Exploring Fructan Utilisation by Members of the Human Intestinal Microbiota

Sarah A. Shapiro

A Thesis Submitted for the Degree of Doctor of Philosophy

2012 – 2015

Institute for Cell and Molecular Biosciences

Newcastle University

Acknowledgements

Foremost, I would like to acknowledge my supervisor, Dr. David Bolam. Dave has entertained a barrage of unanswerable questions and somewhat unconventional experimental designs with boundless patience. Without his guidance, expertise, patience and support, this thesis would not have been possible.

I would like to extend my thanks to Professor Harry Gilbert, who has provided me with invaluable advice, support and encouragement throughout my project.

I am extremely grateful to all of my colleagues, for their advice, training and friendship.

I would also like to acknowledge my collaborators who have contributed to my research; Mr. Carl Morland who provided excellent technical support throughout my work, Professor Glenn Gibson for allowing me access and training with his facilities within the University of Reading and Dr. Arnaud Baslé with whom I have collaborated extensively to obtain x-ray crystal structures.

To my Fiancé, Jonathan Ward, I am forever grateful for all the unwavering support and encouragement over the years.

Abstract

The human gut microbiota contributes to host health and wellbeing in numerous ways, including through polysaccharide fermentation and the production of beneficial short chain fatty acids. Microbiota accessible carbohydrates (MACs) pass through the early digestive tract intact to nourish the microbiota.

Fructans are plant derived fructose polymers found in the diet which act as MACs and elevate short chain fatty acid production. Two linkage types are known between fructose units within fructans, β2-1 and β2-6; homopolymers of these are inulin and levan respectively. Fructan structures containing both β2-1 and β2-6 linkages are common within the human diet, but the role of these as MACs have not been studied. Fructan, particularly inulin and fructo-oligosaccharides (FOS) have been used as prebiotics to selectively support beneficial members of the microbiota. Despite widespread use of inulin as a prebiotic there is a paucity of data regarding the mechanisms employed by the microbiota to recognise, import and degrade this glycan.

In this thesis, several plant derived fructan extracts were visualised using high performance anion exchange chromatography and thin layer chromatography to confirm these contained non-linear structures. We show that these non-linear structures may support both inulin- and levan- utilising *Bacteroides* species, despite previously identified mutually exclusivity for one of the two homopolymers.

Two key inulin using species were chosen as models to probe inulin utilisation within the microbiota; the Gram negative *Bacteroides ovatus*, and the Gram positive *Bifidobacterium adolescentis*.

B. ovatus is a prominent member of the healthy human microbiota, and contains a large number of polysaccharide utilisation loci within the genome, each tightly regulated and specific for a selected glycan. Here we show that *B. ovatus* encodes an inulin utilisation system which includes an endo-acting surface located glycoside hydrolase family 91 enzyme comprised of two gene products (BACOVA_04502 and BACOVA_04503). This enzyme has an appended Carbohydrate Binding Module (CBM) which recognises inulin. SusC/D-homologue pairs are carbohydrate binding and import proteins commonly found within the Bacteroidetes phylum. The structure of a SusD-homologue, BACOVA_04504 was solved and we demonstrate that this protein recognises sucrose terminated FOS and inulin, an unusual specificity which likely assists in the rapid import of desirable short chain FOS though a SusC-homologue. High molecular weight inulin is imported through a SusC-homologue for periplasmic degradation by a glycoside hydrolase family 32 enzyme (GH32). A waste sugar, di-fructose anhydride (DFA), is produced by this system from inulin and released into the intestinal environment; we demonstrate that DFA is produced from inulin by faecal microbiota from three healthy humans and that DFA is not subsequently broken down.

B. adolescentis is able to rapidly utilise inulin polysaccharide, an unusual phenotype amongst *Bifidobacterium*, which generally utilise only short chain FOS, Indeed it has been shown that *B. adolescentis* is supported well during prebiotic treatment. We demonstrate that *B. adolescentis* contains an additional locus compared with other *Bifidobacterium species* and that this locus is responsible for inulin utilisation. The locus includes an inulin recognising extracellular solute binding protein which undergoes a significant conformational change upon ligand recognition which we show through structural studies. The locus contains a LacI-homologue which is able to recognise fructan through the periplasmic binding domain, likely to up-regulate the system; and inulin is finally processed internally by a GH32.

