5 % wt/vol). It was previously shown that without the addition of a carbohydrate, mMRS does not support growth of *B. breve* (Watson *et al.*, 2013). The carbohydrates used were lactose (Sigma Aldrich, Steinheim, Germany), sialic acid and 3' and 6' sialyllactose (Carbosynth, Compton, Berkshire, United Kingdom). Addition of sialic acid at this concentration to the mMRS resulted in a decrease in pH to 6.0, hence the pH was re-adjusted to 6.8 following addition of the sugar and the medium was subsequently filter sterilised. Bifidobacterial cultures were incubated under anaerobic conditions in a modular atmosphere-controlled system (Davidson and Hardy, Belfast, Ireland) at 37°C. *E. coli* was cultured in Luria Bertani broth (LB) at 37°C with agitation (Sambrook, 1989). *Lactococcus lactis* strains were grown in M17 medium supplemented with 0.5 % (wt/vol) glucose at 30°C (Terzaghi & Sandine, 1975). Where appropriate, growth media contained tetracycline (Tet; 10 µg ml^{-1}), chloramphenicol (Cm; 5 µg ml^{-1} for *L. lactis* and *E. coli*, 2.5 µg ml^{-1} for *B. breve*), erythromycin (Em; 100 µg ml^{-1}) or kanamycin (Kan; 50 µg ml^{-1}). Recombinant *E. coli* cells containing pORI19 were selected on LB agar containing Em and Kan, and

supplemented with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) (40 μg ml^{-1}) and 1 mM IPTG (isopropyl-β-D-galactopyranoside).

In order to determine bacterial growth profiles and final optical densities, 10 ml of a freshly prepared mMRS medium, including a particular carbohydrate source (see above), was inoculated with 100 μl (1 %) of a stationary-phase culture of a particular strain. Un-inoculated mMRS was used as a negative control. Cultures were incubated anaerobically for 24-36 h. Optical density (OD_{600nm}) measurements were recorded unless otherwise stated.

Nucleotide sequence analysis. Sequence data were obtained from the Artemis-mediated genome annotations of *B. breve* UCC2003 (O'Connell Motherway *et al.*, 2011b; Rutherford *et al.*, 2000). Database searches were performed using the non-redundant sequence database accessible at the National Centre for Biotechnology Information website (http://www.ncbi.nlm.nih.gov) using BLAST (Altschul *et al.*, 1990). Sequence analysis was performed using the Seqbuilder and Seqman programs of the DNASTAR software package (DNASTAR, Madison, WI, USA).

DNA manipulations. Chromosomal DNA was isolated from *B. breve* UCC2003 as previously described (Riordan, 1998). Plasmid DNA was isolated from *E. coli*, *L. lactis* and *B. breve* using the Roche High Pure plasmid isolation kit (Roche Diagnostics, Basel, Switzerland). An initial lysis step was performed using 30 mg ml^{-1} of lysozyme for 30 min at 37°C prior to plasmid isolation from *L. lactis* or *B. breve*. Single stranded oligonucleotide primers used in this study were synthesised

by Eurofins (Ebersberg, Germany) (Supplementary table 2.1). Standard PCRs were performed using Taq master mix (Qiagen GmBH, Hilden, Germany). *B. breve* colony PCRs were carried out as described previously (O'Connell Motherway *et al.*, 2009). PCR fragments were purified using the Roche High Pure PCR purification kit (Roche Diagnostics). Electroporation of plasmid DNA into *E. coli*, *L. lactis* or *B. breve* was performed as previously described (Maze *et al.*, 2007; Sambrook, 1989; Wells *et al.*, 1993).

Analysis of global gene expression using B. breve DNA microarrays. Global gene expression was determined during log-phase growth of *B. breve* UCC2003 in mMRS supplemented with 0.5 % (wt/vol) sialic acid. The obtained transcriptome was compared to that obtained from log-phase *B. breve* UCC2003 cells when grown in mMRS supplemented with 0.5 % (wt/vol) ribose. DNA microarrays containing oligonucleotide primers representing each of the 1864 identified open reading frames on the genome of *B. breve* UCC2003 were designed and obtained from Agilent Technologies (Palo Alto, Ca., USA). Methods for cell disruption, RNA isolation, RNA quality control, complementary DNA (cDNA) synthesis and labelling were performed as described previously (Zomer *et al.*, 2009). Labelled cDNA was hybridised using the Agilent Gene Expression hybridisation kit (part number 5188-5242) as described in the Agilent Two-Colour Microarray-Based Gene Expression Analysis v4.0 manual (publication number G4140-90050). Following hybridisation, microarrays were washed in accordance with Agilent's standard procedures and scanned using an Agilent DNA microarray scanner (model G2565A). Generated scans were converted to data files with Agilent's Feature Extraction software (Version 9.5). DNA-microarray data were processed as previously described (García

de la Nava *et al.*, 2003; van Hijum *et al.*, 2003; van Hijum *et al.*, 2005). Differential expression tests were performed with the Cyber-T implementation of a variant of the *t*-test (Long *et al.*, 2001).

Construction of B. breve UCC2003 insertion mutants. An internal fragment of Bbr_0161, designated *nanK* (370 base pairs (bp), representing codon numbers 148 through to 271 of the 336 codons of this gene), Bbr_0164, designated *nanB* (504 bp, representing codon numbers 226 through to 394 of the 521 codons of this gene), Bbr_0165, designated *nanC* (355 bp, representing codon numbers 27 through to 145 of the 319 codons of this gene), Bbr_0168, designated *nanA* (301 bp, representing codon numbers 116 through to 216 of the 321 codons of this gene), and Bbr_1247, designated *nagA2* (402 bp, representing codon numbers 107 through to 241 of the 426 codons of this gene) were amplified by PCR using *B. breve* UCC2003 chromosomal DNA as a template and primer pairs NanKF and NanKR, NanBF and NanBR, NanCF and NanCR, NanAF and NanAR, or NagA2F and NagA2R, respectively. The insertion mutants were constructed as described previously (O'Connell Motherway *et al.*, 2009). Site-specific recombination of potential Tet-resistant mutant isolates was confirmed by colony PCR using primer combinations TetWF and TetWR to verify *tetW* gene integration, and primers NanKconfirm, NanBconfirm, NanCconfirm, NanAconfirm and NagA2confirm (positioned upstream of the selected internal fragments of *nanK*, *nanB*, *nanC*, *nanA* and *nagA2*, respectively) in combination with primer TetWF to confirm integration at the correct chromosomal location.

Complementation of B. breve UCC2003-nanA. A DNA fragment encompassing *nanA* was generated by PCR amplification from *B. breve* UCC2003 chromosomal DNA using PFUII DNA polymerase and the primer pair NanApNZ44F and NanApNZ44R. The resulting fragment was digested with PstI and XbaI, and ligated to the similarly digested pNZ44 (McGrath *et al.*, 2001). The ligation mixture was introduced into *L. lactis* NZ9000 by electrotransformation and transformants selected based on Cm resistance. The plasmid content of a number of Cm^r transformants was screened by restriction analysis. The integrity of the cloned insert of one of the recombinant plasmids, designated pNZ44-nanA, was confirmed by sequencing. The *nanA* coding sequence, together with the constitutive p44 lactococcal promoter, specified by pNZ44, was amplified by PCR from a representative pNZ44-nanA plasmid using PFUII DNA polymerase and the primer combination P44FORWARD and NanApNZ44R. The resulting DNA fragment was digested with BamHI and XbaI and ligated to the similarly digested pBC1.2 (Álvarez-Martín *et al.*, 2007). The ligation mixture was introduced into *E. coli* XL1-blue by electrotransformation and transformants were selected based on Tet and Cm resistance. Transformants were checked for plasmid content using colony PCR and restriction analysis of plasmid DNA. Positively identified clones were verified by sequencing and one of the thus identified plasmids was designated pBC1.2-nanA. Plasmid pBC1.2-nanA was introduced into *B. breve* UCC2003-nanA by electrotransformation and transformants were selected based on Tet and Cm resistance.

Plasmid Constructions. For the construction of the plasmids pNZ-nanK, pNZ-nanE, pNZ-nanA, pNZ-nagB1 and pNZ-nagA2, DNA fragments encompassing the

predicted *N*-acetylmannosamine kinase-encoding gene *nanK* (Bbr_0161), *N*-acetylmannosamine-6-phosphate epimerase-encoding gene *nanE* (Bbr_0162), *N*-acetylneuraminate lyase-encoding gene *nanA* (Bbr_0168), glucosamine-6-phosphate deaminase-encoding gene *nagB1* (Bbr_0169) and *N*-acetylglucosamine-6-phosphate deacetylase-encoding gene *nagA2* (Bbr_1247) were generated by PCR amplification from chromosomal DNA of *B. breve* UCC2003 using PFUII DNA polymerase and the primer combinations nanKFOR and nanKREV, nanEFOR and nanEREV, nanAFOR and nanAREV, nagB1FOR and nagB1REV, or nagA2FOR and nagA2REV, respectively. An in-frame His_{10}-encoding sequence was incorporated into the forward primers nanKFOR, nanEFOR and nagA2FOR and the reverse primers nanAREV and nagB1REV, to facilitate downstream protein purification. The generated amplicons were digested with EcoRV and XbaI, and ligated into the ScaI and XbaI digested, nisin-inducible translational fusion plasmid pNZ8150 (Mierau & Kleerebezem, 2005). The ligation mixtures were introduced into *L. lactis* NZ9000 by electrotransformation and transformants were selected based on Cm resistance. The plasmid content of a number of Cm^r transformants was screened by restriction analysis and the integrity of positively identified clones was verified by sequencing.

Protein overproduction and purification. In order to (over)express and purify proteins, 400 ml of M17 broth supplemented with 0.5 % (wt/vol) glucose was inoculated with a 2 % inoculum of a particular *L. lactis* NZ9000 strain, followed by incubation at 30°C until an OD_{600nm} of 0.5 was reached, at which point protein expression was induced by addition of cell-free supernatant of a nisin-producing

strain (de Ruyter *et al.*, 1996), followed by continued incubation for a further 2 h. Cells were harvested by centrifugation and protein purification achieved using the PrepEase histidine-tagged protein purification maxi kit (USB, Germany) according to the manufacturer's instructions. Elution fractions were analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). After electrophoresis the gels were fixed and stained with Coomassie Brilliant blue to identify fractions containing the purified protein. The molecular weights of the proteins were estimated by comparison with rainbow-prestained, low molecular weight protein markers (New England BioLabs, Herefordshire, United Kingdom). Protein concentrations were determined using the Bradford method (Bradford, 1976).

High Performance Thin Layer Chromatography (HPTLC). HPTLC analysis was used for the qualitative determination of the activity of each of the purified enzymes. Purified $NanA_{His}$, $NanK_{His}$, $NanE_{His}$, $NagA2_{His}$ or $NagB1_{His}$ (final concentration 20 µg ml^{-1}), or combinations thereof, were incubated with sialic acid, *N*-acetylmannosamine, *N*-acetylmannosamine-6-phosphate, *N*-acetylglucosamine-6-phosphate or glucosamine-6-phosphate, respectively. Enzymatic assays were performed at 37°C in a total volume of 1 ml in a reaction buffer containing 20 mM Tris-HCl pH8.0, 10 mM KCl and 5 mM $MgSO_4.7H_2O$ with a 5 mg ml^{-1} concentration of a particular carbohydrate substrate. In reactions including $NanK_{His}$, 100 µl of 100 mM ATP was added to the reaction mixture. The reactions were terminated after 24 h by incubation at 65°C for 10 min. HPTLC analysis was performed as previously described (O'Connell Motherway *et al.*, 2011a). An aliquot

(1 µl) of the reaction mixture was spotted onto a Silica Gel 60 plate (10 x 10 cm, Merck) with a Nanomat 4 (Camag, Switzerland). The chromatogram was developed with a butanol-isopropanol-water (3:12:4, vol/vol/vol) solvent system in a horizontal developing chamber. Ascending development was repeated twice at room temperature. The plate was allowed to dry in a fume hood and then sprayed evenly with 20 % (vol/vol) sulfuric acid in ethanol. The plate was dried and heated to 120°C for 10 min to visualise sugar-representing spots. Reaction products were identified by comparison with relevant carbohydrate standards.

Growth of B. bifidum PRL2010 and B. breve UCC2003 or B. breve UCC2003-nanA on 3' sialyllactose. An overnight culture of *B. bifidum* PRL2010 (1 %) was used to inoculate mMRS broth supplemented with 0.05 % (wt/vol) L-cysteine HCl and 0.5 % (wt/vol) 3' sialyllactose and cultivated for 24 h at 37°C under anaerobic conditions. The cells were removed by centrifugation at 9,000 x *g* for 5 min. Growth of *B. bifidum* PRL2010 resulted in a pH decrease to 5.1, therefore the pH of cell free supernatant (CFS) was re-adjusted to 6.8 and the CFS was filter sterilised. This medium was supplemented with 0.01 % (wt/vol) lactose and 0.05 % (wt/vol) L-cysteine-HCl, prior to inoculation with an overnight culture (0.01 %) of either *B. breve* UCC2003 or *B. breve* UCC2003-nanA. Growth was monitored over 72 h, with samples taken at 6 or 12 h intervals. All samples collected were serially diluted in sterile Ringers solution and plated on Reinforced Clostridial Agar (RCA). Viable counts were determined by counting colonies on agar plates using dilutions that yielded between 30 and 300 colony forming units (CFU).

High-Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). Carbohydrate analysis by HPAEC-PAD was performed on CFS samples taken at 6 or 12 h intervals during the 72 h growth experiment, using a Dionex ICS-3000 system (Dionex, Sunnyvale, CA). Oligosaccharide fractions (25 µl aliquots) were separated on a CarboPac PA1 analytical-exchange column (250 mm x 4 mm) with a CarboPac PA1 guard column (50 mm x 4 mm) and a pulsed electrochemical detector in the pulsed amperometric detection (PAD) mode. The elution was performed at a constant flow-rate of 1.0 ml min^{-1} at 30°C using the following eluents for the analysis (A) 200 mM NaOH, (B) 100 mM NaOH, 550 mM sodium acetate (NaAC), and (C) Milli-Q water. The following linear gradient of NaAC was used: 100 mM NaOH: 0-50 min, 0 mM; 50-51 min, 16 mM; 51-56 min, 100 mM; 56.1-61 min, 0 mM. Detection was achieved using a Dionex ED40 detector in PAD mode. The Chromeleon software-version 6.70 (Dionex Corporation) was used for the integration and evaluation of the chromatograms obtained. The chromatographic profiles corresponding to 3' sialyllactose, sialic acid and lactose were qualitatively examined to evaluate 3' sialyllactose hydrolysis, and subsequent metabolism of lactose and/or sialic acid.

Microarray data accession number. The microarray data obtained in this study have been deposited in NCBI's Gene Expression Omnibus database and are accessible through GEO series accession number GSE56291.

2.4 RESULTS

Growth of B. breve strains on sialic acid. In order to determine if *B. breve* strains are capable of sialic acid metabolism, growth in mMRS supplemented with either 0.5 % (wt/vol) lactose or sialic acid was assessed for fourteen *B. breve* strains by measuring the OD_{600nm} following 24 h of anaerobic growth at 37°C. All *B. breve* strains were shown to grow well on lactose (OD_{600nm} >1.2). Eleven out of the fourteen examined *B. breve* strains, including the prototypical strain *B. breve* UCC2003, were capable of moderate growth on sialic acid, reaching OD_{600nm} values between 0.6-0.9, while the three remaining strains were unable to metabolise sialic acid (Fig. 2.2).

Genome response of B. breve UCC2003 to growth on sialic acid. In order to investigate which genes are involved in sialic acid metabolism by *B. breve* UCC2003, global gene expression was determined by microarray analysis during growth of the strain in mMRS supplemented with sialic acid and compared with gene expression when grown in mMRS supplemented with ribose. The pentose sugar ribose was considered an appropriate carbohydrate for comparative transcriptome analysis as the genes involved in ribose metabolism are known (Pokusaeva *et al.*, 2010), and it has been used in several transcriptome studies in *B. breve* UCC2003 (O'Connell *et al.*, 2013; O'Connell *et al.*, 2014; O'Connell Motherway *et al.*, 2011a; O'Connell Motherway *et al.*, 2013). Analysis of the DNA microarray data revealed that the expression of a gene cluster, comprising of twelve open reading frames, corresponding to locus tags Bbr_0160 to Bbr_0172, was significantly up-regulated when *B. breve* UCC2003 was grown on sialic acid relative to when the strain was

grown on ribose (fold change >3.0, $P<0.001$). In addition, two adjacent genes, corresponding to Bbr_1247 (*nagA2*) and Bbr_1248 (*nagB3*), were also up-regulated (see below; Table 2.2).

Genetic organisation of the predicted nan/nag gene clusters. Our presumption, based on the microarray results, was that the proteins encoded by the Bbr_0160-0172 gene cluster, and those of Bbr_1247-1248, schematically outlined in Figure 2.3, are involved in the metabolism of sialic acid in *B. breve* UCC2003. The first gene of this up-regulated cluster, Bbr_0160, encodes a hypothetical protein of unknown function. Bbr_0161, designated *nanK* here, encodes a putative *N*-acetylmannosamine kinase and is predicted to be a member of the ROK protein family. Members of this family include transcriptional regulators and sugar kinases (Titgemeyer *et al.*, 1994). Predicted ROK kinases possess a conserved N-terminal ATP-binding motif, DxGxT, which can be identified in the *nanK* gene of *B. breve* UCC2003 and which is absent in ROK transcriptional regulators (Conejo *et al.*, 2010). The predicted function of NanK is to phosphorylate *N*-acetylmannosamine, producing *N*-acetylmannosamine-6-phosphate. Bbr_0162, designated here as *nanE*, is predicted to encode an *N*-acetylmannosamine-6-phosphate epimerase, the predicted function of which is the conversion of *N*-acetylmannosamine-6-phosphate to *N*-acetylglucosamine-6-phosphate. Bbr_0163 encodes a putative hydrolase protein, the function of which, if any, in relation to sialic acid metabolism is unknown. The deduced protein products of Bbr_0164-0167 are predicted to represent an ABC transport system, including a putative solute-binding protein, permease and two ATP-binding proteins, and were designated *nanB*, *nanC*, *nanD* and *nanF*, respectively. Bbr_0168, designated *nanA*, encodes a putative *N*-acetylneuraminate lyase. The predicted function of NanA is the

cleavage of sialic acid, producing *N*-acetylmannosamine and pyruvate. Bbr_0169, designated *nagB1,* encodes a putative glucosamine-6-phosphate deaminase, the predicted function of which is the conversion of glucosamine-6-phosphate to fructose-6-phosphate, with the concomitant release of ammonia. Bbr_0170 is a pseudogene which is predicted to encode a truncated, non-functional NagA protein. Bbr_0171 encodes a predicted sialidase, designated *nanH,* which due to the lack of a signal sequence at the N-terminal is presumed to be intracellular. Sialidases are responsible for the removal of sialic acid residues from glycans. The *nanH* gene is followed by Bbr_0172, a putative ATPase from the AAA+ superfamily of ATPases, a large and diverse family of enzymes which are found in all kingdoms of living organisms where they are essential in various cellular processes such as proteolysis and DNA replication (Ogura & Wilkinson, 2001). Its function, if any, regarding sialic acid metabolism in *B. breve* UCC2003 is obscure. The Bbr_0160-0172 gene cluster was renamed as the *nan/nag* cluster, as it includes a complete *nan* system, as defined in 2004 as "one that minimally includes orthologues of genes encoding NanA, NanE and NanK" (Vimr *et al.*, 2004), as well as a predicted transport system and the *nagB1*-encoded glucosamine-6-phosphate deaminase. In addition to the *nan/nag* cluster, the Bbr_1247/Bbr_1248 locus is also predicted to be involved in sialic acid metabolism. Bbr_1247 (designated *nagA2*), encodes a putative *N*-acetylglucosamine-6-phosphate deacetylase, the predicted function of which is the removal of acetate from *N*-acetylglucosamine-6-phosphate, producing glucosamine-6-phosphate. Bbr_1248 (designated *nagB3*), encodes an additional glucosamine-6-phosphate deaminase (see above). Therefore, within these two loci, whose encompassing genes are transcriptionally induced when *B. breve* UCC2003 is grown on sialic acid, we can identify elements of a complete sialic acid catabolic pathway

corresponding to that previously outlined in *E. coli* (Fig. 2.1), with the sole difference being a *nanBCDF*-encoded ABC-type transport system instead of a *nanT* encoded transporter in *E. coli* (Plumbridge & Vimr, 1999). Interestingly, the genome of *B. breve* UCC2003 contains two adjacent genes, corresponding to locus tags Bbr_0846 and Bbr_0847, designated here as *nagA1* and *nagB2*, which are predicted to encode additional *N*-acetylglucosamine-6-phosphate deacetylase and glucosamine-6-phosphate deaminase activities respectively, although these genes were not shown to be significantly up-regulated when *B. breve* UCC2003 was grown on sialic acid. The NagA1 protein shares a 74 % identity with NagA2, while the NagB1 protein shares an 84 % identity with NagB2 and an 89 % identity with NagB3. Based on the comparative genome analysis presented in Figure 2.3, the *nan/nag* cluster is highly conserved among the *B. breve* strains whose genomes were recently published (Bottacini *et al.*, 2014). The same genes can also be identified in *B. longum* subsp. *infantis* ATCC15697, another *Bifidobacterium* which can utilise sialic acid, although the order of the genes of this particular *nan/nag* cluster has been slightly rearranged relative to that of *B. breve* UCC2003 (Sela *et al.*, 2008) (Fig. 2.3). There are no other homologs of this cluster within the completed bifidobacterial genomes currently available.

Disruption of the nanK, nanB, nanC, nanA and nagA2 genes in B. breve UCC2003. In order to investigate if disruption of individual genes from the *nan* cluster would affect the ability of *B. breve* UCC2003 to utilise sialic acid, insertion mutants were constructed in the *nanK, nanB, nanC, nanA* and *nagA2* genes, resulting in strains *B. breve* UCC2003-nanK, *B. breve* UCC2003-nanB, *B. breve* UCC2003-nanC, *B. breve* UCC2003-nanA and *B. breve* UCC2003-nagA2, respectively. These

strains were analysed for their ability to grow in mMRS supplemented with sialic acid as compared to *B. breve* UCC2003. As expected, and in contrast to the wild type, there was a complete lack of growth of *B. breve* UCC2003-nanA in media containing sialic acid (Fig. 2.4A). In order to demonstrate that the protein product of *nanA* is uniquely required for sialic acid metabolism in *B. breve* UCC2003, a complementation experiment was performed. The *nanA* gene was cloned into pBC1.2 under the control of the constitutive p44 promoter (Alvarez-Martin *et al.*, 2012) and introduced into *B. breve* UCC2003-nanA. Expression of NanA in *B. breve* UCC2003-nanA-pBC1.2-nanA allowed this strain to regain its ability to grow on sialic acid to an OD_{600nm} level comparable to that reached by the wild type (Fig. 2.4B). The insertion mutants, *B. breve* UCC2003-nanK, *B. breve* UCC2003-nanB and *B. breve* UCC2003-nanC also failed to grow in media containing sialic acid. However, the insertion mutant *B. breve* UCC2003-nagA2 reached OD_{600nm} levels comparable to the wild type strain during growth on sialic acid. Growth of the insertion mutants was not impaired on lactose, where all strains reached final OD_{600nm} levels comparable to that reached by the wild type strain (Fig. 2.4A).

Purification and biochemical characterisation of NanA, NanK, NanE, NagA2 and NagB1. In order to investigate the predicted enzymatic activities of NanA, NanK, NanE, NagA2 and NagB1, the genes encoding the respective enzymes were cloned into *L. lactis* NZ9000 and subsequently overproduced and purified. The thus produced $NanA_{His}$, $NanK_{His}$, $NanE_{His}$, $NagA2_{His}$ and $NagB1_{His}$ proteins exhibited their expected molecular masses of 34 kDa, 33 kDa, 23 kDa, 45 kDa and 29 kDa, respectively (results not shown). Biochemical characterisation was performed by incubating an individual enzyme or combination of enzymes with a particular

substrate and resulting products were analysed by HPTLC. The results are consistent with the predicted sialic acid breakdown pathway as outlined above (Fig. 2.1) (Plumbridge & Vimr, 1999). Under the HPTLC conditions used, it was not possible to visualise sialic acid, however, as expected, $NanA_{His}$, when incubated with its assumed substrate, was shown to produce a product with similar mobility properties as *N*-acetylmannosamine. The expected by-product of the reaction, pyruvate, was not visible under these conditions. $NanK_{His}$ was, only in the presence of ATP, shown to use *N*-acetylmannosamine as a substrate, generating a product with HPTLC mobility properties consistent with its expected product *N*-acetylmannosamine-6-phosphate. Furthermore, when both $NanA_{His}$ and $NanK_{His}$ were incubated with sialic acid and ATP, this also allowed the production of presumed *N*-acetylmannosamine-6-phosphate. Since it was not possible to visualise *N*-acetylglucosamine-6-phosphate or glucosamine-6-phosphate under the HPTLC conditions used, $NanE_{His}$, $NagB1_{His}$ and $NagA2_{His}$ were incubated together with *N*-acetylmannosamine-6-phosphate and, as expected, the combined activities of these three enzymes generated a product exhibiting HPTLC mobility properties consistent with fructose-6-phosphate. If any one or two of these enzymes was not included, this product was not generated, thus demonstrating the contribution of each of these three enzymes to this biochemical conversion (results not shown). Finally, when the five enzymes, $NanA_{His}$, $NanK_{His}$, $NanE_{His}$, $NagA2_{His}$ and $NagB1_{His}$, were incubated with only sialic acid as the substrate (plus ATP), a product was generated having HPTLC mobility properties consistent with the expected end product fructose-6-phosphate (Fig. 2.5).

Growth of B. bifidum PRL2010 and B. breve UCC2003 or B. breve UCC2003-nanA on 3' sialyllactose. Growth of *B. bifidum* PRL2010 and *B. breve* UCC2003 in

mMRS supplemented with 0.5 % (wt/vol) 3' sialyllactose, a trisaccharide consisting of sialic acid and lactose bound by an α2-3 linkage, was first assessed by measuring the OD_{600nm} for 36 h under anaerobic conditions. 3' sialyllactose was found to support growth of *B. bifidum* PRL2010, reaching OD_{600nm} values of approximately 0.8 (Fig. 2.6A). Analysis of the CFS after 24 h growth by HPAEC-PAD and comparison with non-fermented standards of 3' sialyllactose, sialic acid and lactose, showed that 3' sialyllactose was consumed from the medium, yet that consumption of this trisaccharide coincided with the accumulation of a carbohydrate with HPAEC-PAD properties consistent with sialic acid in the medium. It has previously been shown that *B. bifidum* PRL2010 encodes two putative extracellular exo-sialidases on its genome, each presumed to be specific for α2-3 or α2-6 linkages (Turroni *et al.*, 2010). From this information, we assumed that the extracellular sialidase of *B. bifidum* PRL2010 cleaves 3' sialyllactose to produce sialic acid and lactose, of which it can utilise lactose, but not sialic acid (i.e. *B. bifidum* PRL2010 does not exhibit growth in a medium containing sialic acid as a sole carbon source (Turroni *et al.*, 2012a)), resulting in an accumulation of sialic acid in the media. Consistent with this observation, the genome of *B. bifidum* PRL2010 lacks a sialic acid utilisation cluster (Turroni *et al.*, 2010). In contrast, it was found that 3' sialyllactose, and its α2-6 linked counterpart, 6' sialyllactose, were unable to support growth of *B. breve* UCC2003 (results not shown). In order to establish if *B. breve* UCC2003 could grow in the 3' sialyllactose-supplemented mMRS medium in which *B. bifidum* PRL2010 was previously grown for 24 h (spent medium), a small amount of lactose, 0.01 % (wt/vol), was added to the spent medium prior to inoculation to initiate growth . A low inoculum (0.01 %) of *B. breve* UCC2003 was used to allow the strain to undergo multiple growth generations. Growth on lactose was required

for the initial increase in cell numbers during the first 12 h, however, HPAEC-PAD analysis confirmed that lactose was fully utilised by *B. breve* UCC2003 after 12 h of incubation, after which growth and viability was dependent on the ability to utilise the sialic acid in the spent medium. Between 12 and 30 h, *B. breve* UCC2003 was able to reach a viable count of almost 10^9 CFU ml^{-1} from a starting point of 10^8 CFU ml^{-1}, and was able to maintain these viable cell numbers until 72 h (Fig. 2.6B). HPAEC-PAD analysis of the CFS after 30 h of growth shows the absence of the sialic acid-associated peak, implying that the sialic acid released as a result of the exo-sialidase activity of *B. bifidum* PRL2010 was entirely utilised to support growth of *B. breve* UCC2003 (Fig. 2.6C). Consistent with this scenario, *B. breve* UCC2003-nanA (which is incapable of growth in sialic acid, see above) was shown to be incapable of growth in this spent medium with viable counts dropping from their initial inoculation number of 10^5 CFU ml^{-1} after 72 h (Fig. 2.6B). As expected, HPAEC-PAD analysis demonstrated that sialic acid is not metabolised by *B. breve* UCC2003-nanA (Fig. 2.6C).

2.5 DISCUSSION

B. breve strains represent a dominant commensal group in the breast-fed infant gut microbiota, but it is only recently that the factors that contribute to this dominance have become a subject of scientific scrutiny. It has already been shown that *B. breve* strains can liberate *N*-glycans from host glycoproteins (Garrido *et al.*, 2012) and a more recent study has demonstrated that *B. breve* strains can grow on the HMOs lacto-*N*-tetraose, lacto-*N*-neotetraose and sialylated lacto-*N*-tetraose (Ruiz-Moyano *et al.*, 2013). In the current study, we describe the functional characterisation of a locus dedicated to the uptake and utilisation of sialic acid. To our knowledge, this is the first study to adopt a functional genomic approach to gain an understanding of sialic acid metabolism by a *Bifidobacterium* species.

A range of *B. breve* strains were tested for growth on sialic acid as a sole carbon source and of the fourteen strains tested, eleven were shown to be capable of growth. DNA microarray analysis of *B. breve* UCC2003 grown on sialic acid revealed a locus of twelve genes predicted to be involved in sialic acid uptake and catabolism, as well as two genes specifying a putative *N*-acetylglucosamine-6-phosphate deacetylase and glucosamine-6-phosphate deaminase, designated *nagA2* and *nagB3* respectively, located at an unlinked position on the genome. The genome of *B. breve* UCC2003 also contains additional *nagA* and *nagB* genes at the Bbr_0846-0847 locus, which were designated *nagA1* and *nagB2*, respectively. Since the enzymatic activities encoded by *nagAB* genes are also required for the metabolism of amino sugars such as *N*-acetylglucosamine and *N*-acetylgalactosamine (Reizer *et al.*, 1996; White, 1968), the presence of multiple *nagAB* copies on the *B. breve* UCC2003 genome may reflect its metabolic versatility to utilise such amino sugars when they are released from host-derived carbohydrates such as mucin and HMO (Kunz *et al.*,

2000; Podolsky, 1985). Previously performed comparative genomic hybridisation (CGH) analysis revealed that all tested *B. breve* strains harbour the *nan*/*nag* genes on their genome, thus indicating that the ability to utilise sialic acid is well conserved in the *B. breve* species (O'Connell Motherway *et al.*, 2011b). An apparent inconsistency is the presence of these genes on the genomes of *B. breve* NIZO 658, *B. breve* LMG13208 and *B. breve* UCC2007, yet exhibiting an inability to grow on sialic acid; however, it must be considered that the nature of CGH analysis means that only the presence or absence of genes is revealed, but not any mutations in a gene or promoter which may result in loss of function.

The uptake of sialic acid is likely to be facilitated by an ABC transport system, encoded by the *nanBCDF* genes. Disruption of *nanB* or *nanC* in *B. breve* UCC2003 was shown to result in an impairment of growth on sialic acid, thus proving that this predicted transport system is solely responsible for sialic acid uptake. Previous studies have shown that transport of sialic acid into bacterial cells can be achieved by ABC transport systems, as seen in *Haemophilus ducreyi* (Post *et al.*, 2005), transporters of the major facilitator superfamily of proteins, as in *E. coli* (Martinez *et al.*, 1995; Vimr & Troy, 1985b) and *Ba. fragilis* (Brigham *et al.*, 2009) or a tripartite ATP-independent periplasmic (TRAP) transporter, as seen in *Haemophilus influenzae* (Allen *et al.*, 2005) and *V. cholerae* (Chowdhury *et al.*, 2012). ABC transport systems have previously been shown to be involved in carbohydrate uptake in *B. breve* UCC2003 (O'Connell *et al.*, 2013; O'Connell Motherway *et al.*, 2011a; Pokusaeva *et al.*, 2010; Pokusaeva *et al.*, 2011).

Mutations in *nanA* and *nanK* were also shown to cause impaired growth on sialic acid, implying that the enzymes encoded by these genes are essential for sialic acid metabolism. In contrast, the insertion mutant, *B. breve* UCC2003-nagA2, did not

display impaired growth on sialic acid, suggesting that either this gene is not involved in sialic acid metabolism or that another gene is able to compensate for the mutation, a possible candidate for which would be the *nagA1* gene mentioned above.

Analysis of the substrate specificities of the enzymes NanA, NanK, NanE, NagA2 and NagB1 indicates that sialic acid metabolism by *B. breve* UCC2003 occurs by means of a five-step pathway, similar to that previously characterised in *E. coli* (Fig. 2.1) (Plumbridge & Vimr, 1999). An analogous pathway has been described in *H. influenzae* (Vimr *et al.*, 2000), *Staphylococcus aureus* (Olson *et al.*, 2013) and *Lb. sakei* (Anba-Mondoloni *et al.*, 2013), yet it differs slightly from that found in *Ba. fragilis*. It was shown that *Ba. fragilis* does not encode a *nanK* gene, but instead encodes a novel *N*-acetylmannosamine epimerase, designated *nanE*, which converts *N*-acetylmannosamine to *N*-acetylglucosamine, which is then phosphorylated by a *rokA*-encoded kinase (Brigham *et al.*, 2009). While individual enzymes of this pathway have been purified and characterised previously, such as NanA from *C. perfringens* (Krüger *et al.*, 2001), *H. influenzae* (Lilley *et al.*, 1998) and *Lactobacillus plantarum* (Sánchez-Carrón *et al.*, 2011) and the previously mentioned NanE from *Ba. fragilis* (Brigham *et al.*, 2009), to our knowledge this is the first time that the full catabolic pathway of sialic acid has been reconstituted *in vitro*.

In order for *B. breve* UCC2003 to utilise sialic acid, it must first be released from glyconjugates such as mucin or HMO, in which it is most commonly found. Sialidases, which release sialic acid from glycans, have previously been described for infant-derived bifidobacteria, including two intracellular sialidases from *B. longum* subsp. *infantis* ATCC15697 (Sela *et al.*, 2011) and an extracellular sialidase from *B. bifidum* JCM1254 (Kiyohara *et al.*, 2011). Two predicted extracellular exo-α-sialidases were also identified on the genome of *B. bifidum* PRL2010 (Turroni *et*

al., 2010). A putative intracellular sialidase is located in the *nan*/*nag* locus of *B. breve* UCC2003, but as yet its substrate is unknown. Its cognate substrate may be sialylated lacto-*N*-tetraose, which has previously been shown to support growth of various *B. breve* strains (Ruiz-Moyano *et al.*, 2013).

Interestingly, growth experiments of *B. breve* UCC2003 in spent medium in which *B. bifidum* PRL2010 had previously grown on 3' sialyllactose (as the sole carbon source) showed that *B. breve* UCC2003 can cross feed on the sialic acid released by the exo-sialidase activity of *B. bifidum* PRL2010. These results suggest that *B. bifidum* PRL2010 uses its sialidase activity to gain access to the lactose component of this HMO and leaving the sialic acid for *B. breve* UCC2003 (and others) to forage. These results show that a lack of sialidase activity would not necessarily disadvantage the proliferation of a cross-feeding bifidobacterial strain such as *B. breve* UCC2003 in the infant gut. Similarly, a recent study showed that two pathogens, *Salmonella enterica* serovar *Typhimurium* (*S. typhimurium*) and *Clostridium difficile* can scavenge sialic acid released by the sialidase activity of *Bacteroides thetaiotaomicron* in a gnotobiotic mouse (Ng *et al.*, 2013). Within the same study, it was shown that antibiotic-treated mice had elevated levels of free sialic acid in the caecum, as compared to their untreated counterparts. The authors proposed that in the untreated gut, sialic acid is utilised by members of the gut microbiota. Antibiotic treatment causes acute disturbance in this complex microbiota, yet enough sialidase-producing bacteria remain, resulting in elevated levels of sialic acid, which can be exploited by a sialic acid-utilising pathogen, such as *C. difficile* (Ng *et al.*, 2013).

Our results, combined with previous knowledge on the utilisation of sialic acid by pathogens, e.g. *C. difficile,* suggest that the metabolism of sialic acid by certain

bifidobacterial strains such as *B. breve* UCC2003 may provide competition for opportunistic pathogens in a healthy gut environment, thus inhibiting or moderating their proliferation. It also suggests a role for sialic acid-utilising *B. breve* strains as probiotic prophylaxis, particularly during antibiotic treatment, or in probiotic treatment of *C. difficile* infection, as recently suggested (Ley, 2014). Recently, *B. breve* strains were identified as suitable probiotics for the treatment of enteric disorders in infants (Aloisio *et al.*, 2012). Our results suggest that a combination of (extracellular) sialidase-encoding *B. bifidum* strains and sialic acid-utilising *B. breve* strains can improve colonisation and persistence of *B. breve* strains in the infant gut, thus potentially conferring the aforementioned benefits to its infant host.

Previous research has shown that *B. breve* UCC2003, as well as other *B. breve* strains, are capable of utilising a number of constituents of an adult, plant-based diet (O'Connell *et al.*, 2013; O'Connell Motherway *et al.*, 2008; O'Connell Motherway *et al.*, 2011a; O'Connell Motherway *et al.*, 2013; Pokusaeva *et al.*, 2011; Ryan *et al.*, 2006), yet this data did not explain the prevalence of this species in the infant gut. The current study, which demonstrates how *B. breve* UCC2003 can utilise sialic acid, advances our understanding on the prevalence of *B. breve* in the infant gut, where cross-feeding may play an important role in the establishment of a microbiota with a high abundance and diversity of bifidobacterial species.

2.6 ACKNOWLEDGEMENTS.

The Alimentary Pharmabiotic Centre is a research centre funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan. The authors and their work were supported by SFI (Grant Nos. 07/CE/B1368 and SFI/12/RC/2273) and a HRB postdoctoral fellowship (Grant No. PDTM/20011/9) awarded to MOCM.

2.7 TABLES AND FIGURES

Table 2.1: Bacterial strains and plasmids used in this study.

Strains and plasmids	Relevant features	Reference or source
Strains		
***Escherichia coli* strains**		
***E.coli* XL1-blue**	(*supE44 hsdR17 recA1 gyrA96 thi relA1 lac* F′ [*proAB*$^+$ *lacl*^{q} *lacZΔM15* Tn*10*(Tet^r)])	Stratagene
***E.coli* XL1-blue-pBC1.2-nanA**	XL1-blue containing pBC1.2-nanA	This study
***E.coli* EC101**	Cloning host; *repA*$^+$ *kmr*	(Law *et al.*, 1995)
***E.coli* EC101-pNZ-M.Bbrll+Bbr11**	EC101 harboring a pNZ8048 derivative containing bbrllM and bbrlllM	(O'Connell Motherway *et al.*, 2009)
***Lactococcus lactis* strains**		
***L. lactis* NZ9000**	MG1363, *pep*N::*nis*RK, nisin inducible overexpression host	(de Ruyter *et al.*, 1996)
***L. lactis* NZ9000-pNZ-nanA**	NZ9000 containing pNZ-nanA	This study
***L. lactis* NZ9000-pNZ-nanK**	NZ9000 containing pNZ-nanK	This study
***L. lactis* NZ9000-pNZ-nanE**	NZ9000 containing pNZ-nanE	This study
***L. lactis* NZ9000-pNZ-nagA2**	NZ9000 containing pNZ-nagA2	This study
***L. lactis* NZ9000-pNZ-nagB1**	NZ9000 containing pNZ-nagB1	This study
***L. lactis* NZ9000-pNZ44-nanA**	NZ9000 containing pNZ44-nanA	This study
***L. lactis* NZ97000**	Nisin-A producing strain	(de Ruyter *et al.*, 1996)
***Bifidobacterium* sp. strains**		
***B. breve* UCC2003**	Isolate from a nursling stool	(Maze *et al.*, 2007)
***B. breve* UCC2003-nanA**	pORI19-tetW-nanA insertion mutant of UCC2003	This study
***B. breve* UCC2003-nanK**	pORI19-tetW-nanK insertion mutant of UCC2003	This study
***B. breve* UCC2003-nanB**	pORI19-tetW-nanB insertion mutant of UCC2003	This study
***B. breve* UCC2003-nanC**	pORI19-tetW-nanC insertion mutant of UCC2003	This study
***B. breve* UCC2003-nagA2**	pORI19-tetW-nagA2 insertion mutant of UCC2003	This study
***B. breve* UCC2003-nanA-pBC1.2-nanA**	pORI19-tetW-nanA insertion mutant of UCC2003 harbouring complementation construct pBC1.2-nanA	This study
***B. breve* JCM7017**	Isolate from human faeces	JCM
***B. breve* JCM7019**	Isolate from adult faeces	JCM
***B. breve* NCFB2257**	Isolate from infant intestine	NCFB
***B. breve* NCFB2258**	Isolate from infant intestine	NCFB
***B. breve* NCIMB8815**	Isolate from human faeces	NCIMB
***B. breve* NIZO658**	Isolate from human faeces	NIZO

***B. breve* LMG13208**	Isolate from infant intestine	LMG
***B. breve* UCC2007**	Isolate from nursling stool	UCC
***B. breve* UCC2005**	Isolate from infant intestine	UCC
***B. breve* NCTC11815**	Isolate from infant intestine	NCTC
***B. breve* 461B**	Isolate from infant/adult faeces	PRL
***B. breve* 689B**	Isolate from infant faeces	(Bottacini *et al.*, 2014)
***B. breve* 12L**	Mother's milk	(Bottacini *et al.*, 2014)
***B. bifidum* PRL2010**	Isolate from infant faeces	(Turroni *et al.*, 2010)
Plasmids		
pAM5	pBC1-puC19-Tet^r	(Álvarez-Martín *et al.*, 2007)
pNZ44	pNZ8048 containing constitutive p44 promoter from Lactococcal chromosome	(McGrath *et al.*, 2001)
pNZ44-nanA	pNZ44 harbouring *nanA* downstream of p44 promoter	This study
pBC1.2	pBC1-pSC101-Cm^r	(Álvarez-Martín *et al.*, 2007)
pBC1.2-nanA	pBC1-pSC101-Cm^r harbouring *nanA* downstream of p44 promoter	This study
pORI19	Em^r, $repA^-$, ori^+, cloning vector	(Law *et al.*, 1995)
pORI19-tetW-nanA	Internal 301 bp fragment of *nanA* and *tetW* cloned in pORI19	This study
pORI19-tetW-nank	Internal 370 bp fragment of *nanK* and *tetW* cloned in pORI19	This study
pORI19-tetW-nanB	Internal 504 bp fragment of *nanB* and *tetW* cloned in pORI19	This study
pORI19-tetW-nanC	Internal 355 bp fragment of *nanC* and *tetW* cloned in pORI19	This study
pORI19-tetW-nagA2	Internal 402 bp fragment of *nagA2* and *tetW* cloned in pORI19	This study
pNZ8150	Cm^r, nisin-inducible translational fusion vector	(Mierau & Kleerebezem, 2005)
pNZ-nanA	*nanA* cloned downstream of nisin-inducible promoter on pNZ8150	This study
pNZ-nanK	*nanK* cloned downstream of nisin-inducible promoter on pNZ8150	This study
pNZ-nanE	*nanE* cloned downstream of nisin-inducible promoter on pNZ8150	This study
pNZ-nagA2	*nagA2* cloned downstream of nisin-inducible promoter on pNZ8150	This study
pNZ-nagB1	*nagB1* cloned downstream of nisin-inducible promoter on pNZ8150	This study

JCM: Japanese Collection of Microorganisms; NCFB: National Collection of Food Bacteria; NCIMB: National Collection of Industrial and Marine Bacteria; NIZO: Netherlands Institute for Dairy Research; LMG: Belgian Co-ordinated Collection of Microorganisms; UCC: University College Cork, Ireland; NCTC: National Collection of Type Cultures; PRL: Culture collection of probiogenomics, University of Parma.

Table 2.2: Effect of sialic acid on the transcriptome of *B. breve* UCC2003

Locus tag (Gene name)	Predicted Function	Level of upregulation
Bbr_0160	Conserved hypothetical protein	37.32
Bbr_0161 (*nanK*)	ROK family kinase	59.01
Bbr_0162 (*nanE*)	*N*-acetylmannosamine-6-P 2-epimerase	49.91
Bbr_0163	Hydrolase	51.55
Bbr_0164 (*nanB*)	Substrate binding protein	367.68
Bbr_0165 (*nanC*)	ABC transport system permease protein	451.08
Bbr_0166 (*nanD*)	ABC transport system ATP-binding protein	453.7
Bbr_0167 (*nanF*)	ABC transport system ATP-binding protein	435.43
Bbr_0168 (*nanA*)	*N*-acetylneuraminate lyase	435.04
Bbr_0169 (*nagB1*)	Glucosamine-6-P deaminase	275.38
Bbr_0171 (*nanH*)	Sialidase A	204.69
Bbr_0172	ATPase	6.7
Bbr_1247 (*nagA2*)	*N*-acetylglucosamine-6-P deacetylase	3.07
Bbr_1248 (*nagB3*)	Glucosamine-6-P deaminase	4.11

The cut-off point is 3-fold with a *P* value of <0.001

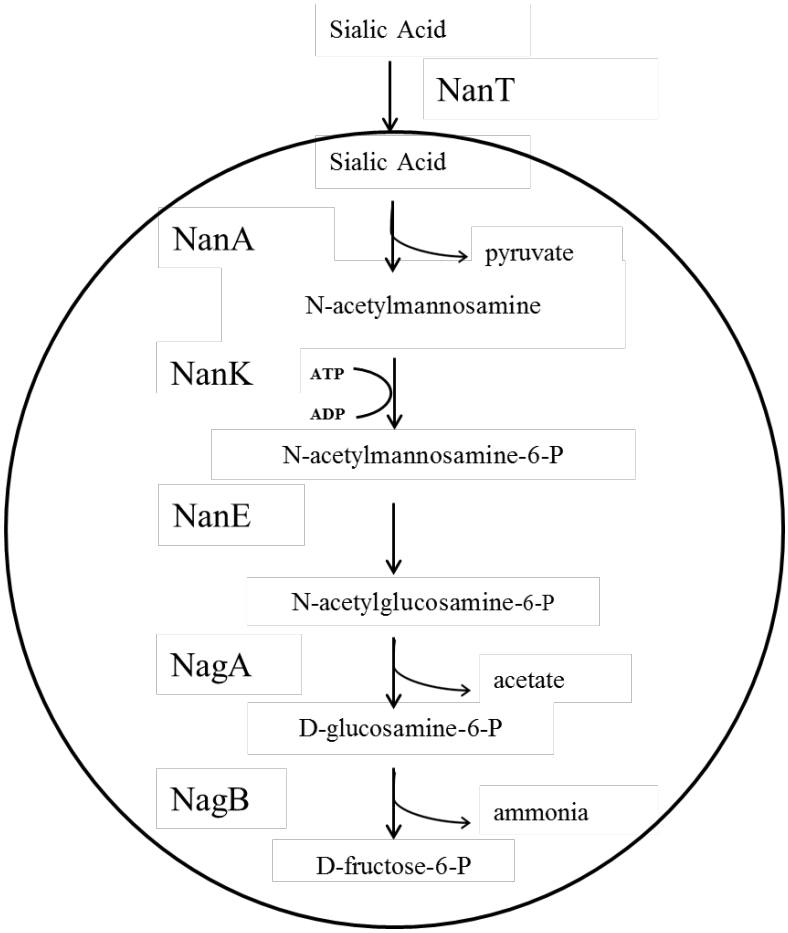


Figure 2.1: Metabolism of sialic acid by *E. coli,* as previously described (Plumbridge & Vimr, 1999). Sialic acid enters the cell through a transporter, encoded by the *nanT* gene. Intracellular sialic acid is cleaved by the *nanA*- encoded *N*-acetylneuraminate lyase, producing pyruvate and *N*-acetylmannosamine. *N*-acetylmannosamine is phosphorylated by the *nanK*-encoded *N*-acetylmannosamine kinase forming *N*-acetylmannosamine-6-phosphate, which is in turn converted to *N*-acetylglucosamine-6-phosphate by the *nanE*-encoded *N*-acetylmannosamine-6-phosphate epimerase. *N*-acetylglucosamine-6-phosphate then enters the amino sugar degradation pathway, where this compound is first deacetylated by a *nagA*-encoded *N*-acetylglucosamine-6-phosphate deacetylase, and subsequently converted to fructose-6-phosphate with the concomitant release of ammonia by the *nagB*-encoded glucosamine-6-phosphate deaminase.

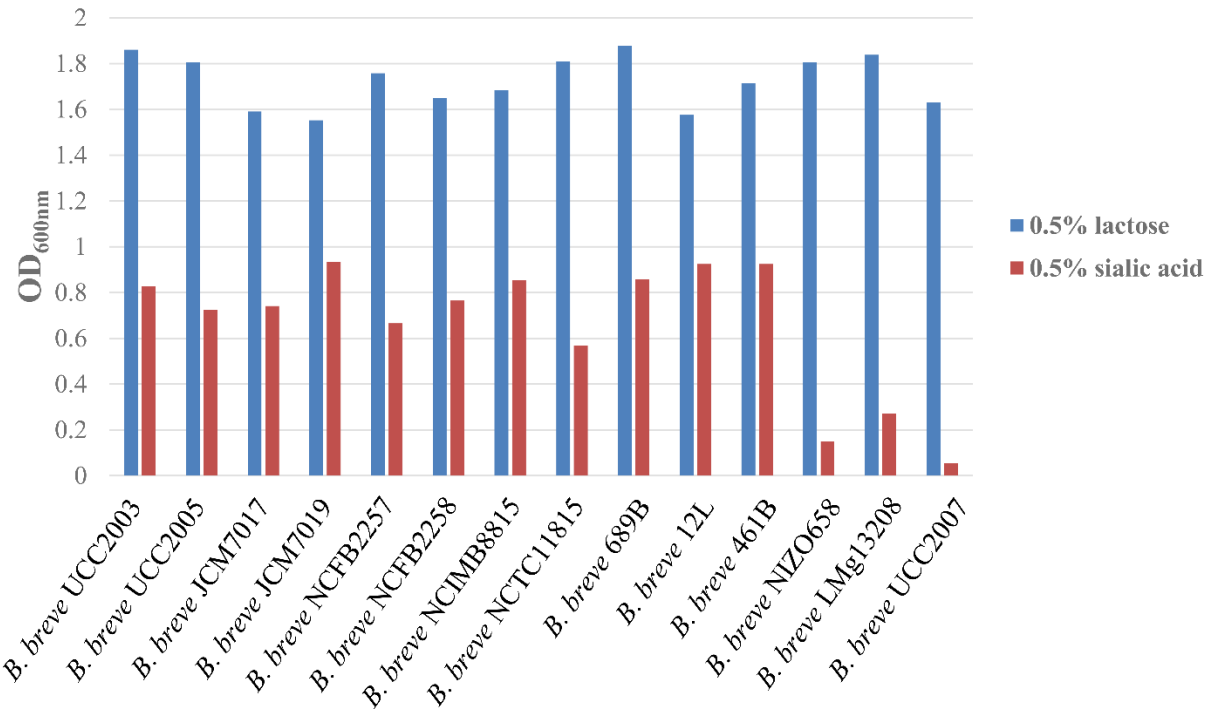


Figure 2.2: Final OD_{600nm} values obtained following 24 h growth of 14 different *B. breve* strains in mMRS containing 0.5 % (wt/vol) lactose or sialic acid as the sole carbon source.

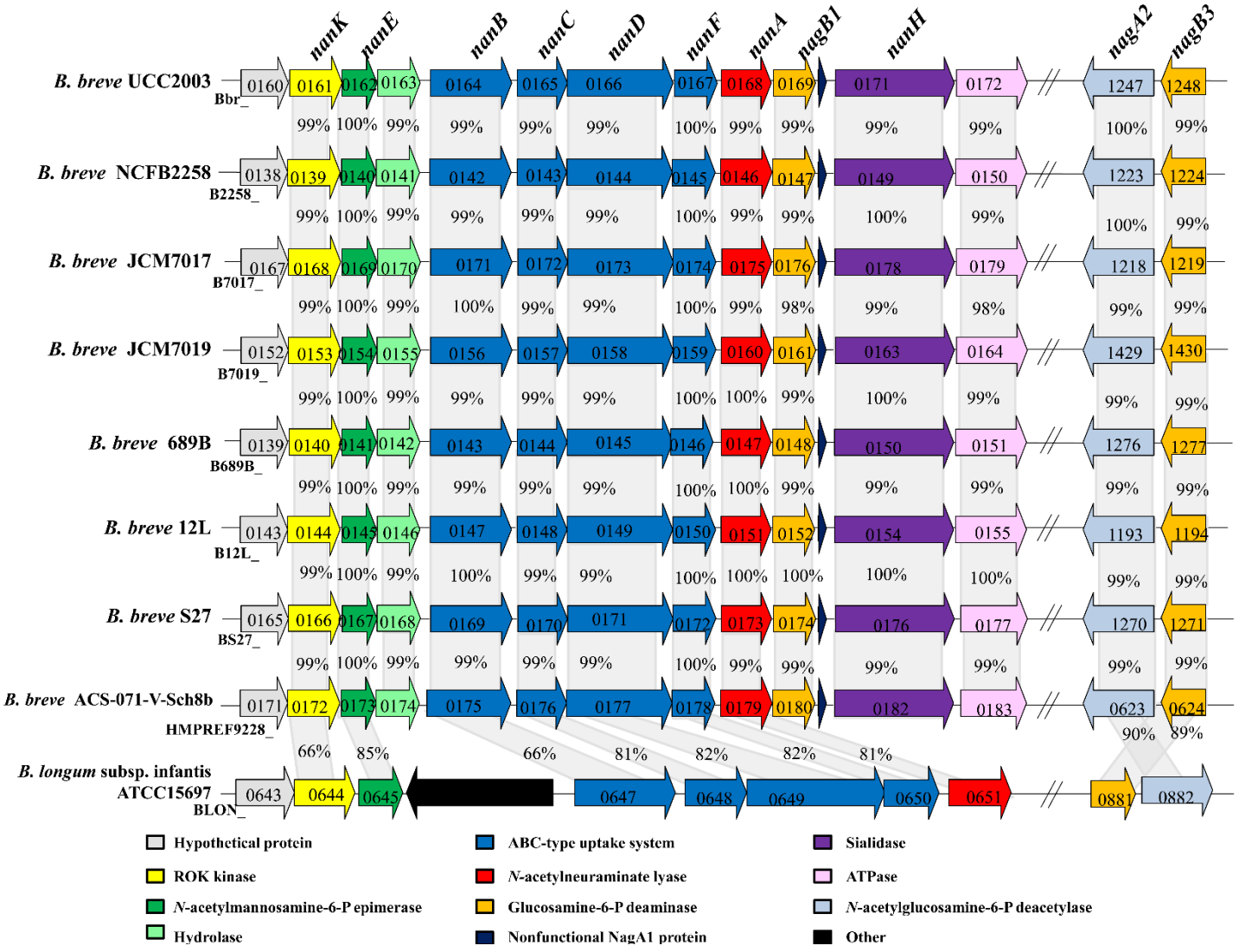
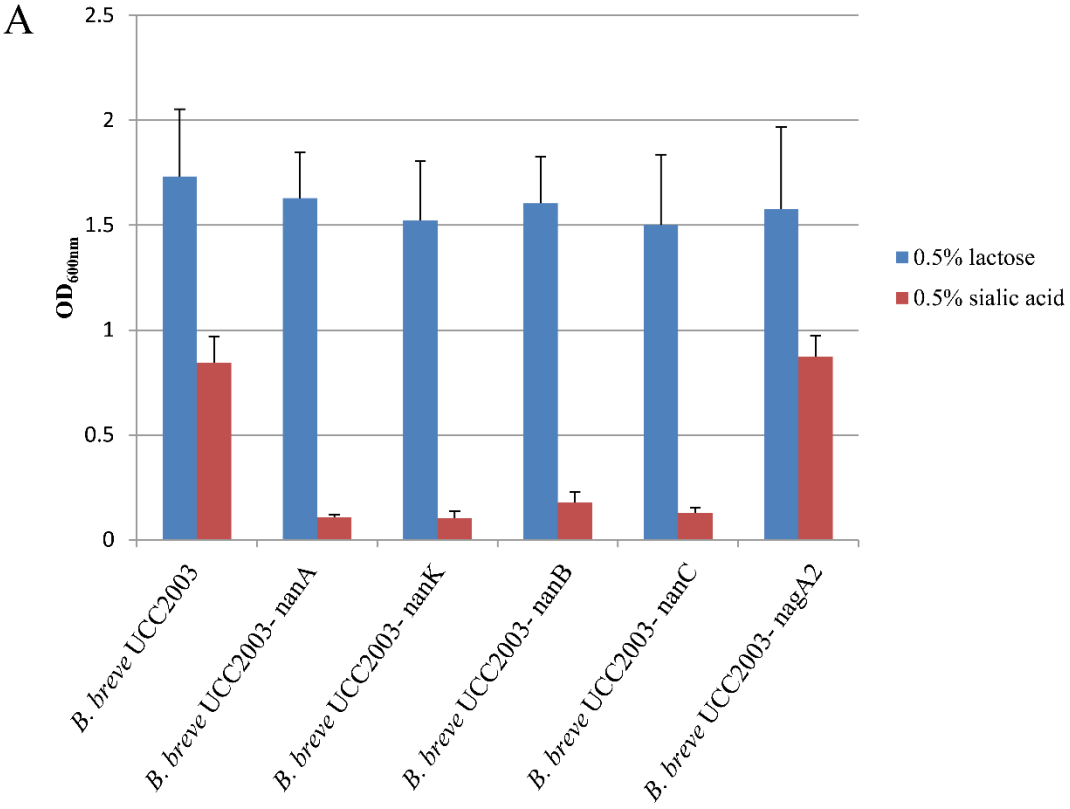

nanK
nanE
nanB
nanC
nanD
nanF
nanA
nagB1
nanH
nagA2
nagB3
B. breve UCC2003
Bbr_
B. breve NCFB2258
B2258_
B. breve JCM7017
B7017_
B. breve JCM7019
B7019_
B. breve 689B
B689B_
B. breve 12L
B12L_
B. breve S27
BS27_
B. breve ACS-071-V-Sch8b
HMPREF9228_
B. longum subsp. infantis ATCC15697
BLON_
Hypothetical protein
ROK kinase
N-acetylmannosamine-6-P epimerase
Hydrolase
ABC-type uptake system
N-acetylneuraminate lyase
Glucosamine-6-P deaminase
Nonfunctional NagA1 protein
Sialidase
ATPase
N-acetylglucosamine-6-P deacetylase
Other

Figure 2.3: Comparison of the sialic acid gene cluster of *B. breve* UCC2003 with corresponding putative sialic acid utilisation loci of other bifidobacteria. Each solid arrow represents an open reading frame. The length of the arrows (which contain the locus tag number) is proportional to the size of the open reading frame and the gene locus name, which is indicative of its putative function, is given at the top. Orthologs are marked with the same colour. The amino acid identity of each predicted protein to its equivalent protein encoded by *B. breve* UCC2003, expressed as a percentage, is given above each arrow.

A


2.5
2
1.5
1
0.5
0
OD600nm
0.5% lactose
0.5% sialic acid
B. breve UCC2003
B. breve UCC2003- nanA
B. breve UCC2003- nanK
B. breve UCC2003- nanB
B. breve UCC2003- nanC
B. breve UCC2003- nagA2

B

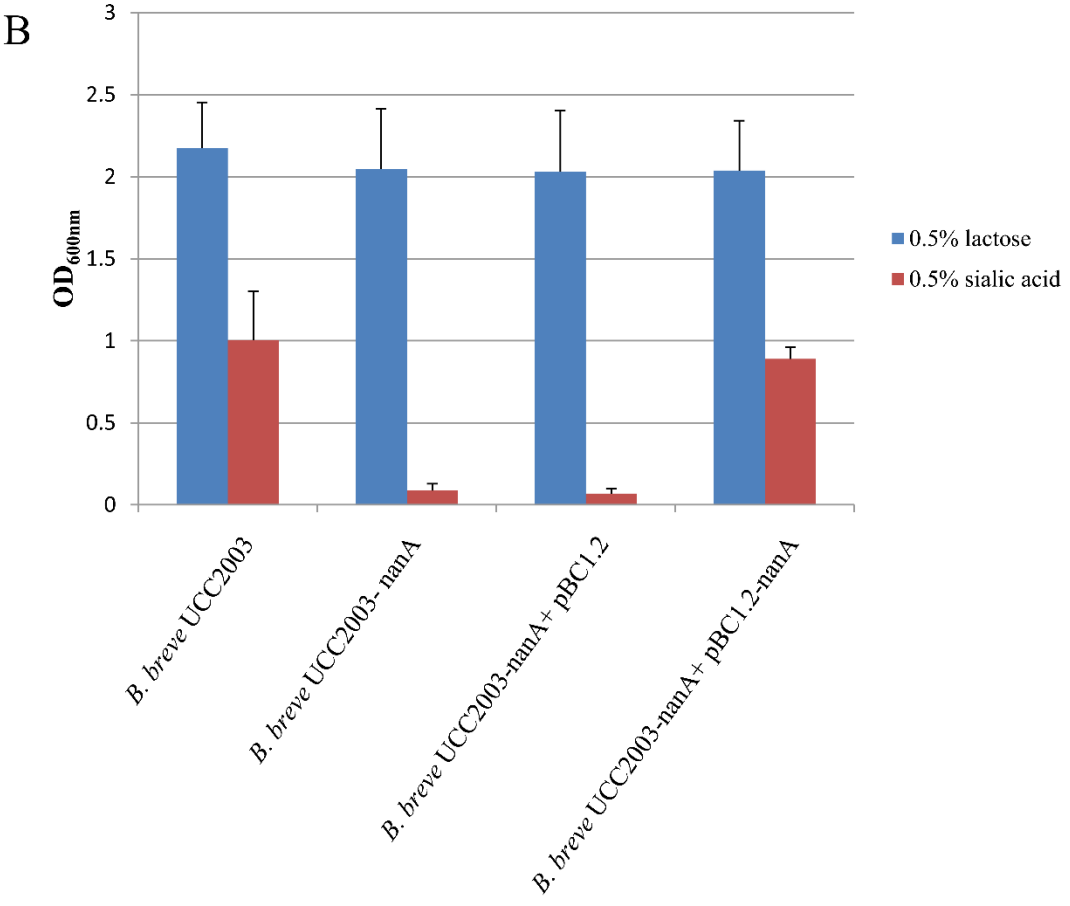

3
2.5
2
1.5
1
0.5
0
OD600nm
0.5% lactose
0.5% sialic acid
B. breve UCC2003
B. breve UCC2003- nanA
B. breve UCC2003-nanA+ pBC1.2
B. breve UCC2003-nanA+ pBC1.2-nanA

Figure 2.4: (A) Final OD_{600nm} values after 24 h of growth of *B. breve* UCC2003 and the insertion mutants, *B. breve* UCC2003-nanA, *B. breve* UCC2003-nanK, *B. breve* UCC2003-nanB, *B. breve* UCC2003-nanC and *B. breve* UCC2003-nagA2 in mMRS containing 0.5 % (wt/vol) lactose or sialic acid as the sole carbon source. **(B)** Final OD_{600nm} values after 24 h of growth of *B. breve* UCC2003, the insertion mutants *B. breve* UCC2003-nanA and *B. breve* UCC2003-nanA-pBC1.2, and the complemented insertion mutant *B. breve* UCC2003-nanA-pBC1.2-nanA in mMRS containing 0.5 % (wt/vol) lactose or sialic acid as the sole carbon source. The results are the mean values obtained from three separate experiments. Error bars represent the standard deviation.

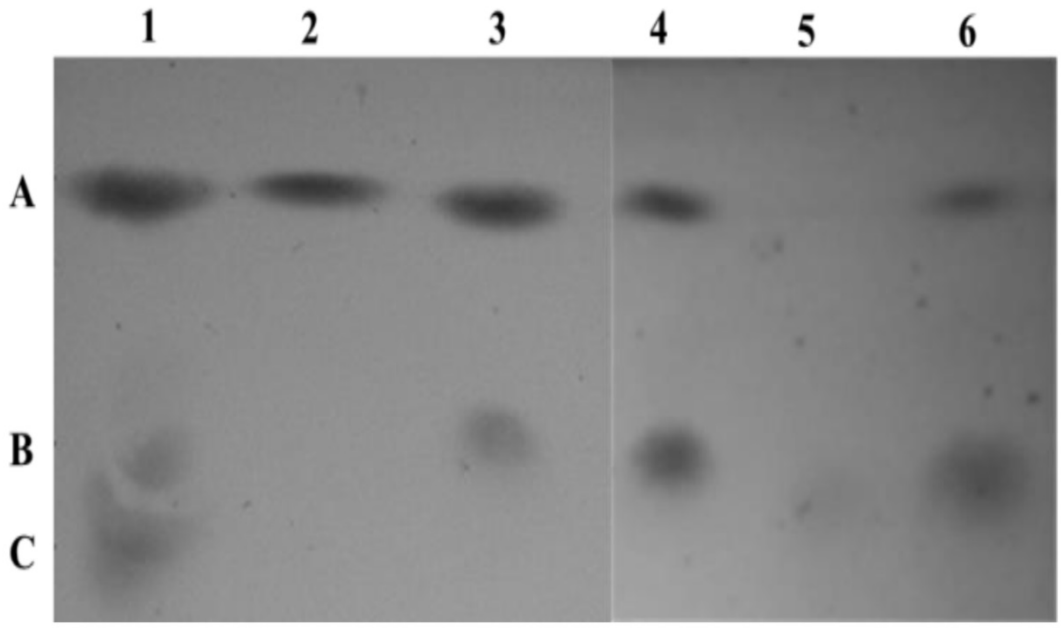
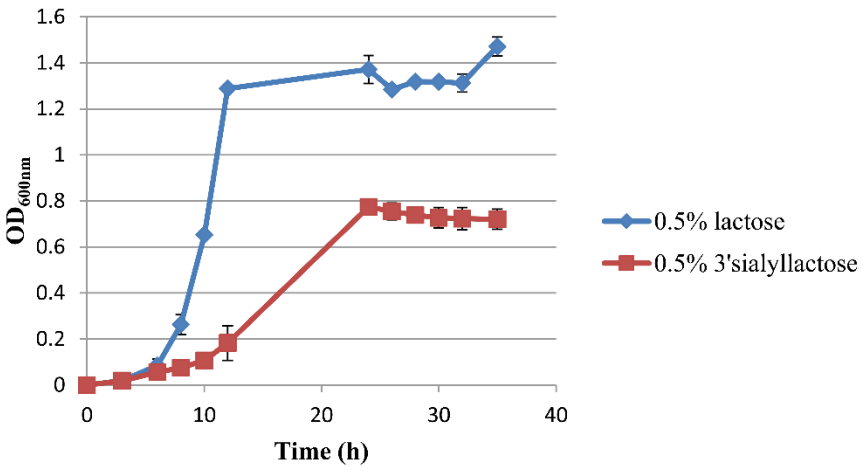
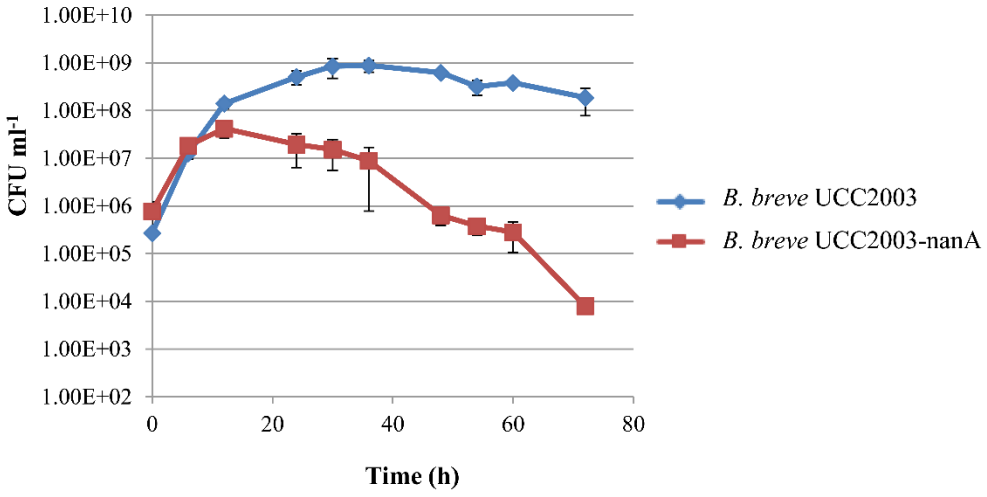


Figure 2.5: Hydrolysis of sialic acid to fructose-6-phosphate by the combined activities of purified $NanA_{His}$, $NanK_{His}$, $NanE_{His}$, $NagA2_{His}$ and $NagB1_{His}$. Lane 1: 5 mg ml^{-1} (wt/vol) standards in kinase buffer: (A) *N*-acetylmannosamine, (B) *N*-acetylmannosamine-6-phosphate, (C) fructose-6-phosphate. Lane 2: $NanA_{His}$ + sialic acid. Lane 3: $NanK_{His}$ + *N*-acetylmannosamine. Lane 4: $NanA_{His}$ and $NanK_{His}$ + sialic acid. Lane 5: $NanE_{His}$, $NagA2_{His}$ and $NagB1_{His}$ + *N*-acetylmannosamine-6-phosphate. Lane 6: $NanA_{His}$, $NanK_{His}$, $NanE_{His}$, $NagA2_{His}$ and $NagB1_{His}$ + sialic acid.

A


1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
OD600nm
0
10
20
30
40
Time (h)
0.5% lactose
0.5% 3'sialyllactose

B


1.00E+10
1.00E+09
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
1.00E+03
1.00E+02
CFU ml-1
0
20
40
60
80
Time (h)
B. breve UCC2003
B. breve UCC2003-nanA

C

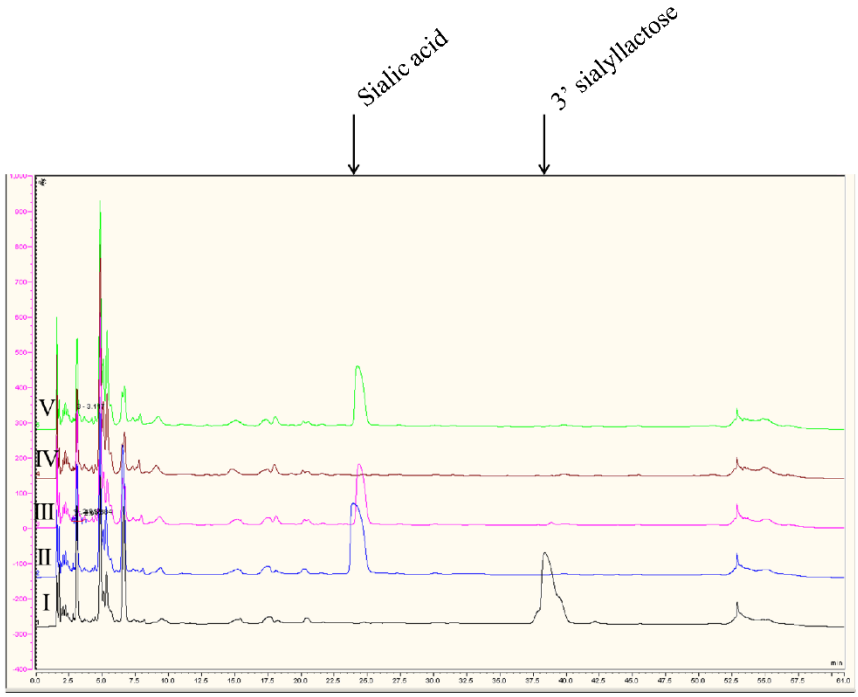

Sialic acid
3' sialyllactose
V
IV
III
II
I

Figure 2.6: (A) Growth profiles of *B. bifidum* PRL2010 in mMRS containing 0.5 % (wt/vol) lactose or 3' sialyllactose for 36 h. The results are the mean values obtained from two separate experiments. Error bars represent the standard deviation. **(B)** Growth profiles of *B. breve* UCC2003 and *B. breve* UCC2003-nanA in mMRS containing 0.5 % (wt/vol) 3'sialyllactose (previously fermented by *B. bifidum* PRL2010 for 24 h). The results are the mean values obtained from two separate experiments. Error bars represent the standard deviation. **(C)** HPAEC profiles of (I) mMRS + 0.5 % (wt/vol) 3' sialyllactose; (II) mMRS + 0.5 % (wt/vol) sialic acid; (III) mMRS + 0.5 % (wt/vol) 3'sialyllactose after 24 h growth of *B. bifidum* PRL2010; (IV) the media from (III) after 30 h growth of *B. breve* UCC2003 (V) the media from (III) after 30 h growth of *B. breve* UCC2003-nanA.

Table S2.1: Oligonucleotide primers used in this study.

Purpose	Primer	Sequence
Cloning of 301 bp fragment of *nanA* in pORI19	NanAF	TAGCAT*CTGCAG*GAGATCGAGCGCCACTTC
	NanAR	TGGCAT*TCTAGA*CACGTTGGCGAGTCCCGGC
Cloning of 370 bp fragment of *nanK* in pORI19	NanKF	GCTATT*AAGCTT*GTAGTGCTCTCGATCGGCGTG
	NanKR	GATCTA*TCTAGA*CGTCCGGGTCGATGAGGTTG
Cloning of 504 bp fragment of *nanB* in pORI19	NanBF	TACGCT*GCATGC*GTGGAACGTGAATGTGGATGAC
	NanBR	GTATGC*TCTAGA*GGGCGGAGGACTTCATGGAC
Cloning of 355 bp fragment of *nanC* in pORI19	NanCF	TAGCAT*AAGCTT*CTTCCTGATGTCGTTCTCCC
	NanCR	GACTAG*TCTAGA*GGTAGCGATGGCGATGATG
Cloning of 402 bp fragment of *nagA2* in pORI19	NagA2F	TAGC*TACTGC*AGGCATTGATGTGGCTCGTG
	NagA2R	GCATCG*TCTAGA*CCTTGGCGGTCTCGTAGTC
Amplification of *tetW*	TetWF	TCAGCT*GTCGAC*ATGCTCATGTACGGTAAG
	TetWR	GCGACG*GTCGAC*CATTACCTTCTGAAACATA
Confirmation of site-specific homologous recombination	NanAconfirm	ACCGTCATGGAGATTGTCGA
	NanKconfirm	GTGACCGACCTGCCAGTGC
	NanBconfirm	GCGGACCTCAAGAAGATGCC
	NanCconfirm	CACAAGTCATTCGCCATCAC
	NagA2confirm	CCGTAATGCAGTGGTGGATG
Cloning of *nanA* into pNZ44	NanApNZ44F	TAGCAT*TCTAGA*GATGTCAGGCCTCGGTGGA
	NanApNZ44R	TAGCAT*TCTAGA*CAACGAAGGGAACCGGGAC
Cloning of *p44-nanA* in pBC1.2	P44FORWARD	TCAGCA*GGATCC*GAGATAATGCCGACTGTAC
Cloning of *nanA* into pNZ8150	NanAFOR	TAGCAT*GATATC*ATGACAACACAATTCCGCG
	NanAREV	CATTAG*TCTAGA*TCA**ATGGTGATGGTGATGGTGATGGTGATGGTG**GGTGGAG GCGGAGACTTC
Cloning of *nanK* into pNZ8150	NanKFOR	TAGCGT*GATATC*ATG**CACCATCACCATCACCATCACCATCACCAT**CCCACCTA CCTCGCCTTCGACATC
	NanKREV	TAACGT*TCTAGA*CTTGTCTATCGCGCGTCGTCCGGC
Cloning of *nanE* into pNZ8150	NanEFOR	TAGCAT*GATATC*ATG**CACCATCACCATCACCATCACCATCACCAT**CATCAGGC AATCGAGAAGATC
	NanEREV	CGTTCA*TCTAGA*CTCACAGCTTCGCCTCG
Cloning of *nagA2* in pNZ8150	NagA2FOR	TAGCAT*GATATC*ATG**CACCATCACCATCACCATCACCATCACCAT**TCTGATAG GGAAGAAATTGTCTC
	NagA2REV	CGTTCA*TCTAGA*GCCATTGCAGCTATTTC

Cloning of *nagB1* in pNZ8150	NagB1FOR	TAGCAT*GATATC*ATGGCAGAAGTCATCATCGTCAAG
	NagB1REV	TGCACG*TCTAGA*TCA**ATGGTGATGGTGATGGTGATGGTGATGGTG**GATCGTC TGCCAGGCCGG

Restriction sites incorporated into oligonucleotide primer sequences are indicated in italics, 10xHis tags are indicated in bold

2.9 REFERENCES

Allen, S., Zaleski, A., Johnston, J. W., Gibson, B. W. & Apicella, M. A. (2005). Novel sialic acid transporter of *Haemophilus influenzae*. *Infect Immun* **73**, 5291 - 5300.

Almagro-Moreno, S. & Boyd, E. F. (2009). Sialic acid catabolism confers a competitive advantage to pathogenic *Vibrio cholerae* in the mouse intestine. *Infect Immun* **77**, 3807-3816.

Aloisio, I., Santini, C., Biavati, B., Dinelli, G., Cencič, A., Chingwaru, W., Mogna, L. & Gioia, D. (2012). Characterization of *Bifidobacterium* spp. strains for the treatment of enteric disorders in newborns. *Appl Microbiol Biotechnol* **96**, 1561-1576.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Álvarez-Martín, P., O'Connell-Motherway, M., van Sinderen, D. & Mayo, B. (2007). Functional analysis of the pBC1 replicon from *Bifidobacterium catenulatum* L48. *Appl Microbiol Biotechnol* **76**, 1395-1402.

Alvarez-Martin, P., O'Connell Motherway, M., Turroni, F., Foroni, E., Ventura, M. & van Sinderen, D. (2012). A two-component regulatory system controls autoregulated serpin expression in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **78**, 7032-7041.

Anba-Mondoloni, J., Chaillou, S., Zagorec, M. & Champomier-Vergès, M. C. (2013). Catabolism of *N*-acetylneuraminic acid, a fitness function of the food-borne lactic acid bacterium *Lactobacillus sakei*, involves two newly characterized proteins. *Appl Environ Microbiol* **79**, 2012-2018.

Bao, Y., Zhu, L. & Newburg, D. S. (2007). Simultaneous quantification of sialyloligosaccharides from human milk by capillary electrophoresis. *Anal Biochem* **370**, 206-214.

Bottacini, F., O'Connell Motherway, M., Kuczynski, J., O'Connell, K., Serafini, F., Duranti, S., Milani, C., Turroni, F., Lugli, G. & other authors (2014). Comparative genomics of the *Bifidobacterium breve* taxon. *BMC Genomics* **15**, 170.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **7**, 248-254.

Brigham, C., Caughlan, R., Gallegos, R., Dallas, M. B., Godoy, V. G. & Malamy, M. H. (2009). Sialic acid (*N*-acetylneuraminic acid) utilization by *Bacteroides fragilis* requires a novel *N*-acetylmannosamine epimerase. *J Bacteriol* **191**, 3629-3638.

Chowdhury, N., Norris, J., McAlister, E., Lau, S. Y. K., Thomas, G. H. & Boyd, E. F. (2012). The VC1777–VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen *Vibrio cholerae*. *Microbiology* **158**, 2158-2167.

Conejo, M., Thompson, S. & Miller, B. (2010). Evolutionary bases of carbohydrate recognition and substrate discrimination in the ROK protein family. *J Mol Evol* **70**, 545-556.

Corfield, A. P., Myerscough, N., Warren, B. F., Durdey, P., Paraskeva, C. & Schauer, R. (1999). Reduction of sialic acid O-acetylation in human colonic mucins in the adenoma-carcinoma sequence. *Glycoconjugate J* **16**, 307-317.

De Man, J. C., Rogosa, M. & Sharpe, M. E. (1960). A medium for the cultivation of lactobacilli. *J Appl Bacteriol* **23**, 130-135.

de Ruyter, P. G., Kuipers, O. P. & de Vos, W. M. (1996). Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl Environ Microbiol* **62**, 3662-3667.

García de la Nava, J., Santaella, D. F., Alba, J. C., Carazo, J. M., Trelles, O. & Pascual-Montano, A. (2003). Engene: the processing and exploratory analysis of gene expression data. *Bioinformatics* **19**, 657-658.

Garrido, D., Nwosu, C., Ruiz-Moyano, S., Aldredge, D., German, J. B., Lebrilla, C. B. & Mills, D. A. (2012). Endo-β-*N*-acetylglucosaminidases from infant gut-associated bifidobacteria release complex *N*-glycans from human milk glycoproteins. *Mol Cell Proteomics* **11**, 775-785.

Kiyohara, M., Tanigawa, K., Chaiwangsri, T., Katayama, T., Ashida, H. & Yamamoto, K. (2011). An exo-α-sialidase from bifidobacteria involved in the

degradation of sialyloligosaccharides in human milk and intestinal glycoconjugates. *Glycobiology* **21**, 437-447.

Krüger, D., Schauer, R. & Traving, C. (2001). Characterization and mutagenesis of the recombinant *N*-acetylneuraminate lyase from *Clostridium perfringens*. *Eur J Biochem* **268**, 3831-3839.

Kunz, C., Rudloff, S., Baier, W., Klein, N. & Strobel, S. (2000). Oligosaccharides in human milk: Structural, functional, and metabolic aspects. *Annu Rev Nutr* **20**, 699-722.

Law, J., Buist, G., Haandrikman, A., Kok, J., Venema, G. & Leenhouts, K. (1995). A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. *J Bacteriol* **177**, 7011-7018.

LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D. & Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr Opin Biotechnol* **24**, 160-168.

Ley, R. (2014). Harnessing microbiota to kill a pathogen: The sweet tooth of *Clostridium difficile*. *Nat Med* **20**, 248-249.

Lilley, G. G., Barbosa, J. A. R. G. & Pearce, L. A. (1998). Expression in *Escherichia coli* of the putative *N*-acetylneuraminate lyase gene (*nanA*) from *Haemophilus influenzae:* Overproduction, purification, and crystallization. *Protein Expression Purif* **12**, 295-304.

Long, A. D., Mangalam, H. J., Chan, B. Y. P., Tolleri, L., Hatfield, G. W. & Baldi, P. (2001). Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework: Analysis of global gene expression in *Escherichia coli* K12. *J Biol Chem* **276**, 19937-19944.

Martin-Sosa, S., Martin, M. J., Garcia-Pardo, L. A. & Hueso, P. (2003). Sialyloligosaccharides in human and bovine milk and in infant formulas: variations with the progression of lactation. *J Dairy Sci* **86**, 52-59.

Martinez, J., Steenbergen, S. & Vimr, E. (1995). Derived structure of the putative sialic acid transporter from *Escherichia coli* predicts a novel sugar permease domain. *J Bacteriol* **177**, 6005 - 6010.

Maze, A., O'Connell Motherway, M., Fitzgerald, G., Deutscher, J. & van Sinderen, D. (2007). Identification and characterization of a fructose phosphotransferase system in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **73**, 545-553.

McGrath, S., Fitzgerald, G. F. & van Sinderen, D. (2001). Improvement and optimization of two engineered phage resistance mechanisms in *Lactococcus lactis*. *Appl Environ Microbiol* **67**, 608-616.

Mierau, I. & Kleerebezem, M. (2005). 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. *Appl Microbiol Biotechnol* **68**, 705-717.

Nees, S. & Schauer, R. (1974). Induction of neuraminidase from *Clostridium perfringens* and the co-operation of this enzyme with acylneuraminate pyruvate lyase. *Behring Inst Mitt* **55**, 68-78.

Nees, S., Schauer, R., Mayer, F. & Ehrlich, K. (1976). Purification and characterization of *N*-acetylneuraminate lyase from *Clostridium perfringens*. *Hoppe-Seyler Zeitschrift physiologische Chemie* **357**, 839-854.

Ng, K. M., Ferreyra, J. A., Higginbottom, S. K., Lynch, J. B., Kashyap, P. C., Gopinath, S., Naidu, N., Choudhury, B., Weimer, B. C. & other authors (2013). Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96-99.

Ninonuevo, M., Park, Y., Yin, H., Zhang, J., Ward, R., Clowers, B., German, B., Freeman, S., Killeen, K. & other authors (2006). A strategy for annotating the human milk glycome. *J Agric Food Chem* **54**, 7471-7480.

O'Connell, K. J., O'Connell Motherway, M., O'Callaghan, J., Fitzgerald, G. F., Ross, R. P., Ventura, M., Stanton, C. & van Sinderen, D. (2013). Metabolism of four α-glycosidic linkage-containing oligosaccharides by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **79**, 6280-6292.

O'Connell, K. J., O'Connell Motherway, M., Liedtke, A., Fitzgerald, G. F., Ross, R. P., Stanton, C., Zomer, A. & van Sinderen, D. (2014). Transcription of two adjacent carbohydrate utilization gene clusters in *Bifidobacterium breve*

UCC2003 is controlled by LacI- and Repressor Open reading frame Kinase (ROK)-type regulators. *Appl Environ Microbiol* **80**, 3604-3614.

O'Connell Motherway, M., Fitzgerald, G., Neirynck, S., Ryan, S., Steidler, L. & van Sinderen, D. (2008). Characterization of ApuB, an extracellular type II amylopullulanase from *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **74**, 6271-6279.

O'Connell Motherway, M., O'Driscoll, J., Fitzgerald, G. F. & Van Sinderen, D. (2009). Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **2**, 321-332.

O'Connell Motherway, M., Fitzgerald, G. F. & van Sinderen, D. (2011a). Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **4**, 403-416.

O'Connell Motherway, M., Zomer, A., Leahy, S. C., Reunanen, J., Bottacini, F., Claesson, M. J., O'Brien, F., Flynn, K., Casey, P. G. & other authors (2011b). Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc Natl Acad Sci USA* **108**, 11217-11222.

O'Connell Motherway, M., Kinsella, M., Fitzgerald, G. F. & van Sinderen, D. (2013). Transcriptional and functional characterization of genetic elements involved in galacto-oligosaccharide utilization by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **6**, 67-79.

Ogura, T. & Wilkinson, A. J. (2001). AAA+ superfamily ATPases: common structure–diverse function. *Genes Cells* **6**, 575-597.

Olson, M. E., King, J. M., Yahr, T. L. & Horswill, A. R. (2013). Sialic acid catabolism in *Staphylococcus aureus*. *J Bacteriol* **195**, 1779-1788.

Plumbridge, J. & Vimr, E. (1999). Convergent pathways for utilization of the amino sugars *N*-acetylglucosamine, *N*-acetylmannosamine, and *N*-acetylneuraminic acid by *Escherichia coli*. *J Bacteriol* **181**, 47-54.

Podolsky, D. K. (1985). Oligosaccharide structures of human colonic mucin. *J Biol Chem* **260**, 8262-8271.

Pokusaeva, K., Neves, A. R., Zomer, A., O'Connell-Motherway, M., MacSharry, J., Curley, P., Fitzgerald, G. F. & van Sinderen, D. (2010). Ribose utilization by the human commensal *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **3**, 311-323.

Pokusaeva, K., O'Connell-Motherway, M., Zomer, A., MacSharry, J., Fitzgerald, G. F. & van Sinderen, D. (2011). Cellodextrin utilization by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **77**, 1681-1690.

Post, D. M., Mungur, R., Gibson, B. W. & Munson, R. S. (2005). Identification of a novel sialic acid transporter in *Haemophilus ducreyi*. *Infect Immun* **73**, 6727 - 6735.

Reizer, J., Ramseier, T. M., Reizer, A., Charbit, A. & Saier, M. H. (1996). Novel phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative *N*-acetylgalactosamine metabolic pathway in *Escherichia coli*. *Microbiology* **142**, 231-250.

Riordan, O. (1998). *Studies on antimicrobial activity and genetic diversity of Bifidobacterium species: molecular characterization of a 5.75 kb plasmid and a chromosomally encoded recA gene homologue from Bifidobacterium breve.*, National University of Ireland, Cork, Cork.

Round, J. L. & Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313-323.

Ruiz-Moyano, S., Totten, S. M., Garrido, D. A., Smilowitz, J. T., German, J. B., Lebrilla, C. B. & Mills, D. A. (2013). Variation in consumption of human milk oligosaccharides by infant gut-associated strains of *Bifidobacterium breve*. *Appl Environ Microbiol* **79**, 6040-6049.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Ryan, S., Fitzgerald, G. & van Sinderen, D. (2006). Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains. *Appl Environ Microbiol* **72**, 5289-5296.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning a laboratory manual*, 2nd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Sánchez-Carrón, G., García-García, M. I., López-Rodríguez, A. B., Jiménez-García, S., Sola-Carvajal, A., García-Carmona, F. & Sánchez-Ferrer, A. (2011). Molecular characterization of a novel *N*-acetylneuraminate lyase from *Lactobacillus plantarum* WCFS1. *Appl Environ Microbiol* **77**, 2471-2478.

Scardovi, V. & Trovatelli, I. D. (1965). The fructose-6-phosphate shunt as a peculiar pattern of hexose degradation in the genus *Bifidobacterium*. *Ann Microbiol*, 19-29.

Schell, M., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M.-C., Desiere, F., Bork, P. & other authors (2002). The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci U S A* **99**, 14422-14427.

Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., Lapidus, A., Rokhsar, D. S., Lebrilla, C. B. & other authors (2008). The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A* **105**, 18964-18969.

Sela, D. A., Li, Y., Lerno, L., Wu, S., Marcobal, A. M., German, J. B., Chen, X., Lebrilla, C. B. & Mills, D. A. (2011). An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides. *J Biol Chem* **286**, 11909-11918.

Servin, A. L. (2004). Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiol Rev* **28**, 405-440.

Terzaghi, B. E. & Sandine, W. E. (1975). Improved medium for lactic streptococci and their bacteriophages. *Appl Microbiol* **29**, 807-813.

Tissier, H. (1900). *Recherchers sur la flora intestinale normale pathologique du nourisson.*, The University of Paris, Paris.

Titgemeyer, F., Reizer, J., Reizer, A. & Saier, M. H. (1994). Evolutionary relationships between sugar kinases and transcriptional repressors in bacteria. *Microbiology* **140**, 2349-2354.

Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J.-H., Zomer, A., Sanchez, B., Bidossi, A., Ferrarini, A. & other authors (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., Strati, F., Foroni, E., Serafini, F., Duranti, S., van Sinderen, D. & Ventura, M. (2012a). Analysis of predicted carbohydrate transport systems encoded by *Bifidobacterium bifidum* PRL2010. *Appl Environ Microbiol* **78**, 5002-5012.

Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., Kerr, C., Hourihane, J., Murray, D. & other authors (2012b). Diversity of bifidobacteria within the infant gut microbiota. *PLoS ONE* **7**, e36957.

van Hijum, S. A. F. T., Garcia De La Nava, J., Trelles, O., Kok, J. & Kuipers, O. P. (2003). MicroPreP: a cDNA microarray data pre-processing framework. *Appl Bioinformatics*, 241-244.

van Hijum, S., de Jong, A., Baerends, R., Karsens, H., Kramer, N., Larsen, R., den Hengst, C., Albers, C., Kok, J. & other authors (2005). A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC Genomics* **6**, 77.

Vaughan, E., Heilig, H., Ben-Amor, K. & De Vos, W. (2005). Diversity, vitality and activities of intestinal lactic acid bacteria and bifidobacteria assessed by molecular approaches. *FEMS Microbiol Rev* **29**, 477-490.

Ventura, M., Canchaya, C., Fitzgerald, G., Gupta, R. & Sinderen, D. (2007). Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria. *Antonie van Leeuwenhoek* **92**, 265-265.

Ventura, M., Turroni, F., O'Connell Motherway, M., MacSharry, J. & van Sinderen, D. (2012). Host–microbe interactions that facilitate gut colonization by commensal bifidobacteria. *Trends Microbiol* **20**, 467-476.

Vimr, E. R. & Troy, F. A. (1985a). Regulation of sialic acid metabolism in *Escherichia coli*: role of *N*-acylneuraminate pyruvate-lyase. *J Bacteriol* **164**, 854 - 860.

Vimr, E. R. & Troy, F. A. (1985b). Identification of an inducible catabolic system for sialic acids (nan) in *Escherichia coli*. *J Bacteriol* **164**, 845-853.

Vimr, E.R., Lichtensteiger, C. & Steenbergen, S. (2000). Sialic acid metabolism's dual function in *Haemophilus influenzae*. *Mol Microbiol* **36**, 1113-1123.

Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. (2004). Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev* **68**, 132-153.

Wang, B., Brand-Miller, J., McVeagh, P. & Petocz, P. (2001). Concentration and distribution of sialic acid in human milk and infant formulas. *Am J Clin Nutr* **74**, 510-515.

Ward, R., Ninonuevo, M., Mills, D., Lebrilla, C. & German, B. (2007). In vitro fermentability of human milk oligosaccharides by several strains of bifidobacteria. *Mol Nutr Food Res* **51**, 1398-1405.

Watson, D., O'Connell Motherway, M., Schoterman, M. H. C., van Neerven, R. J. J., Nauta, A. & van Sinderen, D. (2013). Selective carbohydrate utilization by lactobacilli and bifidobacteria. *J Appl Microbiol* **114**, 1132-1146.

Wells, J. M., Wilson, P. W. & Le Page, R. W. F. (1993). Improved cloning vectors and transformation procedure for *Lactococcus lactis*. *J Appl Bacteriol* **74**, 629-636.

White, R. (1968). Control of amino sugar metabolism in *Escherichia coli* and isolation of mutants unable to degrade amino sugars. *Biochem J* **106**, 847-858.

Zomer, A., Fernandez, M., Kearney, B., Fitzgerald, G. F., Ventura, M. & van Sinderen, D. (2009). An interactive regulatory network controls stress response in *Bifidobacterium breve* UCC2003. *J Bacteriol* **191**, 7039-7049.

Chapter III

A GntR-type transcriptional repressor controls sialic acid utilisation in *Bifidobacterium breve* UCC2003.

This chapter was published in:

Egan M., O' Connell Motherway M., van Sinderen, D. A GntR-type transcriptional repressor controls sialic acid utilisation in *Bifidobacterium breve* UCC2003. *FEMS Microbiol Lett* **362,** 1-9

3.1 ABSTRACT

Bifidobacterium breve strains are numerically prevalent among the gut microbiota of healthy, breast-fed infants. The metabolism of sialic acid, a ubiquitous monosaccharide in the infant and adult gut, by *B. breve* UCC2003 is dependent on a large gene cluster, designated the *nan/nag* cluster. The aim of this study was to analyse the transcriptional regulation of the *nan/nag* cluster and thus sialic acid metabolism in *B. breve* UCC2003. Insertion mutagenesis and transcriptome analysis revealed that the *nan/nag* cluster is regulated by a GntR family transcriptional repressor, designated NanR. NanR was shown to bind to two promoter regions within this cluster, each of which containing an imperfect inverted repeat that is believed to act as the operator sequence. Formation of the DNA-NanR complex is prevented in the presence of sialic acid, which we had previously shown to induce transcription of this gene cluster. NanR represents the first GntR-type transcriptional regulator to be characterised from a member of the genus *Bifidobacterium*.