We explore the niches occupied by each species in the intestine, and predict that *B. ovatus* "shares" glycan with other microbiota, including members of the *bifidobacterium* genus. Our data enable deeper understanding of how fructans interact with the intestinal microbiota, potentially underpinning research into novel and personalised prebiotic therapies.

Contents

List of Figures

List of Tables

Glossary of Terms and Abbreviations

BACOVA_#:
Locus tag within the *Bacteroides ovatus* genome

BAD_#:
Locus tag within the *Bifidobacterium adolescentis* genome

Cazyme:
Carbohydrate Active Enzyme

CBM:
Carbohydrate binding module

Cq:
Quantification cycle, the cycle at which signal exceeds noise in qPCR experiments.

DP:
Degree of polymerisation (e.g. chain length)

DFA:
Di-fructose anhydride, a small sugar

DFA-FOS:
Fructooligosaccharide with a di-fructose anhydride terminus

ESBP:
Extracellular Solute Binding Protein

FPLC:
Fast protein liquid chromatography

FOS:
Fructooligosaccharide

Fructotriose:
β2-1 linked fructose trisaccharide ($F_{\beta 2\text{-}1}F_{\beta 2\text{-}1}F$)

GH:
Glycoside hydrolase

G.I. Tract:
Gastro-intestinal tract

HEPES:
4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid, used as a buffering agent.

HPAEC-PAD:
High performance anion exchange chromatography with pulsed amperometric detection

HMW:
High molecular weight (generally greater than DP > 8)

ITC:
Isothermal titration calorimetry; used for analysing protein-protein or protein-substrate binding.

Kestose:
Trisaccharide with a terminal Sucrose and a β2-1 linked fructose ($G_{\alpha 1\text{-}2\beta}F_{\beta 2\text{-}1}F$)

Kestotetraose:
Tetrasaccharide with a terminal Sucrose and two β2-1 linked fructose units
($G_{\alpha 1\text{-}2\beta}F_{\beta 2\text{-}1}F_{\beta 2\text{-}1}F$)

Kestopentaose:
Pentasaccharide with a terminal Sucrose and three β2-1 linked fructose units
($G_{\alpha 1\text{-}2\beta}F_{\beta 2\text{-}1}F_{\beta 2\text{-}1}F_{\beta 2\text{-}1}F$)

***K_a*:**
Association constant

***K_{cat}*:**
Enzyme turnover; the number of molecules of product formed per unit of time.

***K_d*:**
Dissociation constant

***K_m*:**
The Michaelis constant, which provides an indication of affinity

LMW:
Low molecular weight (generally fewer than DP < 8)

MACs:
Microbiota Accessible Carbohydrates

Metabolomics:
Study of the total metabolite content of a system

Microbiome:
The study of the collective genes, genomes or an aspect of a microbial community, particularly associated with humans (e.g. the gut microbiome).

Microbiota:
A community/population of microbes within a specific environment.

MALDI-TOF MS:
Matrix-assisted laser desorption/ionization-time of flight mass spectrometry, used to measure the mass of small compounds

OD:
Optical density

PBP:
Periplasmic Binding Protein

PCR:
Polymerase chain reaction, used to amplify DNA

PUL:
Polysaccharide utilisation loci

Prebiotic:
Non-living agent used to modulate microbiota (usually a carbohydrate)

Probiotic:
Living agent used to modulate the microbiota (usually a native microbial species)

qPCR:
quantitative real-time polymerase chain reaction, used to detect and quantify specific DNA regions

SDS-PAGE:
Sodium dodecyl sulphate polyacrylamide gel electrophoresis; used to separate and visualise proteins

SUS:
Starch utilisation system

SusA-G:
Proteins within the prototypic SUS

TLC:
Thin layer chromatography; used to separate and visualise sugars

V_{max}:
The maximum rate (e.g. of an enzyme)

Chapter 1. Introduction

1.1. Exploring the Healthy Human Gut Microbiota

1.1.1. Introduction

Humans did not arise from within a sterile environment; our relationship with microbes is not passive, but intricate and integral to our health and well-being. Viable micro-environments on and inside of the human body, such as the skin, urogenital tract and digestive tract, are colonised by specialised microbial communities (**Figure 1.1;** Lasken & McLean, 2014). The largest microbial populations are established within the human large intestine, estimated to harbour 10^{11}-10^{12} cells/ml (Qin *et al.*, 2010; Ley *et al.*, 2006).