3.2 INTRODUCTION

Bifidobacteria are Gram-positive, strictly anaerobic, non-motile, non-sporulating bacteria, which are members of the *Actinobacteria* phylum and the *Bifidobacteriaceae* family. They were first isolated in 1900 by Henri Tissier, who recommended the administration of bifidobacteria to infants for the treatment of diarrhoea (Tissier, 1900). Bifidobacteria are typically found in the large intestine of humans and other mammals and represent the dominant bacterial group among the intestinal microbiota of breast-fed infants (Turroni *et al.*, 2012; Ventura *et al.*, 2007). Bifidobacteria have been implicated in a number of health-promoting activities including vitamin provision (LeBlanc *et al.*, 2013), serum cholesterol reduction (Bordoni *et al.*, 2013), the development and modulation of the immune system (Round & Mazmanian, 2009; Ventura *et al.*, 2012), and protection against pathogenic bacteria (Servin, 2004). For this reason bifidobacteria have been used as probiotic bacterial ingredients in certain foods (Stanton *et al.*, 2005).

Bifidobacteria are saccharolytic bacteria whose survival and colonisation of the large intestine requires the ability to utilise some of the complex diet- and host-derived carbohydrates present in this environment. Approximately 8 % of the predicted proteins in a typical bifidobacterial genome are assigned to the so-called carbohydrate-transport and metabolism COG (Cluster of Orthologous Groups) category, with this proportion rising to as high as 12 % in certain bifidobacterial genomes (Milani *et al.*, 2014; Schell *et al.*, 2002; Ventura *et al.*, 2007).

When bacteria are presented with a number of carbon sources, they will want to utilise the substrate that yields the highest amount of energy, while inhibiting the expression or activity of proteins involved in the uptake and catabolism of other

substrates (Gorke & Stulke, 2008). This preferential carbon source utilisation has also been described in various bifidobacteria, such as *Bifidobacterium animalis* subsp. *lactis* (Trindade *et al.*, 2003), *Bifidobacterium breve* UCC2003 (Pokusaeva *et al.*, 2010; Ryan *et al.*, 2005) and *Bifidobacterium longum* subsp. *longum* NCC2705 (Parche *et al.*, 2006).

One of the primary methods of regulation of carbohydrate utilisation pathways employed by bacteria is the transcriptional control of genes involved in the uptake or utilisation of less-favoured substrates (Cohn & Horibata, 1959). This involves transcriptional regulators, the best known of which is the LacI repressor from *Escherichia coli* (Jacob & Monod, 1961; Lewis, 2005 ; Matthews & Nichols, 1998; Wilson *et al.*, 2007). The characterised bifidobacterial transcriptional regulators have typically been LacI-family repressors, with the sole exception being RafR, a ROK-family transcriptional activator from *B. breve* UCC2003 (O'Connell *et al.*, 2014). Among the LacI-family transcriptional repressors characterised from *B. breve* UCC2003 are GalR which controls transcription of the galactan utilisation cluster (O'Connell Motherway *et al.*, 2011a), RbsR which regulates transcription of the *rbsACBDK* cluster for ribose utilisation (Pokusaeva *et al.*, 2010), CldR which controls transcription of the *cldEFGC* cluster for cellodextrin metabolism (Pokusaeva *et al.*, 2011), and MelR1 and MelR2 which regulate transcription of the melezitose utilisation cluster (O'Connell *et al.*, 2014).

The first characterised GntR-type transcriptional regulator, GntR from *Bacillus subtilis*, represses transcription of the *gnt* operon required for gluconate metabolism (Fujita & Fujita, 1987). In 1991, Haydon and Guest presented the GntR family as transcription factors that possess a conserved N-terminal DNA binding domain, yet

exhibit considerable divergence in their C-terminal effector binding and oligomerisation domains (Haydon & Guest, 1991). Binding of the effector molecule to the C-terminal domain causes a conformational change of the N-terminal domain and thereby its associated DNA-binding properties (Haydon & Guest, 1991). The variable C-terminal portion is used to define the GntR subfamilies, of which the FadR subfamily is the largest (Rigali *et al.*, 2002). GntR-type regulators are typically found in *Actinobacteria*, *Proteobacteria* and *Firmicutes* (Hoskisson & Rigali, 2009), and can regulate various cellular processes, including, but not limited to, mobility in *Vibrio parahaemolyticus* (Jaques & McCarter, 2006), virulence in *Mycobacterium tuberculosis* (Casali *et al.*, 2006) and citrate metabolism in *Enterococcus faecalis* (Blancato *et al.*, 2008).

A GntR-type transcriptional repressor was previously shown to control sialic acid metabolism in *E. coli* (Kalivoda *et al.*, 2003). Sialic acid metabolism in *B. breve* UCC2003 requires a gene cluster comprising of twelve genes (Bbr_0160-0172), designated the *nan/nag* cluster, as well as two additional genes, *nagA2* and *nagB3* (Bbr_1247-1248), located elsewhere on the genome (Egan *et al.*, 2014). The current study describes the characterisation of NanR, the transcriptional regulator of the *nan/nag* cluster in *B. breve* UCC2003, and the first bifidobacterial GntR-type regulator to be characterised.

3.3 MATERIALS AND METHODS

Bacterial strains, plasmids, media and culture conditions. Bacterial strains and plasmids used in this study are listed in Table 3.1. *B. breve* UCC2003 was routinely cultured in Reinforced Clostridial Medium (RCM; Oxoid Ltd., Basingstoke, Hampshire, United Kingdom). Carbohydrate utilisation by bifidobacteria was examined in modified deMan Rogosa Sharpe (mMRS) medium made from first principles (De Man *et al.*, 1960), without the inclusion of the standard carbohydrate source, and supplemented with 0.05 % (wt/vol) L-cysteine HCl and a particular carbohydrate source (0.5 % wt/vol). The carbohydrates used were lactose (Sigma Aldrich, Steinheim, Germany) and sialic acid (Carbosynth, Compton, Berkshire, United Kingdom). Addition of sialic acid to mMRS resulted in a decrease in pH to 6.0, and hence, the pH was re-adjusted to 6.8 following addition of the sugar and subsequently filter sterilised. In order to determine bacterial growth profiles and final optical densities, 10 ml of freshly prepared mMRS medium, including a particular carbohydrate source, was inoculated with 100 µl (1 %) of a stationary-phase culture of a particular strain. Un-inoculated mMRS was used as a negative control. Cultures were incubated anaerobically for 24 h and the optical density (OD_{600nm}) was recorded. Bifidobacterial cultures were incubated under anaerobic conditions in a modular atmosphere-controlled system (Davidson and Hardy, Belfast, Ireland) at 37°C. *E. coli* was cultured in Luria Bertani (LB) broth at 37°C with agitation (Sambrook, 1989). Where appropriate, growth media contained tetracycline (Tet; 10 µg ml^{-1}), chloramphenicol (Cm; 5 µg ml^{-1} for *E. coli*, 2.5 µg ml^{-1} for *B. breve*), ampicillin (Amp; 100 µg ml^{-1}), erythromycin (Em; 100 µg ml^{-1}) or kanamycin (Kan; 50 µg ml^{-1}). Recombinant *E. coli* cells containing pORI19 were selected on LB agar containing Em and Kan, and supplemented with 40 µg ml^{-1} X-gal (5-bromo-4-

chloro-3-indolyl-β-D-galactopyranoside) and 1 mM IPTG (isopropyl-β-D-galactopyranoside).

Nucleotide sequence analysis. Sequence data were obtained from the Artemis-mediated genome annotations of *B. breve* UCC2003 (O'Connell Motherway *et al.*, 2011b; Rutherford *et al.*, 2000). Database searches were performed using the non-redundant sequence database accessible at the National Centre for Biotechnology Information website (http://www.ncbi.nlm.nih.gov) using BLAST (Altschul *et al.*, 1990). Sequence analysis was performed using the Seqbuilder and Seqman programs of the DNASTAR software package (DNASTAR, Madison, WI, USA). Inverted repeats in the Bbr_0160 and *nanB* promoter regions were identified using the PrimerSelect program of the DNASTAR software package and a graphical representation of the putative NanR-binding motif was obtained using WebLogo software (Crooks *et al.*, 2004). Multiple sequence alignments were performed using CLUSTAL W (Goujon *et al.*, 2010).

DNA manipulations. Chromosomal DNA was isolated from *B. breve* UCC2003 as previously described (Riordan, 1998). Plasmid DNA was isolated from *E. coli* and *B. breve* using the Roche High Pure plasmid isolation kit (Roche Diagnostics, Basel, Switzerland). An initial lysis step was performed using 30 mg ml^{-1} of lysozyme for 30 min at 37°C prior to plasmid isolation from *B. breve*. Single stranded oligonucleotide primers used in this study were synthesised by Eurofins (Ebersberg, Germany) (Table 3.2). Standard PCRs were performed using Taq master mix (Qiagen GmBH, Hilden, Germany). *B. breve* colony PCRs were carried out as

described previously (O'Connell Motherway *et al.*, 2009). PCR fragments were purified using the Roche High Pure PCR purification kit (Roche Diagnostics). Electrotransformation of *E. coli* or *B. breve* was performed as previously described (Maze *et al.*, 2007; Sambrook, 1989).

Construction of B. breve UCC2003-nanR. An internal fragment of 347 base pairs (bp), representing codon numbers 102 through to 217 of the 284 codons of gene Bbr_0173 (designated *nanR*), was amplified by PCR using *B. breve* UCC2003 chromosomal DNA as a template and the primer pair NanRIMF and NanRIMR. The insertion mutant was constructed as described previously (O'Connell Motherway *et al.*, 2009). Site-specific recombination of potential Tet-resistant mutant isolates was confirmed by colony PCR using primer combinations TetWF and TetWR to verify *tetW* gene integration, and the primer NanRconfirm (positioned upstream of the selected internal fragment of *nanR*) in combination with primer TetWF to confirm integration at the correct chromosomal location.

***Analysis of global gene expression using B. breve* UCC2003 *DNA microarrays*.** Global gene expression was determined during log-phase growth of *B. breve* UCC2003-nanR in mMRS supplemented with 0.5 % ribose and the obtained transcriptome was compared to that obtained from log-phase *B. breve* UCC2003 cells that had also been grown on ribose as the sole carbon source. DNA microarrays containing oligonucleotide primers representing each of the 1864 identified open reading frames on the genome of *B. breve* UCC2003 were designed and obtained from Agilent Technologies (Palo Alto, Ca., USA). Methods for cell disruption, RNA

isolation, RNA quality control, complementary DNA (cDNA) synthesis and labelling were performed as described previously (Zomer *et al.*, 2009). Labelled cDNA was hybridised using the Agilent Gene Expression hybridisation kit (part number 5188-5242) as described in the Agilent Two-Colour Microarray-Based Gene Expression Analysis v4.0 manual (publication number G4140-90050). Following hybridisation, microarrays were washed in accordance with Agilent's standard procedures and scanned using an Agilent DNA microarray scanner (model G2565A). Generated scans were converted to data files with Agilent's Feature Extraction software (Version 9.5). DNA-microarray data were processed as previously described (García de la Nava *et al.*, 2003; van Hijum *et al.*, 2003; van Hijum *et al.*, 2005). Differential expression tests were performed with the Cyber-T implementation of a variant of the *t*-test (Long *et al.*, 2001).

Plasmid constructions. For the construction of the plasmid pQE-60-nanR, a DNA fragment encompassing the predicted transcriptional repressor-encoding gene, *nanR* (Bbr_0173), was amplified by PCR using PFUII DNA polymerase and the primers NanRFORWARD and NanRREVERSE. Plasmid pQE-60 includes a C-terminal His_6-tag on the plasmid to facilitate downstream purification, to prevent the addition of this His_6-tag to the protein, a stop codon was included on the reverse primer. The generated amplicon was digested with AflIII and BglII, and ligated to NcoI and BglII-digested pQE-60. The ligated mixture was introduced into *E. coli* EC101 (Law *et al.*, 1995) by electrotransformation and transformants were selected based on ampicillin resistance (Amp^r). The plasmid content of a number of Amp^r

transformants was screened by restriction analysis and the integrity of positively identified clones was verified by sequencing.

In order to clone the Bbr_0160 promoter region, a DNA fragment encompassing the intergenic region between the Bbr_0159-0160 genes was generated by PCR amplification from *B. breve* UCC2003 chromosomal DNA using PFUII DNA polymerase and the primer pair, nanP1F and nanP1R. The PCR product was digested with HindIII and XbaI, and ligated to the similarly digested pBC1.2 (Álvarez-Martín *et al.*, 2007). The ligation mixture was introduced into *E. coli* XL1-blue by electrotransformation and transformants were selected based on Tet and Cm resistance. Transformants were checked for plasmid content by restriction analysis and the integrity of positively identified clones was verified by sequencing. Plasmid pBC1.2-nanP1 was introduced into *B. breve* UCC2003 by electrotransformation and transformants were selected based on Cm resistance.

Induced protein production. 25 ml of LB broth was inoculated with a 2 % inoculum of *E. coli* EC101 harboring either pQE-60-nanR or the empty vector, pQE-60 (used as a negative control), followed by incubation at 37°C with agitation until an OD_{600nm} of 0.5 was reached, at which point protein expression was induced by the addition of 1 mM IPTG, followed by further incubation for 2 h. Cells were harvested by centrifugation, resuspended in Tris-HCl (pH 8.0) and disrupted with glass beads in a mini-bead beater (BioSpec Products, Bartlesville, OK). Cellular debris was removed by centrifugation to produce a crude cell extract.

Electrophoretic mobility shift assays (EMSAs). DNA fragments representing the promoter regions upstream of the Bbr_0160 and *nanB* genes were prepared by PCR using Extensor DNA polymerase and IRD-labelled primer pairs (Table 3.3). EMSAs were essentially performed as described previously (Hamoen *et al.*, 1998). In all cases, the binding reactions were performed in a final reaction volume of 20 µl in the presence of poly (dI-dC) in binding buffer (25 mM HEPES, 50 mM sodium glutamate, 50 mM KCl, 1 mM phenylmethanesulfonyl fluoride and 10 % glycerol at pH 7.0). Various amounts of crude cell extract containing pQE-60-nanR or pQE-60 were mixed on ice with a fixed amount of DNA probe (0.1 pmol) and subsequently incubated for 30 min at 37°C. To determine the effector molecule of NanR, sialic acid or *N*-acetylmannosamine was added to the binding reaction in concentrations ranging from 1 mM to 10 mM. Following addition of sialic acid to the binding buffer, the pH was re-adjusted to 7.0. Samples were loaded on a 6 % non-denaturing polyacrylamide (PAA) gel prepared in TAE buffer (40 mM Tris acetate [pH 8.0], 2 mM EDTA) and run in a 0.5 to 2.0 x gradient of TAE at 100 V for 2 h in an Atto Mini PAGE system (Atto Bioscience and Biotechnology, Tokyo, Japan). Signals were detected using an Odyssey Infrared Imaging System (Li-Cor Biosciences, United Kingdom Ltd., Cambridge, United Kingdom) and images were captured using the supplied Odyssey software v3.0.

Primer extension analysis. Total RNA was isolated from exponentially growing cells of *B. breve* UCC2003 and *B. breve* UCC2003-pBC1.2-nanP1 in mMRS supplemented with 0.5 % sialic acid. Primer extension was performed by annealing 1 pmol of an IRD-labelled synthetic oligonucleotide to 20 µg of RNA as previously

described (Ventura *et al.*, 2005), using the primers 0160-R1Rev and nanB-B1Rev (Table 3.3). Sequence ladders of the presumed Bbr_0160 and *nanB* promoter regions were produced using the same primer as in the primer extension reaction and a DNA cycle-sequencing kit (Jena Bioscience, Germany) and were run alongside the primer extension products to allow precise alignment of the transcriptional start site with the corresponding DNA sequence. Separation was achieved on a 6.5 % Li-Cor Matrix KB Plus acrylamide gel. Signal detection and image capture were performed with a Li-Cor sequencing instrument (Li-Cor Biosciences).

RT-PCR. 10 µg of RNA isolated from log phase *B. breve* UCC2003 cells grown in mMRS with 0.5 % sialic acid were treated with DNase (Roche) according to the manufacturer's instructions and used as a template in a 100 µl reaction mixture containing 20 ng of random primers, 0.5 mM of each deoxyribonucleoside triphosphate and Superscript III reverse transcriptase (Invitrogen, Paisley, United Kingdom) to produce cDNA. The generated cDNA was then used as a template for reverse transcription (RT)-PCRs using Extensor DNA polymerase and the primers listed in Table 3.2 in a 20 µl PCR reaction. PCRs were performed under the following conditions: initial denaturation at 95°C for 2 min was followed by 30 cycles of denaturation at 95°C for 30 s, primer annealing at 52°C for 30 s and elongation at 68°C for 30 s. The PCR reactions were terminated with a final elongation at 68°C for 10 min.

Microarray data accession number. The microarray data obtained in this study have been deposited in NCBI's Gene Expression Omnibus database and are accessible through GEO series accession number GSE62885.

3.4 RESULTS AND DISCUSSION.

Identification of a GntR-type regulator in B. breve UCC2003. We previously demonstrated that transcription of the *nan/nag* cluster (corresponding to locus tags Bbr_0160 through to Bbr_0172), as well as the *nagA2* and *nagB3* genes, corresponding to Bbr_1247 and Bbr_1248, are induced in the presence of sialic acid, indicative of the presence of carbohydrate-dependent transcriptional regulation (Egan *et al.*, 2014). Bbr_0173, designated here as *nanR* and located immediately downstream of the *nan/nag* cluster, specifies a 283 amino acid protein (~31 kDa) that belongs to the FadR subfamily of the GntR-type transcriptional regulators (Pfam family: PF00392) (Finn *et al.*, 2010; Rigali *et al.*, 2002). BLASTP analysis revealed that the NanR protein is highly conserved among other sequenced members of the *B. breve* species, perhaps unsurprisingly given that the *nan/nag* cluster is also highly conserved within the species (Egan *et al.*, 2014). NanR also shares high identity with predicted GntR-type regulators in *Bifidobacterium longum* subsp. *infantis* ATCC15697 and *Bifidobacterium dentium* ATCC27679. Analysis of the genes flanking these predicted GntR-type regulators in *B. longum* subsp. *infantis* ATCC15697 and *B. dentium* ATCC27679 revealed the presence of genes predicted to be involved in the metabolism of sialic acid, suggesting that all of these identified bifidobacterial GntR-type regulators are associated with the control of sialic acid metabolism. The previously characterised NanR protein from *E. coli*, also from the FadR subfamily (Kalivoda *et al.*, 2003; Rigali *et al.*, 2002), shares 43 % similarity with the *B. breve* UCC2003 NanR protein. This can be considered significant ($P < 0.001$) considering that *Bifidobacterium* species and *E. coli* represent very different bacteria, on both a phylogenetic and phenotypic level. Multiple sequence alignment of the aforementioned NanR proteins, as well as the FadR sequence from *E. coli* is

presented in Figure 3.1. The winged helix-turn-helix (HTH) DNA-binding domain of NanR (57 amino acids) is similar in size to that of the average DNA-binding domain of GntR-type regulators (62 amino acids) (Hoskisson & Rigali, 2009). Within this domain some of the residues found to be involved in DNA-binding of FadR could also be identified in *B. breve* UCC2003 (Fig. 3.1) (van Aalten *et al.*, 2000; Xu *et al.*, 2001). The deduced ligand binding domain of NanR (124 amino acids) appears to be smaller than the average length of other C-terminal domains of members of the FadR subfamily (150-170 amino acids) (Rigali *et al.*, 2002).

***Transcriptome of B. breve UCC2003-nanR*.** In order to determine if the protein product of *nanR* is responsible for the transcriptional regulation of the *nan/nag* cluster, and/or *nagA2* and *nagB3*, an insertion mutant was constructed and the transcriptome was analysed. It was hypothesised that if NanR was to act as a repressor, mutation of the corresponding gene should lead to increased transcription of the genes it controls, even in the absence of sialic acid. Microarray data revealed that, as compared to *B. breve* UCC2003, the *nan*/*nag* locus (with the exception of Bbr_0172), was significantly up-regulated (>2.5 fold change; P <0.001) in the mutant strain (Table 3.4). These data are consistent with our prediction that *nanR* acts as a transcriptional repressor of the *nan/nag* cluster. Interestingly, the *nagA2* and *nagB3* genes were not up-regulated in *B. breve* UCC2003-nanR. As speculated in Chapter VI this thesis, these genes may also be involved in the degradation of other *N*-acetylglucosamine and/or *N*-acetylgalactosamine-containing oligosaccharides, and may therefore be under the control of a separate and as yet unknown regulatory system.

B. breve UCC2003-nanR was also investigated for the ability to utilise sialic acid. As expected, the mutant did not display impaired growth on sialic acid as it attained final OD_{600nm} values comparable to the wild type strain (Fig. 3.2). This is consistent with the notion of NanR being a transcriptional repressor.

Electrophoretic mobility shift assays. In order to determine if the NanR protein directly interacts with operator sequences within the *nan/nag* cluster, the *nanR* gene product was overexpressed in *E. coli* EC101 using the pQE-60 protein overexpression system. The recombinant protein NanR-His lacked binding ability (data not shown) and it was hypothesised that the His-tag interferes with the folding or binding of the protein. For this reason, *nanR* was cloned into the pQE-60 vector, heterologously expressed without a His-tag and crude cell extracts were used to perform EMSAs. Unfortunately, due to the apparent intrinsic instability of the protein, it was not possible to produce enough active protein to attain a full shift of the DNA fragment. Further extensive efforts were made to improve the activity of hetorologously expressed NanR, including cloning the gene into *Lactococcus lactis* NZ9000 with an N-terminal or C-terminal His-tag as previously described for other bifidobacterial transcriptional regulators (O'Connell *et al.*, 2014; O'Connell Motherway *et al.*, 2011a; Pokusaeva *et al.*, 2010) and cloning the gene under the constitutive p44 promoter in *B. breve* UCC2003 (Alvarez-Martin *et al.*, 2012), yet only a partial mobility shift of the target DNA fragment could ever be achieved. However, the results obtained with crude cell extract overexpressing NanR demonstrate that NanR specifically binds to DNA fragments encompassing the intergenic regions upstream of Bbr_0160 and Bbr_0164 (*nanB*) (Fig. 3.3A and

3.3B). NanR failed to bind to the fragments encompassing the intergenic regions upstream of *nanC, nanA* or *nanH* (data not shown). Crude cell extracts of *E. coli* EC101 containing empty pQE-60 was used as a negative control and, as expected, did not alter the electrophoretic behaviour of the DNA fragments (Fig. 3.3B).

Dissection of the Bbr_0160 promoter region showed that NanR binding required a 211 bp region, within which a 21 bp imperfect inverted repeat was identified (5'-ATTAGACATC>A<GACGTCTGAT-3') Similarly, dissection of the *nanB* promoter region showed that NanR binding required a 196 bp region which also includes a 21 bp imperfect inverted repeat (5'-ATCAGACATC>A<GATGTCATAT-3'), and which is similar to the imperfect inverted repeat identified upstream of Bbr_0160 (Fig. 3.3A, 3.3B). Further fragmentation of both promoter regions to the point where the inverted repeats were excluded resulted in the absence of binding, consistent with the notion that these inverted repeats indeed represent operator sequences of NanR (Fig. 3.3A, 3.3B). This conserved inverted repeat was also identified upstream of homologues of the Bbr_0160 and *nanB* genes in of all the currently available complete *B. breve* genomes (Bottacini *et al.*, 2014) and the genome of *B. longum* subsp. *infantis* ATCC15697 (Sela *et al.*, 2008), although it could not be identified at the expected locations on the genome of *B. dentium* ATCC27679.

Operator binding sequences of GntR-type regulators can be in the form of inverted or direct repeats (Rigali *et al.*, 2002). A consensus palindromic operator sequence for members of the GntR-type family has been described, comprising 5'-$(N)_yGT(N)_xAC(N)_y$-3', with the number of residues (y) and the nature of the central residues (x) varying (Rigali *et al.*, 2002). The operator sequence of NanR as deduced

in this study includes the conserved GT and AC residues, although in the opposite order as in the described consensus sequence.

Sialic acid is the effector molecule for the NanR repressor from *E. coli* (Kalivoda *et al.*, 2003), whereas in *Vibrio vulnificus* and *Staphylococcus aureus*, RpiR-type regulators control transcription of sialic acid-utilising gene clusters and the effector molecule is *N*-acetylmannosamine, an intermediate of the sialic acid breakdown pathway (Hwang *et al.*, 2013; Kim *et al.*, 2011; Olson *et al.*, 2013; Plumbridge & Vimr, 1999). At a concentration of 1 mM sialic acid, a significant reduction of binding of NanR to the *nanB* promoter region was observed, whereas *N*-acetylmannosamine had no effect on the binding ability of NanR. A similar effect was observed for the Bbr_0160 promoter region (Fig. 3.3C).

Identification of the transcription start sites of Bbr_0160 and nanB. In order to determine the transcription start site of the presumed promoters upstream of Bbr_0160 and *nanB*, primer extension analysis was performed using RNA extracted from *B. breve* UCC2003 and *B. breve* UCC2003-pBC1.2-nanP1 grown in mMRS containing 0.5 % sialic acid. For the Bbr_0160 promoter region, initial attempts to attain a primer extension product from mRNA isolated from *B. breve* UCC2003 wild type cells during growth on sialic acid were unsuccessful, therefore to increase the amount of mRNA transcripts of this promoter region, a DNA fragment encompassing the promoter region was cloned into pBC1.2 and introduced into *B. breve* UCC2003. Single extension products were identified upstream of the Bbr_0160 and *nanB* genes, and in both cases potential promoter recognition sequences resembling consensus -10 and -35 hexamers could be identified upstream

of the transcription start sites (Fig. 3.4A). Hence, these two promoters were designated *nanP1* and *nanP2*. In both cases, the predicted NanR operator sequence overlaps the -35 sequence, a location consistent with the characterisation of NanR as a transcriptional repressor (Fig. 3.4A). Binding of NanR is expected to interfere with the interaction of RNA polymerase with the -10 and -35 promoter sequences, thus preventing transcription of the operon.

RT-PCRs. In order to determine if the *nan/nag* locus exists as a single or multiple transcriptional unit, transcriptional analysis by RT-PCR was performed using cDNA templates derived from mRNA transcripts isolated from *B. breve* UCC2003 during growth on sialic acid. The obtained cDNA was used as a template in various PCR reactions that targeted the intergenic regions between each gene of the *nan/nag* cluster (NB. some genes of the cluster overlap or are separated by a single base pair and were therefore not assessed). As expected, when chromosomal DNA isolated from *B. breve* UCC2003 was used as a positive control, PCR products were obtained for each intergenic region. Using cDNA as a template, PCR products were obtained for each intergenic region, with the exception of that upstream of Bbr_0160, being the first gene of the *nan/nag* cluster (Fig. 3.4B). Using these results, in combination with data obtained from EMSAs and primer extension analysis, we deduce that the *nan/nag* cluster consists of two transcriptional units. The first transcriptional unit, driven by the *nanP1* promoter upstream of the Bbr_0160 gene, overlaps with the second transcriptional unit, which is driven by the *nanP2* promoter upstream of the *nanB* gene. Bbr_0173, the gene encoding NanR, is oriented in the opposite direction and is therefore not co-transcribed with the *nan/nag* gene cluster. Accordingly, a

predicted *rho*-independent terminator sequence was identified downstream of the Bbr_0172 gene.

At over fifteen thousand base pairs, the sialic acid utilisation cluster is the largest carbohydrate metabolic cluster to be characterised from *B. breve* UCC2003, followed by the galactan utilisation cluster at almost ten thousand base pairs (O'Connell Motherway *et al.*, 2011a). That such a large gene cluster can be driven by just two promoters reflects the tightly organised nature of the gene cluster, particularly between the Bbr_0160-0169 genes. Interestingly, despite the presence of a pseudogene (Bbr_0170) between Bbr_0169 (*nagB1*) and Bbr_0171 (*nanH*), RT-PCR and EMSA results indicated the absence of a promoter in this region, thus indicating that either or both of the *nanP1* and *nanP2*-driven transcripts terminates at the terminator sequence (mentioned above) downstrean of Bbr_0172.

3.5 CONCLUDING REMARKS

The current study investigated the regulation of sialic acid metabolism in *B. breve* UCC2003 by NanR, a GntR-type transcriptional repressor. To our knowledge, NanR is the first GntR-family regulator characterised from bifidobacteria, and thus represents a diversion from the customary LacI-mediated regulation of carbohydrate metabolism in this genus (Lee & O'Sullivan, 2010; O'Connell *et al.*, 2014; O'Connell Motherway *et al.*, 2011a; Pokusaeva *et al.*, 2010; Pokusaeva *et al.*, 2011; Schell *et al.*, 2002). Although GntR-type regulators are commonly found in the *Actinobacteria* phylum (Hoskisson & Rigali, 2009), it was previously noted that the number of such regulators on the genome of *B. longum* subsp. *longum* NCC2705 is rather low compared to other bacterial genomes (Schell *et al.*, 2002). The genome of *B. longum* subsp. *longum* NCC2705 encodes just one predicted GntR-family transcriptional regulator (Schell *et al.*, 2002), while the *B. breve* UCC2003 genome is predicted to encode three members of this regulator family, including NanR (O'Connell Motherway *et al.*, 2011b). We identified homologues of NanR in all currently available complete *B. breve* genome sequences, as well as *B. longum* subsp. *infantis* ATCC15697 and *B. dentium* ATCC27679. Each of the NanR-encoding genes was located adjacent to genes predicted to be involved in the uptake and metabolism of sialic acid.

The *nan/nag* cluster was previously identified as required for the uptake and metabolism of sialic acid (Egan *et al.*, 2014). The data assembled in this study provides significant information on the remarkably efficient method by which *B. breve* UCC2003 regulates this particularly large gene cluster. It was previously hypothesised that negative regulation allows bifidobacteria to react faster to

fluctuations in the carbohydrate availability in the large intestine (Schell *et al.*, 2002), and the NanR-mediated control of the *nan/nag* cluster represents another example of such adaptation.

3.6 ACKNOWLEDGEMENTS

The Alimentary Pharmabiotic Centre is a research centre funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan. The authors and their work were supported by SFI (Grant Nos. 07/CE/B1368 and SFI/12/RC/2273) and a HRB postdoctoral fellowship (Grant No. PDTM/20011/9) awarded to MOCM.

3.7 TABLES AND FIGURES

Table 3.1: Bacterial strains and plasmids used in this study.

Strains and plasmids	Relevant features	Reference or source
Strains		
***Escherichia coli* strains**		
***E.coli* EC101**	Cloning host; *repA*$^+$ *kmr*	(Law *et al.*, 1995)
***E.coli* EC101-pNZ-M.Bbrll+Bbr11**	EC101 harboring a pNZ8048 derivative containing bbrllM and bbrlllM	(O'Connell Motherway *et al.*, 2009)
***E. coli* EC101-pQE-60**	EC101 containing pQE-60	This study
***E. coli* EC101-pQE-60-nanR**	EC101 containing pQE-60-nanR	This study
***E. coli* XL1-blue**	(*supE44 hsdR17 recA1 gyrA96 thi relA1 lac* F′ [*proAB*$^+$ *lacI*^{q} *lacZΔM15* Tn*10*(Tet^r)])	Stratagene
***E. coli* XL1-blue-pBC1.2-nanP1**	XL1-blue containing pBC1.2-nanP1	This study
***Bifidobacterium* sp. strains**		
***B. breve* UCC2003**	Isolate from a nursling stool	(Maze *et al.*, 2007)
***B. breve* UCC2003-nanR**	pORI19-tetW-nanR insertion mutant of UCC2003	This study
***B. breve* UCC2003-pBC1.2-nanP1**	*B. breve* UCC2003 containing pBC1.2-nanP1	This study
Plasmids		
pORI19	Em^r, repA$^-$, ori$^+$, cloning vector	(Law *et al.*, 1995)
pORI19-tetW-nanR	Internal 347 bp fragment of *nanR* and *tetW* cloned in pORI19	This study
pQE-60	Amp^r, IPTG inducible vector	Qiagen
pQE-60-nanR	*nanR* cloned into the IPTG inducible vector, pQE-60	This study
pBC1.2	pBC1-pSC101-Cm^r	(Álvarez-Martín *et al.*, 2007)
pBC1.2-nanP1	Bbr_0160 promoter region cloned in pBC1.2	This study

Table 3.2: Oligonucleotide primers used in this study.

Purpose	Primer	Sequence
Cloning of 347 bp fragment of *nanR* in pORI19	NanRIMF	GCTATT*AAGCTT*GAAGCGCTCAACATCGTCAAAG
	NanRIMR	GATCTA*TCTAGA*GAACGGACCATCTGCTTGACG
Amplification of *tetW*	TetWF	TCAGCT*GTCGAC*ATGCTCATGTACGGTAAG
	TetWR	GCGACG*GTCGAC*CATTACCTTCTGAAACATA
Confirmation of site-specific homologous recombination	NanRconfirm	CGCTGTGACGAGACGATGG
Cloning of *nanR* in pQE-60	NanRFORWARD	GACTAG*ACATGT*GCAACCTATCCGACTCG
	NanRREVERSE	GCATGC*AGATCT*TCACTCGGTACCTCCCTGAATG
Cloning of nanP1 in pBC1.2	nanP1F	TCGATC*AAGCTT*CGAGCTGCCTGATACGCC
	nanP1R	TCAGCT*TCTAGA*CTCGCGCAGCTGCTTCATG
RT-PCRs	0160FOR-RT-PCR	CGAGCTGCCTGATACGCC
	0160REV-RT-PCR	CTCGCGCAGCTGCTTCATG
	0162FOR-RT-PCR	CCTCATCGACCCGGACGTG
	0162REV-RT-PCR	CGCTCACCTGGCCCTTG
	0163FOR-RT-PCR	GTCGCATCCACACGCCTG
	0163REV-RT-PCR	GAATGGATGTCGGAGTTGAGC
	0164FOR-RT-PCR	CGGTTACGCCCATACTGC
	0164REV-RT-PCR	CTCCCTTGGCGTTCGTCG
	0165FOR-RT-PCR	CCACCGCCGTCAAGAAGAAC
	0165REV-RT-PCR	GACGCCGAATGCCACAAAC
	0168FOR-RT-PCR	CATTGTGGAGCGCGGAGC
	0168REV-RT-PCR	GCGTCGATCATGCGGTTG
	0169FOR-RT-PCR	CACCAACCAGATGCCCAAC
	0169REV-RT-PCR	GAGTCCCAGCACCGCGTTC
	0171FOR-RT-PCR	CTCGTGCTGCTCGCGTTC
	0171REV-RT-PCR	GCGCACTGTTCCATGTCTG
	0172FOR-RT-PCR	CATTGGCCTGTCTTGGGAG
	0172REV-RT-PCR	GCCATCATCGTCTTGCCTAC

Restriction sites incorporated into oligonucleotide primer sequences are indicated in italics.

Table 3.3: IRD-labelled primers used to generate PCR products and primer extension products in this study.

Primer	Sequence
0160-R1For	CGAGCTGCCTGATACGCC
0160-R1Rev	CTCGATCCACGGCTGCTC
0160-R2For	CGATCACACACACACCGTTG
0160-R2Rev	CTCGCGCAGCTGCTTCATG
0160-R3For	CAAGGAGGAAACCAGCATGC
0160-R3Rev	GATCTCGTCCTCAATCGGCAG
nanB-B1For	CATACTGCACCGCTGGGC
nanB-B1Rev	CTCCCTTGGCGTTCGTCG
nanB-B2For	GTTTGAGACCGTTGCATTGC
nanB-B2Rev	GGACGTGGACAGCGGATTG
nanB-B3For	GACCCGAATGGCCAAGACC
nanB-B3Rev	GACGCGCTCGAACGAGGTC

Table 3.4: Transcriptome analysis of *B. breve* UCC2003-nanR compared to *B. breve* UCC2003 grown on 0.5 % ribose

Locus tag (gene name)	Predicted Function	Level of upregulation
Bbr_0160	Conserved hypothetical protein	6.58
Bbr_0161 (*nanK*)	ROK family kinase	3.08
Bbr_0162 (*nanE*)	*N*-acetylmannosamine-6-phosphate epimerase	2.74
Bbr_0163	Hydrolase	4.66
Bbr_0164 (*nanB*)	Substrate binding protein	14.27
Bbr_0165 (*nanC*)	ABC transport system permease protein	8.21
Bbr_0166 (*nanD*)	ABC transport system ATP-binding protein	3.46
Bbr_0167 (*nanF*)	ABC transport system ATP-binding protein	3.27
Bbr_0168 (*nanA*)	*N*-acetylneuraminate lyase	6.85
Bbr_0169 (*nagB1*)	Glucosamine-6-phosphate deaminase	4.83
Bbr_0171 (*nanH*)	Sialidase A	4.25

The cut-off point is 2.5-fold with a *P* value of <0.001.

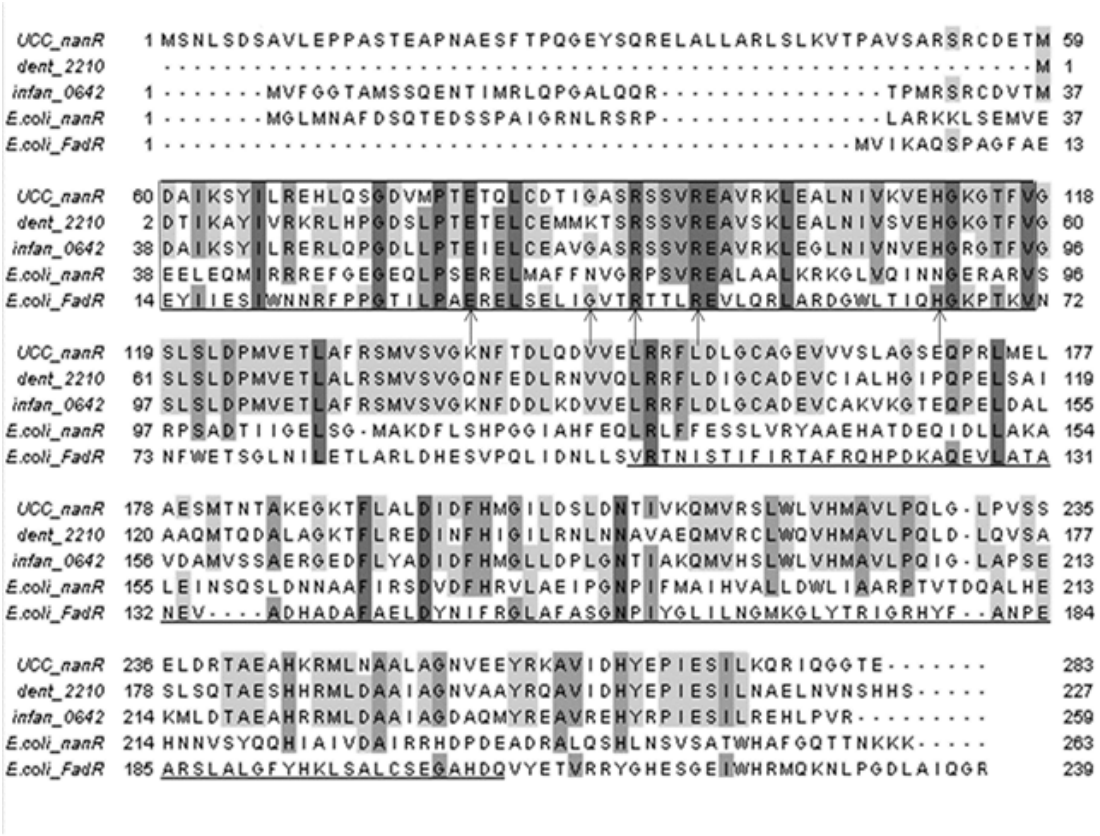


Figure 3.1: Multiple sequence alignment of the NanR protein from *B. breve* UCC2003 compared to predicted GntR-type regulators from *B. dentium* ATCC27679 and *B. longum* subsp. *infantis* ATCC15697, the previously characterised NanR protein from *E. coli* (Kalivoda *et al.*, 2003) and the FadR transcriptional regulator from *E. coli* (Xu *et al.*, 2001) using CLUSTAL X (Goujon *et al.*, 2010). Conserved residues are shaded based on conservation in 60 % or more of the sequences used. The predicted N-terminal DNA binding domain of NanR is indicated by a black box. Arrows below the sequence point at residues previously shown to be involved in DNA-binding of FadR that are conserved in *B. breve* sequences (van Aalten *et al.*, 2000; Xu *et al.*, 2001). The predicted C-terminal domain of NanR is underlined.

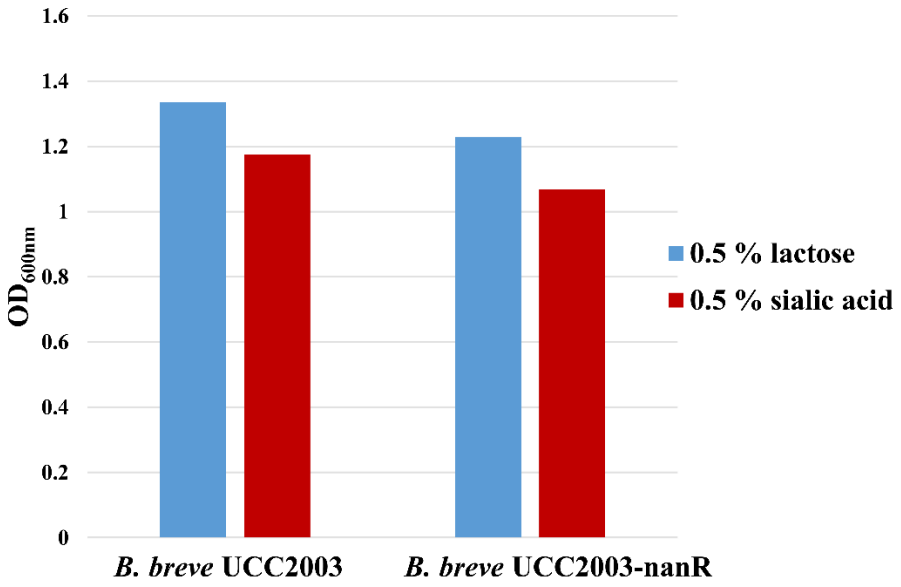


Figure 3.2: Final OD_{600nm} values obtained following 24 h growth of *B. breve* UCC2003 and *B. breve* UCC2003-nanR in mMRS containing 0.5 % (wt/vol) lactose or sialic acid as the sole carbon source. The results are the mean values obtained from two different experiments.

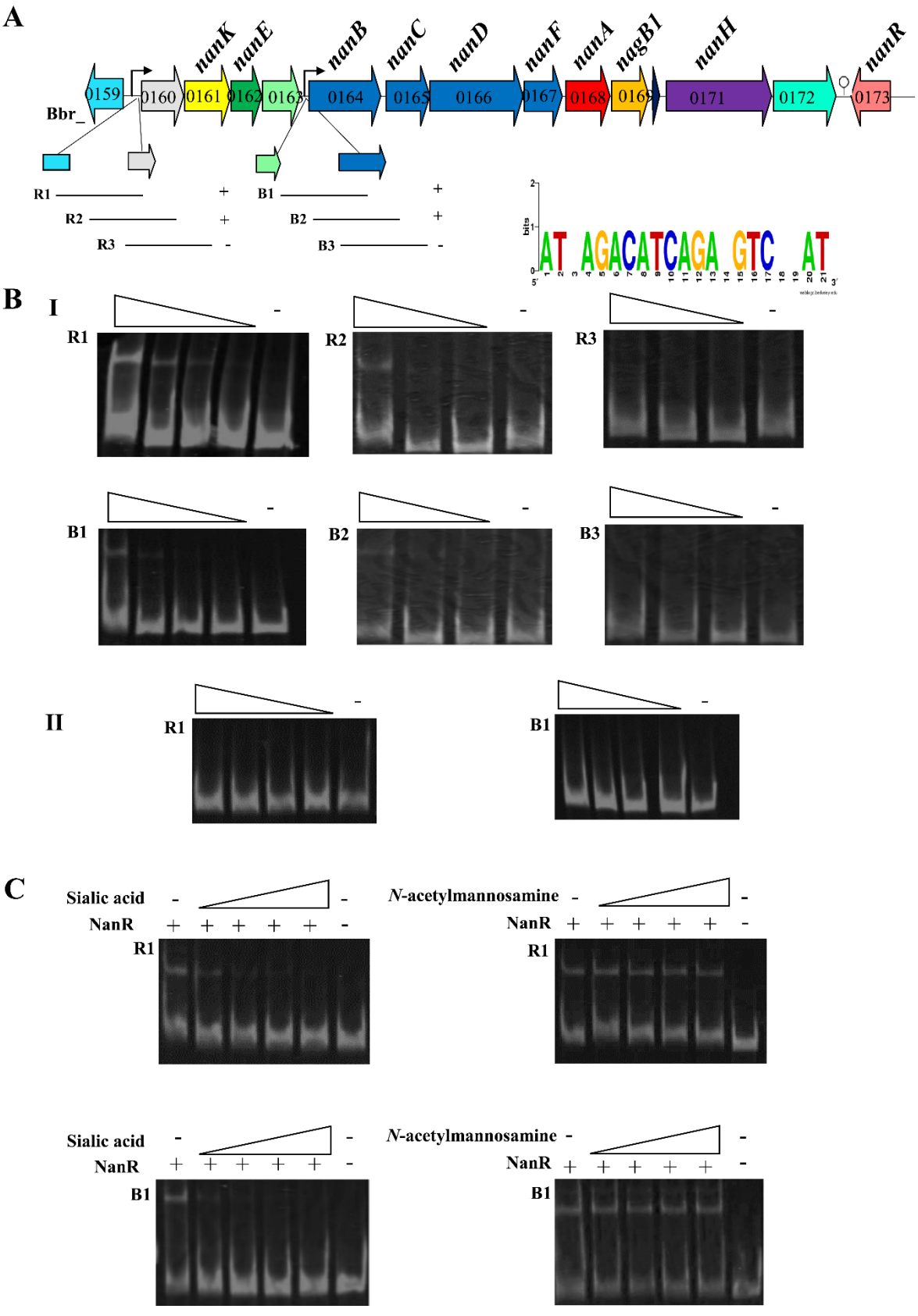

A
nanK
nanE
nanB
nanC
nanD
nanF
nanA
nagB1
nanH
nanR
Bbr_
0159
0160
0161
0162
0163
0164
0165
0166
0167
0168
0169
0171
0172
0173
R1
R2
R3
B1
B2
B3
bits
B
I
II
C
Sialic acid
NanR
N-acetylmannosamine

Figure 3.3: (A) Schematic representation of the sialic acid utilisation cluster of *B. breve* UCC2003 and DNA fragments used in electrophoretic mobility shift assays (EMSAs) for the Bbr_0160 and *nanB* promoter regions together with Weblogo representation of the predicted operator of NanR. Plus or minus signs indicate ability or inability of NanR to bind to the DNA fragment. The bent arrows represent the proven promoter regions and the lollipop sign represents a putative *rho*-independent terminator region. **(B)** EMSAs showing the interactions of (I) crude cell extract containing pQE-60-NanR with the DNA fragments R1, R2 and R3 and B1, B2 and B3 and (II) the interactions of crude cell extract containing pQE-60 (negative control) with the DNA fragments R1 and B1. The minus symbol indicates reactions to which no crude cell extract was added, while the remaining lanes represent binding reactions with the respective DNA probes incubated with increasing amounts of crude cell extract. Each successive lane from right to left represents a doubling of the amount of crude cell extract. **(C)** EMSAs showing NanR interaction with the R1 and B1 DNA fragments with the addition of sialic acid or *N*-acetylmannosamine in concentrations ranging from 1 mM to 10 mM.

A

I

CACCGTTGAGGTGCCCTCGGAAGGCGGAGTAGATGTCCGCAGCGGCGGA

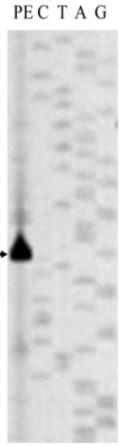
AATCCGCCCTCGGAAAGCCAAGTGGCGGTCGAAGCGCGGGCCGATGGAA

CACGCTCGTTGCGCGGCGTTAATAGAT**TAGACA**TCAGACGTCTGATAAT

AG**TAGAAT**GGAAATGCTCAGGCACCGGCATTTCAAGGAGGAAACCAGC

ATG

-35, -10, *, RBS



II

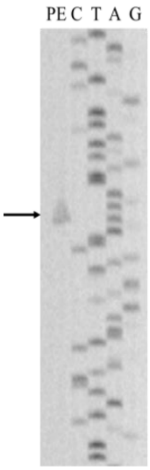
TACCGCGGTTTGAGACCGTTGCATTGCAAGGAAAAAACGTGGTAGCTGCA

CTCATAAAACGTCTGATGAAAC**TTGCAA**TCAGACATCAGATGTCA**TATGA**

TTAAGAAAGATTCGATTGGATGGAGAATCGAATCCTGATTCAGTTATCAC

TCAACGAAGAGGAGAAACCGATG

-35, -10, *, RBS



B

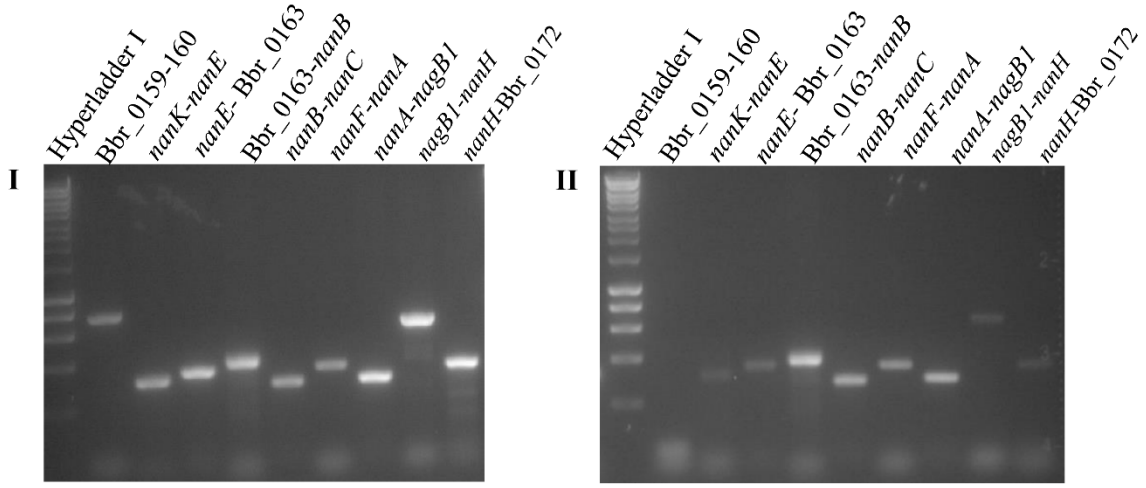


Figure 3.4: (A) Schematic representation of the (I) Bbr_0160 and (II) *nanB* promoter regions. Boldface type and underlining indicate -10 and -35 hexamers (as deduced from the primer extension results) and ribosomal binding site (RBS); the transcriptional start site is indicated by an asterisk. The arrows under sequences indicate the inverted repeat that represents the putative NanR-binding sequence. The arrows in the right panels indicate the primer extension products. **(B)** PCR products obtained using (I) *B. breve* UCC2003 chromosomal DNA and (II) cDNA obtained from mRNA isolated from log-phase *B. breve* UCC2003 cells during growth in sialic acid using oligonucleotide primers spanning the intergenic regions between the genes of the *nan/nag* cluster.

3.8 REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410

Álvarez-Martín, P., O'Connell-Motherway, M., van Sinderen, D. & Mayo, B. (2007). Functional analysis of the pBC1 replicon from *Bifidobacterium catenulatum* L48. *Appl Microbiol Biotechnol* **76**, 1395-1402.

Alvarez-Martin, P., O'Connell Motherway, M., Turroni, F., Foroni, E., Ventura, M. & van Sinderen, D. (2012). A two-component regulatory system controls autoregulated serpin expression in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **78**, 7032-7041.

Blancato, V. S., Repizo, G. D., Suárez, C. A. & Magni, C. (2008). Transcriptional regulation of the citrate gene cluster of *Enterococcus faecalis* involves the GntR family transcriptional activator CitO. *J Bacteriol* **190**, 7419-7430.

Bordoni, A., Amaretti, A., Leonardi, A., Boschetti, E., Danesi, F., Matteuzzi, D., Roncaglia, L., Raimondi, S. & Rossi, M. (2013). Cholesterol-lowering probiotics: in vitro selection and in vivo testing of bifidobacteria. *Appl Microbiol Biotechnol* **97**, 8273-8281.

Bottacini, F., O'Connell Motherway, M., Kuczynski, J., OConnell, K., Serafini, F., Duranti, S., Milani, C., Turroni, F., Lugli, G. & other authors (2014). Comparative genomics of the *Bifidobacterium breve* taxon. *BMC Genomics* **15**, 170.

Casali, N., White, A. M. & Riley, L. W. (2006). Regulation of the *Mycobacterium tuberculosis mce1* operon. *J Bacteriol* **188**, 441-449.

Cohn, M. & Horibata, K. (1959). Inhibition by glucose of the induced synthesis of the beta-galactoside-enzyme system of *Escherichia coli*. Analysis of maintenance. *J Bacteriol* **78**, 601-612.

Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Res* **14**, 1188-1190.

De Man, J. C., Rogosa, M. & Sharpe, M. E. (1960). A medium for the cultivation of lactobacilli. *J Appl Bacteriol* **23**, 130-135.

Egan, M., O'Connell Motherway, M., Ventura, M. & van Sinderen, D. (2014). Metabolism of sialic acid by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **80**, 4414-4426.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G. & other authors (2010). The Pfam protein families database. *Nucleic Acids Res* **38**, D211-D222.

Fujita, Y. & Fujita, T. (1987). The gluconate operon gnt of *Bacillus subtilis* encodes its own transcriptional negative regulator. *Proc Natl Acad Sci U S A* **84**, 4524-4528.

García de la Nava, J., Santaella, D. F., Alba, J. C., Carazo, J. M., Trelles, O. & Pascual-Montano, A. (2003). Engene: the processing and exploratory analysis of gene expression data. *Bioinformatics* **19**, 657-658.

Gorke, B. & Stulke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Micro* **6**, 613-624.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res* **38**, W695-W699.

Hamoen, L. W., Van Werkhoven, A. F., Bijlsma, J. J. E., Dubnau, D. & Venema, G. (1998). The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev* **12**, 1539-1550.

Haydon, D. J. & Guest, J. R. (1991). A new family of bacterial regulatory proteins. *FEMS Microbiol Lett* **79**, 291-296.

Hoskisson, P. A. & Rigali, S. (2009). Variation in form and function: the helix-turn-helix regulators of the GntR superfamily. *Adv Appl Microbiol* **69**, 1-22.

Hwang, J., Kim, B. S., Jang, S. Y., Lim, J. G., You, D. J., Jung, H. S., Oh, T. K., Lee, J. O., Choi, S. H. & other authors (2013). Structural insights into the regulation of sialic acid catabolism by the *Vibrio vulnificus* transcriptional repressor NanR. *Proc Natl Acad Sci U S A* **110**, E2829-E2837.

Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**, 318-356.

Jaques, S. & McCarter, L. L. (2006). Three new regulators of swarming in *Vibrio parahaemolyticus*. *J Bacteriol* **188**, 2625-2635.

Kalivoda, K. A., Steenbergen, S. M., Vimr, E. R. & Plumbridge, J. (2003). Regulation of sialic acid catabolism by the DNA binding protein NanR in *Escherichia coli*. *J Bacteriol* **185**, 4806-4815.

Kim, B. S., Hwang, J., Kim, M. H. & Choi, S. H. (2011). Cooperative regulation of the *Vibrio vulnificus nan* gene cluster by NanR protein, cAMP receptor protein, and *N*-acetylmannosamine 6-phosphate. *J Biol Chem* **286**, 40889-40899.

Law, J., Buist, G., Haandrikman, A., Kok, J., Venema, G. & Leenhouts, K. (1995). A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. *J Bacteriol* **177**, 7011-7018.

LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D. & Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr Opin Biotechnol* **24**, 160-168.

Lee, J. H. & O'Sullivan, D. (2010). Genomic insights into bifidobacteria. *Microbiol Mol Biol Rev* **74**, 378-416.

Lewis, M. (2005). The lac repressor. *Comptes rendus biologies* **328**, 521-548.

Long, A. D., Mangalam, H. J., Chan, B. Y. P., Tolleri, L., Hatfield, G. W. & Baldi, P. (2001). Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework: Analysis of global gene expression in *Escherichia coli* K12. *J Biol Chem* **276**, 19937-19944.

Matthews, K. S. & Nichols, J. C. (1998). Lactose repressor protein: functional properties and structure. *Prog Nucleic Acid Res* **58**, 127-164.

Maze, A., O'Connell Motherway, M., Fitzgerald, G., Deutscher, J. & van Sinderen, D. (2007). Identification and characterization of a fructose phosphotransferase system in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **73**, 545-553.

Milani, C., Lugli, G. A., Duranti, S., Turroni, F., Bottacini, F., Mangifesta, M., Sanchez, B., Viappiani, A., Mancabelli, L. & other authors (2014). Genome encyclopaedia of type strains of the genus *Bifidobacterium*. *Appl Environ Microbiol* **80**, 6290-6302.

O'Connell, K. J., O'Connell Motherway, M., Liedtke, A., Fitzgerald, G. F., Ross, R. P., Stanton, C., Zomer, A. & van Sinderen, D. (2014). Transcription of two adjacent carbohydrate utilization gene clusters in *Bifidobacterium breve* UCC2003 is controlled by LacI- and Repressor Open reading frame Kinase (ROK)-type regulators. *Appl Environ Microbiol* **80**, 3604-3614.

O'Connell Motherway, M., O'Driscoll, J., Fitzgerald, G. F. & Van Sinderen, D. (2009). Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **2**, 321-332.

O'Connell Motherway, M., Fitzgerald, G. F. & van Sinderen, D. (2011a). Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **4**, 403-416.

O'Connell Motherway, M., Zomer, A., Leahy, S. C., Reunanen, J., Bottacini, F., Claesson, M. J., O'Brien, F., Flynn, K., Casey, P. G. & other authors (2011b). Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc Natl Acad Sci USA* **108**, 11217-11222.

Olson, M. E., King, J. M., Yahr, T. L. & Horswill, A. R. (2013). Sialic acid catabolism in *Staphylococcus aureus*. *J Bacteriol* **195**, 1779-1788.

Parche, S., Beleut, M., Rezzonico, E., Jacobs, D., Arigoni, F., Titgemeyer, F. & Jankovic, I. (2006). Lactose-over-glucose preference in *Bifidobacterium longum* NCC2705: *glcP*, encoding a glucose transporter, is subject to lactose repression. *J Bacteriol* **188**, 1260-1265.

Plumbridge, J. & Vimr, E. (1999). Convergent pathways for utilization of the amino sugars *N*-acetylglucosamine, *N*-acetylmannosamine, and *N*-acetylneuraminic acid by *Escherichia coli*. *J Bacteriol* **181**, 47-54.

Pokusaeva, K., Neves, A. R., Zomer, A., O'Connell Motherway, M., MacSharry, J., Curley, P., Fitzgerald, G. F. & van Sinderen, D. (2010). Ribose utilization by the human commensal *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **3**, 311-323.

Pokusaeva, K., O'Connell-Motherway, M., Zomer, A., MacSharry, J., Fitzgerald, G. F. & van Sinderen, D. (2011). Cellodextrin utilization by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **77**, 1681-1690.

Rigali, S., Derouaux, A., Giannotta, F. & Dusart, J. (2002). Subdivision of the Helix-Turn-Helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem* **277**, 12507-12515.

Riordan, O. (1998). *Studies on antimicrobial activity and genetic diversity of Bifidobacterium species: molecular characterization of a 5.75 kb plasmid and a chromosomally encoded recA gene homologue from Bifidobacterium breve.*, National University of Ireland, Cork, Cork.

Round, J. L. & Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313-323.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Ryan, S., Fitzgerald, G. & van Sinderen, D. (2005). Transcriptional regulation and characterization of a novel β-fructofuranosidase-encoding gene from *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **71**, 3475-3482.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning a laboratory manual*, 2nd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Schell, M., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M. C., Desiere, F., Bork, P. & other authors (2002). The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci U S A* **99**, 14422-14427.

Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., Lapidus, A., Rokhsar, D. S., Lebrilla, C. B. & other authors (2008). The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A* **105**, 18964-18969.

Servin, A. L. (2004). Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiol Rev* **28**, 405-440.

Stanton, C., Ross, R. P., Fitzgerald, G. F. & Sinderen, D. V. (2005). Fermented functional foods based on probiotics and their biogenic metabolites. *Curr Opin Biotechnol* **16**, 198-203.

Tissier, H. (1900). *Recherchers sur la flora intestinale normale pathologique du nourisson.,* The University of Paris, Paris.

Trindade, M. I., Abratt, V. R. & Reid, S. J. (2003). Induction of sucrose utilization genes from *Bifidobacterium lactis* by sucrose and raffinose. *Appl Environ Microbiol* **69**, 24-32.

Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., Kerr, C., Hourihane, J., Murray, D. & other authors (2012). Diversity of bifidobacteria within the infant gut microbiota. *PLoS ONE* **7**, e36957.

van Aalten, D. M. F., DiRusso, C. C., Knudsen, J. & Wierenga, R. K. (2000). Crystal structure of FadR, a fatty acid-responsive transcription factor with a novel acyl coenzyme A-binding fold. *EMBO J* **19**, 5167-5177.

van Hijum, S. A. F. T., Garcia De La Nava, J., Trelles, O., Kok, J. & Kuipers, O. P. (2003). MicroPreP: a cDNA microarray data pre-processing framework. *Appl Bioinformatics*, 241-244.

van Hijum, S., de Jong, A., Baerends, R., Karsens, H., Kramer, N., Larsen, R., den Hengst, C., Albers, C., Kok, J. & other authors (2005). A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC Genomics* **6**, 77.

Ventura, M., Zink, R., Fitzgerald, G. F. & van Sinderen, D. (2005). Gene structure and transcriptional organization of the *dnaK* operon of *Bifidobacterium*

breve UCC2003 and application of the operon in bifidobacterial tracing. *Appl Environ Microbiol* **71**, 487-500.

Ventura, M., Canchaya, C., Fitzgerald, G., Gupta, R. & Sinderen, D. (2007). Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria. *Antonie van Leeuwenhoek* **92**, 265-265.

Ventura, M., Turroni, F., Motherway, M. O. C., MacSharry, J. & van Sinderen, D. (2012). Host–microbe interactions that facilitate gut colonization by commensal bifidobacteria. *Trends Microbiol* **20**, 467-476.

Wilson, C. J., Zhan, H., Swint-Kruse, L. & Matthews, K. S. (2007). The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. *Cell Mol Life Sci* **64**, 3-16.

Xu, Y., Heath, R. J., Li, Z., Rock, C. O. & White, S. W. (2001). The FadR·DNA complex: Transcriptional control of fatty acid metabolism in *Escherichia coli*. *J Biol Chem* **276**, 17373-17379.

Zomer, A., Fernandez, M., Kearney, B., Fitzgerald, G. F., Ventura, M. & van Sinderen, D. (2009). An interactive regulatory network controls stress response in *Bifidobacterium breve* UCC2003. *J Bacteriol* **191**, 7039-7049.

CHAPTER III

Cross-feeding by *Bifidobacterium breve* UCC2003 during co-cultivation with *Bifidobacterium bifidum* PRL2010 in a mucin-based medium.

Dr. Mary O' Connell Motherway constructed the plasmid-containing derivatives of *Bifidobacterium breve* UCC2003 and *Bifidobacterium bifidum* PLR2010.

Dr. Michelle Kilcoyne performed quantitative HPAEC-PAD and HPLC analyses.

This chapter was published in:

Egan, M., O' Connell Motherway, M., Kilcoyne, M., Kane, M., Joshi, L., Ventura, M. & van Sinderen, D. (2014). Cross-feeding by *Bifidobacterium breve* UCC2003 during co-cultivation with *Bifidobacterium bifidum* PRL2010 in a mucin-based medium**.** *BMC Microbiol* **14,** 282**.**

4.1 ABSTRACT

Bifidobacteria constitute a specific group of commensal bacteria that commonly inhabit the mammalian gastrointestinal tract. *Bifidobacterium breve* UCC2003 is known to utilise a variety of plant/diet/host-derived carbohydrates, including cellodextrins, starch and galactan, as well as the mucin and HMO-derived, monosaccharide, sialic acid. In the current study, we investigated the ability of this strain to utilise parts of a host-derived source of carbohydrate, namely the mucin glycoprotein, when grown in co-culture with the mucin-degrading *Bifidobacterium bifidum* PRL2010. *B. breve* UCC2003 was shown to exhibit growth properties in a mucin-based medium, but only when grown in the presence of *B. bifidum* PRL2010, which is known to metabolise mucin. A combination of HPAEC-PAD and transcriptome analyses identified some of the possible monosaccharides and oligosaccharides which support this enhanced co-cultivation growth/viability phenotype. This study provides evidence of a gut commensal relationship between two bifidobacterial species, thus advancing our knowledge on the metabolic adaptability which allows the former strain to colonise the (infant) gut by its extensive metabolic abilities to (co-)utilise available carbohydrate sources.

4.2 INTRODUCTION

Bifidobacteria are Gram positive, anaerobic, Y-shaped bacteria that have been found in the gastrointestinal tract (GIT) of mammals, birds and insects and have also been isolated from the human oral cavity and sewage (Ventura *et al.*, 2007). In recent years, bifidobacteria have attracted attention due to the purported health benefits associated with their presence in the gut, including development of the immune system (Round & Mazmanian, 2009), provision of vitamins (LeBlanc *et al.*, 2013), and protection against pathogenic bacteria (Servin, 2004). Bifidobacteria rapidly colonise the infant gut in the first days and weeks of life. In a recent study on the microbial diversity of the infant gut, it was found that members of the *Actinobacteria* phylum were dominant, with various bifidobacterial species present in high abundance, in particular *Bifidobacterium longum*, *Bifidobacterium bifidum*, *Bifidobacterium breve* and *Bifidobacterium catenulatum* (Turroni *et al.*, 2012b).

Survival and growth of bifidobacteria in the gastrointestinal tract requires them to employ an arsenal of enzymes to metabolise the complex carbohydrates prevalent in this environment (Pokusaeva *et al.*, 2011a). *B. breve* UCC2003 is well-adapted to utilise various diet/plant-derived oligo- and poly-saccharides, including melezitose, raffinose, cellodextrins, GOS, starch and galactan (O'Connell *et al.*, 2013; O'Connell Motherway *et al.*, 2011a; O'Connell Motherway *et al.*, 2013; O'Connell Motherway *et al.*, 2008; Pokusaeva *et al.*, 2011b; Ryan *et al.*, 2006). *B. breve* UCC2003 also metabolises the mucin- and human milk oligosaccharide (HMO)-derived monosaccharide sialic acid (Egan *et al.*, 2014), which is more consistent with this strain's origin as a nursling stool isolate from a breast-fed infant, where it would be expected to metabolise HMOs and/or the structurally similar oligosaccharides found in mucin glycoproteins. Along with dietary components, host-derived

oligosaccharides are believed to form part of the nutrient resource for certain intestinal bacteria.

The original investigations on the (bio)chemical composition of human colonic mucin described twenty-one discrete oligosaccharide structures (Podolsky, 1985). Since then, it has been estimated that carbohydrate constitutes approximately 80 % of the total mucin mass, and that MUC2, the prominent secretory mucin in the colon, may contain more than 100 different *O*-linked glycans (Larsson *et al.*, 2009; Tytgat *et al.*, 1994). Glycosylation of the peptide backbone can be *N*-linked to an asparagine residue or *O*-linked to serine or threonine residues via *N*-acetylgalactosamine (GalNAc). Subsequent elongation of this structure results in a number of distinct core structures. While eight mucin type core structures are known, only four regularly occur in human mucins. Core 1 is formed by the addition of galactose (Gal) in a β1-3 linkage to GalNAc to produce galacto-*N*-biose (GNB, also known as the T antigen). Core 3, the most common core structure in the human colon, is formed by the addition of *N*-acetylglucosamine (GlcNAc) in a β1-3 linkage to GalNAc. Core 2 and core 4 structures are formed by the addition of GlcNAc in a β1-6 linkage to core 1 and core 3, respectively. Each of the core structures can be further elongated by the addition of Gal, GalNAc and GlcNAc. The oligosaccharide chains can also be substituted with sialic acid, fucose (Fuc) or sulfate residues in terminal or branched positions (Capon *et al.*, 2001; Larsson *et al.*, 2009; Podolsky, 1985; Robbe *et al.*, 2004).

Due to the high complexity and variability of the mucin oligosaccharide chains, only a small proportion of the culturable intestinal microbiota is believed to encode enzymes required for (partial) degradation of mucin into free sugars. These include

members of the *Bifidobacterium, Bacteroides* and *Ruminococcus* genera, as well as a more recently characterised bacterium isolated from human faeces, *Akkermansia muciniphila* (Crost *et al.*, 2013; Derrien *et al.*, 2004; Sonnenburg *et al.*, 2005; Turroni *et al.*, 2010b). As mentioned previously, *B. bifidum* is one of the most abundant species in the infant intestine (Turroni *et al.*, 2012b). A study in 2010 revealed that some 60 % of the glycosyl hydrolases identified on the genome of *B. bifidum* PRL2010 are linked to the degradation of mucin, many of which are predicted to be extracellular (Turroni *et al.*, 2010a; Turroni *et al.*, 2010b). Comparative genome hybridisation analysis revealed that most of these genes encoding these glycosyl hydrolases are conserved within the examined members of the *B. bifidum* species (Turroni *et al.*, 2010a; Turroni *et al.*, 2010b).

It has been suggested that the degradation of mucin by a small number of extracellular glycosidase-producing bacteria may provide nutritional support to other enteric bacteria (Hoskins *et al.*, 1985). It has been shown that a number of HMO-derived degradation products remain in the media during vegetative growth of *B. bifidum* (Asakuma *et al.*, 2011). *B. breve* UCC2003 can cross-feed on sialic acid derived from the metabolism of 3' sialyllactose, an abundant HMO, by *B. bifidum* PRL2010 (Egan *et al.*, 2014). The aim of the current study was to establish if *B. breve* UCC2003 is able to cross-feed on the oligosaccharides released by *B. bifidum* PRL2010 extracellular activity on mucin, and secondly, to investigate which, if any, particular components of mucin *B. breve* UCC2003 utilises under such circumstances.

4.3 MATERIALS AND METHODS

***Bacterial strains, plasmids, media and culture conditions*.** Bacterial strains and plasmids used in this study are listed in Table 4.1. *B. breve* UCC2003 and its derivatives were routinely cultured in Reinforced Clostridial Medium (RCM; Oxoid Ltd, Basingstoke, Hampshire, United Kingdom). *B. bifidum* PRL2010 was routinely cultured in modified deMan Rogosa Sharpe (mMRS) medium made from first principles (but excluding a carbohydrate source) (De Man *et al.*, 1960), and supplemented with 0.05 % (wt/vol) L-cysteine HCl and 1 % (wt/vol) lactose (unless otherwise stated). All carbohydrates used in this study were purchased from Sigma Aldrich and were of the highest purity available. To prepare mucin-containing media, a 0.8 % (wt/vol) concentration of mucin from porcine stomach (Type III) was prepared in water and autoclaved at 115°C for 10 minutes, in order to optimise mucin dissolution, yet minimising glycosidic hydrolysis. This was added to an equal volume of twice-concentrated mMRS, resulting in a final concentration of 0.4 % mucin in mMRS. Bifidobacterial cultures were incubated under anaerobic conditions in a modular atmosphere-controlled system (Davidson and Hardy, Belfast, Ireland) at 37°C. *Escherichia coli* was cultured in Luria Bertani (LB) broth at 37 °C with agitation (Sambrook, 1989). Where appropriate, growth media contained tetracycline (Tet; 10 μg ml^{-1}), chloramphenicol (Cm; 5 μg ml^{-1} for *E. coli*, 2.5 μg ml^{-1} for *B. bifidum*), erythromycin (Em; 100 μg ml^{-1}) or kanamycin (Kan; 50 μg ml^{-1}). Recombinant *E. coli* cells containing pORI19 were selected on LB agar containing Em and Kan, and supplemented with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) (40 μg ml^{-1}) and 1 mM IPTG (isopropyl-β-D-galactopyranoside).

Nucleotide sequence analysis. Sequence data were obtained from the Artemis-mediated genome annotations of *B. breve* UCC2003 (O'Connell Motherway *et al.*, 2011b; Rutherford *et al.*, 2000). Database searches were performed using non-redundant sequences available at the National Centre for Biotechnology Information internet site (http://www.ncbi.nlm.nih.gov) using BLAST (Altschul *et al.*, 1990). Sequence analysis was performed using the Seqbuilder and Seqman programs of the DNASTAR software package (DNASTAR, Madison, WI, USA).

DNA manipulations. Chromosomal DNA was isolated from *B. breve* UCC2003 as previously described (Riordan, 1998). Plasmid DNA was isolated from *E. coli*, *B. breve* and *B. bifidum* using the Roche High Pure plasmid isolation kit (Roche Diagnostics, Basel, Switzerland). An initial lysis step was performed using 30 mg ml^{-1} of lysozyme for 30 min at 37°C prior to plasmid isolation from bifidobacteria. Single stranded oligonucleotide primers used in this study were synthesised by Eurofins (Ebersberg, Germany) (Table 4.2). Standard PCRs were performed using Taq PCR master mix (Qiagen GmbH, Hilden, Germany). *B. breve* colony PCRs were carried out as described previously (O'Connell Motherway *et al.*, 2009). PCR products were purified using the Roche High Pure PCR purification kit (Roche Diagnostics). Electroporation of plasmid DNA into *E. coli* or bifidobacteria was performed as described previously (Maze *et al.*, 2007; Sambrook, 1989).

Construction of B. breve UCC2003 insertion mutants. An internal fragment of Bbr_1742, designated here as *fucP* (400 base pairs (bp), representing codon numbers 137 through to 270 out of the 421 codons of *fucP*), Bbr_1587, designated here as

lnbP (479 bp, representing codon numbers 108 through to 267 of the 751 codons of *lnbP*) and Bbr_1833, designated here as *lacZ7* (568 bp, representing codon numbers 123 through 312 of the 699 codons of *lacZ7*), were amplified by PCR using *B. breve* UCC2003 chromosomal DNA as a template and primer pair FucPF and FucPR, LnbPF and LnbPR and LacZ7F and LacZ7R, respectively. The insertion mutants were constructed as described previously (O'Connell Motherway *et al.*, 2009). Site-specific recombination in potential Tet-resistant mutant isolates was confirmed by colony PCR using primer combinations TetWF and TetWR to confirm *tetW* gene integration, and the primers FucPconfirm, LnbPconfirm and LacZ7confirm (positioned upstream of the selected internal fragment of *fucP*, *lnbP* and *lacZ7*, respectively) and TetWF to confirm integration at the correct chromosomal location.

***Evaluation of B. breve UCC2003 growth on mucin*.** Growth of *B. breve* UCC2003 using mucin as the sole carbon source was determined both independently and in co-culture with *B. bifidum* PRL2010. Plasmid-containing derivatives of the wild type strains were used, namely *B. breve* UCC2003-pAM5, which contains the pAM5 plasmid and is therefore Tet resistant (Álvarez-Martín *et al.*, 2007) and *B. bifidum* PRL2010-pPKCM7, which contains the pPKCM7 plasmid and is Cm resistant (Cronin *et al.*, 2007). Use of plasmid-containing derivatives allowed for selection and enumeration of either *B. breve* UCC2003 or *B. bifidum* PRL2010 colonies on RCA containing the appropriate antibiotic. All *B. breve* UCC2003 mutant strains were also Tet resistant. A 0.001 % inoculum of a stationary phase culture of *B. breve* UCC2003 strains (see Table 1; *B. breve* UCC2003-pAM5, or the mutant strains, *B. breve* UCC2003-nanA (Egan *et al.*, 2014), *B. breve* UCC2003-fucP, *B. breve* UCC2003-lnbP, *B. breve* UCC2003-lacZ7 or *B. breve* UCC2003-galT (Ruiz *et al.*,

2013)) and/or a 0.01 % inoculum of *B. bifidum* PRL2010-pPKCM7, were added to mMRS medium supplemented with 0.05 % (wt/vol) L-cysteine HCl and 0.4 % (wt/vol) mucin (see above). Growth of the cultures was measured over 72 h, with samples taken every 6 or 12 h. All samples collected were serially 10-fold diluted in sterile Ringers solution and plated onto RCA supplemented with 1 % (wt/vol) lactose and the appropriate antibiotic. Viable counts were determined by counting colonies on agar plates using dilutions that yielded between 30 and 300 colony forming units (CFU).

Transcriptome analysis using B. breve UCC2003-based microarrays. *B. breve* UCC2003-pAM5 was grown in mMRS medium supplemented with 0.05 % (wt/vol) L-cysteine HCl and 0.4 % ribose to an OD_{600nm} of 0.5 and then harvested by centrifugation at 9,000 x *g* for 2 min at room temperature. *B. breve* UCC2003-pAM5 grown in mucin in co-culture with *B. bifidum* PRL2010-pPKCM7 (see above) was similarly harvested after 30 h of growth. DNA microarrays containing oligonucleotide primers representing each of the 1864 annotated genes on the genome of *B. breve* UCC2003 were designed by and obtained from Agilent Technologies (Palo Alto, CA, USA). Methods for cell disruption, RNA isolation, RNA quality control, complementary DNA (cDNA) synthesis and labelling were performed as described previously (Zomer *et al.*, 2009). Labelled cDNA was hybridised using the Agilent Gene Expression hybridisation kit (part number 5188-5242) as described in the Agilent Two-Color Microarray-Based Gene Expression Analysis (v4.0) manual (publication number G4140-90050). Following hybridisation, microarrays were washed as described in the manual and scanned using Agilent's DNA microarray scanner G2565A. The scanning results were

converted to data files with Agilent's Feature Extraction software (version 9.5). DNA-microarray data were processed as previously described (García de la Nava *et al.*, 2003; van Hijum *et al.*, 2005; van Hijum *et al.*, 2003). Differential expression tests were performed with the Cyber-T implementation of a variant of the *t*-test (Long *et al.*, 2001).

***Analysis of the monosaccharide composition of fermented and non-fermented mucin*.** Identification and quantification of the free monosaccharides in non-fermented mMRS supplemented with 0.4 % mucin and the media following 30 h fermentation by *B. bifidum* PRL2010-pPKCM7 was performed according to Dionex technical note 40 (http://www.dionex.com/en-us/webdocs/5052-TN40-IC-Glycoprotein-Monosaccharide-23May2012-LPN1632-01.pdf) and as previously described (Kilcoyne *et al.*, 2009), using high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD). In brief, media was pelleted, sterilised through a 0.2 μm filter and then lyophilised to dryness. The lyophilised powder was dissolved at 1 mg ml^{-1} in purified water. Dilutions were injected on to a Dionex ICS 3000 system (Dionex, Sunnyvale, CA) equipped with a CarboPac PA20 analytical column (150 mm x 3 mm) with an Amino-trap column (30 mm x 3 mm) and separated as previously described (Kilcoyne *et al.*, 2012). Resulting peaks were identified and quantified by comparison to a standard curve with a mixture containing the common mammalian residues Fuc, glucosamine (GlcN), galactosamine (GalN), Gal, glucose (Glc) and mannose (Man). Samples were injected three times and the average value is reported for the concentration. Samples were then spiked with a known concentration of Fuc and Gal standards and re-injected to confirm the identification of these residues in the sample.

To confirm the identity of Fuc in spent medium, samples were labelled with 2-aminobenzamide (2-AB) according to the published method (Bigge *et al.*, 1995). The 2-AB labelled samples were cleaned on Glycoclean S cartridges (Prozyme) according to manufacturer's instructions and vacuum centrifuged dry. Samples were analysed by reverse phase-high performance liquid chromatography (RP-HPLC) injected onto a Waters Alliance 2695 instrument and separated on a Phenomenex Luna 3u C18(2) (150 mm x 4.6 mm) column using previously described conditions (Hardy, 1997).

During co-culture experiments samples of cell free supernatants were taken every 6 or 12 h to qualitatively analyse the carbohydrate composition of the media by HPAEC-PAD. Samples (25 µl aliquots) were separated on a CarboPac PA1 analytical-exchange column (250 mm x 4 mm) with a CarboPac PA1 guard column (50 mm x 4 mm). Elution was performed at a constant flow-rate of 1.0 ml min^{-1} at 30 °C using the following eluents for the analysis (A) 200 mM NaOH, (B) 100 mM NaOH, 550 mM sodium acetate, and (C) purified water. The following linear gradient of sodium acetate was used with 100 mM NaOH: from 0 to 50 min, 0 mM; from 50 to 51 min, 16 mM; from 51 to 56 min, 100 mM; from 56 to 61 min, 0 mM. Eluate was monitored with a Dionex ED40 detector in the PAD mode. The chromatogram of non-fermented mucin was used to evaluate mucin utilisation by *B. bifidum* PRL2010-pPKCM7 and *B. breve* UCC2003-pAM5, and the Chromeleon software v. 6.70 (Dionex Corporation) was used for the integration and evaluation of the chromatograms obtained.

***Microarray data accession number*.** The microarray data obtained in this study have been deposited in NCBI's Gene Expression Omnibus database and are accessible through GEO series accession number GSE59013.

4.4 RESULTS.

Growth of B. breve UCC2003 on mucin. Growth of *B. breve* UCC2003-pAM5 and *B. bifidum* PRL2010-pPKCM7 in mMRS supplemented with 0.4 % (wt/vol) mucin, independently or in co-culture, was measured by viable plate counts over 72 h. Monitoring of growth in co-culture required the use of plasmid-containing derivatives of the wild type strains. The plasmids used, pAM5 for *B. breve* UCC2003 or pPKCM7 for *B. bifidum* PRL2010, conferred Tet or Cm resistance to the respective strains, thus allowing selection and enumeration of each strain by viable plate count on RCA containing the corresponding antibiotic. Both *B. breve* UCC2003-pAM5 and *B. bifidum* PRL2010-pPKCM7 were also cultured in mMRS supplemented with 0.5 % lactose, a substrate on which both strains achieve a high level of growth (Turroni *et al.*, 2012a; Watson *et al.*, 2013), and it was found that presence of the plasmid did not impair growth of the strain on this substrate (results not shown). A low inoculum of 0.001 % and/or 0.01 % of *B. breve* UCC2003-pAM5 and *B. bifidum* PRL2010-pPKCM7, respectively, was used to allow the strains to undergo multiple generations of growth. In the absence of an added carbohydrate, no growth was observed for either *B. breve* UCC2003-pAM5 or *B. bifidum* PRL2010-pPKCM7 in mMRS medium (results not shown). As expected, the positive control, *B. bifidum* PRL2010-pPKCM7, was capable of growth in mucin-containing mMRS medium, from a low inoculum of 10^4 CFU ml^{-1} it attained cell numbers of approximately 10^8 CFU ml^{-1} following 30 h of incubation (Turroni *et al.*, 2010b) (Fig. 4.1). However, a notably different growth profile was observed for *B. breve* UCC2003-pAM5 on this medium. During the first 12 h of incubation in this medium, *B. breve* UCC2003-pAM5 viable cell numbers increased from 10^4 CFU ml^{-1} to almost 10^7 CFU ml^{-1}, followed by a decline in viable cells to below 10^4 CFU ml^{-}

[1] (Fig. 4.1). HPAEC-PAD analysis (see below) revealed the presence of contaminating monosaccharides, namely Glc and Gal, in the non-fermented mucin preparation. It is unknown whether these carbohydrates were released from partial degradation of the mucin during autoclaving, despite efforts made to reduce such degradation (see above), or were present as contaminating monosaccharides in the mucin preparation. Given that *B. breve* UCC2003 can utilise both of these contaminants as a sole carbon source (Ruiz *et al.*, 2013; Watson *et al.*, 2013), it is hypothesised that they are responsible for the observed growth during the first 12 h following inoculation. Once these carbon sources were utilised, and as *B. breve* UCC2003-pAM5 apparently does not have the required enzymes to degrade intact mucin, a substantial drop in viable count was observed. Interestingly, when *B. bifidum* PRL2010-pPKCM7 was included as a potential carbohydrate-releasing bacterium in co-culture with *B. breve* UCC2003-pAM5, an improvement in growth and survival abilities of *B. breve* UCC2003-pAM5 was observed. In co-culture, *B. breve* UCC2003-pAM5 reached cell numbers of almost 10^8 CFU ml^{-1} after 24 h and maintained a viable count of over 10^6 CFU ml^{-1} after 72 h (Fig. 4.1). This suggests that *B. breve* UCC2003-pAM5 is cross-feeding on (some of) the sugars released following mucin breakdown by *B. bifidum* PRL2010-pPKCM7. The cell numbers of *B. bifidum* PRL2010-pPKCM7 were moderately lower in co-culture as compared to when the strain was grown independently in this medium and this was attributed to the presence of another acid-producing in the culture. However, the growth profile of *B. bifidum* PRL2010-pPKCM7 in co-culture followed a similar trend to that of the strain growing individually, suggesting that the two strains do not compete for limiting amounts of mucin-derived carbohydrates.

***Gene expression analysis using B. breve UCC2003 DNA microarrays*.** In order to investigate how the transcriptome of *B. breve* UCC2003-pAM5 is affected by growth in co-culture with *B. bifidum* PRL2010-pPKCM7 in mucin-containing mMRS, global gene expression was determined by microarray analysis during growth of the strain in co-culture and compared with gene expression when grown on ribose as the sole carbon source. Ribose was considered an appropriate carbohydrate for comparative transcriptome analysis as the genes involved in ribose metabolism are unique to this sugar (Pokusaeva *et al.*, 2010), thus making it an ideal substrate for gene expression analysis on other carbohydrates. It has previously been used as a reference condition in several transcriptome studies in *B. breve* UCC2003 (Egan *et al.*, 2014; O'Connell *et al.*, 2013; O'Connell *et al.*, 2014; O'Connell Motherway *et al.*, 2011a; O'Connell Motherway *et al.*, 2013). Furthermore, as a pentose sugar, ribose is suitable for gene expression analysis in mucin, given that the monosaccharide components of mucin are all hexose sugars (Glc, Gal, GalNAc, GlcNAc or Fuc), except sialic acid which is a nine-carbon monosaccharide. In the current study, gene expression was determined following 30 h growth in co-culture. This was to ensure that any observed changes in the transcriptome of *B. breve* UCC2003-pAM5 could be attributed solely to growth in co-culture with *B. bifidum* PRL2010-pPKCM7, as opposed to growth on free monosaccharides in the mucin preparation which appeared to be responsible for the initial 12 h of growth of *B. breve* UCC2003-pAM5 in mucin (see above). Analysis of DNA microarray data showed that a number of stress-related genes were up-regulated, which was attributed to the cells being in co-culture with another bacterium or to the possibility that the cells had entered the stationary phase. Nonetheless, the majority of genes that were shown to be significantly up-regulated (fold change >2.5, $P<0.001$) in co-

culture are predicted to be involved in the transport and metabolism of carbohydrates (Table 4.3). One such cluster, Bbr_1740-1742, is believed to be involved in Fuc metabolism and includes genes predicted to encode a dihydrodipicolinate synthase (Bbr_1740), an enzyme which was previously shown to be involved in Fuc metabolism in *Campylobacter jejuni* (Stahl *et al.*, 2011), a hypothetical protein (Bbr_1741) and a Fuc permease (*fucP*, Bbr_1742). Transcription of a predicted β-galactosidase-encoding gene (Bbr_1833, designated *lacZ7*) was also significantly increased, this enzyme is possibly involved in the removal of Gal from mucin oligosaccharides, suggesting that *B. breve* UCC2003 may utilise one or more Gal-containing fractions of mucin. The Bbr_1585-1590 gene cluster was also up-regulated during growth in co-culture. Bbr_1585 encodes a predicted UDP-Glc-4-epimerase, designated *galE*. Previously, UDP-Glc-4-epimerase enzymes have been shown to be involved in the Leloir pathway for Gal metabolism (Frey, 1996), as well as the GNB/LNB pathway for the metabolism of (ga)lacto-*N*-biose (GNB/LNB) derived from HMO or mucin (Kitaoka *et al.*, 2005; Nishimoto & Kitaoka, 2007). Bbr_1586 encodes a predicted *N*-acetylhexosamine-1-kinase, designated *nahK*. In *B. longum* JCM 1217, an *N*-acetylhexosamine-1-kinase, which shares 90 % identity to NahK, was shown to have a role in the previously mentioned GNB/LNB pathway (Nishimoto & Kitaoka, 2007). Bbr_1587 encodes a predicted LNB phosphorylase, designated *lnbP*, an enzyme which is also involved in the GNB/LNB pathway (Derensy-Dron *et al.*, 1999; Kitaoka *et al.*, 2005; Nishimoto & Kitaoka, 2007). The LnbP protein from *B. breve* UCC2003 shares 90 % identity with the previously characterised LnbP protein from *B. bifidum* JCM 1254 (Kitaoka *et al.*, 2005). Finally, as regards to this upregulated cluster, Bbr_1588-1590 encode a predicted ABC transport system, including two predicted permease proteins (represented by

Bbr_1588 and Bbr_1589), and a solute binding protein (Bbr_1590), which shares 98 % identity with the GNB/LNB-specific binding protein of *B. longum* JCM1217 (Suzuki *et al.*, 2008). Bbr_1884 (designated *galT2*), encoding a predicted Gal-1-phosphate uridyltransferase, another enzyme required for the Leloir pathway and the GNB/LNB pathway (Frey, 1996; Kitaoka *et al.*, 2005; Nishimoto & Kitaoka, 2007), was also up-regulated when *B. breve* UCC2003 was grown in mucin in co-culture with *B. bifidum* PRL2010. Finally, two gene clusters responsible for the transport and metabolism of sialic acid were significantly up-regulated under these growth conditions (Bbr_0160-0171 and Bbr_1247-1248) (Egan *et al.*, 2014) (Table 4.3).

***Analysis of the monosaccharide composition of fermented and non-fermented mucin*.** HPAEC-PAD was used to quantitatively analyse the carbohydrate profile of non-fermented mucin, compared to the cell free supernatant (CFS) of *B. bifidum* PRL2010-pPKCM7 grown in mucin for 30 h. Non-fermented mucin was shown to contain 0.13 nmol mg^{-1} of free Gal and 3.96 nmol mg^{-1} of free Glc, both of which support growth of *B. breve* UCC2003 (Ruiz *et al.*, 2013; Watson *et al.*, 2013), thus presenting a plausible explanation for the initial increase in *B. breve* UCC2003-pAM5 cell numbers during the first 12 h of incubation in mucin-containing medium (Table 4.4). However, after 30 h growth of *B. bifidum* PRL2010-pPKCM7, the carbohydrate profile of the medium was significantly altered. Glc was no longer detected, indicating its uptake by *B. bifidum* PRL2010-pPKCM7, but the amount of free Gal increased to 51.73 nmol mg^{-1}, while 34.94 nmol mg^{-1} of free Fuc was also detected, indicative of the extracellular glycosidase activity of *B. bifidum* PRL2010-pPKCM7 (Table 4.4). The presence of Fuc was further verified by fluorescent labelling of the CFS samples and analysis by RP-HPLC (Fig. 4.2A). RP-HPLC

analysis also confirmed that free Fuc was absent in non-fermented mMRS supplemented with 0.4 % mucin, which indicated that its presence was a result of *B. bifidum* PRL2010-pPKCM7 extracellular activity (Fig. 4.2A). To investigate whether these released monosaccharides from mucin support the observed growth and viability of *B. breve* UCC2003-pAM5 in co-culture, the carbohydrate profile of the CFS of the co-culture was compared to that of *B. bifidum* PRL2010-pPKCM7 grown independently in mucin (hence any difference observed could be attributed to the presence of *B. breve* UCC2003-pAM5). The results support the hypothesis that *B. breve* UCC2003-pAM5 is cross-feeding on at least one carbohydrate released following degradation of mucin by *B. bifidum* PRL2010-pPKCM7. After 30 h Fuc was present in the carbohydrate profile of *B. bifidum* PRL2010-pPKCM7, but absent (or at least below the detection level) in that of the co-culture, indicative of its uptake and utilisation by *B. breve* UCC2003-pAM5 (Fig. 4.2B).

Growth of B. breve UCC2003 insertion mutants in co-culture in mucin with B. bifidum PRL2010-pPKCM7. HPAEC-PAD and transcriptome data suggested that the improved growth and viability of *B. breve* UCC2003-pAM5 in co-culture was a result of the strain cross-feeding on Fuc, while microarray data suggested sialic acid, Gal, GNB and/or another of the Gal-containing constituents of mucin may also be utilised. To determine which, if any, of these sugars was primarily responsible for growth of the strain in co-culture, insertion mutants were constructed in the *fucP, lnbP* and *lacZ7* genes, resulting in strains *B. breve* UCC2003-fucP, *B. breve* UCC2003-lnbP and *B. breve* UCC2003-lacZ7. Two additional mutants, *B. breve* UCC2003-galT, which harbours a transposon in a predicted Gal-1-phosphate uridyltransferase-encoding gene, designated *galT1* and was shown to be incapable of

growth in Gal (Ruiz *et al.*, 2013), and the previously described insertion mutant, *B. breve* UCC2003-nanA, which cannot utilise sialic acid (Egan *et al.*, 2014), were also tested. The insertion mutants were individually analysed for their ability to grow in mucin in the presence or absence of *B. bifidum* PRL2010-pPKCM7. In the absence of *B. bifidum* PRL2010-pPKCM7, the mutant strains behaved similarly to *B. breve* UCC2003-pAM5, as initial growth over the first 12 h was followed by a decline in viable cells to 10^4 CFU ml^{-1} or lower at 72 h (Fig. 4.3). It was also found that removing the ability of the strain to utilise a particular sugar did not significantly affect its ability to grow in co-culture with *B. bifidum* PRL2010-pPKCM7. Similar to *B. breve* UCC2003-pAM5, all mutant strains attained a viable count of between 10^7 and 10^8 CFU ml^{-1} between 12 h and 30 h, and maintained a viable count of 10^6 CFU ml^{-1} until 72 h, with the exception of *B. breve* UCC2003-fucP which dropped to 10^5 CFU ml^{-1} (Fig. 4.3). The carbohydrate profiles of each mutant grown in co-culture were also analysed by HPAEC-PAD and it was found that each mutant, with the exception of *B. breve* UCC2003-fucP, produced an identical carbohydrate profile to that produced by the *B. breve* UCC2003-pAM5. *B. breve* UCC2003-fucP, as expected, was not capable of utilising/transporting Fuc, as evident from the presence of Fuc in its carbohydrate profile (Fig. 4.4).