The advent of cheaper, high throughput sequencing techniques in the early 2000s led to an explosion of available genome sequences; for the first time, detailed exploration of the human microbiota as a whole could be undertaken, where previously only microbes which could be cultured could be studied. Indeed the cost of sequencing still declines as automated methods improve (Metzker, 2010). In addition to the identification of thousands of novel genes from intestinal microbes, sequencing can now be used to evaluate the bacterial diversity and abundance within complex samples without the need for culturing, through shotgun metagenomic sequencing, multiple displacement amplification and 16S classification. It is these pioneering techniques which have facilitated the expansion of research into understanding the microbiota. Whilst powerful, these techniques are still developing and researchers must be vigilant and employ well-informed sampling techniques to avoid artificially skewing the data (Walker *et al.*, 2015).

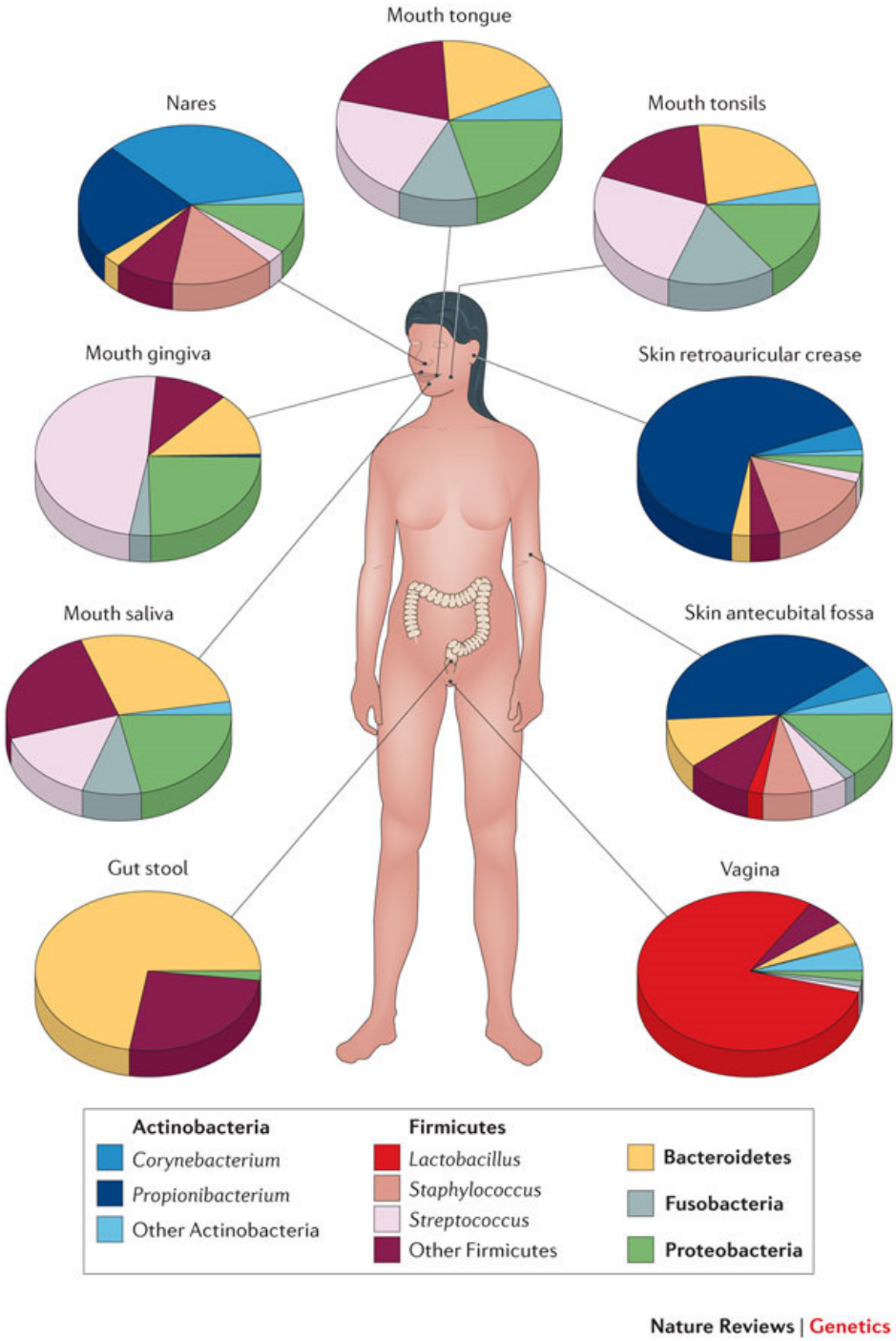


Figure 1.1. The human body is colonised by specialised microbial communities (Lasken & McLean, 2014). Bacterial diversity (relative abundance) across body sites was determined using single cell multiple displacement amplification (a PCR independent amplification technique) and subsequent classification of resultant 16S sequences (Lasken & McLean, 2014).