4.5 DISCUSSION

Early investigations into the degradation of mucin established the role of extracellular glycosidases produced by a sub-population of bacteria in the gut, suggesting that the remaining enteric bacteria may cross feed on the released oligosaccharides (Boulding & Hoskins, 1981; Hoskins *et al.*, 1985). Since then, much of the research into bacterial degradation of mucin has focused on the characterisation of such extracellular glycosidases, such as two α-L-fucosidases and an exo-α-sialidase from *B. bifidum* JCM1254, and an endo-α-*N*-acetylgalactosaminidase from *B. longum* JCM1217 (Ashida *et al.*, 2009; Katayama *et al.*, 2005; Katayama *et al.*, 2004; Kiyohara *et al.*, 2011). Another study identified, by a genomic as well as a transcriptomic and proteomic approach, a number of extracellular enzymes involved in mucin degradation by *B. bifidum* PRL2010, including a predicted cell wall-anchored endo-α-*N*-acetylgalactosamine, two α-L-fucosidases, two exo-α-sialidases, a β-galactosidase and two putative *N*-acetyl-β-hexosaminidases, all of which contained a signal peptide (Turroni *et al.*, 2010b).

Extracellular enzymes similar to those outlined above were not found in the genome of *B. breve* UCC2003 (O'Connell Motherway *et al.*, 2011b). Therefore, it was not surprising that *B. breve* UCC2003-pAM5 was incapable of high density growth in a medium containing mucin as the sole carbon source. However, when *B. bifidum* PRL2010-pPKCM7 was included as a co-cultivating, mucin-degrading bacterium, the growth and viability of *B. breve* UCC2003-pAM5 was improved compared to the control situation. A combination of HPAEC-PAD and transcriptome analyses identified some of the possible monosaccharides and oligosaccharides which could support this enhanced co-cultivation growth/viability phenotype of *B. breve*

UCC2003-pAM5, represented by sialic acid, Fuc, Gal and/or Gal-containing constituents of mucin.

HPAEC-PAD analysis identified two monosaccharides, namely Fuc and Gal, which were released from mucin by *B. bifidum* PRL2010-pPKCM7 activity. Fuc was shown to be internalised by *B. breve* UCC2003-pAM5 during growth in co-culture, suggesting a role in the enhanced growth/viability phenotype of the strain in co-culture. Transcriptome and mutagenesis data supports this hypothesis, as shown by increased transcription of a cluster (Bbr_1740-1742) involved in the uptake and utilization of Fuc and by the inability of a mutant in the *fucP* gene to internalise Fuc from the growth medium.

Gal has previously been shown to support the growth of *B. breve* UCC2003 (Ruiz *et al.*, 2013). However, given that Gal is ubiquitous in mucin, it may be internalised by *B. breve* UCC2003-pAM5 as a monosaccharide and/or as a constituent of a larger oligosaccharide. Interestingly, transcriptome data revealed increased transcription of the Bbr_1585-1590 gene cluster, which includes predicted GalE, NahK and LnbP-encoding genes, all of which are required for the metabolism of GNB (Nishimoto & Kitaoka, 2007), as well as an adjacent predicted ABC transport system. Outside of this cluster, the predicted *galT2* gene (Bbr_1884) was also up-regulated. A previous study in *B. bifidum* hypothesised that GalT2 is specifically required for the metabolism of GNB from mucin (De Bruyn *et al.*, 2013). Up-regulation of each of the required genes for the complete GNB pathway, as well as a predicted transport system, suggests that GNB might be the preferred source of Gal for *B. breve* UCC2003 in co-culture. A predicted β-galactosidase-encoding gene, *lacZ7*, was also up-regulated in co-culture, implying that *B. breve* UCC2003 may utilise another Gal-containing oligosaccharide. However, the precise substrate or preferred linkage of

this enzyme is still a matter of speculation. Gal-containing oligosaccharides and GNB were not identified in the HPAEC-PAD data. However, the presence of four predicted β-galactosidases on the genome of *B. bifidum* PRL2010, as well as a clear homolog of the GNB pathway (Turroni *et al.*, 2010b), suggests that *B. bifidum* PRL2010 also utilises these sugars, and it is therefore possible that in co-culture the two strains compete for such sugars.

Interestingly, transcriptome data also revealed increased transcription of two clusters previously shown to be up-regulated in the presence of sialic acid (Egan *et al.*, 2014). This is inconsistent with the HPAEC-PAD data, in which sialic acid was not identified; however, it should be noted that the mucin from porcine stomach used in this study contains only 0.5 %- 1.5 % bound sialic acid, and it is thus possible that the released sialic acid is below the detection level of the HPAEC-PAD system. However, since *B. bifidum* PRL2010 has been shown to be incapable of using sialic acid (Egan *et al.*, 2014; Turroni *et al.*, 2010b), it seems likely that *B. breve* UCC2003 would utilise this sugar, even if available in very low quantities.

While accepting the HPAEC-PAD and transcriptome data are not definitive, the results suggest that *B. breve* UCC2003-pAM5 utilises a combination of sugars in co-culture, a suggestion supported by the fact that none of the mutants tested displayed a different phenotype to *B. breve* UCC2003-pAM5 in co-culture. Construction of double or triple mutant strains is expected to result in strains that are impaired in growth/viability under co-culture conditions, however, at present such multiple mutant construction is not technically feasible for *B. breve* UCC2003 (and have to the best of our knowledge not been described for any bifidobacterial species/strain).

Our results highlight the different approaches taken by two species of bifidobacteria to proliferate and survive in the gut. *B. bifidum* PRL2010 has been shown to utilise mucin oligosaccharides (and structurally similar HMOs), a highly complex, yet ubiquitous carbohydrate source in the gut. Aside from host glycans such as mucin and HMO, however, the fermentation ability of *B. bifidum* PRL2010 is rather limited (Turroni *et al.*, 2012a; Turroni *et al.*, 2010b). *B. breve* UCC2003, on the other hand, is a more versatile bacterium from a metabolic perspective, capable of utilising a number of host and diet-derived carbohydrates. As seen in this study it can scavenge constituents of mucin released by the extracellular glycosidase activity of other bifidobacteria, as well as cross-feed on certain HMO (Egan *et al.*, 2014), a characteristic which reflects the abundance of representatives of this species in the infant gut (Turroni *et al.*, 2012b). However, *B. breve* UCC2003 also has the ability to utilise a number of dietary-derived carbohydrates such as starch, galactan, cellodextrins and raffinose (O'Connell *et al.*, 2013; O'Connell Motherway *et al.*, 2011a; O'Connell Motherway *et al.*, 2008; Pokusaeva *et al.*, 2011b) and accordingly *B. breve* strains have been identified in the adult human bifidobacterial population, although the relative abundance is lower than in infants (Turroni *et al.*, 2009; Turroni *et al.*, 2012b). The similar phenotype of *B. breve* UCC2003-pAM5 and the mutant strains in co-culture also highlights the adaptability of the strain, emphasising its ability to switch to a different carbon source, depending on the carbohydrates available. This versatility is reflected on the genome of *B. breve* UCC2003, which contains a large number of carbohydrate utilisation clusters (O'Connell Motherway *et al.*, 2011b), suggesting an ability to alternate between and/or co-utilise diet- and host- derived carbohydrates depending on availability. The gut commensal *Bacteroides thetaiotaomicron*, displays a remarkable metabolic flexibility, whereby

in the absence of a dietary-derived fibre, this bacterium will shift its metabolic activities towards the degradation of mucin (Sonnenburg *et al.*, 2005). These results suggest that *B. breve* UCC2003 may also be capable of such flexibility, although further study is required.

The ability to degrade mucin seems to be limited to particular gut commensals, such as certain species of *Bacteroides, Bifidobacterium, Ruminococcus* and *Akkermansia* (Crost *et al.*, 2013; Derrien *et al.*, 2004; Salyers *et al.*, 1978; Salyers *et al.*, 1977; Sonnenburg *et al.*, 2005; Turroni *et al.*, 2010b). Pathogenic bacteria appear to be poorly adapted to mucin degradation (Marcobal *et al.*, 2013), however, multiple studies have shown pathogens utilising constituents of mucin released by commensal glycosidases. For example, *C. jejuni* has been shown to utilise Fuc as a substrate for growth (Stahl *et al.*, 2011), while enterohaemorrhagic *E. coli* uses Fuc as a signal to induce virulence (Pacheco *et al.*, 2012). Similarly, *Salmonella typhimurium* and *Clostridium difficile* have been shown to utilise sialic acid released by the sialidase activity of *Bacteroides thetaotaomicron* in a gnotobiotic mouse (Ng *et al.*, 2013). The uptake of mucin-derived Fuc by *B. breve* UCC2003, as well as the presumable utilisation of mucin-derived sialic acid, suggests a potential role for *B. breve* UCC2003 in “mopping up” the released constituents of mucin, providing competition to potential pathogens and inhibiting or limiting their proliferation, although further study is required to expand on this hypothesis.

The mutually beneficial relationship between the host and the intestinal bacteria has been well established (Bäckhed *et al.*, 2005), however, it is of equal importance to acknowledge the symbiotic relationships formed between genera and species of the intestinal bacteria. Cross-feeding between *B. breve* UCC2003 and *B. bifidum* PRL2010 has now been shown for mucin in the present work and previously for 3’

sialyllactose (Egan *et al.*, 2014), and given the structural similarity between mucin oligosaccharides and HMOs, it may be assumed that *B. breve* UCC2003 has the ability to cross feed on other HMO-derived constituents that are released by *B. bifidum* PRL2010. These results highlight the compatibility of these two species of bifidobacteria, especially when viewed in contrast to the strategy employed by *B. longum* subsp. *infantis,* which assimilates HMOs in their intact form, leaving none for potential cross-feeding with another species (Sela & Mills, 2010). Interestingly, another type of cross-feeding has been described, whereby *Eubacterium halii* utilises lactate produced by *Bifidobacterium adolescentis* during growth on starch, resulting in butyrate production by *Eu. halii* (Belenguer *et al.*, 2006). Butyrate is a major source of energy for colonocytes and has also been implicated in protection against colonic carcinogenesis (Roediger, 1980; Velázquez *et al.*, 1996). Understanding such complex interactions between different members of the intestinal bacteria is of crucial importance when attempting to influence the activity or composition of the intestinal microbiota, such as in the use of probiotics.

This study provides *in vitro* proof for the existence of a commensal relationship between two species of bifidobacteria in the large intestine, namely *B. breve* and *B. bifidum,* in which *B. breve* UCC2003 benefits from the carbohydrates released by the extracellular glycosidase activities of *B. bifidum* PRL2010. To our knowledge, this is the first study to describe the molecular details of such cross-feeding, with particular emphasis on the carbohydrate components which support the improved growth and survival in co-culture. The results shown here improve our knowledge on how *B. breve* UCC2003 colonises the (infant) gut in the absence of dietary-derived carbohydrates, and also emphasise this strain's ability to switch between carbohydrate sources depending on availability. This is an advantageous

characteristic in terms of enhanced survival and colonisation ability in both the infant and adult gut, and may present an advantage to the host in limiting opportunities for pathogenic microbes to proliferate in the gut.

4.6 ACKNOWLEDGMENTS

The Alimentary Pharmabiotic Centre is a research centre funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan. The authors and their work were supported by SFI (Grant Nos. 07/CE/B1368, and SFI/12/RC/2273 and 08/SRC/B1393), the EU FP7 programme (Grant no. 260600) and a HRB postdoctoral fellowship (Grant No. PDTM/20011/9) awarded to MOCM.

4.7 TABLES AND FIGURES

Table 4.1: Bacterial strains and plasmids used in this study.

Strains and plasmids	Relevant features	Reference or source
Strains		
***Escherichia coli* strains**		
***E.coli* EC101**	Cloning host; *repA*$^+$ *kmr*	(Law *et al.*, 1995)
***E.coli* EC101-pNZ-M.Bbrll+Bbr11**	EC101 harboring a pNZ8048 derivative containing bbrllM and bbrlllM	(O'Connell Motherway *et al.*, 2009)
***Bifidobacterium* sp. strains**		
***B. breve* UCC2003**	Isolate from a nursling stool	(Maze *et al.*, 2007)
***B. breve* UCC2003-pAM5**	*B. breve* UCC2003 harboring pAM5	This study
***B. breve* UCC2003-nanA**	pORI19-Tet-nanA (Bbr_0168) insertion mutant of *B. breve* UCC2003	(Egan *et al.*, 2014)
***B. breve* UCC2003-fucP**	pORI19-Tet-fucP (Bbr_1742) insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-lnbP**	pORI19-Tet-lnbP (Bbr_1587) insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-lacZ7**	pORI19-Tet-lacZ7 (Bbr_1833) insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-galT**	Tet^r transposon mutant of *B. breve* UCC2003	(Ruiz *et al.*, 2013)
***B. bifidum* PRL2010**	Isolate from infant faeces	(Turroni *et al.*, 2010b)
***B. bifidum* PRL2010-pPKCM7**	*B. bifidum* PRL2010 harboring pPKCM7	This study
Plasmids		
pAM5	pBC1-puC19-Tc^r	(Álvarez-Martín *et al.*, 2007)
pPKCM7	pblueCm harboring rep pCIBA089	(Cronin *et al.*, 2007)
pORI19	Em^r, repA$^-$, ori$^+$, cloning vector	(Law *et al.*, 1995)
pORI19-tetW-fucP	Internal 400bp fragment of *fucP* and *tetW* cloned in pORI19	This study
pORI19-tetW-lnbP	Internal 479bp fragment of *lnbP* and *tetW* cloned in pORI19	This study
pORI19-tetW-lacZ7	Internal 568bp fragment of *lacZ7* and *tetW* cloned in pORI19	This study

Table 4.2: Oligonucleotide primers used in this study

Purpose	Primer	Sequence
Cloning of 400bp fragment of *fucP* (Bbr_1742) in pORI19	FucPF	TAGCAT*AAGCTT*GGCGAATCGTTCGTATCA
	FucPR	GATATC*TCTAGA*GCGCCCCAGTGCTTGAGC
Cloning of 479bp fragment of *lnbP* (Bbr_1587) in pORI19	LnbPF	TAGCAT*AAGCTT*CACACAGGTATTGGGAGGTTG
	LnbPR	CTAGTC*TCTAGA*GTTGTAGGCGCCACCATCC
Cloning of 568bp fragment of *lacZ7* (Bbr_1833) in pORI19	LacZ7F	TAGCAT*AAGCTT*CCAGGCCAAGAACTCCAGTG
	LacZ7R	CATGAT*TCTAGA*CAGCTTGGGCAGGTTGAACG
Amplification of *tetW*	TetWF	TCAGCT*GTCGAC*ATGCTCATGTACGGTAAG
	TetWR	GCGACG*GTCGAC*CATTACCTTCTGAAACATA
Confirmation of site-specific homologous recombination	FucPconfirm	TGTTCGCCATGTTCGTTATC
	LnbPconfirm	GATCACTCTGCATATGGACG
	LacZ7confirm	GTACCGACATCGACGCGTTC

Restriction sites incorporated into oligonucleotide primer sequences are indicated in italics.

Table 4.3: Effect of mucin (in co-culture with *B. bifidum* PRL2010) on the transcriptome of *B. breve* UCC2003

Locus Tag (Gene Name)	Predicted Function	Level of upregulation (*P*<0.001)
Bbr_0161 (*nanK*)	Conserved hypothetical protein in ROK family	4.22
Bbr_0162 (*nanE*)	*N*-acetylmannosamine-6-phosphate 2-epimerase	5.46
Bbr_0164 (*nanB*)	ABC transport system, solute binding protein	14.08
Bbr_0165 (*nanC*)	ABC transport system, permease protein	11.82
Bbr_0166 (*nanD*)	ABC transport system, ATP-binding protein	18.92
Bbr_0167 (*nanF*)	ABC transport system, ATP-binding protein	22.14
Bbr_0168 (*nanA*)	*N*-acetylneuraminate lyase	15.17
Bbr_0169 (*nagB1*)	Glucosamine-6-phosphate isomerase	17.77
Bbr_0171 (*nanH*)	Sialidase	7.23
Bbr_0173 (*nanR*)	Transcriptional regulator, GntR family	4.80
Bbr_1247 (*nagA2*)	*N*-acetylglucosamine-6-phosphate deacetylase	4.59
Bbr_1248 (*nagB3*)	Glucosamine-6-phosphate isomerase	6.08
Bbr_1585 (*galE*)	UDP-Glc-4-epimerase	3.96
Bbr_1586 (*nahK*)	*N*-acetylhexosamine kinase	3.67
Bbr_1587 (*lnbP*)	Lacto-*N*-biose phosphorylase	3.19
Bbr_1588	ABC transport system, permease protein	3.34
Bbr_1589	ABC transport system, permease protein	2.88
Bbr_1590	ABC transport system, solute binding protein	8.95
Bbr_1740 (*dapA4*)	Dihydrodipicolinate synthase	9.91
Bbr_1741	Conserved hypothetical protein	8.83
Bbr_1742 (*fucP*)	L-fucose permease	7.81
Bbr_1833 (*lacZ7*)	Beta-galactosidase	4.31
Bbr_1879	PTS system, glucose-specific IIABC component	6.97
Bbr_1880	PTS system, GlcNAc-specific IIBC component	18.15
Bbr_1884 (*galT2*)	Gal-1-phosphate uridylyltransferase	3.86

The cut-off point for the level of up-regulation is 2.5-fold with a *P*-value of <0.001.

Table 4.4: Quantification of Fuc, Gal and Glc by HPAEC-PAD.

	Fuc	**Gal**	**Glc**
mMRS + 0.4 % mucin (non-fermented)	n.d.	0.13	3.96
mMRS + 0.4 % mucin (30 h growth of *B. bifidum* PRL2010)	34.94	51.73	n.d.

Concentrations given in nmol mg^{-1}. n.d.= not detected.

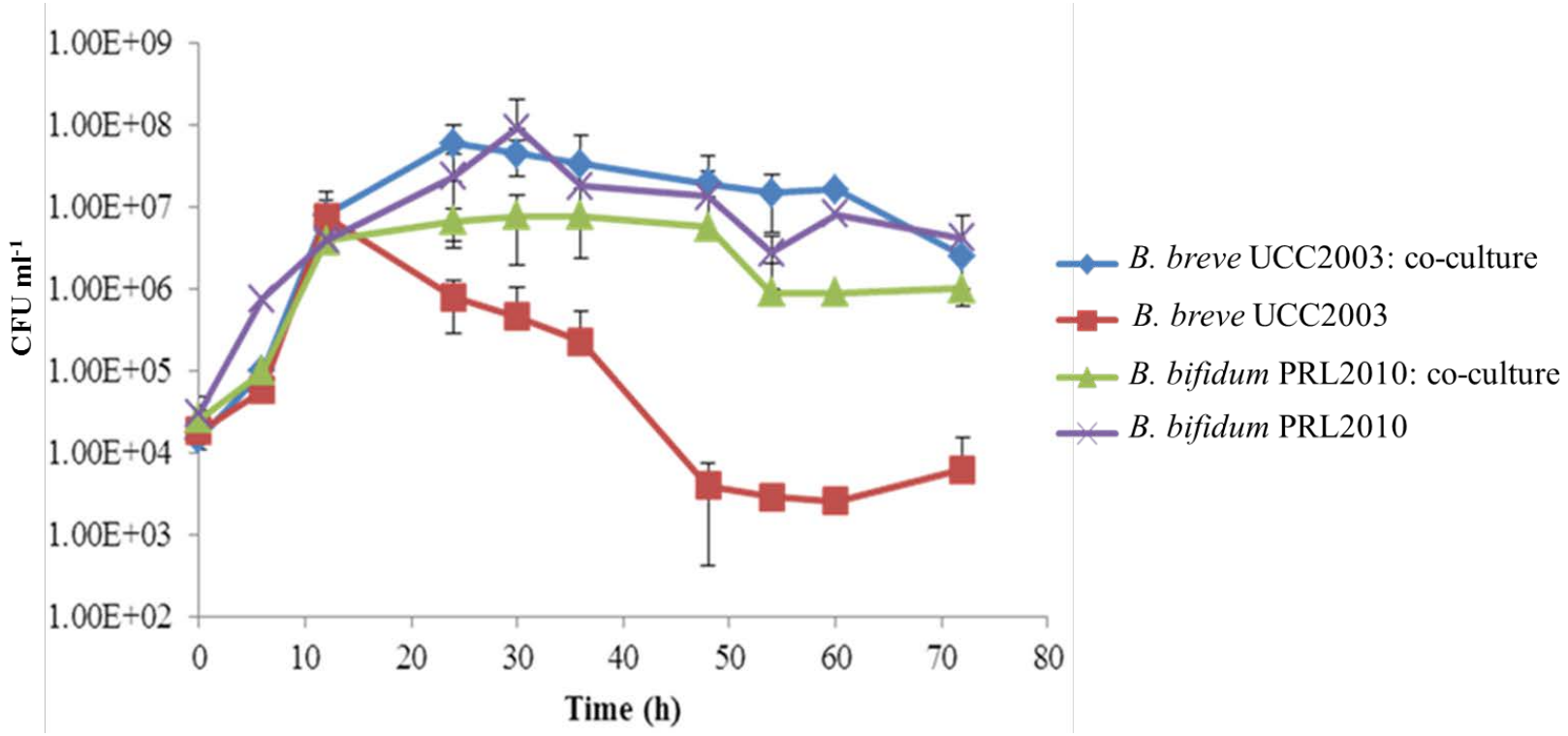


Figure 4.1: Individual and co-culture growth profiles of *B. breve* UCC2003-pAM5 and *B. bifidum* PRL2010-pPKCM7 in mucin. All growth experiments were performed in mMRS supplemented with 0.4 % mucin from porcine stomach for 72 h. The results presented are the mean values of duplicate experiments. Error bars represent the standard deviation.

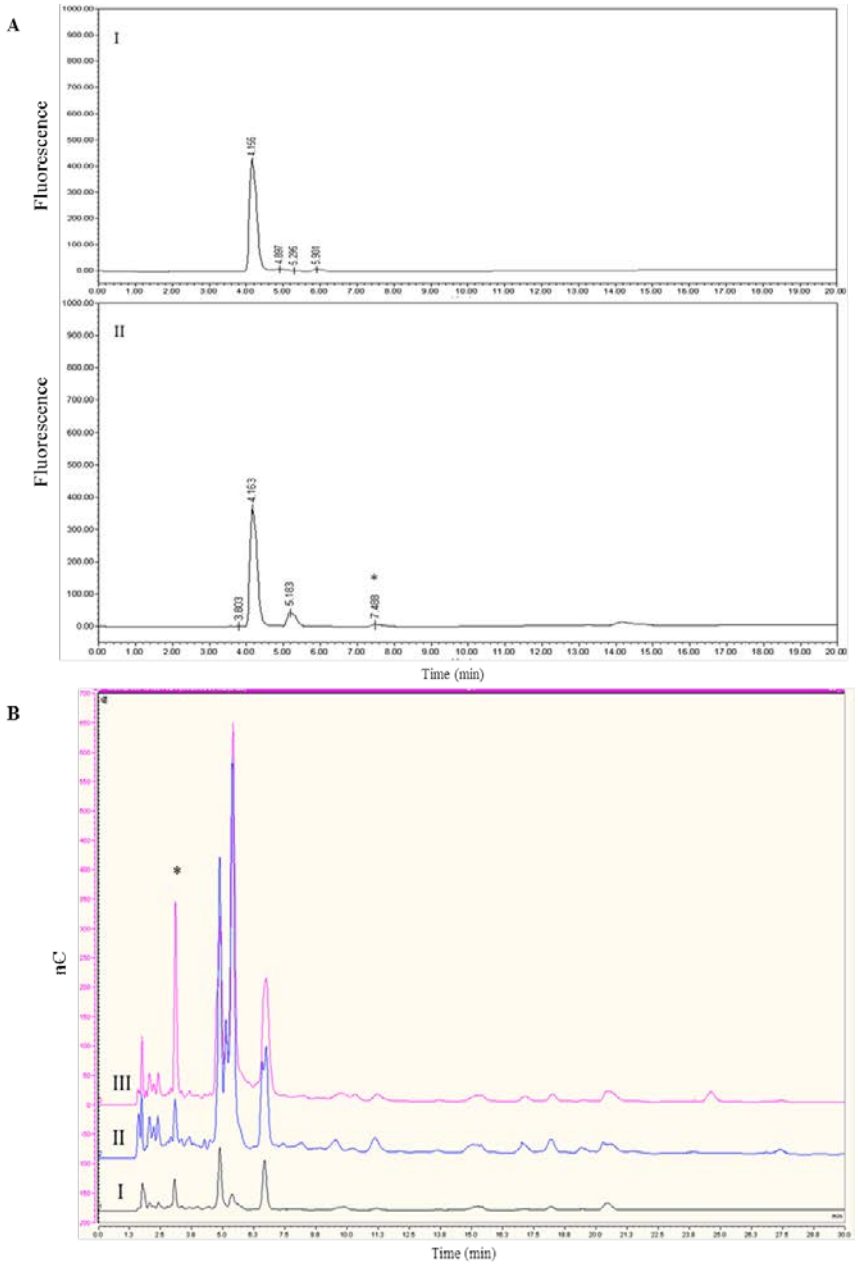


Figure 4.2: (A) HPLC profile of 2-AB -labelled (I) mMRS supplemented with 0.4 % mucin and (II) the media from (I) after 30 h growth of *B. bifidum* PRL2010-pPKCM7. The peak for Fuc is marked with *. **(B)** Qualitative HPAEC-PAD analysis of (I) mMRS supplemented with 0.4 % mucin, (II) the media from (I) following 30 h growth of *B. bifidum* PRL2010-pPKCM7 and *B. breve* UCC2003-pAM5 in co-culture, (III) the media from (I) after 30 h growth of *B. bifidum* PRL2010-pPCM7. The peak for Fuc is marked with *. nC, nanocoulombs.

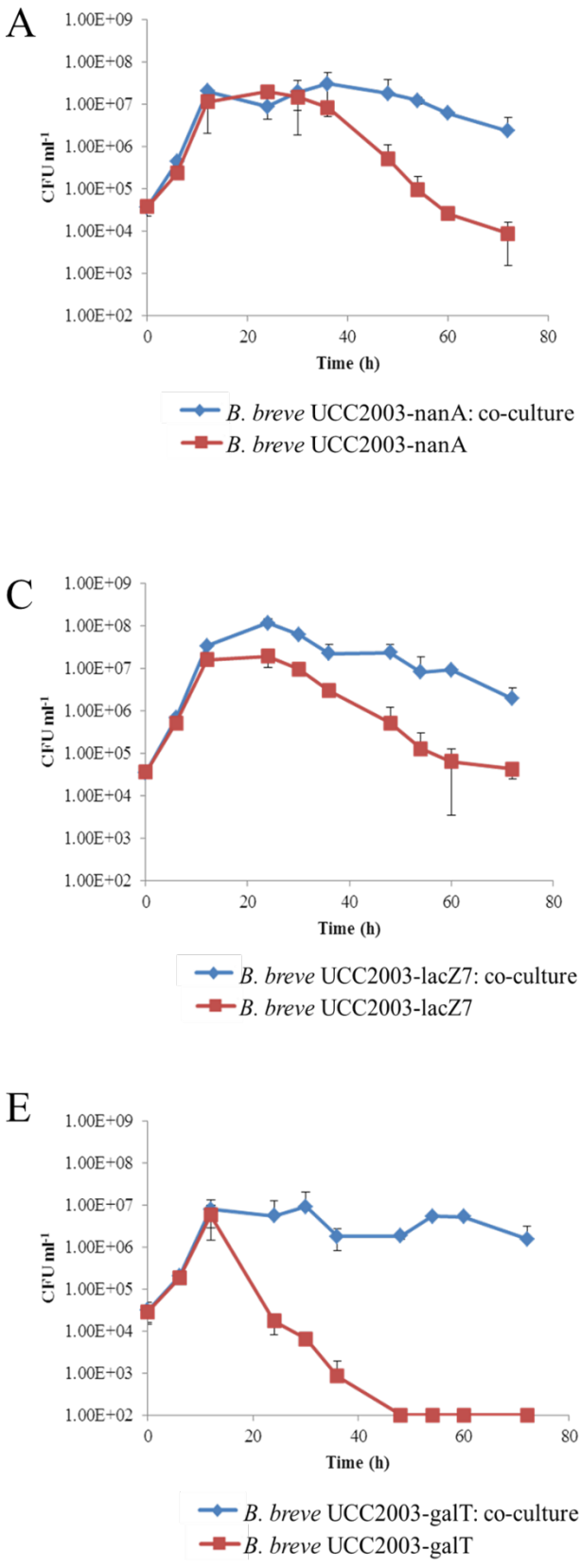

A
B
C
D
E
CFU ml-1
Time (h)
B. breve UCC2003-nanA: co-culture
B. breve UCC2003-nanA
B. breve UCC2003-fucP: co-culture
B. breve UCC2003-fucP
B. breve UCC2003-lacZ7: co-culture
B. breve UCC2003-lacZ7
B. breve UCC2003-lnbP: co-culture
B. breve UCC2003-lnbP
B. breve UCC2003-galT: co-culture
B. breve UCC2003-galT

Figure 4.3: Growth profiles of the mutants **(A)** *B. breve* UCC2003-nanA, **(B)** *B. breve* UCC2003-fucP, **(C)** *B. breve* UCC2003-lacZ7, **(D)** *B. breve* UCC2003-lnbP and **(E)** *B. breve* UCC2003-galT in mMRS supplemented with 0.4 % mucin, separately or in co-culture with *B. bifidum* PRL2010-pPKCM7, over 72 h. The results presented are the mean values of duplicate experiments. Error bars represent standard deviation.

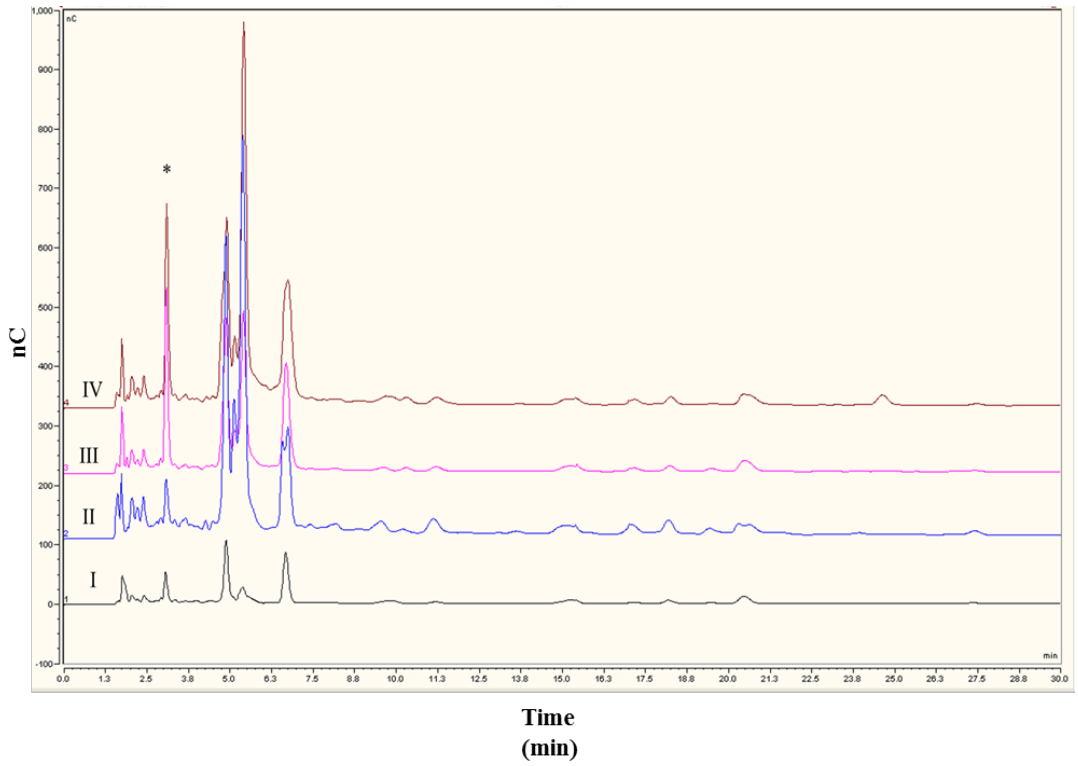


Figure 4.4: Qualitative HPAEC-PAD analysis of (I) mMRS supplemented with 0.4 % mucin, (II) the media from (I) following 36 h growth of *B. breve* UCC2003-pAM5 and *B. bifidum* PRL2010-pPKCM7 in co-culture, (III) the media from (I) following 36 h growth of *B. bifidum* PRL2010-pPKCM7, (IV) the media from (I) following 36 h growth of *B. breve* UCC2003-fucP and *B. bifidum* PRL2010-pPKCM7 in co-culture. The peak for Fuc is marked with *.

4.8 REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Álvarez-Martín, P., O'Connell-Motherway, M., van Sinderen, D. & Mayo, B. (2007). Functional analysis of the pBC1 replicon from *Bifidobacterium catenulatum* L48. *Appl Microbiol Biotechnol* **76**, 1395-1402.

Asakuma, S., Hatakeyama, E., Urashima, T., Yoshida, E., Katayama, T., Yamamoto, K., Kumagai, H., Ashida, H., Hirose, J. & other authors (2011). Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. *J Biol Chem* **286**, 34583-34592.

Ashida, H., Miyake, A., Kiyohara, M., Wada, J., Yoshida, E., Kumagai, H., Katayama, T. & Yamamoto, K. (2009). Two distinct α-L-fucosidases from *Bifidobacterium bifidum* are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. *Glycobiology* **19**, 1010-1017.

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920.

Belenguer, A., Duncan, S. H., Calder, A. G., Holtrop, G., Louis, P., Lobley, G. E. & Flint, H. J. (2006). Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. *Appl Environ Microbiol* **72**, 3593-3599.

Bigge, J. C., Patel, T. P., Bruce, J. A., Goulding, P. N., Charles, S. M. & Parekh, R. B. (1995). Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal Biochem* **230**, 229-238.

Boulding, E. & Hoskins, L. (1981). Mucin degradation in human colon ecosystems. Evidence for the existence and role of bacterial subpopulations producing glycosidases as extracellular enzymes. *J Clin Invest* **67**, 163–172.

Capon, C., Maes, E., Michalski, J., Leffler, H. & Kim, Y. S. (2001). Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. *Biochem J* **358**, 657-664.

Cronin, M., Knobel, M., O'Connell-Motherway, M., Fitzgerald, G. F. & van Sinderen, D. (2007). Molecular dissection of a bifidobacterial replicon. *Appl Environ Microbiol* **73**, 7858-7866.

Crost, E. H., Tailford, L. E., Le Gall, G., Fons, M., Henrissat, B. & Juge, N. (2013). Utilisation of mucin glycans by the human gut symbiont *Ruminococcus gnavus* is strain-dependent. *PLoS ONE* **8**, e76341.

De Bruyn, F., Beauprez, J., Maertens, J., Soetaert, W. & De Mey, M. (2013). Unraveling the Leloir pathway of *Bifidobacterium bifidum*: Significance of the uridylyltransferases. *Appl Environ Microbiol* **79**, 7028-7035.

De Man, J. C., Rogosa, M. & Sharpe, M. E. (1960). A medium for the cultivation of lactobacilli. *J Appl Bacteriol* **23**, 130-135.

Derensy-Dron, D., Krzewinski, F., Brassart, C. & Bouquelet, S. (1999). β-1,3-galactosyl-*N*-acetylhexosamine phosphorylase from *Bifidobacterium bifidum* DSM 20082: characterization, partial purification and relation to mucin degradation. *Biotechnol Appl Biochem* **29**, 3-10.

Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* **54**, 1469-1476.

Egan, M., O'Connell Motherway, M., Ventura, M. & van Sinderen, D. (2014). Metabolism of sialic acid by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **80**, 4414-4426.

Frey, P. A. (1996). The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *FASEB J* **10**, 461-470.

García de la Nava, J., Santaella, D. F., Alba, J. C., Carazo, J. M., Trelles, O. & Pascual-Montano, A. (2003). Engene: the processing and exploratory analysis of gene expression data. *Bioinformatics* **19**, 657-658.

Hardy, M. R. (1997). *Glycan labeling with the fluorophores 2-aminobenzamide and anthranilic acid*. New York: Marcel Dekker Inc.

Hoskins, L. C., Agustines, M., McKee, W. B., Boulding, E. T., Kriaris, M. & Niedermeyer, G. (1985). Mucin degradation in human colon ecosystems. Isolation

and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *J Clin Invest* **75**, 944-953.

Katayama, T., Fujita, K. & Yamamoto, K. (2005). Novel bifidobacterial glycosidases acting on sugar chains of mucin glycoproteins. *J Biosci Bioeng* **99**, 457-465.

Katayama, T., Sakuma, A., Kimura, T., Makimura, Y., Hiratake, J., Sakata, K., Yamanoi, T., Kumagai, H. & Yamamoto, K. (2004). Molecular cloning and characterization of *Bifidobacterium bifidum* 1,2-α-l-fucosidase (*afcA*), a novel inverting glycosidase (glycoside hydrolase family 95). *J Bacteriol* **186**, 4885-4893.

Kilcoyne, M., Gerlach, J. Q., Gough, R., Gallagher, M. E., Kane, M., Carrington, S. D. & Joshi, L. (2012). Construction of a natural mucin microarray and interrogation for biologically relevant glyco-epitopes. *Anal Chem* **84**, 3330-3338.

Kilcoyne, M., Shah, M., Gerlach, J. Q., Bhavanandan, V., Nagaraj, V., Smith, A. D., Fujiyama, K., Sommer, U., Costello, C. E. & other authors (2009). O-glycosylation of protein subpopulations in alcohol-extracted rice proteins. *J Plant Physiol* **166**, 219-232.

Kitaoka, M., Tian, J. & Nishimoto, M. (2005). Novel putative galactose operon involving lacto-*N*-biose phosphorylase in *Bifidobacterium longum*. *Appl Environ Microbiol* **71**, 3158-3162.

Kiyohara, M., Tanigawa, K., Chaiwangsri, T., Katayama, T., Ashida, H. & Yamamoto, K. (2011). An exo-α-sialidase from bifidobacteria involved in the degradation of sialyloligosaccharides in human milk and intestinal glycoconjugates. *Glycobiology* **21**, 437-447.

Larsson, J. M., Karlsson, H., Sjovall, H. & Hansson, G. C. (2009). A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19**, 756-766.

Law, J., Buist, G., Haandrikman, A., Kok, J., Venema, G. & Leenhouts, K. (1995). A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. *J Bacteriol* **177**, 7011-7018.

LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D. & Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr Opin Biotechnol* **24**, 160-168.

Long, A. D., Mangalam, H. J., Chan, B. Y. P., Tolleri, L., Hatfield, G. W. & Baldi, P. (2001). Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework: Analysis of global gene expression in *Escherichia coli* K12. *J Biol Chem* **276**, 19937-19944.

Marcobal, A., Southwick, A. M., Earle, K. A. & Sonnenburg, J. L. (2013). A refined palate: Bacterial consumption of host glycans in the gut. *Glycobiology* **23**, 1038-1046.

Maze, A., O'Connell-Motherway, M., Fitzgerald, G., Deutscher, J. & van Sinderen, D. (2007). Identification and characterization of a fructose phosphotransferase system in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **73**, 545-553.

Ng, K. M., Ferreyra, J. A., Higginbottom, S. K., Lynch, J. B., Kashyap, P. C., Gopinath, S., Naidu, N., Choudhury, B., Weimer, B. C. & other authors (2013). Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96-99.

Nishimoto, M. & Kitaoka, M. (2007). Identification of *N*-acetylhexosamine 1-kinase in the complete lacto-*N*-biose I/galacto-*N*-biose metabolic pathway in *Bifidobacterium longum*. *Appl Environ Microbiol* **73**, 6444-6449.

O'Connell, K. J., O'Connell Motherway, M., O'Callaghan, J., Fitzgerald, G. F., Ross, R. P., Ventura, M., Stanton, C. & van Sinderen, D. (2013). Metabolism of four α-glycosidic linkage-containing oligosaccharides by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **79**, 6280-6292.

O'Connell, K. J., O'Connell Motherway, M., Liedtke, A., Fitzgerald, G. F., Ross, R. P., Stanton, C., Zomer, A. & van Sinderen, D. (2014). Transcription of two adjacent carbohydrate utilization gene clusters in *Bifidobacterium breve* UCC2003 is controlled by LacI- and Repressor Open reading frame Kinase (ROK)-type regulators. *Appl Environ Microbiol* **80**, 3604-3614.

O'Connell Motherway, M., Fitzgerald, G. F. & van Sinderen, D. (2011a). Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **4**, 403-416.

O'Connell Motherway, M., O'Driscoll, J., Fitzgerald, G. F. & Van Sinderen, D. (2009). Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **2**, 321-332.

O'Connell Motherway, M., Kinsella, M., Fitzgerald, G. F. & van Sinderen, D. (2013). Transcriptional and functional characterization of genetic elements involved in galacto-oligosaccharide utilization by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **6**, 67-79.

O'Connell Motherway, M., Fitzgerald, G., Neirynck, S., Ryan, S., Steidler, L. & van Sinderen, D. (2008). Characterization of ApuB, an extracellular type II amylopullulanase from *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **74**, 6271-6279.

O'Connell Motherway, M., Zomer, A., Leahy, S. C., Reunanen, J., Bottacini, F., Claesson, M. J., O'Brien, F., Flynn, K., Casey, P. G. & other authors (2011b). Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc Natl Acad Sci USA* **108**, 11217-11222.

Pacheco, A. R., Curtis, M. M., Ritchie, J. M., Munera, D., Waldor, M. K., Moreira, C. G. & Sperandio, V. (2012). Fucose sensing regulates bacterial intestinal colonization. *Nature* **492**, 113-117.

Podolsky, D. K. (1985). Oligosaccharide structures of human colonic mucin. *J Biol Chem* **260**, 8262-8271.

Pokusaeva, K., Fitzgerald, G. & Sinderen, D. (2011a). Carbohydrate metabolism in Bifidobacteria. *Genes Nutr* **6**, 285-306.

Pokusaeva, K., O'Connell-Motherway, M., Zomer, A., MacSharry, J., Fitzgerald, G. F. & van Sinderen, D. (2011b). Cellodextrin utilization by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **77**, 1681-1690.

Pokusaeva, K., Neves, A. R., Zomer, A., O'Connell-Motherway, M., MacSharry, J., Curley, P., Fitzgerald, G. F. & van Sinderen, D. (2010). Ribose utilization by the human commensal *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **3**, 311-323.

Riordan, O. (1998). *Studies on antimicrobial activity and genetic diversity of Bifidobacterium species: molecular characterization of a 5.75 kb plasmid and a chromosomally encoded recA gene homologue from Bifidobacterium breve.*, National University of Ireland, Cork, Cork.

Robbe, C., Capon, C., Coddeville, B. & Michalski, J. (2004). Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem J* 307-316.

Roediger, W. E. (1980). Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. *Gut* **21**, 793-798.

Round, J. L. & Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313-323.

Ruiz, L., Motherway, M. O. C., Lanigan, N. & van Sinderen, D. (2013). Transposon mutagenesis in *Bifidobacterium breve*: Construction and characterization of a Tn5 transposon mutant library for *Bifidobacterium breve* UCC2003. *PLoS ONE* **8**, e64699.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Ryan, S., Fitzgerald, G. & van Sinderen, D. (2006). Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains. *Appl Environ Microbiol* **72**, 5289-5296.

Salyers, A. A., Palmer, J. K. & Wilkins, T. D. (1978). Degradation of polysaccharides by intestinal bacterial enzymes. *Am J Clin Nutr* **31**, S128-S130.

Salyers, A. A., West, S. E., Vercellotti, J. R. & Wilkins, T. D. (1977). Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Appl Environ Microbiol* **34**, 529-533.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning a laboratory manual, 2nd ed.* , 2nd Edition edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Sela, D. & Mills, D. (2010). Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol* **18**, 298-307.

Servin, A. L. (2004). Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiol Rev* **28**, 405-440.

Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C.-H., Westover, B. P., Weatherford, J., Buhler, J. D. & Gordon, J. I. (2005). Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307**, 1955-1959.

Stahl, M., Friis, L. M., Nothaft, H., Liu, X., Li, J., Szymanski, C. M. & Stintzi, A. (2011). L-Fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proc Natl Acad Sci U S A* **108**, 7194-7199.

Suzuki, R., Wada, J., Katayama, T., Fushinobu, S., Wakagi, T., Shoun, H., Sugimoto, H., Tanaka, A., Kumagai, H. & other authors (2008). Structural and thermodynamic analyses of solute-binding protein from *Bifidobacterium longum* specific for core 1 disaccharide and lacto-*N*-biose I. *J Biol Chem* **283**, 13165-13173.

Turroni, F., Milani, C., van Sinderen, D. & Ventura, M. (2010a). Genetic strategies for mucin metabolism in *Bifidobacterium bifidum* PRL2010. *aging* **107**, 19514-19519.

Turroni, F., Strati, F., Foroni, E., Serafini, F., Duranti, S., van Sinderen, D. & Ventura, M. (2012a). Analysis of predicted carbohydrate transport systems encoded by *Bifidobacterium bifidum* PRL2010. *Appl Environ Microbiol* **78**, 5002-5012.

Turroni, F., Foroni, E., Pizzetti, P., Giubellini, V., Ribbera, A., Merusi, P., Cagnasso, P., Bizzarri, B., de'Angelis, G. L. & other authors (2009). Exploring the diversity of the bifidobacterial population in the human intestinal tract. *Appl Environ Microbiol* **75**, 1534-1545.

Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., Kerr, C., Hourihane, J., Murray, D. & other authors (2012b). Diversity of bifidobacteria within the infant gut microbiota. *PLoS ONE* **7**, e36957.

Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J.-H., Zomer, A., Sanchez, B., Bidossi, A., Ferrarini, A. & other authors (2010b). Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc Natl Acad Sci U S A* **107**, 19514-19519.

Tytgat, K. M., Buller, H. A., Opdam, F. J., Kim, Y. S., Einerhand, A. W. & Dekker, J. (1994). Biosynthesis of human colonic mucin: Muc2 is the prominent secretory mucin. *Gastroenterology* **107**, 1352-1363.

van Hijum, S., de Jong, A., Baerends, R., Karsens, H., Kramer, N., Larsen, R., den Hengst, C., Albers, C., Kok, J. & other authors (2005). A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC Genomics* **6**, 77.

van Hijum, S. A. F. T., Garcia De La Nava, J., Trelles, O., Kok, J. & Kuipers, O. P. (2003). MicroPreP: a cDNA microarray data pre-processing framework. *Appl Bioinformatics*, 241-244.

Velázquez, O., Lederer, H. & Rombeau, J. (1996). Butyrate and the colonocyte. *Digest Dis Sci* **41**, 727-739.

Ventura, M., Canchaya, C., Fitzgerald, G., Gupta, R. & Sinderen, D. (2007). Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria. *Antonie van Leeuwenhoek* **92**, 265-265.

Watson, D., O'Connell Motherway, M., Schoterman, M. H. C., van Neerven, R. J. J., Nauta, A. & van Sinderen, D. (2013). Selective carbohydrate utilization by lactobacilli and bifidobacteria. *J Appl Microbiol* **114**, 1132-1146.

Zomer, A., Fernandez, M., Kearney, B., Fitzgerald, G. F., Ventura, M. & van Sinderen, D. (2009). An interactive regulatory network controls stress response in *Bifidobacterium breve* UCC2003. *J Bacteriol* **191**, 7039-7049.