The microbiome refers to the collective microbial genomes within an individual or body site and may also be used to describe the collective action of these genes (e.g. during fermentation). The Human Microbiome Project (HMP) was established in 2008 with the general aim of characterising the human microbiome, including the gut microbiome.

One of the original goals of the HMP is to discover whether there is a core set of genes in all healthy individuals, referred to as the "core microbiome" (**Figure 1.2**; Turnbaugh *et al.,* 2007).

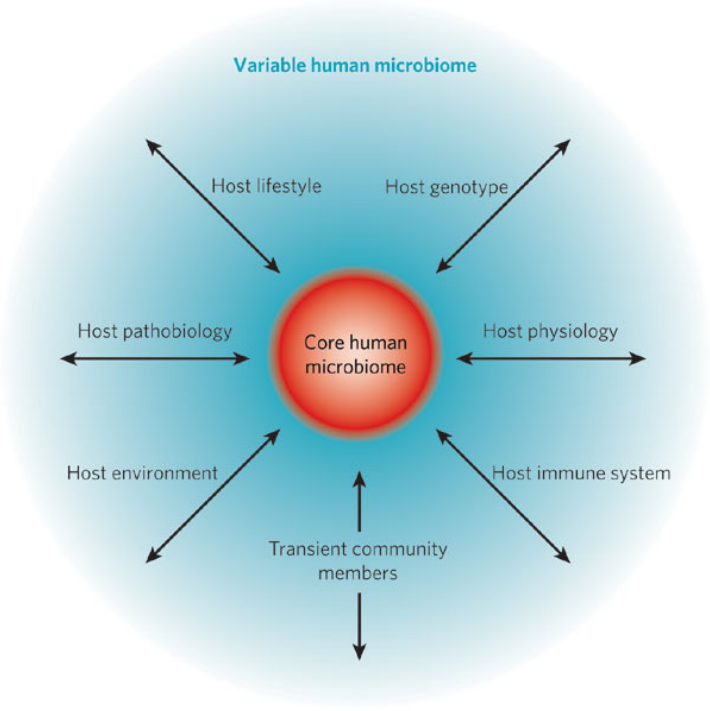


Figure 1.2. The concept of the core human microbiome (Turnbaugh *et al.*, 2007). This diagram demonstrates the concept of the core human microbiome (red), the collective subset of genes encoded by microbiota across all or the majority of healthy humans. Depending on the personalised microbial community of each individual, which is influenced by the environmental and internal factors listed, a variable or personalised microbiome will arise (blue). The gradient and two-way arrows indicate the relative plasticity of the human microbiome compared to the human genome – core genes may become superfluous, and genes from the variable microbiome may become part of the core subset as humans and their endogenous populations evolve (Turnbaugh *et al.*, 2007).

The concept of a core microbiome allows for the identification of genes or pathways which are vital to human health and therefore could underpin novel therapeutics aimed at maintaining or establishing healthy microbial communities.

The search for a core microbiome has proven difficult however, with healthy individuals displaying highly diverse microbial communities. The MetaHIT study in 2011 identified three global varieties, or enterotypes, of the healthy human gut microbiota. These enterotypes (1-3) were identified by variations in the prominence of three genera, *Bacteroides*, *Prevotella* and *Ruminococcus* respectively. The authors concluded that microbial diversity in healthy individuals was therefore stratified rather than continuous (**Figure 1.3;** Arumugam *et al*., 2011); however, it is argued that classifying the intestinal microbiota in this manner is an oversimplification of continuous diversity between individuals (Knights *et al.*, 2014).

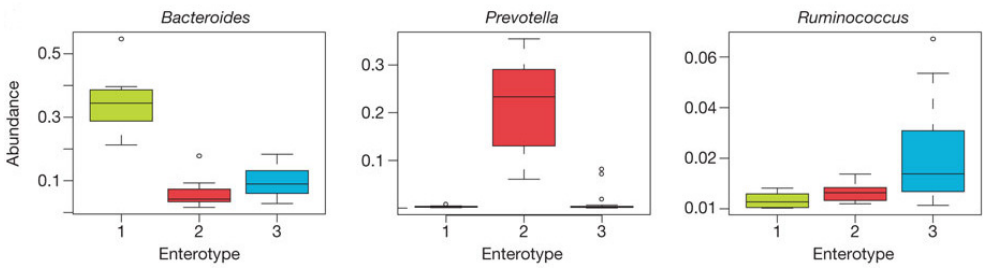


Figure 1.3. Three enterotypes of the human gut microbiota (Arumugam *et al.*, 2011). The three enterotypes identified by Arumugam *et al.* can be identified by variations in the abundance of three genera *Bacteroides* (green), *Prevotella* (red) and *Ruminococcus* (blue). Enterotype 1 is rich in *Bacteroides,* enterotype 2 in *Prevotella* and enterotype 3 in *Ruminococcus.* Enterotype 2 displays reduced representation of *Bacteroides* and *Ruminococcus* whilst the other two enterotypes maintain larger populations of all three genera (Arumugam *et al.*, 2011).

1.1.2. Production of Short Chain Fatty Acids and Other Metabolites

As the gut microbiota ferment nutrients, metabolites are released which interact with other microbes and the host. The study of the total metabolite content across a whole system is known as “metabolomics”.

Metabolites can affect the host in a systemic manner; conventionally raised mice contained more blood plasma metabolites than germ free mice, and in over 10% of metabolites significantly different concentration were observed between the germ free and conventional mice (Wikoff *et al.*, 2009). Since the establishment of metabolomics as a useful research tool further studies have sought to identify metabolite biomarkers for health or disease, and to further understand how the metabolome changes over the course of life or during an illness.

Short Chain Fatty Acids (SCFAs) are the main metabolites produced by the microbiota upon glycan fermentation. Three SCFAs; acetate, butyrate and propionate (typically in a 3:1:1 ratio) are found with a combined concentration of 50-150 mM in the bowel (Louis *et al.*, 2014). Butyrate producers include members of the Firmicutes such as *Roseburia spp.;* propionate is mainly produced by Bacteroidetes (e.g. *Bacteroides spp.*) during glycan degradation and by Firmicutes from succinate or lactate. However two distinct propionate producing pathways (the acrylate and propanediol pathways) have also recently been identified in gut bacteria (Reichardt *et al.*, 2014). Acetate is the most abundant SCFA and is produced by either acetogenic bacteria such as *Blautia hydrogenotrophica* from H_2 and CO_2 or from formate (Miller & Wolin, 1996) or as an additional fermentation product by most of the microbiota (Louis *et al.*, 2014). SCFAs beneficially affect host physiology in a number of ways. Acetate was identified as a factor in the prevention of enteropathogenic infection through stimulation of epithelial cells (Fukuda *et al.*, 2011). Butyrate is the preferred source of energy for the intestinal epithelium (Roediger, 1980).