Chapter V

Identification of glycosulfatases encoded by *Bifidobacterium breve* UCC2003

5.1 ABSTRACT

A bacterial nursling stool isolate, *Bifidobacterium breve* UCC2003, encodes two putative sulfatases. Among members of the gut microbiota sulfatase activity has previously been thoroughly characterised in the *Bacteroides* genus, but has yet to be described in bifidobacteria. Using a combination of transcriptomic and functional genomic approaches, we identified a gene cluster, designated *ats2*, dedicated to the uptake and metabolism of the sulfated monosaccharide, *N*-acetylglucosamine-6-sulfate. A transporter of the major facilitator superfamily is presumed to be required for uptake of this sulfated carbohydrate, while biochemical analysis showed that co-expression of the *atsA2*-encoded sulfatase and the *atsB2*-encoded anSME (anaerobic sulfatase maturation enzyme) is required for sulfatase activity. Transcription of the *ats2* cluster is regulated by a ROK-family transcriptional repressor. The substrate of the other presumed sulfatase cluster, *ats1*, has yet to be determined, however, using insertion mutagenesis, electrophoretic mobility shift assays and primer extension analysis the promoter regions within this cluster were characterised.

5.2 INTRODUCTION

The *Bifidobacterium* genus represents the dominant bacterial group among the intestinal microbiota of infants (Turroni *et al.*, 2012b; Vaishampayan *et al.*, 2010), while also typically constituting between 2 % and 10 % of the adult intestinal microbiota (Agans *et al.*, 2011; Andersson *et al.*, 2008; Turroni *et al.*, 2008; Zwielehner *et al.*, 2009). Bifidobacteria are saccharolytic organisms whose ability to colonise and survive in the large intestine is presumed to depend on the ability to metabolise complex carbohydrates present in this environment. Certain bifidobacterial species including *Bifidobacterium longum* subsp. *longum*, *Bifidobacterium adolescentis* and *Bifidobacterium breve* utilise a range of plant/diet derived oligosaccharides such as raffinose, arbinoxylan, galactan and cellodextrins (O'Connell *et al.*, 2013; O'Connell Motherway *et al.*, 2011a; Pastell *et al.*, 2009; Pokusaeva *et al.*, 2011a; Pokusaeva *et al.*, 2011b; Van Laere *et al.*, 2000). Bifidobacterial metabolism of host-derived carbohydrates such as mucin and human milk oligosaccharides (HMOs) is also well-described, with the typically infant-derived species *B. longum* subsp. *infantis* and *Bifidobacterium bifidum* particularly well-adapted to utilize these ubiquitous carbon sources in the infant gut (Sela *et al.*, 2008; Turroni *et al.*, 2010). Extensive research has identified and characterised both transport systems and glycosyl hydrolases required for the uptake and breakdown (of components) of these host-derived carbohydrates (Ashida *et al.*, 2009; Garrido *et al.*, 2011; Sela *et al.*, 2011; Sela *et al.*, 2012; Turroni *et al.*, 2012a).

Host-derived glycoproteins such as mucin and proteoglycans (e.g. chondroitin sulfate and heparan sulfate), which are found in the colonic mucosa and/or human milk, are often highly sulfated (Eliakim *et al.*, 1986; Larsson *et al.*, 2009; Newburg *et al.*, 1995; Oshiro *et al.*, 2001). Three sulfated oligosaccharide structures have also

been isolated from HMO, although they are present in only minute quantities (>100 μg ml^{-1}) (Guérardel *et al.*, 1999). Human colonic mucin is heavily sulfated, which is in contrast to mucin from the stomach or small intestine, the presumed purpose of which is to protect mucin against degradation by bacterial glycosidases (Brockhausen, 2003; Filipe, 1978). Despite this apparent protective measure, glycosulfatase activity has been identified in various members of the gut microbiota, e.g. *Bacteroides thetaiotaomicron, Bacteroides ovatus* and *Prevotella* strain RS2 (Roberton *et al.*, 1993; Salyers *et al.*, 1977).

Sulfatases constitute a family of hydrolases that are highly conserved between prokaryotes and eukaryotes, and whose substrate specificity can range from glycoproteins and glycolipids to hydroxysteroids (von Figura *et al.*, 1998). Sulfatases uniquely require a 3-oxoalanine (typically called Cα-formylglycine or FGly) residue at their active site, the absence of which results in dramatically decreased catalytic activity (Bond *et al.*, 1997; Lukatela *et al.*, 1998; Schmidt *et al.*, 1995). Eukaryotic sulfatases carry a conserved cysteine (Cys) residue, which is post-translationally modified to FGly in the endoplasmic reticulum (Dierks *et al.*, 1997; Schmidt *et al.*, 1995). Prokaryotic sulfatases carry either a Cys or a serine (Ser) residue which requires post-translational conversion to FGly in the cytosol in order to convert the enzyme to an active state (Beil *et al.*, 1995; Miech *et al.*, 1998; Szameit *et al.*, 1999).

In bacteria, two distinct systems have been described for the post-translational modification of sulfatase enzymes. In *Mycobacterium tuberculosis* H37Rv, the conversion of the Cys_{58} residue to FGly is catalysed by an FGly-generating enzyme (FGE) which requires oxygen as a co-factor (Carlson *et al.*, 2008; Dierks *et al.*, 2003). In *Klebsiella pneumoniae* DSM 681, the conversion of the Ser_{72} residue of

the *atsA*-encoded sulfatase is catalysed by the AtsB enzyme, which is a member of the *S*-adenosyl-L-methionine (AdoMet)-dependent family of radical enzymes (Fang *et al.*, 2004; Szameit *et al.*, 1999). Similar enzymes were also characterised from *Clostridium perfringens* ATCC13124 and *Ba. thetaiotaomicron* VPI-5481 which are active on both Cys and Ser-type sulfatases (Benjdia *et al.*, 2008; Berteau *et al.*, 2006). Crucially these enzymes are active under anaerobic conditions and were thus designated anaerobic sulfatase maturing enzymes (anSME) (Berteau *et al.*, 2006). Metagenomic sequence analysis has revealed that genes that are predicted to encode sulfatases and anSME enzymes are enriched in the gut microbiomes of humans compared to non-gut microbial communities (Kurokawa *et al.*, 2007). However, despite extensive research into the metabolism of mucin and HMO, sulfatase activity has yet to be described in bifidobacteria. In the current study, we identify two predicted sulfatase and anSME-encoding gene clusters in *B. breve* UCC2003 (and other *B. breve* strains), and demonstrate that one such cluster is required for the metabolism of the sulfated monosaccharide *N*-acetylglucosamine-6-sulfate (GlcNAc-6-S).

5.3 MATERIALS AND METHODS

***Bacterial strains, plasmids, media and culture conditions*.** Bacterial strains and plasmids used in this study are listed in Table 5.1. *B. breve* UCC2003 was routinely cultured in Reinforced Clostridial Medium (RCM; Oxoid Ltd., Basingstoke, Hampshire, United Kingdom). Carbohydrate utilisation by *B. breve* UCC2003 was examined in modified deMan Rogosa Sharpe (mMRS) medium made from first principles (De Man *et al.*, 1960), excluding a carbohydrate source, supplemented with 0.05 % (wt/vol) L-cysteine HCl and a particular carbohydrate source (0.5 % wt/vol). The carbohydrates used were lactose (Sigma Aldrich, Steinheim, Germany) and GlcNAc-6-S (Dextra Laboratories, Reading, United Kingdom). In order to determine bacterial growth profiles and final optical densities, 10 ml of freshly prepared mMRS medium, supplemented with a particular carbohydrate, was inoculated with 100 µl (1 %) of a stationary-phase culture of a particular strain. Un-inoculated mMRS was used as a negative control. Cultures were incubated anaerobically for 24 h and the optical density (OD_{600nm}) was recorded. Bifidobacterial cultures were incubated under anaerobic conditions in a modular atmosphere-controlled system (Davidson and Hardy, Belfast, Ireland) at 37°C. *Escherichia coli* was cultured in Luria Bertani broth (LB) at 37°C with agitation (Sambrook, 1989). *Lactococcus lactis* strains were grown in M17 medium supplemented with 0.5 % (wt/vol) glucose at 30°C (Terzaghi & Sandine, 1975). Where appropriate, growth media contained tetracycline (Tet; 10 µg ml^{-1}), chloramphenicol (Cm; 5 µg ml^{-1} for *E. coli* and *L. lactis*, 2.5 µg ml^{-1} for *B. breve*), ampicillin (Amp; 100 µg ml^{-1}), erythromycin (Em; 100 µg ml^{-1}) or kanamycin (Kan; 50 µg ml^{-1}). Recombinant *E. coli* cells containing pORI19 were selected on LB agar containing Em and Kan, and supplemented with 40 µg ml^{-1} X-gal (5-bromo-4-

chloro-3-indolyl-β-D-galactopyranoside) and 1 mM IPTG (isopropyl-β-D-galactopyranoside).

Nucleotide sequence analysis. Sequence data were obtained from the Artemis-mediated genome annotations of *B. breve* UCC2003 (O'Connell Motherway *et al.*, 2011b; Rutherford *et al.*, 2000). Database searches were performed using the non-redundant sequence database accessible at the National Centre for Biotechnology Information website (http://www.ncbi.nlm.nih.gov) using BLAST (Altschul *et al.*, 1990). Sequence analysis was performed using the Seqbuilder and Seqman programs of the DNASTAR software package (DNASTAR, Madison, WI, USA). Inverted repeats were identified using the PrimerSelect program of the DNASTAR software package and a graphical representation of the identified motifs was obtained using WebLogo software (Crooks *et al.*, 2004).

***DNA manipulations*.** Chromosomal DNA was isolated from *B. breve* UCC2003 as previously described (Riordan, 1998). Plasmid DNA was isolated from *E. coli*, *L. lactis* and *B. breve* using the Roche High Pure plasmid isolation kit (Roche Diagnostics, Basel, Switzerland). An initial lysis step was performed using 30 mg ml^{-1} of lysozyme for 30 min at 37°C prior to plasmid isolation from *L. lactis* or *B. breve*. Single stranded oligonucleotide primers used in this study were synthesised by Eurofins (Ebersberg, Germany) (Table 5.2). Standard PCRs were performed using Taq PCR master mix (Qiagen GmBH, Hilden, Germany). *B. breve* colony PCRs were carried out as described previously (O'Connell Motherway *et al.*, 2009). PCR fragments were purified using the Roche High Pure PCR purification kit

(Roche Diagnostics). Electroporation of plasmid DNA into *E. coli*, *L. lactis* or *B. breve* was performed as previously described (Maze *et al.*, 2007; Sambrook, 1989; Wells *et al.*, 1993).

***Construction of B. breve UCC2003 insertion mutants*.** Internal fragments of Bbr_0351, designated here as *atsR1* (this fragment encompassed 404 base pairs [bp], and represented codon numbers 102 through to 236 of the 354 codons of this gene), Bbr_0849, designated here as *atsR2* (fragment encompassed 408 bp, representing codon numbers 134 through to 271 of the 395 codons of this gene), Bbr_0851, designated *atsT* (fragment encompassed 416 bp, representing codon numbers 149 through to 288 of the 476 codons of this gene) and Bbr_0852, designated *atsA2* (fragment encompassed 402 bp, representing codon numbers 148 through to 281 of the 509 codons of this gene) were amplified by PCR using *B. breve* UCC2003 chromosomal DNA as a template and primer pairs atsR1F and atsR1R, atsR2F and atsR2R, atsTF and atsTR and atsA2F and atsA2R, respectively (Table 5.2). The insertion mutants were constructed as described previously (O'Connell Motherway *et al.*, 2009). Site-specific recombination of potential Tet-resistant mutant isolates was confirmed by colony PCR using primer combinations TetWF and TetWR to verify *tetW* gene integration, and the primers atsR1confirm, atsR2confirm, atsTconfirm and atsA2confirm (positioned upstream of the selected internal fragments of *atsR1*, *atsR2*, *atsT* and *atsA2*, respectively) in combination with primer TetWF to confirm integration at the correct chromosomal location.

Analysis of global gene expression using B. breve DNA microarrays. Global gene expression was determined during log-phase growth of *B. breve* UCC2003 in mMRS supplemented with 0.5 % GlcNAc-6-S and the obtained transcriptome was compared to that obtained from *B. breve* UCC2003 grown in mMRS supplemented with 0.5 % ribose. Similarly, global gene expression of the insertion mutants *B. breve* UCC2003-atsR1 and *B. breve* UCC2003-atsR2 was determined during log-phase growth of the mutants in mMRS supplemented with 0.5 % ribose and these transcriptomes were also compared to that from *B. breve* UCC2003 grown in 0.5 % ribose. DNA microarrays containing oligonucleotide primers representing each of the 1864 identified open reading frames on the genome of *B. breve* UCC2003 were designed and obtained from Agilent Technologies (Palo Alto, Ca., USA). Methods for cell disruption, RNA isolation, RNA quality control, complementary DNA (cDNA) synthesis and labelling were performed as described previously (Zomer *et al.*, 2009). Labelled cDNA was hybridised using the Agilent Gene Expression hybridisation kit (part number 5188-5242) as described in the Agilent Two-Colour Microarray-Based Gene Expression Analysis v4.0 manual (publication number G4140-90050). Following hybridisation, microarrays were washed in accordance with Agilent's standard procedures and scanned using an Agilent DNA microarray scanner (model G2565A). Generated scans were converted to data files with Agilent's Feature Extraction software (Version 9.5). DNA-microarray data were processed as previously described (García de la Nava *et al.*, 2003; van Hijum *et al.*, 2003; van Hijum *et al.*, 2005). Differential expression tests were performed with the Cyber-T implementation of a variant of the *t*-test (Long *et al.*, 2001).

***Plasmid Constructions*.** For the construction of plasmids pNZ-atsR1 and pNZ-atsR2, DNA fragments encompassing the complete coding region of the predicted transcriptional regulators *atsR1* (Bbr_0351) and *atsR2* (Bbr_0849) were generated by PCR amplification from chromosomal DNA of *B. breve* UCC2003 using PFUII DNA polymerase and the primer combinations atsR1FOR and atsR1REV, and atsR2FOR and atsR2REV, respectively (Table 5.2). The generated amplicons were digested with NcoI and XbaI, and ligated into the similarly digested, nisin-inducible translational fusion plasmid pNZ8048 (Mierau & Kleerebezem, 2005). The ligation mixtures were introduced into *L. lactis* NZ9000 by electrotransformation and transformants were selected based on Cm resistance. The plasmid content of a number of Cm^r transformants was screened by restriction analysis and the integrity of positively identified clones was verified by sequencing.

For the construction of the plasmids pQE-60-atsA2 and pBC1.2-atsB2, DNA fragments encompassing the complete coding region of the predicted sulfatase-encoding gene *atsA2* (Bbr_0852) and the predicted anSME-encoding gene *atsB2* (Bbr_0853) were generated by PCR amplification from *B. breve* UCC2003 chromosomal DNA using PFUII DNA polymerase and the primer pairs AtsA2FOR and AtsA2REV, and AtsB2FOR and AtsB2REV, respectively (Table 5.2). The amplified 1.5 kb *atsA2*-encompassing fragment was digested with NcoI and BglII, and ligated to the similarly digested pQE-60 plasmid. The ligation mixtures were introduced into *E. coli* XL1-blue by electrotransformation and transformants were selected based on Tet and Amp resistance. The amplified 1.4 kb *atsB2*-encompassing fragment was digested with HindIII and XbaI, and ligated to the similarly digested pBC1.2 plasmid (Álvarez-Martín *et al.*, 2007). The ligation mixtures were introduced into *E. coli* XL1-blue by electrotransformation and transformants were

selected based on Tet and Cm resistance. Transformants were screened by restriction analysis and the integrity of positively identified clones was verified by sequencing. A representative pBC1.2-atsB2 plasmid was introduced into *E. coli* XL1-blue-pQE-60-atsA2 by electrotransformation and transformants were selected based on Tet, Amp and Cm resistance.

To clone the Bbr_0849 promoter region, a DNA fragment encompassing the intergenic region between the Bbr_0849 and Bbr_0850 genes was generated by PCR amplification from *B. breve* UCC2003 chromosomal DNA using PFUII DNA polymerase and the primer pair, atsPromF and atsPromR (Table 5.2). The PCR product was digested with HindIII and XbaI, and ligated to the similarly digested pBC1.2 (Álvarez-Martín *et al.*, 2007). The ligation mixture was introduced into *E. coli* XL1-blue by electrotransformation and transformants were selected based on Tet and Cm resistance. Transformants were checked for plasmid content by restriction analysis and the integrity of positively identified clones was verified by sequencing. Plasmid pBC1.2-atsProm was introduced into *B. breve* UCC2003-atsR2 by electrotransformation and transformants were selected based on Tet and Cm resistance.

Protein overproduction and purification. For the heterologous expression of AtsR1 and AtsR2, 25 ml of M17 broth supplemented with 0.5 % (wt/vol) glucose was inoculated with a 2 % inoculum of *L. lactis* NZ9000 harbouring either pNZ-atsR1, pNZ-atsR2 or the empty vector pNZ8048 (used as a negative control), followed by incubation at 30°C until an OD_{600nm} of 0.5 was reached, at which point protein expression was induced by addition of cell-free supernatant of a nisin-producing

strain (de Ruyter *et al.*, 1996), followed by continued incubation for a further 2 h. Cells were harvested by centrifugation, resuspended in 10 mM Tris-HCl (pH 8.0), and disrupted with glass beads in a mini-bead beater (BioSpec Products, Bartlesville, OK). Cellular debris was removed by centrifugation to produce a crude cell extract.

For the heterologous expression of AtsA2, AtsB2, or the co-expression of AtsA2 and AtsB2, 2 % inocula of the generated *E. coli* XL1-blue derivatives harbouring either pQE-60-atsA2, pBC1.2-atsB2 or pQE-60-atsA2-pBC1.2-atsB2 (see above) were used to inoculate a 25 ml volume of LB supplemented with the appropriate antibiotics. Cultures were incubated at 37°C until an OD_{600nm} of 0.5 was reached, at which point protein expression was induced by the addition of 1 mM IPTG, followed by further incubation for 2 h. Preparation of crude cell extract was performed as described above.

***Enzyme assays*.** The predicted sulfatase activity of AtsA2 was analysed following IPTG-induced expression in the presence or absence of the predicted anSME, AtsB2. A 50 µl volume of crude cell extract of *E. coli* XL1-blue derivatives containing either pQE-60-atsA2, pBC1.2-atsB2 (used as a negative control) or pQE-60-atsA2-pBC1.2-atsB2 was incubated with 0.1 mg ml^{-1} of GlcNAc-6-S. Enzymatic assays were performed at 37°C in a total volume of 500 µl in a reaction buffer containing 50 mM morpholinepropanesulfonic acid (MOPS, pH 7.0). The reactions were terminated after 24 h by incubation at 65°C for 10 min and were stored at -20°C prior to high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) analysis (see below).

High-Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). HPAEC-PAD was performed using a Dionex ICS-3000 system (Dionex, Sunnyvale, CA). Carbohydrate fractions (25 µl aliquots) were separated on a CarboPac PA1 (Dionex) analytical-exchange column (250 mm x 4 mm) with a CarboPac PA1 guard column (50 mm x 4 mm) and a pulsed electrochemical detector in the pulsed amperometric detection (PAD) mode. The elution was performed at a constant flow-rate of 1.0 ml min^{-1} at 30°C using the following eluents for the analysis (A) 200 mM NaOH, (B) 100 mM NaOH, 550 mM sodium acetate, and (C) Milli-Q water. The following linear gradient of NaAc was used: 100 mM NaOH: 0-50 min, 0 mM; 50-51 min, 16 mM; 51-56 min, 100 mM; 56.1-61 min, 0 mM. Detection was achieved using a Dionex ED40 detector in PAD mode. The Chromeleon software-version 6.70 (Dionex Corporation) was used for the integration and evaluation of the chromatograms obtained. The chromatographic profiles corresponding to *N*-acetylglucosamine (GlcNAc) and GlcNAc-6-S were qualitatively examined to evaluate sulfatase activity.

Electrophoretic mobility shift assays (EMSA). DNA fragments representing different portions of each of the promoter regions upstream of the *atsA1*, Bbr_0353, *atsR2* and *atsT* genes were prepared by PCR using IRD-labelled primer pairs (Table 5.3). EMSAs were essentially performed as described previously (Hamoen *et al.*, 1998). In all cases, the binding reactions were performed in a final reaction volume of 20 µl in the presence of poly (dI-dC) in binding buffer (20 mM Tris-HCl, 5 mM $MgCl_2$, 0.5 mM dithiothreitol (DTT), 1 mM EDTA, 50 mM KCl, 10 % glycerol at pH 7.0). Various amounts of *L. lactis* NZ9000 crude cell extract containing pNZ-atsR1, pNZ-atsR2 or pNZ8048 were mixed on ice with a fixed amount of DNA

probe (0.1 pmol) and subsequently incubated for 30 min at 37°C. Samples were loaded on a 6 % non-denaturing polyacrylamide (PAA) gel prepared in TAE buffer (40 mM Tris acetate (pH 8.0), 2 mM EDTA) and run in a 0.5 to 2.0 x gradient of TAE at 100 V for 120 min in an Atto Mini PAGE system (Atto Bioscience and Biotechnology, Tokyo, Japan). Signals were detected using an Odyssey Infrared Imaging System (Li-Cor Biosciences, United Kingdom Ltd., Cambridge, United Kingdom) and images were captured using the supplied Odyssey software v3.0. To determine the effector molecule of AtsR2, GlcNAc or GlcNAc-6-S was added to the binding reaction in concentrations ranging from 2.5 mM to 20 mM.

Primer extension analysis. Total RNA was isolated from exponentially growing cells of *B. breve* UCC2003-atsR1, *B. breve* UCC2003-atsR2 or *B. breve* UCC2003-atsR2-pBC1.2-atsProm in mMRS supplemented with 0.5 % ribose, as previously described (Zomer *et al.*, 2009). Primer extension was performed by annealing 1 pmol of an IRD-labelled synthetic oligonucleotide to 20 µg of RNA as previously described (Ventura *et al.*, 2005), using primers AtsR1F1F, AtsR1F1R, AtsR2R1F or AtsR2T1R (Table 5.3). Sequence ladders of the presumed *atsA1*, Bbr_0353, *atsR2* and *atsT* promoter regions were produced using the same primer as in the primer extension reaction and a DNA cycle-sequencing kit (Jena Bioscience, Germany) and were run alongside the primer extension products to allow precise alignment of the transcriptional start site with the corresponding DNA sequence. Separation was achieved on a 6.5 % Li-Cor Matrix KB Plus acrylamide gel. Signal detection and image capture were performed with a Li-Cor sequencing instrument (Li-Cor Biosciences).

5.4 RESULTS

***Genetic organisation of the sulfatase gene clusters in B. breve UCC2003*.** Based on the presence of a sulfatase-associated PFAM domain PF00884 and the previously described N-terminally located sulfatase signature (CxPxR) (Dierks *et al.*, 1999; Sardiello *et al.*, 2005), two putative Cys-type sulfatase-encoding genes were identified on the genome of *B. breve* UCC2003. The first, repesented by the gene with the associated locus tag Bbr_0352 (and here designated as *atsA1*) is located in a cluster of four genes, designated the *ats1* cluster, which also includes a gene encoding a predicted hypothetical membrane spanning protein (Bbr_0349), a gene (Bbr_0350, designated here as *atsB1*) specifying a putative anSME which contains the signature motif CxxxCxxC characteristic of the radical AdoMet-dependent superfamily (Sofia *et al.*, 2001), and a gene specifying a predicted LacI-type transriptional regulator (Bbr_0351, designated *atsR1*). Adjacent to this gene cluster, a gene cluster encoding a predicted ABC-type transport system was identified (corresponding to locus tags Bbr_0353 through to Bbr_0355) (Fig. 5.1).

The second predicted sulfatase-encoding gene, Bbr _0852 (designated *atsA2* here), is located in another cluster of four genes (Bbr_0851 through to Bbr_0854), which was designated the *ats2* gene cluster. Bbr_0851, designated here as *atsT*, encodes a predicted transporter from the major facilitator superfamily. Bbr_0853 (designated *atsB2*) encodes a second putative anSME, which also contains the signature CxxxCxxC motif. Bbr_0854 encodes a predicted membrane spanning protein of unknown function (Fig. 5.1). The AtsA1 and AtsA2 proteins share 28 % amino acid identity, while the AtsB1 and AtsB2 proteins exhibit 74 % identity between each other. The *ats2* gene cluster has a notably different GC content (63.96 %) compared to the *B. breve* UCC2003 genome average (58.73 %), suggesting that this gene

cluster may have been acquired by horizontal transfer, whereas the GC content of the *ats1* cluster (57.6 %) is comparable to that of the genome.

BLASTP analysis revealed that the presence of potential sulfatase and anSME-encoding genes in bifidobacteria is limited to the *B. breve* species. Based on the comparative genome analysis presented in Figure 5.1, we found that the putative sulfatase clusters are well conserved among the *B. breve* strains whose genomes were recently published (Bottacini *et al.*, 2014). Of the currently available complete *B. breve* genomes, *B. breve* NCFB2258, *B. breve* 689B, *B. breve* 12L and *B. breve* S27 encode clear homologues of both identified putative sulfatase gene clusters described above. In contrast, the genomes of *B. breve* JCM7017, *B. breve* JCM7019 and *B. breve* ACS-071-V-Sch8b contain just a single, but variable putative sulfatase cluster (Fig. 5.1).

Growth of B. breve UCC2003 in mMRS supplemented with 0.5 % GlcNAc-6-S.

The presence of two putative sulfatase-encoding clusters on the genome of *B. breve* UCC2003 suggests that this gut commensal is capable of removing a sulfate ester from a sulfated compound, possibly a sulfated carbohydrate. Therefore, in order to determine if *B. breve* UCC2003 is capable of metabolising the sulfated amino sugar GlcNAc-6-S (the only commercially available sulfated monosaccharide that could be obtained in affordable quantities for growth experiments), growth of the strain in mMRS supplemented with either 0.5 % lactose or GlcNAc-6-S as the sole carbon source was assessed by measuring the OD_{600nm} after 24 h of anaerobic growth at 37°C. The strain was shown to grow well on lactose (OD_{600nm} > 1.5), while it was also capable of reasonable growth on GlcNAc-6-S (OD_{600nm} > 0.8) (Fig. 5.2). This

result is consistent with the presence of predicted sulfatase and anSME-encoding genes on the *B. breve* UCC2003 genome.

Genome response of B. breve UCC2003 to growth on GlcNAc-6-S. In order to investigate which genes are responsible for GlcNAc-6-S metabolism in *B. breve* UCC2003, global gene expression was determined by microarray analysis during growth of the strain in mMRS supplemented with GlcNAc-6-S and compared with gene expression when grown in mMRS supplemented with ribose. Ribose was considered an appropriate carbohydrate for comparative transcriptome analysis because the genes involved in ribose metabolism are known (Pokusaeva *et al.*, 2010), while it has furthermore been used in a number of transcriptome studies in this strain (Egan *et al.*, 2014a; Egan *et al.*, 2014b; O'Connell *et al.*, 2013; O'Connell *et al.*, 2014; O'Connell Motherway *et al.*, 2013). Of the two predicted sulfatase and anSME-encoding gene clusters of *B. breve* UCC2003 (see above), the *ats2* gene cluster was significantly up-regulated (fold change >3.0, *P*-value of <0.001) during growth on GlcNAc-6-S, while no (significant) difference in the level of transcription was observed for the *ats1* gene cluster (Table 5.4). Three other gene clusters of interest were also significantly up-regulated (corresponding to locus tags Bbr_0846 through to Bbr_0849, Bbr_1585 through to Bbr_1590, and Bbr_1247 through to Bbr_1249; see Fig. 5.3 and Table 5.4).

Within the Bbr_0846-0849 gene cluster, which is separated from the *ats2* cluster by a single gene (Fig. 5.3), Bbr_0846 (*nagA1*) and Bbr_0847 (*nagB2*) are predicted to encode an *N*-acetylglucosamine-6-phosphate deacetylase and a glucosamine-6-phosphate deaminase, respectively. Bbr_0848 encodes a predicted ROK-family

kinase (designated *nagK*) which contains the characteristic DxGxT motif at the N-terminal (Conejo *et al.*, 2010). It exhibits 42 % similarity at protein level with the previously described ROK-family NagK characterised from *E. coli* K-12 which phosphorylates GlcNAc producing *N*-acetylglucosamine-6-phosphate (GlcNAc-6-P) (Uehara & Park, 2004). Therefore this cluster encodes putative enzymes for the complete GlcNAc catabolic pathway as previously described in *E. coli*, whereby GlcNAc is first phosphorylated by NagK producing GlcNAc-6-P, followed by NagA-mediated deacetylation to produce glucosamine-6-phosphate, and the NagB-mediated deamination and isomerisation to produce fructose-6-phosphate (Uehara & Park, 2004; White, 1968). Bbr_0849 encodes a predicted transcriptional regulator from the ROK family (designated here as *atsR2*).

The Bbr_1585-1590 cluster includes a predicted UDP-glucose-4-epimerase (Bbr_1585, *galE*), a predicted *N*-acetylhexosamine-1-kinase (Bbr_1586, *nahK*) and a predicted lacto-*N*-biose phosphorylase (Bbr_1586, *lnbP*), representing three of the four enzymes required for the degradation of galacto-*N*-biose (Galβ1-3GalNAc; GNB), which is found in mucin, or lacto-*N*-biose (Galβ1-3GlcNAc; LNB), a known HMO (Kitaoka *et al.*, 2005; Nishimoto & Kitaoka, 2007). The final three genes of this cluster encode a predicted ABC transport system, including two predicted permease proteins and a solute binding protein, respectively (Fig. 5.3). This gene cluster was previously shown to be transcriptionally up-regulated when *B. breve* UCC2003 was grown in co-culture with *B. bifidum* PRL2010 in mucin (Egan *et al.*, 2014b).

Finally, the Bbr_1247-1249 cluster contains a gene specifying an *N*-acetylglucosamine-6-phosphate deacetylase (Bbr_1247) and a glucosamine-6-phosphate deaminase (Bbr_1248)-encoding gene, designated *nagA2* and *nagB3*,

respectively. These genes were previously up-regulated during *B. breve* UCC2003 growth on sialic acid (Egan *et al.*, 2014a). The NagA1 protein shares a 74 % identity with NagA2, while the NagB2 protein shares 84 % identity with NagB1 of the *nan/nag* cluster for sialic acid metabolism (Egan *et al.*, 2014a) and 84 % identity with NagB3. Bbr_1249 encodes a predicted transcriptional regulator from the ROK family (Fig. 5.3).

Disruption of the atsT and atsA2 genes. In order to investigate if disruption of individual genes from the *ats2* gene cluster would affect the ability of *B. breve* UCC2003 to utilise GlcNAc-6-S, insertion mutants were constructed in the *atsT* and *atsA2* genes, resulting in the strains *B. breve* UCC2003-atsT and *B. breve* UCC2003-atsA2, respectively (see Materials and Methods). The insertion mutants were analysed for their ability to grow in mMRS supplemented with GlcNAc-6-S as compared to *B. breve* UCC2003. As expected, and in contrast to the wild type, there was a complete lack of growth of *B. breve* UCC2003-atsT and *B. breve* UCC2003-atsA2 in media containing GlcNAc-6-S as the sole carbon source (Fig. 5.2), thus demonstrating the involvement of the disrupted genes in GlcNAc-6-S metabolism. However, it should be noted that construction of the insertion mutants in either of these genes is likely to have a polar effect on the downstream genes *atsB2* and Bbr_0854 (Fig. 5.1), which may also contribute to the lack of growth of the mutant strains. Growth of the insertion mutants was not impaired on lactose, where all strains reached final OD_{600nm} levels comparable to that reached by the wild type strain (Fig. 5.2).

Biochemical analysis. In order to *in vitro* test the presumed sulfatase activity of AtsA2 and the role (if any) of AtsB2 in relation to AtsA2 activity, the predicted sulfatase and anSME-encoding genes *atsA2* and *atsB2* were cloned into separate, compatible plasmids pQE-60 and pBC1.2 to produce pQE-60-atsA2 and pBC1.2-atsB2, with both genes under the control of the IPTG-inducible *lac* promoter (see Materials and Methods). Three *E. coli* XL1-blue derived strains were generated, harbouring pQE-60-atsA2, pBC1.2-atsB2 or pQE-60-atsA2-pBC1.2-atsB2. Following IPTG induction, crude cell extracts of these three strains were obtained and then used to investigate sulfatase activity. Under the HPAEC-PAD conditions used, it was not possible to visualise GlcNAc-6-S, yet the assumed product GlcNAc was identifiable. As expected, accumulation of GlcNAc was not observed when the substrate was incubated with the negative control, pBC1.2-atsB2. Incubation of crude cell extract containing pQE-60-atsA2-pBC1.2-atsB2 with GlcNAc-6-S resulted in an accumulation of GlcNAc, yet no such accumulation was observed when crude cell extract containing pQE-60-atsA2 was used (Fig. 5.4). This implies that *atsA2* must be co-expressed with the anSME-encoding *atsB2* in order to produce an active sulfatase enzyme. This is likely to be achieved through the modification of the Cys_{64} residue of AtsA2 to an FGly residue by AtsB2, as previously described for the *atsB* homologues from *K. pneumoniae, C. perfringens* and *Ba. thetaiotaomicron* (Benjdia *et al.*, 2008; Berteau *et al.*, 2006; Szameit *et al.*, 1999).

Transcriptome of B. breve UCC2003-atsR1 and B. breve UCC2003-atsR2. The Bbr_0846-0849 gene cluster, which is up-regulated when *B. breve* UCC2003 is grown on GlcNAc-6-S, and the *ats2* gene cluster are separated by just a single gene (Fig. 5.1). As mentioned above, the Bbr_0849 gene encodes a predicted ROK-type

transcriptional regulator and was thus designated *atsR2*. To investigate if this predicted transcriptional regulator controls transcription of the *ats2* and/or the Bbr_0846-0849 gene clusters, an insertion mutant was constructed, resulting in the strain *B. breve* UCC2003-atsR2. It was hypothesised that if this gene encoded a repressor, mutation of the gene would lead to increased transcription of the genes it controls even in the absence of the inducing carbohydrate. Microarray data revealed that in comparison to *B. breve* UCC2003, the *ats2* operon was significantly up-regulated (>9.0 fold change; P <0.001) in the mutant strain, thus identifying *atsR2* as a transcriptional repressor (Table 5.5; Fig. 5.5). The Bbr_0846-0849 gene cluster was clearly down-regulated in the mutant strain, however, since the *atsR2* gene represents the first gene of this presumed operon (Fig. 5.5), a polar effect on downstream located genes can be expected. Part of the *B. breve* UCC2003 locus responsible for the production of exopolysaccharide (EPS), Bbr_0435-0441, was also down-regulated in this insertion mutant, most likely due to the deletion of this cluster. This is an occasional occurrence in the construction of insertion mutants in this strain, possibly a stress-related effect of homologous recombination.

As mentioned above, the *ats1* operon includes a predicted LacI-type transcriptional regulator, designated *atsR1* (Fig. 5.1). In order to determine if this predicted regulator controls transcription of the *ats1* operon, an insertion mutant was constructed resulting in strain *B. breve* UCC2003-atsR1. Microarray data revealed that in comparison to *B. breve* UCC2003, the *atsA1* and *atsR1* genes as well as the predicted ABC-transport system encoded by the Bbr_0353-0355 genes, were significantly up-regulated (>9.0 fold change; P <0.001) in the mutant strain (Table 5.5; Fig. 5.5). It is likely that the insertion mutant has a polar effect on the transcription of the two genes downstream of *atsR1*, namely *atsB1* and Bbr_0347

(Fig. 5.5), and it is thus possible that these genes are also co-transcribed with *atsA1* and *atsR1* in the wild type strain.

***Electrophoretic mobility shift assays*.** In order to determine if the AtsR1 and AtsR2 proteins directly interact with promoter regions of the *ats1* and *ats2* gene clusters respectively, crude cell extracts of *L. lactis* NZ9000-pNZ-atsR1 or *L. lactis* NZ9000-pNZ-atsR2 were used to perform EMSAs, with crude cell extracts of *L. lactis* NZ9000-pNZ8048 (empty vector) used as a negative control. As expected, the negative control did not alter the electrophoretic behaviour of any of the tested DNA fragments (Fig 5.6B, 5.7B). The results obtained with crude cell extract overexpressing AtsR2 demonstrate that this presumed regulator binds to DNA fragments encompassing the upstream regions of *atsR2* and *atsT* (Fig. 5.6A, 5.6B). Dissection of the promoter region of *atsR2* showed that AtsR2 binding required a 184 bp region within which a 21 bp imperfect inverted repeat was identified. Similarly, dissection of the *atsT* promoter region revealed that AtsR2 binding required a 192 bp region which also includes a 21 bp imperfect inverted repeat, similar to that identified upstream of *atsR2*. When either of the inverted repeats were excluded, binding of AtsR2 to such DNA fragments was prevented, suggesting that these inverted repeats contained the operator sequence of AtsR2 (Fig. 5.6A, 5.6B).

To demonstrate if AtsR2 binding to its DNA target is affected by the presence of a carbohydrate effector molecule, GlcNAc and GlcNAc-6-S were tested for their effects on the formation of the AtsR2-DNA complex. The ability of AtsR2 to bind to the promoter regions of *atsR2* or *atsT* was abolished in the presence of 2.5 mM GlcNAc-6-S, the lowest concentration used in this assay. The presence of GlcNAc

was shown to inhibit binding of AtsR2 to the *atsR2* and *atsT* promoter regions, yet only at GlcNAc concentrations above 5 mM (Fig. 5.6C). This suggests that while GlcNAc-6-S has the highest affinity for the regulator and is therefore the most likely effector of this repressor protein, the structurally similar GlcNAc is also able to bind this regulator, yet at concentrations that are probably not relevant from a cell physiology perspective.

The results obtained with crude cell extract containing AtsR1 demonstrate that AtsR1 binds to a DNA fragment which encompasses the intergenic region between *atsA1* and Bbr_0353, which is consistent with the transcriptome profile of *B. breve* UCC2003-atsR1 (Table 5.5, Fig. 5.7A, 5.7B). Dissection of this intergenic region revealed that AtsR1 binds to two distinct DNA fragments of 184 bp and 124 bp. Within these two fragments, a well conserved imperfect inverted repeat of 18 bp was identified. Further fragmentations to the point where the inverted repeats were excluded resulted in an absence of binding, suggesting that these inverted repeats represent the operator sequence of AtsR1 (Fig. 5.7A, 5.7B).

Identification of the transcription start sites of atsA1, Bbr_0353, atsR2 and atsT.
Based on the EMSA results and the transcriptomes of *B. breve* UCC2003-atsR1 and *B. breve* UCC2003-atsR2, it was deduced that the intergenic region between *atsA1* (Bbr_0352) and Bbr_0353 contains two AtsR1-dependent promoters, while an AtsR2-dependent promoter is located upstream of both *atsR2* and *atsT* (Fig. 5.5). In order to determine the transcriptional start site of these presumed promoters, primer extension analysis was performed using RNA extracted from *B. breve* UCC2003-atsR1, *B. breve* UCC2003-atsR2 or *B. breve* UCC2003-atsR2-pBC1.2-atsProm

grown in mMRS supplemented with 0.5 % ribose. Microarray analysis had shown that the expression levels of *atsA1* and Bbr_0353 were high when *B. breve* UCC2003-atsR1 was grown on ribose and similarly, *atsT* was highly up-regulated in the *B. breve* UCC2003-atsR2 insertion mutant. For this reason, and because the inducer of the *ats1* cluster is still unknown, the mutant strains were considered most suitable for primer extension analyis. For the *atsR2* promoter region, initial attempts to attain a primer extension product from mRNA isolated from *B. breve* UCC2003-atsR2 cells were unsuccessful. In an attempt to increase the amount of mRNA transcripts of this promoter region, a DNA fragment encompassing the promoter region was cloned into pBC1.2 and introduced into *B. breve* UCC2003-atsR2, generating the strain *B. breve* UCC2003-atsR2-pBC1.2-atsProm. Single extension products were identified upstream of *atsR2* and *atsT* (Fig. 5.8A, 5.8B). Two extension products, only differing one nucleotide in length, were identified upstream of both *atsA1* and Bbr_0353 (Fig. 5.8C, 5.8D). Potential promoter recognition sequences resembling consensus -10 and -35 hexamers were identified upstream of each of the transcription start sites (Fig. 5.8). In all cases, the deduced operator sequences of AtsR1 or AtsR2 overlap with the respective -35 or -10 sequences, consistent with our findings that both AtsR1 and AtsR2 act as transcriptional repressors.

5.5 DISCUSSION

Metagenomic sequence analysis has revealed that genes predicted to encode sulfatases and anSMEs are enriched in the gut microbiomes of humans compared to non-gut microbial communities (Kurokawa *et al.*, 2007). *Ba. thetaiotaomicron* has been used as a model strain for interesting studies into the importance of sulfatase activity to bacterial fitness *in vivo.* In the earlier studies of chondroitin sulfate and heparan sulfate metabolism by this species, mutagenesis of a gene designated *chuR*, which was then predicted to encode a regulatory protein, resulted in the inability to compete with wild type *Ba. thetaiotaomicron* in germ free mice (Cheng *et al.*, 1992). The *chuR* gene was later found to encode an anSME (Benjdia *et al.*, 2008). Recently, this anSME was again shown to have significant importance in this strain's ability to colonise the gut, as an isogenic derivative of this strain (designated ΔanSME) carrying a deletion in the anSME-encoding gene displayed reduced fitness *in vivo* even in the presence of dietary glycans (Benjdia *et al.*, 2011). This is a particularly interesting observation given that it was previously shown that *Ba. thetaiotaomicron* preferentially utilises plant polysaccharides over host glycans, even when both are available (Sonnenburg *et al.*, 2005). The authors speculated that anSME activity and associated sulfatase activity are important as the bacterium adapts to the gut environment (Benjdia *et al.*, 2011).

The current study describes two gene clusters in *B. breve* UCC2003, each encoding a (predicted) sulfatase and accompanying anSME, as well as an associated transport system and transcriptional regulator. One of these gene clusters, *ats2*, is shown to be required for the metabolism of GlcNAc-6-S, a monosaccharide component of human colonic mucin and heparan sulfate (Hopwood & Elliott, 1983; Larsson *et al.*, 2009). The substrate for the enzymes encoded by the *ats1* gene cluster has yet to be

elucidated. We speculate that this cluster may be involved in the metabolism of other sulfated monosaccharides such as *N*-acetylgalactosamine-6-sulfate, which is found in chondroitin sulfate (Eliakim *et al.*, 1986). It is also possible that the sulfatase enzyme encoded by the *atsA1* gene is specific for other linkages between the sulfate ester and the monosaccharide. Four sulfatase enzymes recently characterised from *Ba. thetaiotaomicron* were shown to have strict specificity for either 2-*O*-sulfated, 4-*O*-sulfated or 6-*O*-sulfated substrates (Ulmer *et al.*, 2014). Therefore it would be interesting to test the growth and transcriptome response of *B. breve* UCC2003 to *N*-acetylglucosamine-3-sulfate or *N*-acetylgalactosamine-3-sulfate.

Interestingly, the two gene clusters, *ats1* and *ats2*, are quite dissimilar in terms of their genetic organsiation. The organisation of the *ats1* cluster resembles that of a typical bifidobacterial carbohydrate utilisation cluster as it includes genes encoding a predicted ABC-type transport system, a LacI-type repressor (*atsR1*) and the potentially carbohydrate-active *atsA1*-encoded sulfatase and *atsB*-encoded anSME, which in this case replace the typical glycosyl hydrolase-encoding gene(s) (Schell *et al.*, 2002). The *atsA1*, *atsR1* and *atsB1* genes, and the genes specifying the ABC-type transport system are located on adjacent, divergent gene clusters, yet their promoters are co-regulated by AtsR1. AtsR1 shares many of the features of a typical LacI-type transcription factor, as (i) it solely regulates adjacent genes representing a single carbohydrate catabolic pathway, (ii) it is autoregulated and (iii) its predicted binding sequence contains the conserved CG pair at its centre and is located within 30 and 140 bp of the start codon (Ravcheev *et al.*, 2014).

In the *ats2* cluster, the *atsT* gene encodes a predicted transporter of the major facilitator superfamily, while the *atsA2* and *atsB2* genes are adjacent, similar to their homologous genes in *K. pneumoniae* and *Prevotella* strain RS2 (Murooka *et al.*,

1990; Wright *et al.*, 2000). We obtained compelling evidence that the *ats2* cluster is co-regulated with the Bbr_0846-0849 cluster by the ROK-family transcriptional repressor AtsR2. The only previously characterised bifidobacterial ROK-family transcriptional regulator is RafA, the transcriptional activator of the raffinose utilisation cluster in *B. breve* UCC2003 (O'Connell *et al.*, 2014). The Bbr_0846-0848 genes are presumably involved in the metabolism of GlcNAc following the removal of the sulfate residue from GlcNAc-6-S. The fructose-6-phosphate produced from GlcNAc by the combined activities of NagK, NagA1 and NagB2 is expected to enter the fructose-6-phosphate phosphoketolase pathway or bifid shunt, the central metabolic pathway of bifidobacteria (Scardovi & Trovatelli, 1965). It is interesting that *B. breve* UCC2003 is capable of growth on GlcNAc-6-S as a sole carbon source but apparently not on GlcNAc (Pokusaeva *et al.*, 2011a). Since the *B. breve* UCC2003 genome seems to encode the enzymes required to metabolise GlcNAc, it suggests that the *atsT* transporter has (high) affinity for only the sulfated form of the sugar.

The function (if any) of the Bbr_0850 gene, which encodes a predicted aldose-6-phosphate epimerase and is located between the co-regulated Bbr_0846-0849 and *ats2* gene clusters has yet to be elucidated. We speculate that this enzyme plays a role in the metabolism of sulfated *N*-acetylgalactosamine (GalNAc), a component of chondroitin sulfate (Eliakim *et al.*, 1986), possibly by isomerisation of GalNAc-6-P to GlcNAc-6-P, similar to the isomerisation of *N*-acetylmannosamine-6-phosphate to GlcNAc-6-P in the sialic acid catabolic pathway by an *N*-acetylmannosamine-6-phosphate epimerase (Egan *et al.*, 2014a). This would also require the putative NagK ROK-family kinase to be active on both GlcNAc and GalNAc, or for a second kinase to be identifed elsewhere on the genome.

As mentioned above, sulfatase activity has not previously been described in the *Bifidobacterium* genus. In terms of the metabolism of host-derived carbohydrates such as mucin and HMO, the *B. breve* species has been described as a scavenger of mono- and oligosaccharides that are released by the activity of specialised mucin and HMO-degraders such as *B. bifidum* (Sela & Mills, 2010). Recent studies have shown that it employs this strategy to great effect, as *B. breve* UCC2003 utilises components of 3' sialyllactose (a HMO) and mucin following the degradation of these sugars by *B. bifidum* PRL2010, whereas in the absence of *B. bifidum* PRL2010, it is not capable of utilising either of these sugars as a sole carbon source (Egan *et al.*, 2014a; Egan *et al.*, 2014b). Nonetheless, it is intruiging that within the *Bifidobacterium* genus, sulfatase-encoding genes appear to be restricted to and conserved within the *B. breve* species, allowing it to utilise sulfated monosaccharides such as GlcNAc-6-S (and possibly others) which are inaccessible to other *Bifidobacterium* species. This, combined with recent knowledge that *B. breve* strains can utilise complex *N*-glycans similar to those found in human milk (Garrido *et al.*, 2012) and more constituents of HMO than was originally imagined (Ruiz-Moyano *et al.*, 2013), suggests that the description as a scavenger may be underselling the metabolic flexibility of this species. Instead, it possible that this species employs a multifaceted approach to carbohydrate metabolism in the (infant) gut.