All major SCFAs and lactate have been implicated in reducing inflammation in the intestinal epithelium through extensive interactions with the host immune system (**Figure 1.4;** Hoeppli *et al.*, 2015; Iraporda *et al.*, 2015; Chang *et al.*, 2014; Smith *et al.*, 2013) and both butyrate and propionate have been implicated in reducing the risk of colorectal cancer (Singh *et al.*, 2014; Fung *et al.*, 2012), regulating intestinal gluconeogenesis (Brüssow & Parkinson, 2014) and improving satiety (Arora *et al.*, 2011; Hosseini *et al.*, 2011).

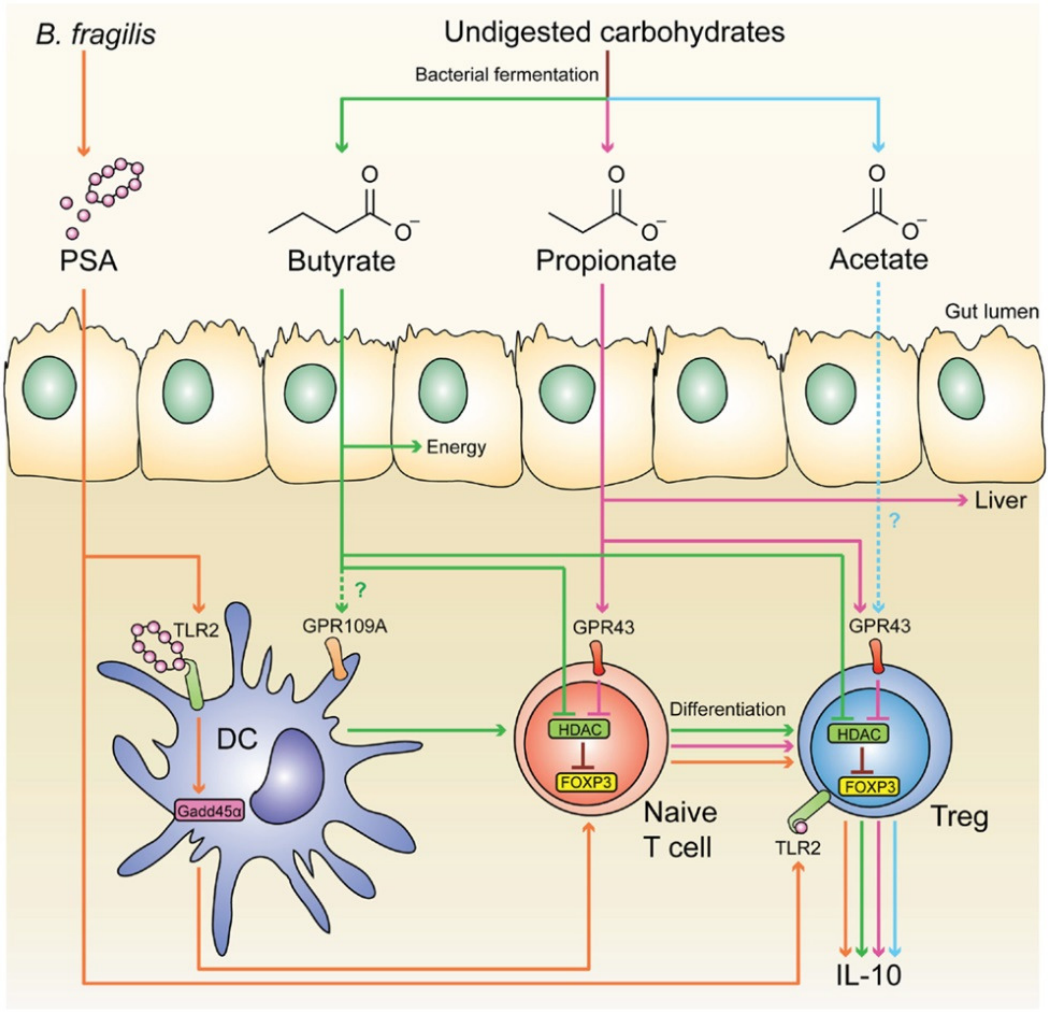


Figure 1.4. SCFA and Polysaccharide A (PSA) produced by microbial fermentation influence the host immune system (Hoeppli *et al.*, 2015). In their recent review Hoeppli and colleagues outline the presently known and suspected (?) interactions between SCFAs and regulatory T cells (Treg). The authors included PSA, a capsular glycan produced by *Bacteroides fragilis* with established therapeutic effects (Mazmanian *et al.*, 2008; Hoeppli *et al.*, 2015)

Not all metabolites are beneficial however; the degradation of protein, fat and ethanol by the microbiota can produce harmful reactive oxygen species (ROS) and other toxic or carcinogenic compounds (**Figure 1.5;** Louis *et al.*, 2014).