Bacterial sulfatase activity has been well-described in the *Bacteroides* genus as well as *Prevotella* strain RS2 (Roberton *et al.*, 1993; Salyers *et al.*, 1977; Tsai *et al.*, 1991; Willis *et al.*, 1996). 28 predicted sulfatase-encoding genes were identified on the genome of *Ba. thetaiotaomicron*, 20 of which are predicted extracellular enzymes, yet the previously described *chuR* gene is the sole anSME-encoding gene

(Benjdia *et al.*, 2011). This gene was predicted to be constitutively expressed and the authors proposed that this would allow maturation of the various sulfatase proteins as they are produced in a substrate dependent manner (Benjdia *et al.*, 2011; Sonnenburg *et al.*, 2005; Sonnenburg *et al.*, 2006). Interestingly, a novel method of desulfating mucin which does not require a sulfatase enzyme was characterised from the *Prevotella* strain, whereby a sulfoglycosidase removes GlcNAc-6-S from purified porcine gastric mucin (Rho *et al.*, 2005). The presence of a signal sequence on this glycosulfatase, thus indicating extracellular activity, is interesting in relation to the current study, as it presents a source of GlcNAc-6-S to *B. breve* strains, suggestive of another cross-feeding opportunity for members of this species. This is particularly noteworthy when it is considered that the sulfatase enzymes produced by *B. breve* UCC2003 are intracellular, implying that *B. breve* UCC2003 is again reliant on the extracellular glycosyl hydrolase activity of other members of the gut microbiota in order to gain access to mucin-derived sulfated monosaccharides.

The current study investigated a previously unexplored area of carbohydrate metabolism in bifidobacteria, namely the metabolism of sulfated carbohydrates. While future studies will first focus on the identification on the substrate(s) of the *ats1* gene cluster, it would also be interesting to perform *in vivo* analysis, similar to that performed using *Ba. thetaiotaomicron*, of sulfatase or anSME insertion mutants of *B. breve* UCC2003 to investigate if these genes contribute to *B. breve* UCC2003 fitness in the gut.

5.6 ACKNOWLEDGEMENTS

The Alimentary Pharmabiotic Centre is a research centre funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan. The authors and their work were supported by SFI (Grant Nos. 07/CE/B1368 and SFI/12/RC/2273) and a HRB postdoctoral fellowship (Grant No. PDTM/20011/9) awarded to MOCM.

5.7 TABLES AND FIGURES

Table 5.1: Bacterial strains and plasmids used in this study.

Strains and plasmids	Relevant features	Reference or source
Strains		
***Escherichia coli* strains**		
***E.coli* EC101**	Cloning host; *repA*$^+$ *kmr*	(Law *et al.*, 1995)
***E.coli* EC101-pNZ-M.Bbrll+Bbr11**	EC101 harboring a pNZ8048 derivative containing bbrllM and bbrlllM	(O'Connell Motherway *et al.*, 2009)
***E. coli* XL1-blue**	*(supE44 hsdR17 recA1 gyrA96 thi relA1 lac* F′ [*proAB*$^+$ *lacI*^{q} *lacZΔM15* Tn*10*(Tet^r)])	Stratagene
***E. coli* XL1-blue-pQE-60-atsA2**	XL1-blue containing pQE-60-atsA2	This study
***E. coli* XL1-blue-pBC1.2-atsB2**	XL1-blue containing pBC1.2-atsB2	This study
***E. coli* XL1-blue-pQE-60-atsA2-pBC1.2-atsB2**	XL1-blue containing pQE-60-atsA2 and pBC1.2-atsB2	This study
***L. lactis* strains**		
***L. lactis* NZ9000**	MG1363, *pep*N::*nis*RK, nisin inducible overexpression host	(de Ruyter *et al.*, 1996)
***L. lactis* NZ9000-pNZ8048**	NZ9000 containing pNZ8048	This study
***L. lactis* NZ9000-pNZ-atsR1**	NZ9000 containing pNZ-atsR1	This study
***L. lactis* NZ9000-pNZ-atsR2**	NZ9000 containing pNZ-atsR2	This study
***L. lactis* NZ97000**	Nisin-A producing strain	(de Ruyter *et al.*, 1996)
***Bifidobacterium* sp. strains**		
***B. breve* UCC2003**	Isolate from a nursling stool	(Maze *et al.*, 2007)
***B. breve* UCC2003-atsR1**	pORI19-tetW-atsR1 insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-atsR2**	pORI19-tetW-atsR2 insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-atsT**	pORI19-tetW-atsT insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-atsA2**	pORI19-tetW-atsA2 insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-atsR2-pBC1.2-atsProm**	pORI19-tetW-atsR2 insertion mutant of UCC2003 containing pBC1.2-atsProm	This study
Plasmids		
pAM5	pBC1-pUC19-Tc^r	(Álvarez-Martín *et al.*, 2007)
pNZ8048	Cm^r, nisin-inducible translational fusion vector	(Mierau & Kleerebezem, 2005)
pNZ-atsR1	*atsR1* cloned downstream of nisin-inducible promoter on pNZ8048	This study
pNZ-atsR2	*atsR2* cloned downstream of nisin-inducible promoter on pNZ8048	This study
pORI19	Em^r, repA$^-$, ori$^+$, cloning vector	(Law *et al.*, 1995)
pORI19-tetW-atsR1	Internal 404 bp fragment of *atsR1* and *tetW* cloned in pORI19	This study
pORI19-tetW-atsR2	Internal 408 bp fragment of *atsR2* and *tetW* cloned in pORI19	This study

pORI19-tetW-atsT	Internal 416 bp fragment of *atsT* and *tetW* cloned in pORI19	This study
pORI19-tetW-atsA2	Internal 402 bp fragment of *atsA2* and *tetW* cloned in pORI19	This study
pQE-60	Amp^r, IPTG inducible vector	Qiagen
pQE-60-atsA2	*atsA2* cloned into the IPTG inducible vector, pQE-60	This study
pBC1.2	pBC1-pSC101-Cm^r	(Álvarez-Martín *et al.*, 2007)
pBC1.2-atsB2	*atsB2* cloned into pBC1.2	This study
pBC1.2-atsProm	*atsR2* promoter region cloned into pBC1.2	This study

Table 5.2: Oligonucleotide primers used in this study.

Purpose	Primer	Sequence
Cloning of 404 bp fragment of *atsR1* in pORI19	AtsR1F	TGATCA*AAGCTT*CGTCCGAGGAGAACAGCATC
	AtsR1R	TAGCTA*TCTAGA*CCACCCGACATCGATCAAC
Cloning of 408 bp fragment of *atsR2* in pORI19	AtsR2F	GACTAG*AAGCTT*GCCATCACGATCGACGACG
	AtsR2R	TAGCAT*TCTAGA*GCATCCCGGACGTCCACAG
Cloning of 416 bp fragment of *atsT* in pORI19	AtsTF	GACTAG*AAGCTT*GATCTCCTTCCGCCAGCTC
	AtsTR	TAGCAT*TCTAGA*CGTTGGTGCCGGTCAGCTG
Cloning of 402 bp fragment of *atsA2* in pORI19	AtsA2F	GACTAG*AAGCTT*GAATACGTCGCCTGGCTCAAG
	AtsA2R	TAGCAT*TCTAGA*CCTCCACTGGTCGTTGTCG
Amplification of *tetW*	TetWF	TCAGCT*GTCGAC*ATGCTCATGTACGGTAAG
	TetWR	GCGACG*GTCGAC*CATTACCTTCTGAAACATA
Confirmation of site-specific homologous recombination	AtsR1confirm	CAGCGGTCTTATCTCATTCGTC
	AtsR2confirm	CATCGACACGGCATACTGG
	AtsTconfirm	CATCTTCGGCGCGTTATG
	AtsA2confirm	GGAAACCGACTGGACCTACAC
Cloning of *atsR1* in pNZ8048	AtsR1FOR	TGATCA*CCATGG*CAACGACGCTCAAAGAC
	AtsR1REV	GCATCG*TCTAGA*CAGCGCCGTCAGACATTG
Cloning of *atsR2* in pNZ8048	AtsR2FOR	TACGTA*CCATGG*TGCATTTCGCATCGG
	AtsR2REV	GCTAGC*TCTAGA*GTGGAATATGCGGTGCGTG
Cloning of *atsA2* in pQE-60	AtsA2FOR	GATTAG*CCATGG*ACAGCCAGACCAAACCGC
	AtsA2REV	TAGCAT*AGATCT*CTTGGCGACTTGGTCCC
Cloning of *atsB2* in pBC1.2	AtsB2FOR	TGCCTA*TCTAGA*AAGGAGATGTATCGACGGAATGCGTTG
	AtsB2REV	GCTATA*AAGCTT*CACACGAGGAAAGCGGTCAC
Cloning of *atsProm* in pBC1.2	AtsPromF	GTACTA*AAGCTT*CCAGTATGCCGTGTCGATG
	AtsPromR	TAGCTA*TCTAGA*CGCAATGCCAGAAACTCAGC

Restriction sites incorporated into oligonucleotide primer sequences are indicated in italics.

Table 5.3: IRD-labelled primers used to generate PCR products and primer extension products in this study.

Primer	Sequence
AtsR1F1F	GCGGAATCGTGGGTCATGTC
AtsR1F1R	GTATCTCCGCCGCCGAC
AtsR1F2F	GCCGATTTCAGCTCCTTCG
AtsR1F2R	CCATTTCTAGGCATCTCATCG
AtsR1F3F	GAAGACGACACGATGG
AtsR1F3R	GTAACACAGTATTCGCCG
AtsR1F4F	GGTTCCGTATTCTTCCATGC
AtsR1F4R	GTCATCGCATCTCCATATCG
AtsR1F5F	CGATATGGAGTGCGATGAC
AtsR1F5R	GTAATGATCCGACGACGTGC
AtsR2R1F	CATCGTGTTATTGGCGCGG
AtsR2R1R	GACGCCATATCACAGAGGGTTG
AtsR2R2F	GCATGCGGCGTGAACTCC
AtsR2R2R	CGCAATGCCAGAAACTCAGC
AtsR2R3F	GATGTTGCCTTGCGGTATG
AtsR2R3R	CAACGGCTGCCCACTGG
AtsR2T1F	GGTCCTCCTTCGTCTGTGTGG
AtsR2T1R	GTCGTGGCATATCGTTCGG
AtsR2T2F	GGGCCGACGAAGTTGTTG
AtsR2T2R	CGATGAGACCGCCGATG
AtsR2T3F	CTAGCGGCATTCAGTATCGAG
AtsR2T3R	GCGGCAGAACAGCAGGAAC

Table 5.4: Effect of GlcNAc-6-S on the transcriptome of *B. breve* UCC2003

Locus tag (gene name)	Predicted Function	Level of upregulation
Bbr_0846 (*nagA1*)	*N*-acetylglucosamine-6-phosphate deacetylase	12.63
Bbr_0847 (*nagB2*)	Glucosamine-6-phosphate deaminase	6.17
Bbr_0848 (*nagK*)	Sugar kinase, ROK family	9.85
Bbr_0849 (*atsR2*)	Transcriptional regulator, ROK family	8.58
Bbr_0851 (*atsT*)	Carbohydrate transport protein	96.75
Bbr_0852 (*atsA2*)	Sulfatase	35.36
Bbr_0853 (*atsB2*)	anSME	31.25
Bbr_0854	Hypothetical membrane spanning protein	4.175
Bbr_1247 (*nagA2*)	*N*-acetylglucosamine-6-phosphate deacetylase	10.84
Bbr_1248 (*nagB3*)	Glucosamine-6-phosphate deaminase	11.88
Bbr_1249	Transcriptional regulator, ROK family	3.07
Bbr_1585 (*galE*)	UDP-glucose 4-epimerase	3.09
Bbr_1586 (*nahK*)	*N*-acetylhexosamine kinase	4.96
Bbr_1587 (*lnbP*)	lacto-*N*-biose phosphorylase	6.58
Bbr_1588	Permease protein of ABC transporter system	6.24
Bbr_1589	Permease protein of ABC transporter system	8.27
Bbr_1590	Solute-binding protein of ABC transporter system	23.97

The cutoff point is 3- fold with a *P*-value of <0.001.

Table 5.5: Transcriptome analysis of *B. breve* UCC2003-atsR1 and *B. breve* UCC2003-atsR2 as compared to *B. breve* UCC2003 grown on 0.5 % (wt/vol) ribose.

Locus tag (gene name)	Predicted Function	Level of up-regulation	
		atsR1	***atsR2***
Bbr_0351 (*atsR1*)	Transcriptional regulatory protein	14.95	-
Bbr_0352 (*atsA1*)	Sulfatase family protein	65.29	-
Bbr_0353	Sugar-binding protein	105.0	-
Bbr_0354	Permease of ABC transporter for sugars	86.23	-
Bbr_0355	Permease of ABC transporter for sugars	53.23	-
Bbr_0851 (*atsT*)	Major facilitator transporter	-	106.28
Bbr_0852 (*atsA2*)	Sulfatase family protein	-	59.58
Bbr_0853 (*atsB2*)	anSME	-	15.57
Bbr_0854	Conserved membrane spanning domain	-	9.09

The cutoff point is 9- fold with a *P*-value of <0.001; values below the cutoff are indicated by a minus.

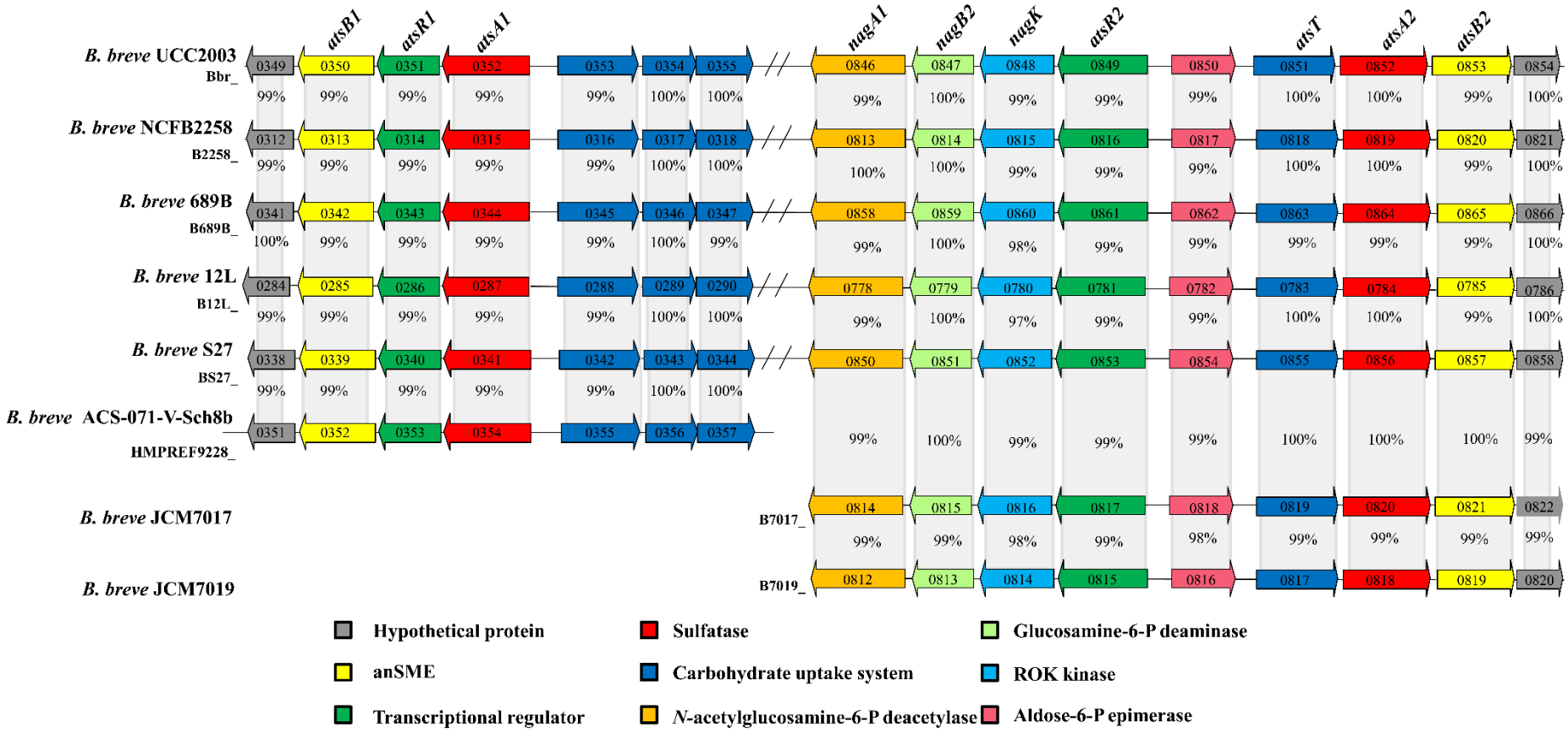

atsB1
atsR1
atsA1
nagA1
nagB2
nagK
atsR2
atsT
atsA2
atsB2
B. breve UCC2003
Bbr_
0349 0350 0351 0352 0353 0354 0355 0846 0847 0848 0849 0850 0851 0852 0853 0854
99% 99% 99% 99% 99% 100% 100% 99% 100% 99% 99% 99% 100% 100% 99% 100%
B. breve NCFB2258
B2258_
0312 0313 0314 0315 0316 0317 0318 0813 0814 0815 0816 0817 0818 0819 0820 0821
99% 99% 99% 99% 99% 100% 100% 100% 100% 99% 99% 99% 100% 100% 99% 100%
B. breve 689B
B689B_
0341 0342 0343 0344 0345 0346 0347 0858 0859 0860 0861 0862 0863 0864 0865 0866
100% 99% 99% 99% 99% 100% 99% 99% 100% 98% 99% 99% 99% 99% 99% 100%
B. breve 12L
B12L_
0284 0285 0286 0287 0288 0289 0290 0778 0779 0780 0781 0782 0783 0784 0785 0786
99% 99% 99% 99% 99% 100% 100% 99% 100% 97% 99% 99% 100% 100% 99% 100%
B. breve S27
BS27_
0338 0339 0340 0341 0342 0343 0344 0850 0851 0852 0853 0854 0855 0856 0857 0858
99% 99% 99% 99% 99% 100% 100%
B. breve ACS-071-V-Sch8b
HMPREF9228_
0351 0352 0353 0354 0355 0356 0357
99% 100% 99% 99% 99% 100% 100% 100% 99%
B. breve JCM7017
B7017_
0814 0815 0816 0817 0818 0819 0820 0821 0822
99% 99% 98% 99% 98% 99% 99% 99% 99%
B. breve JCM7019
B7019_
0812 0813 0814 0815 0816 0817 0818 0819 0820
Hypothetical protein
anSME
Transcriptional regulator
Sulfatase
Carbohydrate uptake system
N-acetylglucosamine-6-P deacetylase
Glucosamine-6-P deaminase
ROK kinase
Aldose-6-P epimerase

Figure 5.1: Comparison of the sulfatase and anSME-encoding gene clusters of *B. breve* UCC2003 with corresponding loci in the currently available complete *B. breve* genome sequences. Each solid arrow represents an open reading frame. The length of the arrows (which contain the locus tag number) is proportional to the size of the open reading frame. The corresponding gene name, which is indicative of putative function, is given above relevant arrows at the top of the figure. Orthologs are marked with the same colour. The amino acid identity of each predicted protein to its equivalent protein encoded by *B. breve* UCC2003, expressed as a percentage, is given above each arrow.

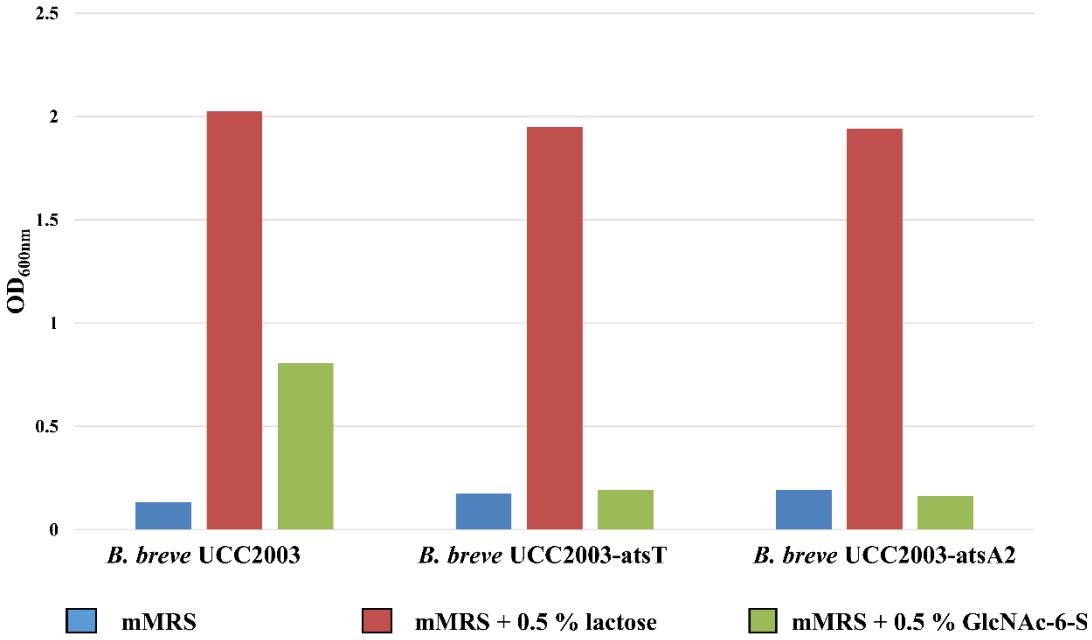


Figure 5.2: Final OD_{600nm} values obtained following 24 h growth of *B. breve* UCC2003, *B. breve* UCC2003-atsT and *B. breve* UCC2003-atsA2 in mMRS without supplementation with a carbon source (negative control) or containing 0.5 % (wt/vol) lactose or GlcNAc-6-S as the sole carbon source. The results presented were obtained from a single experiment.

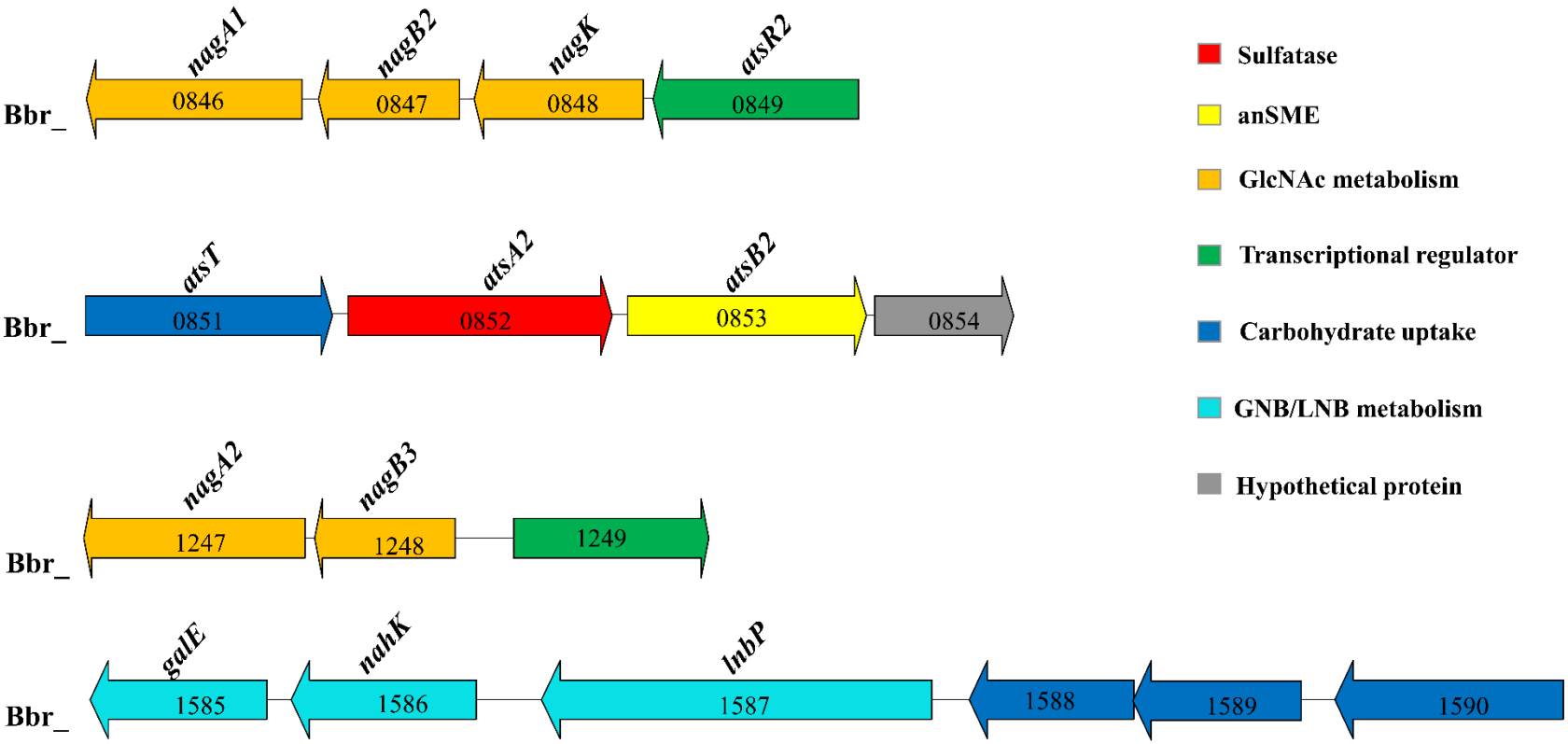


Figure 5.3: Schematic representation of the four *B. breve* UCC2003 gene clusters up-regulated during growth on GlcNAc-6-S as the sole carbon source. The length of the arrows (which contain the locus tag number) is proportional to the size of the open reading frame and the gene locus name, which is indicative of its putative function, is given at the top. Genes are grouped by colour based on their predicted function in carbohydrate metabolism.

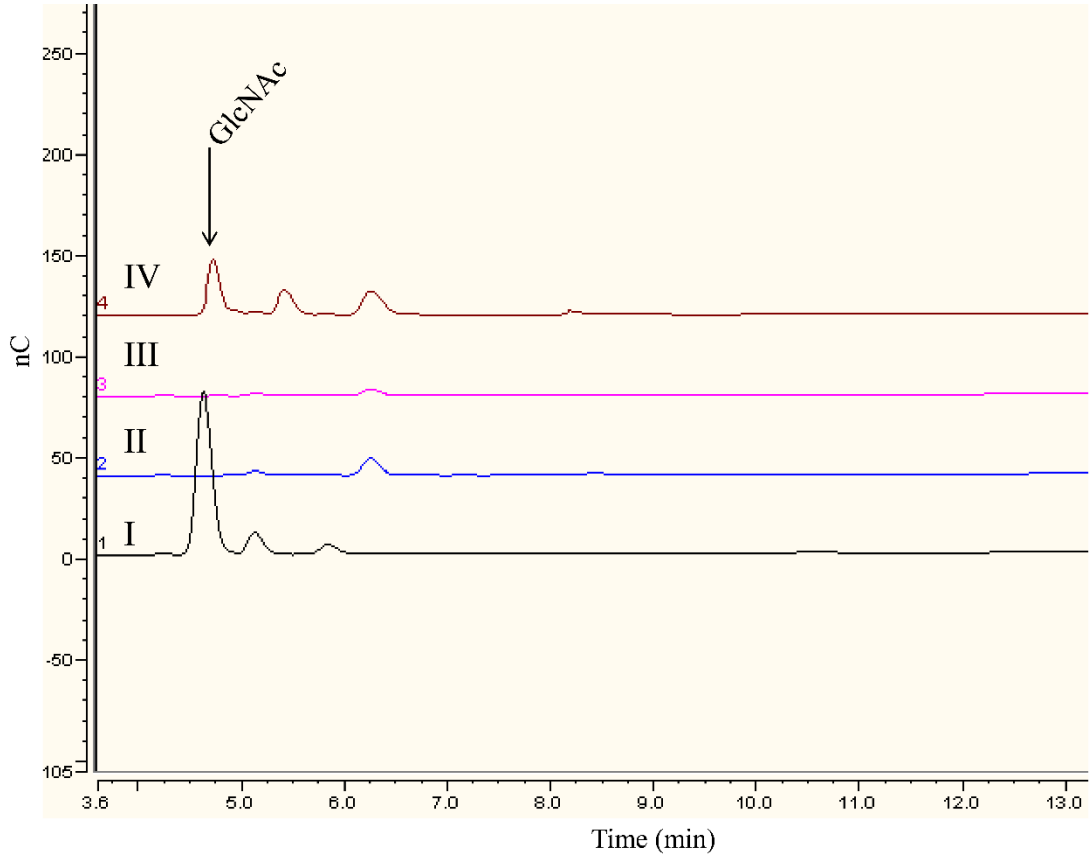


Figure 5.4: HPAEC-PAD analysis of the sulfatase activity of AtsA2, AtsB2 and co-expressed AtsA2 and AtsB2. The chromatographic position of GlcNAc is indicated by an arrow. **(I)** 0.1 mg ml^{-1} GlcNAc standard in 50 mM MOPS buffer **(II)** 0.1 mg ml^{-1} of GlcNAc-6-S incubated with crude cell extract containing pQE-60-AtsA2 **(III)** 0.1 mg ml^{-1} of GlcNAc-6-S incubated with crude cell extract containing pBC1.2-AtsB2 **(IV)** 0.1 mg ml^{-1} of GlcNAc-6-S incubated with crude cell extract containing pQE-60-AtsA2-pBC1.2-AtsB2. nC, nanocoulombs.

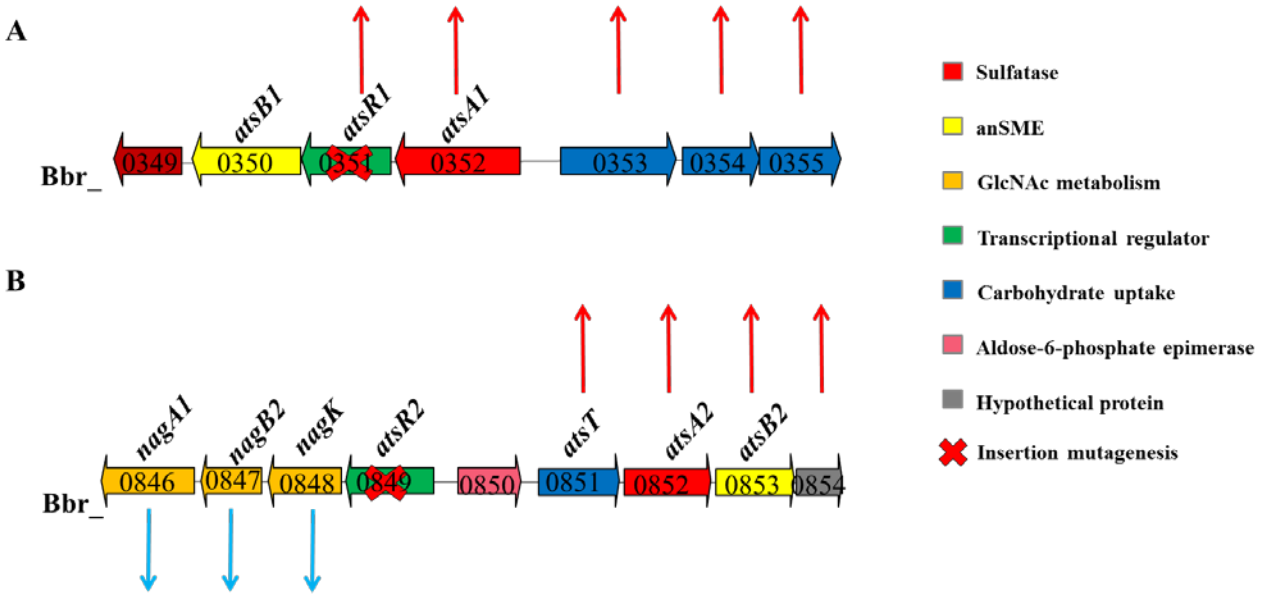
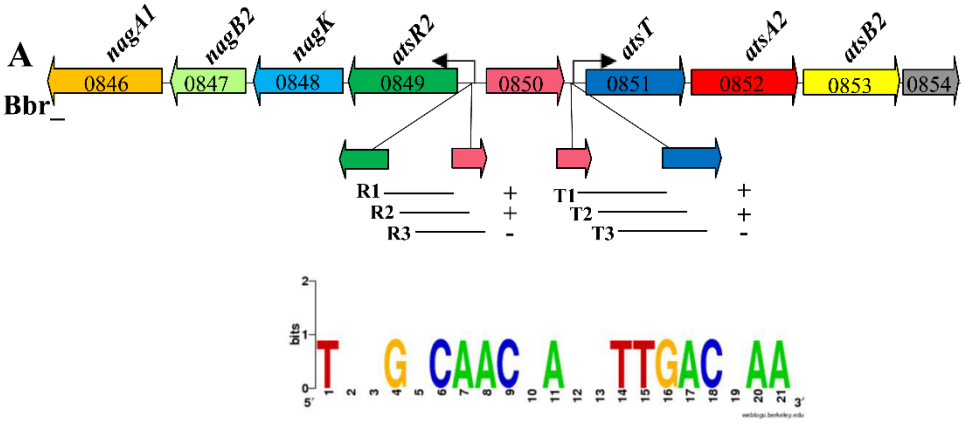
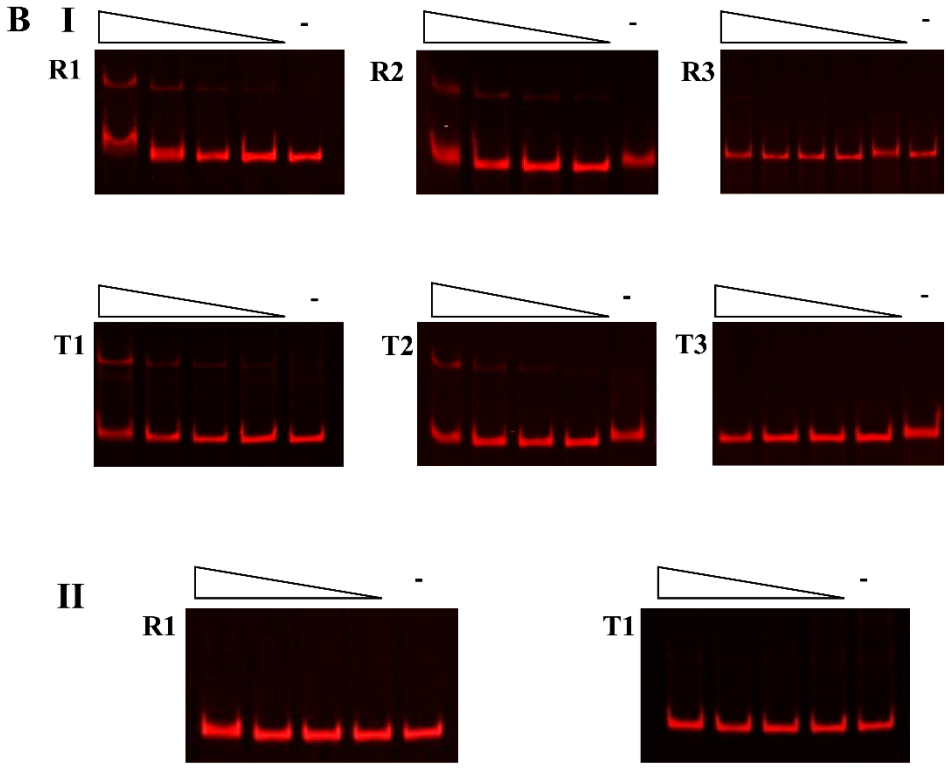


Figure 5.5: Schematic representation of the gene clusters up- or down-regulated in the insertion mutant strains (A) *B. breve* UCC2003-atsR1 and (B) *B. breve* UCC2003-atsR2 as compared to *B. breve* UCC2003 during growth in mMRS supplemented with 0.5% (wt/vol) ribose. The genes in which the insertion mutants were constructed are marked with a red cross symbol. Red and blue arrows represent up- or down-regulation, respectively. The absence of an arrow indicates no change in the level of transcription compared to *B. breve* UCC2003. The gene locus name, which is indicative of its putative function, is given at the top. Genes are grouped by colour based on their predicted function in carbohydrate metabolism.


A
Bbr_
nagA1
nagB2
nagK
atsR2
atsT
atsA2
atsB2
0846
0847
0848
0849
0850
0851
0852
0853
0854
R1 +
R2 +
R3 -
T1 +
T2 +
T3 -

bits
T G CAAC A TTGAC AA


B I
R1
R2
R3
T1
T2
T3
II
R1
T1

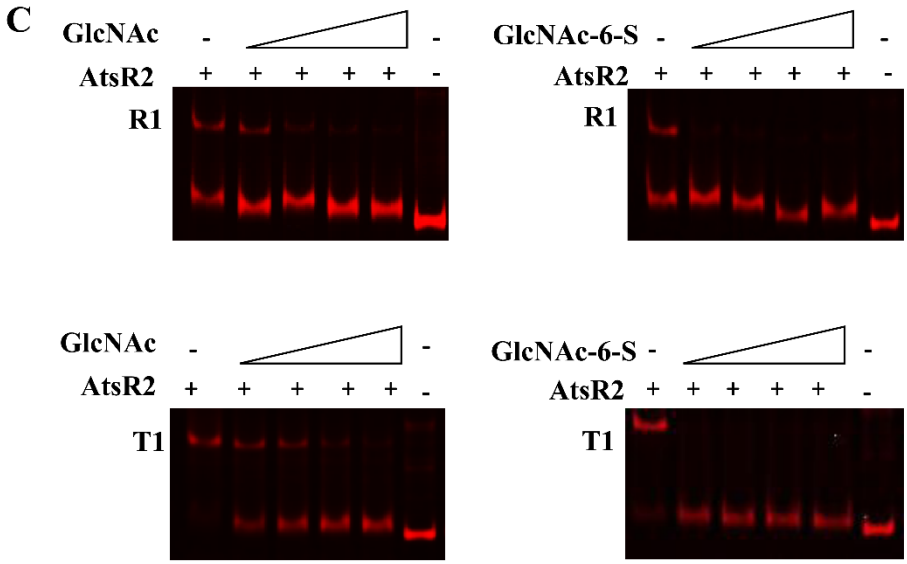

C
GlcNAc
AtsR2
R1
GlcNAc-6-S
T1

Figure 5.6: (A) Schematic representation of the *ats2* gene cluster of *B. breve* UCC2003 and DNA fragments used in EMSAs for the *atsR2* and *atsT* promoter regions, together with Weblogo representation of the predicted operator of AtsR2. Plus or minus signs indicate ability or inability of AtsR2 to bind to the DNA fragment. The bent arrows represent the position and direction of the proven promoter sequences (see Fig. 7). **(B)** EMSAs showing the interactions of (I) crude cell extract containing pNZ-AtsR2 with the DNA fragments R1, R2, R3,T1, T2 and T3 and (II) crude cell extract containing pNZ8048 (empty vector) with the DNA fragments R1 and T1. The minus symbol indicates reactions to which no crude cell extract was added, while the remaining lanes represent binding reactions with the respective DNA probes incubated with increasing amounts of crude cell extract. Each successive lane from right to left represents a doubling of the amount of crude cell extract. **(C)** EMSAs showing AtsR2 interaction with the DNA fragments R1 and T1 with the addition of GlcNAc or GlcNAc-6-S in concentrations ranging from 2.5 mM to 20 mM.

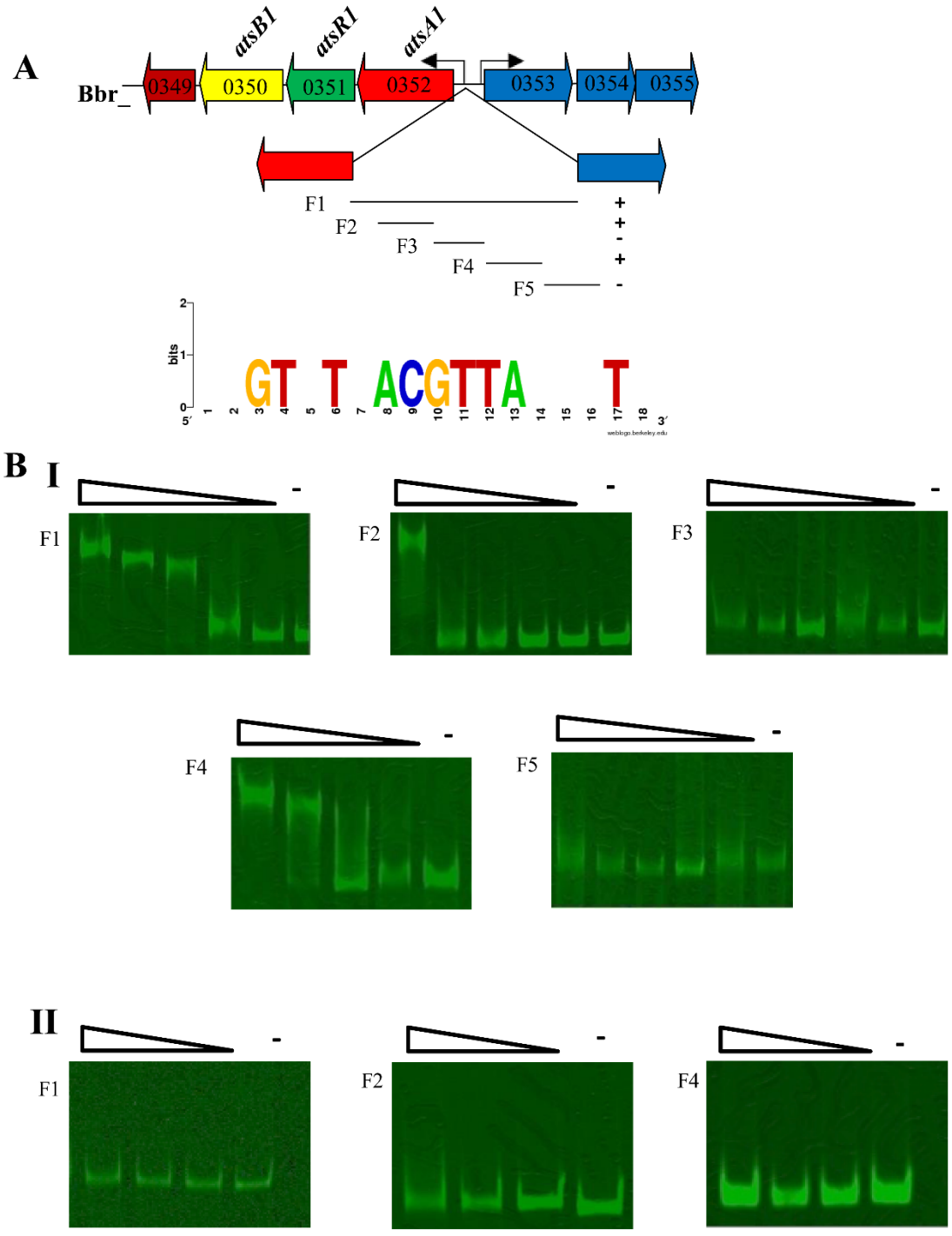

A
Bbr_
0349
0350
0351
0352
0353
0354
0355
atsB1
atsR1
atsA1
F1
F2
F3
F4
F5
+
+
-
+
-
bits
GT T ACGTTA T
B I
F1
F2
F3
F4
F5
II
F1
F2
F4

Figure 5.7: (A) Schematic representation of the *ats1* gene cluster of *B. breve* UCC2003 and DNA fragments used in EMSAs for the *atsA1* and Bbr_0353 promoter regions, together with Weblogo representation of the predicted operator of AtsR1. Plus or minus signs indicate ability or inability of AtsR1 to bind to the DNA fragment. The bent arrows represent the proven promoter regions (see Fig. 7). **(B)** EMSAs showing the interactions of (I) crude cell extract containing pNZ-atsR1 with the DNA fragments F1, F2, F3, F4 and F5 and (II) crude cell extract containing pNZ8048 (empty vector) with the DNA fragments F1, F2 and F4. The minus symbol indicates reactions to which no crude cell extract was added, while the remaining lanes represent binding reactions with the respective DNA probes incubated with increasing amounts of crude cell extract. Each successive lane from right to left represents a doubling of the amount of crude cell extract.

A

AtsR2 -35

ATTGTATCCATCCCGACATATTTAGGCAACCATC**TTGACT**AATTAATTT

-10 *

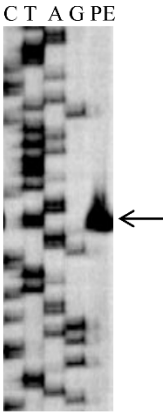
CACCATGC**TATATT**TATCAATTGAAGCCTATCAACGAGCCGATTAGCC

GCGCACAAGACTCGAAAAAGGAGTTCACGCCGCATGCCCGCCCGCA

AAGATGACACAATAGGTATTGGCATCATCGAAAAGCGCCTCTTCAAGC

RBS

GCTGACGGGAGCTTTTCATG



B

TTCTTCACCGTGAACCATTGATATTTCAGTAATCTTTGAGTCTTTTGCG

-35 AtsR2 -10

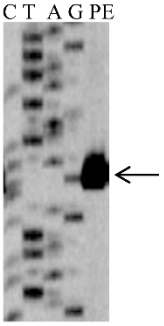
TGCGTTATTCCTAAATATG**TCAACA**AGGTTGACGAAATGATG**TATACT**

*

GAATCACGCCAGCTATCACTAGCGGCATTCAGTATCGAGATTCAAAGG

RBS

AGAAAAGGATACTTTCATG



C

ATAGCGCTGTTCCTCATCGCCACCCCCGCCACATTCGCAGTAGAAGAC

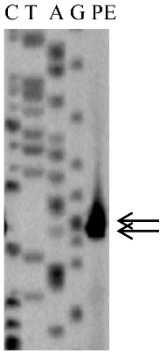
-35 AtsR1 -10 * *

GACACTCACGA**TGGTTT**AACGTTAAGATTTATGC**TATAGT**ACGAAGG

AGCTGAAATCGGCACCGCGGGGGGGCGTCATCACACCGGATTCCCAA

RBS

CCCGCGAACCTATGGAAAAGGAGCTATCGTG



D

-35

TTGACGCAATCGCCGTGTGGCGCGCCGGAGGGGGCTGGGTT**TTGATT**

AtsR1 -10 * *

TCGGCGAATACTGTGT**TACGTT**ATCTTAACGATACGATATGGAGATGC

RBS

GATGACACTGCATTCGTATCGGTCCAGTAAAGAAGCTGACGAAAGAG

GTTCATGTG

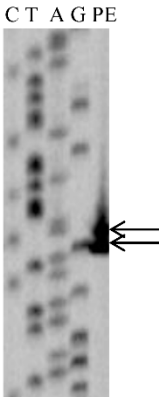


Figure 5.8: Schematic representation of the *atsR2* (panel A), *atsT* (panel B), *atsA1* (panel C) and Bbr_0353 (panel D) promoter regions. Boldface type and underlining indicate -10 and -35 hexamers (as deduced from the primer extension results) and ribosomal binding site (RBS); the transcriptional start site is indicated by an asterisk. The arrows underneath the indicated DNA sequences indicate the inverted repeats that represent the presumed AtsR1 or AtsR2 binding sites as displayed above the arrows. The arrows in the right panels indicate the primer extension products.

5.8 REFERENCES:

Agans, R., Rigsbee, L., Kenche, H., Michail, S., Khamis, H. J. & Paliy, O. (2011). Distal gut microbiota of adolescent children is different from that of adults. *FEMS Microbiol Ecol* **77**, 404-412.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Álvarez-Martín, P., O'Connell-Motherway, M., van Sinderen, D. & Mayo, B. (2007). Functional analysis of the pBC1 replicon from *Bifidobacterium catenulatum* L48. *Appl Microbiol Biotechnol* **76**, 1395-1402.

Andersson, A. F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P. & Engstrand, L. (2008). Comparative analysis of human gut microbiota by barcoded pyrosequencing. *PloS one* **3**, e2836.