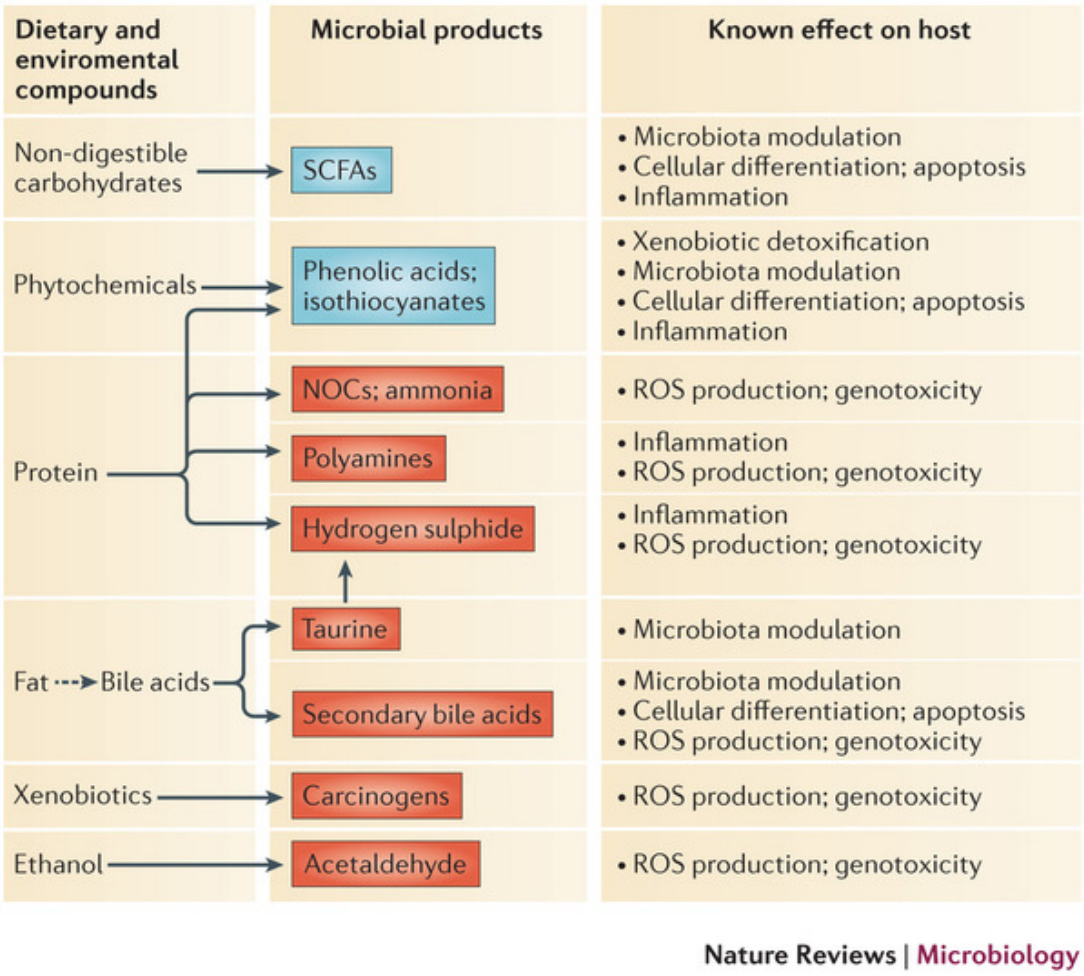


Figure 1.5. A summary of major microbial metabolites and the known effect of these compounds on the host with regards to the initiation and/or progression of colorectal cancer (Louis *et al.*, 2014). Microbial metabolites including SCFAs can have a therapeutic effect on the host with regard to colorectal cancer (shown in blue). Other metabolites may have carcinogenic effects (shown in red). Understanding the interplay between microbial metabolites and cancer will underpin the design of novel therapeutics and the identification of metabolite biomarkers (Louis *et al.*, 2014).