Ashida, H., Miyake, A., Kiyohara, M., Wada, J., Yoshida, E., Kumagai, H., Katayama, T. & Yamamoto, K. (2009). Two distinct α-L-fucosidases from *Bifidobacterium bifidum* are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. *Glycobiology* **19**, 1010-1017.

Beil, S., Kehrli, H., James, P., Staudenmann, W., Cook, A. M., Leisinger, T. & Kertesz, M. A. (1995). Purification and characterization of the arylsulfatase synthesized by *Pseudomonas aeruginosa* PAO during growth in sulfate-free medium and cloning of the arylsulfatase gene (*atsA*). *Eur J Biochem* **229**, 385-394.

Benjdia, A., Subramanian, S., Leprince, J., Vaudry, H., Johnson, M. K. & Berteau, O. (2008). Anaerobic sulfatase-maturating enzymes, first dual substrate radical S-adenosylmethionine enzymes. *J Biol Chem* **283**, 17815-17826.

Benjdia, A., Martens, E. C., Gordon, J. I. & Berteau, O. (2011). Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont, *Bacteroides thetaiotaomicron*. *J Biol Chem* **286**, 25973-25982.

Berteau, O., Guillot, A., Benjdia, A. & Rabot, S. (2006). A new type of bacterial sulfatase reveals a novel maturation pathway in prokaryotes. *J Biol Chem* **281**, 22464-22470.

Bond, C. S., Clements, P. R., Ashby, S. J., Collyer, C. A., Harrop, S. J., Hopwood, J. J. & Guss, J. M. (1997). Structure of a human lysosomal sulfatase. *Structure* **5**, 277-289.

Bottacini, F., OConnell Motherway, M., Kuczynski, J., OConnell, K., Serafini, F., Duranti, S., Milani, C., Turroni, F., Lugli, G. & other authors (2014). Comparative genomics of the *Bifidobacterium breve* taxon. *BMC Genomics* **15**, 170.

Brockhausen, I. (2003). Sulphotransferases acting on mucin-type oligosaccharides. *Biochem Soc Trans* **31**, 318-325.

Carlson, B. L., Ballister, E. R., Skordalakes, E., King, D. S., Breidenbach, M. A., Gilmore, S. A., Berger, J. M. & Bertozzi, C. R. (2008). Function and structure of a prokaryotic formylglycine-generating enzyme. *J Biol Chem* **283**, 20117-20125.

Cheng, Q., Hwa, V. & Salyers, A. A. (1992). A locus that contributes to colonization of the intestinal tract by *Bacteroides thetaiotaomicron* contains a single regulatory gene (*chuR*) that links two polysaccharide utilization pathways. *J Bacteriol* **174**, 7185-7193.

Conejo, M., Thompson, S. & Miller, B. (2010). Evolutionary bases of carbohydrate recognition and substrate discrimination in the ROK protein family. *J Mol Evol* **70**, 545-556.

Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Res* **14**, 1188-1190.

De Man, J. C., Rogosa, M. & Sharpe, M. E. (1960). A medium for the cultivation of lactobacilli. *J Appl Bacteriol* **23**, 130-135.

de Ruyter, P. G., Kuipers, O. P. & de Vos, W. M. (1996). Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl Environ Microbiol* **62**, 3662-3667.

Dierks, T., Schmidt, B. & von Figura, K. (1997). Conversion of cysteine to formylglycine: A protein modification in the endoplasmic reticulum. *Proc Natl Acad Sci U S A* **94**, 11963-11968.

Dierks, T., Lecca, M. R., Schlotterhose, P., Schmidt, B. & von Figura, K. (1999). Sequence determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases. *EMBO J* **18**, 2084-2091.

Dierks, T., Schmidt, B., Borissenko, L. V., Peng, J., Preusser, A., Mariappan, M. & von Figura, K. (2003). Multiple sulfatase deficiency is caused by mutations in the gene encoding the human Cα-formylglycine generating enzyme. *Cell* **113**, 435-444.

Egan, M., O'Connell Motherway, M., Ventura, M. & van Sinderen, D. (2014a). Metabolism of sialic acid by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **80**, 4414-4426.

Egan, M., O' Connell Motherway, M., Kilcoyne, M., Kane, M., Joshi, L., Ventura, M. & van Sinderen, D. (2014b). Cross-feeding by *Bifidobacterium breve* UCC2003 during co-cultivation with *Bifidobacterium bifidum* PRL2010 in a mucin-based medium. *BMC Microbiol* **14**, 282.

Eliakim, R., Gilead, L., Ligumsky, M., Okon, E., Rachmilewitz, D. & Razin, E. (1986). Histamine and chondroitin sulfate E proteoglycan released by cultured human colonic mucosa: indication for possible presence of E mast cells. *Proc Natl Acad Sci U S A* **83**, 461-464.

Fang, Q., Peng, J. & Dierks, T. (2004). Post-translational formylglycine modification of bacterial sulfatases by the radical S-adenosylmethionine protein AtsB. *J Biol Chem* **279**, 14570-14578.

Filipe, M. (1978). Mucins in the human gastrointestinal epithelium: a review. *Invest Cell Pathol* **2**, 195-216.

García de la Nava, J., Santaella, D. F., Alba, J. C., Carazo, J. M., Trelles, O. & Pascual-Montano, A. (2003). Engene: the processing and exploratory analysis of gene expression data. *Bioinformatics* **19**, 657-658.

Garrido, D., Kim, J. H., German, J. B., Raybould, H. E. & Mills, D. A. (2011). Oligosaccharide binding proteins from *Bifidobacterium longum* subsp. *infantis* reveal a preference for host glycans. *PLoS ONE* **6**, e17315.

Garrido, D., Nwosu, C., Ruiz-Moyano, S., Aldredge, D., German, J. B., Lebrilla, C. B. & Mills, D. A. (2012). Endo-β-*N*-acetylglucosaminidases from infant gut-associated bifidobacteria release complex *N*-glycans from human milk glycoproteins. *Mol Cell Proteomics* **11**, 775-785.

Guérardel, Y., Morelle, W., Plancke, Y., Lemoine, J. & Strecker, G. (1999). Structural analysis of three sulfated oligosaccharides isolated from human milk. *Carbohydr Res* **320**, 230-238.

Hamoen, L. W., Van Werkhoven, A. F., Bijlsma, J. J. E., Dubnau, D. & Venema, G. (1998). The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev* **12**, 1539-1550.

Hopwood, J. J. & Elliott, H. (1983). *N*-acetylglucosamine 6-sulfate residues in keratan sulfate and heparan sulfate are desulfated by the same enzyme. *Biochemistry international* **6**, 141-148.

Kitaoka, M., Tian, J. & Nishimoto, M. (2005). Novel putative galactose operon involving lacto-*N*-biose phosphorylase in *Bifidobacterium longum*. *Appl Environ Microbiol* **71**, 3158-3162.

Kurokawa, K., Itoh, T., Kuwahara, T., Oshima, K., Toh, H., Toyoda, A., Takami, H., Morita, H., Sharma, V. K. & other authors (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**, 169-181.

Larsson, J. M., Karlsson, H., Sjovall, H. & Hansson, G. C. (2009). A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19**, 756-766.

Law, J., Buist, G., Haandrikman, A., Kok, J., Venema, G. & Leenhouts, K. (1995). A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. *J Bacteriol* **177**, 7011-7018.

Long, A. D., Mangalam, H. J., Chan, B. Y. P., Tolleri, L., Hatfield, G. W. & Baldi, P. (2001). Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework: Analysis of global gene expression in *Escherichia coli* K12. *J Biol Chem* **276**, 19937-19944.

Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., Von Figura, K. & Saenger, W. (1998). Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* **37**, 3654-3664.

Maze, A., O'Connell-Motherway, M., Fitzgerald, G., Deutscher, J. & van Sinderen, D. (2007). Identification and characterization of a fructose phosphotransferase system in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **73**, 545-553.

Miech, C., Dierks, T., Selmer, T., von Figura, K. & Schmidt, B. (1998). Arylsulfatase from *Klebsiella pneumoniae* carries a formylglycine generated from a serine. *J Biol Chem* **273**, 4835-4837.

Mierau, I. & Kleerebezem, M. (2005). 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. *Appl Microbiol Biotechnol* **68**, 705-717.

Murooka, Y., Ishibashi, K., Yasumoto, M., Sasaki, M., Sugino, H., Azakami, H. & Yamashita, M. (1990). A sulfur- and tyramine-regulated *Klebsiella aerogenes* operon containing the arylsulfatase (*atsA*) gene and the *atsB* gene. *J Bacteriol* **172**, 2131-2140.

Newburg, D., Linhardt, R., Ampofo, S. & Yolken, R. (1995). Human milk glycosaminoglycans inhibit HIV glycoprotein gp120 binding to its host cell CD4. *J Nutr* **125**, 419.

Nishimoto, M. & Kitaoka, M. (2007). Identification of *N*-acetylhexosamine 1-kinase in the complete lacto-*N*-biose I/galacto-*N*-biose metabolic pathway in *Bifidobacterium longum*. *Appl Environ Microbiol* **73**, 6444-6449.

O'Connell, K. J., O'Connell Motherway, M., O'Callaghan, J., Fitzgerald, G. F., Ross, R. P., Ventura, M., Stanton, C. & van Sinderen, D. (2013). Metabolism of four α-glycosidic linkage-containing oligosaccharides by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **79**, 6280-6292.

O'Connell, K. J., O'Connell Motherway, M., Liedtke, A., Fitzgerald, G. F., Ross, R. P., Stanton, C., Zomer, A. & van Sinderen, D. (2014). Transcription of two adjacent carbohydrate utilization gene clusters in *Bifidobacterium breve* UCC2003 is controlled by LacI- and Repressor Open reading frame Kinase (ROK)-type regulators. *Appl Environ Microbiol* **80**, 3604-3614.

O'Connell Motherway, M., O'Driscoll, J., Fitzgerald, G. F. & Van Sinderen, D. (2009). Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **2**, 321-332.

O'Connell Motherway, M., Fitzgerald, G. F. & van Sinderen, D. (2011a). Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **4**, 403-416.

O'Connell Motherway, M., Zomer, A., Leahy, S. C., Reunanen, J., Bottacini, F., Claesson, M. J., O'Brien, F., Flynn, K., Casey, P. G. & other authors (2011b). Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb

tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc Natl Acad Sci USA* **108**, 11217-11222.

O'Connell Motherway, M., Kinsella, M., Fitzgerald, G. F. & van Sinderen, D. (2013). Transcriptional and functional characterization of genetic elements involved in galacto-oligosaccharide utilization by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **6**, 67-79.

Oshiro, M., Ono, K., Suzuki, Y., Ota, H., Katsuyama, T. & Mori, N. (2001). Immunohistochemical localization of heparan sulfate proteoglycan in human gastrointestinal tract. *Histochem Cell Biol* **115**, 373-380.

Pastell, H., Westermann, P., Meyer, A. S., Tuomainen, P. & Tenkanen, M. (2009). *In vitro* fermentation of arabinoxylan-derived carbohydrates by bifidobacteria and mixed fecal microbiota. *J Agric Food Chem* **57**, 8598-8606.

Pokusaeva, K., Neves, A. R., Zomer, A., O'Connell-Motherway, M., MacSharry, J., Curley, P., Fitzgerald, G. F. & van Sinderen, D. (2010). Ribose utilization by the human commensal *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **3**, 311-323.

Pokusaeva, K., Fitzgerald, G. F. & Sinderen, D. (2011a). Carbohydrate metabolism in bifidobacteria. *Genes Nutr* **6**, 285-306.

Pokusaeva, K., O'Connell-Motherway, M., Zomer, A., MacSharry, J., Fitzgerald, G. F. & van Sinderen, D. (2011b). Cellodextrin utilization by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **77**, 1681-1690.

Ravcheev, D. A., Khoroshkin, M. S., Laikova, O. N., Tsoy, O. V., Sernova, N. V., Petrova, S. A., Rakhmaninova, A. B., Novichkov, P. S., Gelfand, M. S. & other authors (2014). Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Front Microbiol* **5**.

Rho, J. H., Wright, D. P., Christie, D. L., Clinch, K., Furneaux, R. H. & Roberton, A. M. (2005). A novel mechanism for desulfation of mucin: identification and cloning of a mucin-desulfating glycosidase (sulfoglycosidase) from *Prevotella* strain RS2. *J Bacteriol* **187**, 1543-1551.

Riordan, O. (1998). *Studies on antimicrobial activity and genetic diversity of Bifidobacterium species: molecular characterization of a 5.75 kb plasmid and a chromosomally encoded recA gene homologue from Bifidobacterium breve.*, National University of Ireland, Cork, Cork.

Roberton, A., McKenzie, C., Sharfe, N. & Stubbs, L. (1993). A glycosulphatase that removes sulphate from mucus glycoprotein. *Biochem J* **293**, 683-689.

Ruiz-Moyano, S., Totten, S. M., Garrido, D. A., Smilowitz, J. T., German, J. B., Lebrilla, C. B. & Mills, D. A. (2013). Variation in consumption of human milk oligosaccharides by infant gut-associated strains of *Bifidobacterium breve*. *Appl Environ Microbiol* **79**, 6040-6049.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945.

Salyers, A. A., Vercellotti, J. R., West, S. E. & Wilkins, T. D. (1977). Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. *Appl Environ Microbiol* **33**, 319-322.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning a laboratory manual*, 2nd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Sardiello, M., Annunziata, I., Roma, G. & Ballabio, A. (2005). Sulfatases and sulfatase modifying factors: an exclusive and promiscuous relationship. *Hum Mol Gen* **14**, 3203-3217.

Scardovi, V. & Trovatelli, I. D. (1965). The fructose-6-phosphate shunt as a peculiar pattern of hexose degradation in the genus *Bifidobacterium*. *Ann Microbiol*, 19-29.

Schell, M., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M.-C., Desiere, F., Bork, P. & other authors (2002). The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci U S A* **99**, 14422-14427.

Schmidt, B., Selmer, T., Ingendoh, A. & Figurat, K. v. (1995). A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* **82**, 271-278.

Sela, D. & Mills, D. (2010). Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol* **18**, 298-307.

Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., Lapidus, A., Rokhsar, D. S., Lebrilla, C. B. & other authors (2008). The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A* **105**, 18964-18969.

Sela, D. A., Li, Y., Lerno, L., Wu, S., Marcobal, A. M., German, J. B., Chen, X., Lebrilla, C. B. & Mills, D. A. (2011). An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides. *J Biol Chem* **286**, 11909-11918.

Sela, D. A., Garrido, D., Lerno, L., Wu, S., Tan, K., Eom, H.-J., Joachimiak, A., Lebrilla, C. B. & Mills, D. A. (2012). *Bifidobacterium longum* subsp. *infantis* ATCC 15697 α-fucosidases are active on fucosylated human milk oligosaccharides. *Appl Environ Microbiol* **78**, 795-803.

Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F. & Miller, N. E. (2001). Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization

using new analysis and information visualization methods. *Nucleic acids research* **29**, 1097-1106.

Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C.-H., Westover, B. P., Weatherford, J., Buhler, J. D. & Gordon, J. I. (2005). Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307**, 1955-1959.

Sonnenburg, E. D., Sonnenburg, J. L., Manchester, J. K., Hansen, E. E., Chiang, H. C. & Gordon, J. I. (2006). A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. *Proc Natl Acad Sci U S A* **103**, 8834-8839.

Szameit, C., Miech, C., Balleininger, M., Schmidt, B., von Figura, K. & Dierks, T. (1999). The iron sulfur protein AtsB is required for posttranslational formation of formylglycine in the *Klebsiella* sulfatase. *J Biol Chem* **274**, 15375-15381.

Terzaghi, B. E. & Sandine, W. E. (1975). Improved medium for lactic streptococci and their bacteriophages. *Appl Microbiol* **29**, 807-813.

Tsai, H. H., Hart, C. A. & Rhodes, J. M. (1991). Production of mucin degrading sulphatase and glycosidases by *Bacteroides thetaiotaomicron*. *Lett Appl Microbiol* **13**, 97-101.

Turroni, F., Ribbera, A., Foroni, E., van Sinderen, D. & Ventura, M. (2008). Human gut microbiota and bifidobacteria: from composition to functionality. *Antonie van Leeuwenhoek* **94**, 35-50.

Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J.-H., Zomer, A., Sanchez, B., Bidossi, A., Ferrarini, A. & other authors (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., Strati, F., Foroni, E., Serafini, F., Duranti, S., van Sinderen, D. & Ventura, M. (2012a). Analysis of predicted carbohydrate transport systems encoded by *Bifidobacterium bifidum* PRL2010. *Appl Environ Microbiol* **78**, 5002-5012.

Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., Kerr, C., Hourihane, J., Murray, D. & other authors (2012b). Diversity of bifidobacteria within the infant gut microbiota. *PLoS ONE* **7**, e36957.

Uehara, T. & Park, J. T. (2004). The *N*-acetyl-D-glucosamine kinase of *Escherichia coli* and its role in murein recycling. *J Bacteriol* **186**, 7273-7279.

Ulmer, J. E., Vilén, E. M., Namburi, R. B., Benjdia, A., Beneteau, J., Malleron, A., Bonnaffé, D., Driguez, P.A., Descroix, K. & other authors (2014). Characterization of glycosaminoglycan (GAG) sulfatases from the human gut symbiont *Bacteroides thetaiotaomicron* reveals the first GAG-specific bacterial endosulfatase. *J Biol Chem* **289**, 24289-24303.

Vaishampayan, P. A., Kuehl, J. V., Froula, J. L., Morgan, J. L., Ochman, H. & Francino, M. P. (2010). Comparative metagenomics and population dynamics of the gut microbiota in mother and infant. *Genome Biol Evol* **2**, 53-66.

van Hijum, S. A. F. T., Garcia De La Nava, J., Trelles, O., Kok, J. & Kuipers, O. P. (2003). MicroPreP: a cDNA microarray data pre-processing framework. *Appl Bioinformatics*, 241-244.

van Hijum, S., de Jong, A., Baerends, R., Karsens, H., Kramer, N., Larsen, R., den Hengst, C., Albers, C., Kok, J. & other authors (2005). A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC Genomics* **6**, 77.

Van Laere, K. M., Hartemink, R., Bosveld, M., Schols, H. A. & Voragen, A. G. (2000). Fermentation of plant cell wall derived polysaccharides and their corresponding oligosaccharides by intestinal bacteria. *J Agric Food Chem* **48**, 1644-1652.

Ventura, M., Zink, R., Fitzgerald, G. F. & van Sinderen, D. (2005). Gene structure and transcriptional organization of the *dnaK* operon of *Bifidobacterium breve* UCC2003 and application of the operon in bifidobacterial tracing. *Appl Environ Microbiol* **71**, 487-500.

von Figura, K., Schmidt, B., Selmer, T. & Dierks, T. (1998). A novel protein modification generating an aldehyde group in sulfatases: its role in catalysis and disease. *BioEssays* **20**, 505-510.

Wells, J. M., Wilson, P. W. & Le Page, R. W. F. (1993). Improved cloning vectors and transformation procedure for *Lactococcus lactis*. *J Appl Bacteriol* **74**, 629-636.

White, R. (1968). Control of amino sugar metabolism in *Escherichia coli* and isolation of mutants unable to degrade amino sugars. *Biochem J* **106**, 847-858.

Willis, C. L., Cummings, J. H., Neale, G. & Gibson, G. R. (1996). *In vitro* effects of mucin fermentation on the growth of human colonic sulphate-reducing bacteria. *Anaerobe* **2**, 117-122.

Wright, D. P., Knight, C. G., Parkar, S. G., Christie, D. L. & Roberton, A. M. (2000). Cloning of a mucin-desulfating sulfatase gene from *Prevotella* strain RS2 and its expression using a *Bacteroides* recombinant system. *J Bacteriol* **182**, 3002-3007.

Zomer, A., Fernandez, M., Kearney, B., Fitzgerald, G. F., Ventura, M. & van Sinderen, D. (2009). An interactive regulatory network controls stress response in *Bifidobacterium breve* UCC2003. *J Bacteriol* **191**, 7039-7049.

Zwielehner, J., Liszt, K., Handschur, M., Lassl, C., Lapin, A. & Haslberger, A. G. (2009). Combined PCR-DGGE fingerprinting and quantitative-PCR indicates shifts in fecal population sizes and diversity of *Bacteroides*, bifidobacteria and *Clostridium* cluster IV in institutionalized elderly. *Exp Geront* **44**, 440-446.

Chapter VI

General Discussion and Future Perspectives

With the ever-increasing number of publicly available bifidobacterial genomes (185 genomes including complete and draft genome sequences as of NCBI, December 2014), there is a wealth of genomic information about this genus at our fingertips. Genomics-based research has provided significant and insightful information into the phylogenomics of the *Bifidobacterium* genus (Lugli *et al.*, 2014; Milani *et al.*, 2014), while pan-genome and core-genome analyses have generated information on the DNA sequences that are common among or different between bifidobacterial species, including genetic traits such as those pertaining to ecological origin, carbohydrate utilisation and host-microbe interactions (Chapter I) (Bottacini *et al.*, 2014; Duranti *et al.*, 2012; Milani *et al.*, 2013; Ventura *et al.*, 2014). However, functional genomics still has a significant role to play in taking the information gleaned from genomic approaches towards the characterisation and understanding of the specific traits that define bifidobacterial physiology and permit its colonisation in the gut. In this thesis, a number of functional genomic approaches were taken to investigate the metabolism of host-derived carbohydrates by a prototypical bifidobacterial isolate from nursling stool, *Bifidobacterium breve* UCC2003.

Other bifidobacterial species of infant gut origin, namely *Bifidobacterium longum* subsp. *infantis* ATCC15697 and *Bifidobacterium bifidum*, were previously shown to be well-adapted to utilise host-derived carbohydrates, particularly human milk oligosaccharides (HMOs) and/or mucin (Sela & Mills, 2010; Turroni *et al.*, 2014). *In silico* analysis of the *B. breve* UCC2003 genome revealed putative gene clusters involved in the metabolism of host derived mono- and oligosaccharides such as sialic acid, lacto-*N*-biose (LNB) and galacto-*N*-biose (GNB). Chapter II of this thesis focused on the

identification and characterisation of two gene clusters required for the metabolism of sialic acid in *B. breve* UCC2003. Metabolism of this nine-carbon monosaccharide was shown to occur by means of a five-step pathway, analogous to that originally described in *Escherichia coli* (Plumbridge & Vimr, 1999) and being well-conserved among commensals and pathogens of the gut microbiota, including *Lactobacillus sakei* (Anba-Mondoloni *et al.*, 2013), *Staphylococcus aureus* (Olson *et al.*, 2013) and *Haemophilus influenzae* (Vimr *et al.*, 2000). It is interesting to observe that among *Bifidobacterium* species, the sialic acid utilisation cluster is well conserved in the typically infant-derived species, *B. breve* and *B. longum* subsp. *infantis* (Egan *et al.*, 2014a; Sela *et al.*, 2008). Breast-fed infants typically have higher numbers of bifidobacteria compared to their formula-fed counterparts (Fanaro *et al.*, 2003; Roger *et al.*, 2010). Although the reason for such disparity is likely to be due to a number of factors, it is interesting to note that infant formula has previously been shown to contain only approximately a quarter of the sialic acid levels compared to that found in human milk. Significantly, 70 % of the sialic acid in infant formula is linked to glycoproteins whereas in human milk the majority of sialic acid is bound to oligosaccharides, possibly making the sialic acid in infant formula less accessible to bacterial sialidases (Wang *et al.*, 2001).

Chapter III is a follow on from Chapter II, and describes the mechanism by which the sialic acid utilisation gene cluster, or *nan/nag* cluster, is regulated. Unlike the sialic acid catabolic pathway, the transcriptional regulation of sialic acid utilisation may vary among bacterial components of the intestinal microbiota. In *B. breve* UCC2003, transcription of the *nan/nag* cluster is controlled by a GntR-family transcriptional

regulator, designated NanR, representing the first regulator of this kind to be characterised from bifidobacteria.

From *in silico* analysis of the *B. breve* UCC2003 genome, it is notable that this strain encodes just three predicted extracellular glycosyl hydrolases (O'Connell Motherway *et al.*, 2011b). Therefore it seems unlikely that *B. breve* UCC2003 follows the same strategy as that employed by *B. bifidum* PRL2010, whereby mucin and HMO are extracellularly degraded to their component mono- and oligosaccharides (Asakuma *et al.*, 2011; Turroni *et al.*, 2010). Instead, as described in Chapter II of this thesis, *B. breve* UCC2003 can take advantage of the extracellular sialidase activity of *B. bifidum* PRL2010 to scavenge sialic acid released from 3' sialyllactose. Following a similar strategy, it was shown in Chapter IV that, while being itself incapable of degrading mucin, *B. breve* UCC2003 is a highly effective scavenger of the constituent mono- and oligosaccharides of mucin released by the extracellular glycosyl hydrolase activity of *B. bifidum* PRL2010. While such cross-feeding between genera and species of the intestinal microbiota had long been speculated (Hoskins *et al.*, 1985; Sela & Mills, 2010; Ward *et al.*, 2007), these results provide conclusive proof that *B. breve* can indeed enhance its growth and survival abilities through the metabolism of mucin and HMO components, which are released by *B. bifidum* extracellular activity.

Finally, Chapter V introduced a previously unexplored area of carbohydrate metabolism in bifidobacteria, namely the metabolism of sulfated carbohydrates. *B. breve* UCC2003 was shown to metabolise the sulfated monosaccharide *N*-acetylglucosamine-6-sulfate (GlcNAc-6-S) through one of two sulfatase-encoding gene clusters identified on the *B. breve* UCC2003 genome, whereas the substrate of the second cluster has yet to be

determined. It is intriguing that within the *Bifidobacterium* genus, sulfatase-encoding genes are restricted to and conserved within the *B. breve* species only, apparently allowing members of this species to utilise sulfated components of mucin, HMOs and other glycans, which are inaccessible even to *B. bifidum* and *B. longum* subsp. *infantis*, representing known specialists of mucin and HMO metabolism.

While each chapter of this thesis deals with a specific aspect of host-derived carbohydrate metabolism, there are connecting aspects of the studied metabolic activities. This became obvious when analysing the transcriptomic response of *B. breve* UCC2003 to growth on GlcNAc-6-S. Within the *ats2* gene cluster and its neighbouring Bbr_0846-049 gene cluster, each of the genes required for the metabolism of this sulfated sugar to fructose-6-phosphate can be identified. We predict that the sulfate residue is removed by AtsA2-mediated sulfatase activity, followed by the catabolism of GlcNAc as previously described in *E. coli* (White, 1968), through the activities of the NagK, NagA1 and NagB2 enzymes, the genes for which are located within the Bbr_0846-0849 gene cluster (Chapter V). If GlcNAc-6-S is indeed metabolised through this pathway, the additional *nagA2* and *nagB3* genes, and the LNB/GNB gene cluster which are also up-regulated during growth on this sulfated sugar, do not appear to add any functionality.

However, by analysis of sialic acid metabolism, as well as the NanR-mediated regulation of the *nan/nag* cluster, some logic for this apparent anomaly can be found. In Chapter II it was shown that by the combined activities of the NanA, NanK and NanE enzymes, *N*-acetylglucosamine-6-phosphate (GlcNAc-6-P) is produced from sialic acid, which is then converted to fructose-6-phosphate by the combined activities of NagA2

and NagB3 (Egan *et al.*, 2014a). However in Chapter III it was found that only the *nanA, nanK* and *nanE* genes, which are part of the *nan/nag* cluster, are under the control of the NanR repressor, whereas transcription of the *nagA2* and *nagB3* genes is not. We hypothesise that the accumulation of the metabolic intermediate GlcNAc-6-P induces expression of the *nagA2* and *nagB3* genes, which are possibly under the transcriptional control of a predicted ROK-family transcriptional regulator, encoded by a gene that is located immediately downstream of the *nagB3* gene. This hypothesis would also explain how the accumulation of GlcNAc-6-P by NagK activity causes the induction of the *nagA2* and *nagB3* genes during growth on GlcNAc-6-S.

As mentioned above, there seems no metabolic reason for the induction of the predicted LNB/GNB gene cluster during growth on GlcNAc-6-S. However, a careful review of the previously performed microarrays of *B. breve* UCC2003 during growth on mucin in co-culture with *B. bifidum* PRL2010 and in sialic acid (when the imposed threshold for up-regulation was lowered slightly) revealed that the LNB/GNB gene cluster was also significantly up-regulated (fold change >2.5, $P<0.001$) during growth under the aforementioned conditions (Chapters II and IV). It is not surprising that these genes are induced during growth in co-culture as GNB is the primary disaccharide of the core 1 and core 2 structures of mucin and it was hypothesised that it represents one of the mucin components supporting growth of *B. breve* UCC2003 under these conditions (Egan *et al.*, 2014b). However, none of the genes encoded by the LNB/GNB gene cluster have a role in relation to sialic acid metabolism (Egan *et al.*, 2014a). It seems that the up-regulation of this gene cluster is concurrent with the up-regulation of the *nagA2* and *nagB3* genes even during growth on substrates such as GlcNAc-6-S and

sialic acid which have no obvious need for the protein products of this gene cluster. This would imply that the *nagA2* and *nagB3* genes and the predicted LNB/GNB gene cluster are co-regulated and if our hypothesis is correct, the metabolic intermediate that controls this regulation is GlcNAc-6-P.

From the information provided in this thesis, we can hypothesise that GlcNAc-6-P or its dephosphorylated form *N*-acetylglucosamine (GlcNAc) play an important role in the regulation of various gene clusters involved in the metabolism of host-derived carbohydrates. GlcNAc is a primary monosaccharide constituent in a wide range of host-derived carbohydrates including mucin (Varki *et al.*, 2009), *N*-linked glycans (Morelle & Michalski, 2007) and in the majority of HMO species (Bode & Jantscher-Krenn, 2012). Even in the metabolism of sugars which do not contain GlcNAc, such as sialic acid and GNB, GlcNAc-6-P is produced as an intermediate (Nishimoto & Kitaoka, 2007; Plumbridge & Vimr, 1999). Therefore it seems that GlcNAc-6-P is a common intermediate in the metabolic pathways of a wide range of host-derived carbohydrates (Fig. 6.1) and perhaps it is unsurprising that it may play a role in the induction of a variety of genes related to host-derived carbohydrate metabolism. The first step towards proving this hypothesis would be the construction of an insertion mutant in the gene (with locus tag Bbr_1249), which is predicted to encode a ROK-family transcriptional regulator and is located immediately downstream of the *nagA2* and *nagB3* genes. If this protein functions as a transcriptional repressor, a microarray would reveal the gene clusters it regulates. If successful, this could then be followed by cloning and overproduction of the protein and electrophoretic mobility shift assays in the presence or

absence of GlcNAc-6-P to determine if it is the effector molecule of the transcriptional regulator.

It was previously shown in *E. coli* that the metabolism of amino sugars such as GlcNAc, *N*-acetylmannosamine as well as sialic acid requires convergent pathways which ultimately result in the production of fructose-6-phosphate (Plumbridge & Vimr, 1999). It seems that a similar system of converging pathways exists in *B. breve* UCC2003. However, possibly due to horizontal gene transfer events, as speculated for the acquisition of the *ats2* and Bbr_0846-0849 gene clusters (Chapter V), some of these pathways are partially redundant, thus resulting in the simultaneous up-regulation of homologous genes. Whilst this can be seen as inefficient (and perhaps explains the lower level of growth of the strain on GlcNAc-6-S compared to lactose), it also has its advantages, as the strain can cope with the loss of a copy of each gene. In fact such an event appeared to already have occurred in *B. breve* UCC2003. Within the sialic acid utilisation cluster, the Bbr_0170 gene encodes a truncated, and therefore a presumed non-functional *nagA* gene, yet the strain can still utilise sialic acid (Egan *et al.*, 2014a). Furthermore, construction of an insertion mutant in the *nagA2* gene did not result in the impairment of growth on sialic acid, presumably because the *nagA1* gene compensates for the *nagA2* mutation (Egan *et al.*, 2014a).

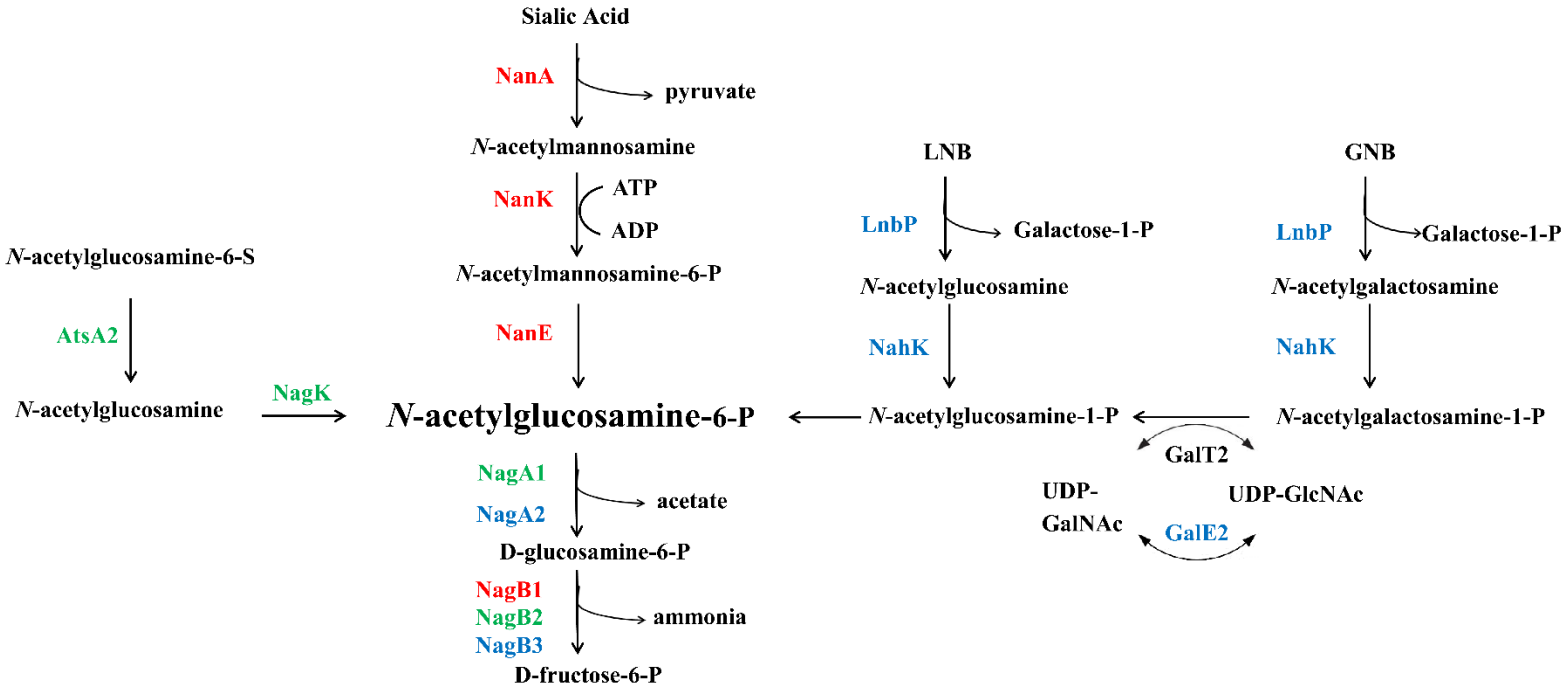


Figure 6.1: Proposed routes for the breakdown of GlcNAc-6-S, LNB and GNB as well as the proven catabolic pathway of sialic acid in *B. breve* UCC2003. Each metabolic pathway has GlcNAc-6-P as a common intermediate, which is converted to fructose-6-phosphate by the combined activities of NagA and NagB and subsequently enters the bifid shunt. The galactose-1-phosphate produced from LnbP activity on LNB and GNB is presumably metabolised via the Leloir pathway. Enzymes are grouped by colour based on the proven or predicted transcriptional regulator of their encoding genes. Red: regulated by NanR (Chapter III); Green: regulated by AtsR2 (Chapter V); Blue: putatively regulated by a ROK family transcriptional repressor encoded by the Bbr_1249 gene.

On the *B. breve* UCC2003 genome, it is interesting to note the difference between the gene clusters involved in the metabolism of plant-derived carbohydrates and those involved in host-derived carbohydrate metabolism. The *cld* cluster for cellodextrin metabolism (Pokusaeva *et al.*, 2011), the *gal* locus involved in galactan metabolism (O'Connell Motherway *et al.*, 2011a), the *raf* cluster for raffinose metabolism and the *melABCDE* cluster for melezitose metabolism (O'Connell *et al.*, 2013) are involved in the catabolism of various plant-derived carbohydrates. These gene clusters mirror the typical carbohydrate utilisation cluster as described in *B. longum* subsp. *longum* as they include a LacI-type transcriptional regulator (with the exception of the RafR ROK-family activator of the *raf* cluster), an ABC-type uptake system and between one and six genes encoding glycosyl hydrolases (Schell *et al.*, 2002). Therefore, the genes required for the uptake, metabolism and regulation of these plant-derived sugars are all specifically confined within these gene clusters. This is in contrast to the metabolism of host-derived carbohydrates, which as described in this thesis can involve multiple unconnected gene clusters and metabolic pathways. Notably, *Bacteroides thetaiotaomicron*, another member of the intestinal microbiota which can utilise both dietary and host glycans, was shown to preferentially utilise plant-derived polysaccharides when both such carbon/energy sources are available (Sonnenburg *et al.*, 2005). It would be interesting to perform similar analyses for *B. breve* UCC2003, the results of which may be useful in the development of new prebiotic or synbiotic products.

From the work presented in this thesis, it can be concluded that *B. breve* UCC2003 is a highly versatile microorganism which employs a very flexible and multifaceted

approach to carbohydrate utilisation in the gut. As we continue to employ a functional genomic approach to increase our understanding of this strategy, it is interesting to observe and investigate the complex, interactive network of metabolic and regulatory pathways that allow *B. breve* UCC2003 to utilise a variety of host-derived carbohydrates.

6.1 REFERENCES

Anba-Mondoloni, J., Chaillou, S., Zagorec, M. & Champomier-Vergès, M.-C. (2013). Catabolism of *N*-acetylneuraminic acid, a fitness function of the food-borne lactic acid bacterium *Lactobacillus sakei*, involves two newly characterized proteins. *Appl Environ Microbiol* **79**, 2012-2018.

Asakuma, S., Hatakeyama, E., Urashima, T., Yoshida, E., Katayama, T., Yamamoto, K., Kumagai, H., Ashida, H., Hirose, J. & other authors (2011). Physiology of consumption of human milk oligosaccharides by infant gut-associated bifidobacteria. *J Biol Chem* **286**, 34583-34592.

Bode, L. & Jantscher-Krenn, E. (2012). Structure-function relationships of human milk oligosaccharides. *Adv Nutr* **3**, 383S-391S.

Bottacini, F., OConnell Motherway, M., Kuczynski, J., OConnell, K., Serafini, F., Duranti, S., Milani, C., Turroni, F., Lugli, G. & other authors (2014). Comparative genomics of the *Bifidobacterium breve* taxon. *BMC Genomics* **15**, 170.

Duranti, S., Turroni, F., Milani, C., Foroni, E., Bottacini, F., Dal Bello, F., Ferrarini, A., Delledonne, M., van Sinderen, D. & other authors (2012). Exploration of genomic diversity and core genome of the *Bifidobacterium adolescentis* phylogenetic group by means of a polyphasic approach. *Appl Environ Microbiol* **79**, 336-346.

Egan, M., O'Connell Motherway, M., Ventura, M. & van Sinderen, D. (2014a). Metabolism of sialic acid by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **80**, 4414-4426.

Egan, M., O' Connell Motherway, M., Kilcoyne, M., Kane, M., Joshi, L., Ventura, M. & van Sinderen, D. (2014b). Cross-feeding by *Bifidobacterium breve* UCC2003 during co-cultivation with *Bifidobacterium bifidum* PRL2010 in a mucin-based medium. *BMC Microbiol* **14**, 282.

Fanaro, S., Chierici, R., Guerrini, P. & Vigi, V. (2003). Intestinal microflora in early infancy: composition and development. *Acta paediatrica* **92**, 48-55.

Hoskins, L. C., Agustines, M., McKee, W. B., Boulding, E. T., Kriaris, M. & Niedermeyer, G. (1985). Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *J Clin Invest* **75**, 944-953.

Lugli, G. A., Milani, C., Turroni, F., Duranti, S., Ferrario, C., Viappiani, A., Mancabelli, L., Mangifesta, M., Taminiau, B. & other authors (2014). Investigation of the evolutionary development of the genus *Bifidobacterium* by comparative genomics. *Appl Environ Microbiol* **80**, 6383-6394.

Milani, C., Duranti, S., Lugli, G. A., Bottacini, F., Strati, F., Arioli, S., Foroni, E., Turroni, F., van Sinderen, D. & other authors (2013). Comparative genomics of

Bifidobacterium animalis subsp. *lactis* reveals a strict monophyletic bifidobacterial taxon. *Appl Environ Microbiol* **79**, 4304-4315.

Milani, C., Lugli, G. A., Duranti, S., Turroni, F., Bottacini, F., Mangifesta, M., Sanchez, B., Viappiani, A., Mancabelli, L. & other authors (2014). Genome encyclopaedia of type strains of the genus *Bifidobacterium*. *Appl Environ Microbiol* **80**, 6290-6302.

Morelle, W. & Michalski, J.-C. (2007). Analysis of protein glycosylation by mass spectrometry. *Nat Protocols* **2**, 1585-1602.

Nishimoto, M. & Kitaoka, M. (2007). Identification of *N*-acetylhexosamine 1-kinase in the complete lacto-*N*-biose I/galacto-*N*-biose metabolic pathway in *Bifidobacterium longum*. *Appl Environ Microbiol* **73**, 6444-6449.

O'Connell, K. J., O'Connell Motherway, M., O'Callaghan, J., Fitzgerald, G. F., Ross, R. P., Ventura, M., Stanton, C. & van Sinderen, D. (2013). Metabolism of four α-glycosidic linkage-containing oligosaccharides by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **79**, 6280-6292.

O'Connell Motherway, M., Fitzgerald, G. F. & van Sinderen, D. (2011a). Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* **4**, 403-416.

O'Connell Motherway, M., Zomer, A., Leahy, S. C., Reunanen, J., Bottacini, F., Claesson, M. J., O'Brien, F., Flynn, K., Casey, P. G. & other authors (2011b). Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc Natl Acad Sci USA* **108**, 11217-11222.

Olson, M. E., King, J. M., Yahr, T. L. & Horswill, A. R. (2013). Sialic acid catabolism in *Staphylococcus aureus*. *J Bacteriol* **195**, 1779-1788.

Plumbridge, J. & Vimr, E. (1999). Convergent pathways for utilization of the amino sugars *N*-acetylglucosamine, *N*-acetylmannosamine, and *N*-acetylneuraminic acid by *Escherichia coli*. *J Bacteriol* **181**, 47-54.

Pokusaeva, K., O'Connell-Motherway, M., Zomer, A., MacSharry, J., Fitzgerald, G. F. & van Sinderen, D. (2011). Cellodextrin utilization by *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* **77**, 1681-1690.

Roger, L. C., Costabile, A., Holland, D. T., Hoyles, L. & McCartney, A. L. (2010). Examination of faecal *Bifidobacterium* populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology* **156**, 3329-3341.

Schell, M., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M.-C., Desiere, F., Bork, P. & other authors (2002). The genome sequence

of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci U S A* **99**, 14422-14427.

Sela, D. & Mills, D. (2010). Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol* **18**, 298-307.

Sela, D. A., Chapman, J., Adeuya, A., Kim, J. H., Chen, F., Whitehead, T. R., Lapidus, A., Rokhsar, D. S., Lebrilla, C. B. & other authors (2008). The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A* **105**, 18964-18969.

Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C.-H., Westover, B. P., Weatherford, J., Buhler, J. D. & Gordon, J. I. (2005). Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307**, 1955-1959.

Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J.-H., Zomer, A., Sanchez, B., Bidossi, A., Ferrarini, A. & other authors (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., Duranti, S., Bottacini, F., Guglielmetti, S., Van Sinderen, D. & Ventura, M. (2014). *Bifidobacterium bifidum* as an example of a specialized human gut commensal. *Front Microbiol* **5**.

Varki, A., Cummings, R. D., Esko, J. D., Freeze, H., Stanley, P., Bertozzi, C. R., Hart, G. W. & Etzler, M. E. (2009). *Essentials of Glycobiology*, 2nd edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Ventura, M., Turroni, F., Lugli, G. A. & van Sinderen, D. (2014). Bifidobacteria and humans: our special friends, from ecological to genomics perspectives. *J Sci Food Agr* **94**, 163-168.

Vimr, E., Lichtensteiger, C. & Steenbergen, S. (2000). Sialic acid metabolism's dual function in *Haemophilus influenzae*. *Mol Microbiol* **36**, 1113-1123.

Wang, B., Brand-Miller, J., McVeagh, P. & Petocz, P. (2001). Concentration and distribution of sialic acid in human milk and infant formulas. *Am J Clin Nutr* **74**, 510-515.

Ward, R., Ninonuevo, M., Mills, D., Lebrilla, C. & German, B. (2007). In vitro fermentability of human milk oligosaccharides by several strains of bifidobacteria. *Mol Nutr Food Res* **51**, 1398-1405.

White, R. (1968). Control of amino sugar metabolism in *Escherichia coli* and isolation of mutants unable to degrade amino sugars. *Biochem J* **106**, 847-858.

ACKNOWLEDGMENTS

It is my pleasure to acknowledge the people who have contributed to the completion of this thesis.

First and foremost, I would like to express my deepest gratitude to my supervisor Professor Douwe van Sinderen. Thank you for your continued encouragement, patience and belief in me, as well as the countless helpful discussions from which I learned so much. I could not have asked for a better mentor for my PhD.

To Dr. Mary O' Connell Motherway, I greatly appreciate all the help and advice you have given me throughout my PhD. Thank you for sharing your knowledge and experience, your guidance was invaluable to me.

I would like to acknowledge Professor Colin Hill who was on my thesis committee, read my reports and attended poster sessions.

The staff of the Microbiology Department, in particular Dan Walsh for performing HPLC analysis and his help with various computer issues, as well as Paddy O' Reilly, Collette Crowley, Margaret Higgins and Pat Higgins, to name but a few.

I would also like to acknowledge people I collaborated with in various parts of this thesis. Dr. Michelle Kilcoyne, Dr. Marian Kane and Professor Lokesh Joshi of the Alimentary Glycoscience Research Centre, NUIG, Ireland, who performed HPAEC-PAD and HPLC analysis for the co-culture paper. I greatly appreciate their help in analysing the data and writing the paper. Professor Marco Ventura of the University of Parma, Italy, who provided the *B. bifidum* PRL2010 strain and whose insight and advice when writing my two papers was hugely appreciated. Dr. Daniel O' Driscoll

from UCC for performing mass spectrometry analysis and finally, Dr. Lorena Ruiz for providing me with the transposon mutant for the co-culture paper.

To my colleagues in Lab 4.25 and to everyone who has passed through over the last four years, thank you for sharing the highs and lows of research with me and for making the lab an enjoyable place to work. I am grateful to all of you and everybody in Lab 340 and Lab 4.23. To my “twin” Amy, I’ve been lucky to work alongside such an awesome friend for the last four years. Thank you for celebrating with me when things went right and for your positive attitude during the more stressful times. My special thanks also to Debbie and Kerry who were there from the beginning, I was lucky to work with such great friends.

To my parents, Tom and Brede, I would never have achieved this without your support and encouragement throughout my education. My motivation to always do the best I can comes from both of you. I cannot thank you enough.

To my sisters Karina and Karole, who have always been there to spur me on, thank you for always showing an interest in my work, even from the other side of the world, it meant a lot to me.

To my aunt Mary Tereasa, the first scientist of our family. Your passion for science and your enthusiasm for always learning more have always been an inspiration to me.

Finally to my dearest friends, Clare and Megan, I am grateful for your support and friendship over the years.