1.1.3. Diversity of the Human Gut Microbiota

Identifying a "core human microbiome" or distinct gut microbiota enterotypes could be a useful tool for categorising microbial diversity in order to distinguish between individuals or between healthy and diseased states, however these classifications do not give a full picture of the normal human microbial variation and remain controversial.

During the study Arumugan and colleagues compared the metagenomes from 39 individuals of 6 nationalities. The abundance of microbes at the genus level were plotted for the 30 most abundance genera (**Figure 1.6.**) showing that *Bacteroides* were highly prominent, and dominated representation of the Bacteroidetes phylum. By contrast, members of the Firmicutes phylum are more diverse at the genus level. Actinobacteria are also prominent, represented by two genera including the third most abundant genera, *Bifidobacterium*.

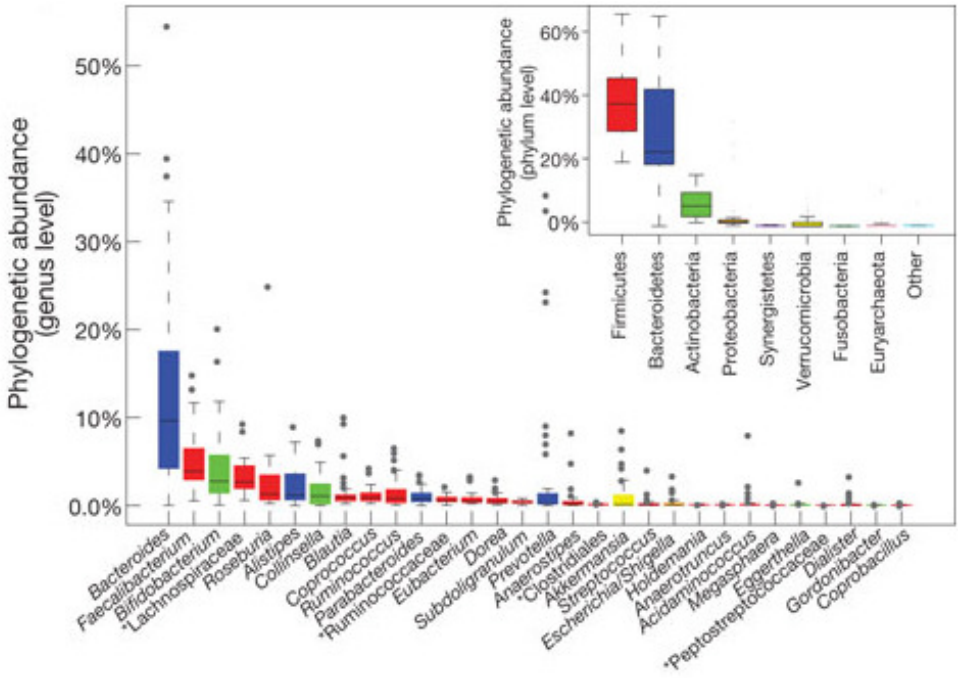


Figure 1.6. Abundance of the thirty most prominent genera across 39 healthy individuals (Arumugan *et al.*, 2011). The thirty most abundant genera by number of reads, shown as box plots. Inset: Abundance breakdown at the phylum level. Genus and phylum level abundances were calculated using reference genome based mapping using 85% and 65% cutoffs, respectively (Arumugan *et al.*, 2011).

Another large scale study performed by the HMP examined multiple body sites across 242 healthy American adults to create microbial diversity profiles for each body habitat. The researchers looked for correlations between microbial variance and host phenotype (e.g. weight, age and ethnicity). Whilst some distinct correlations were shown; age for example appears to affect the microbial composition, particularly on the skin, the variance between subjects could not be satisfactorily explained by the phenotypic metadata alone.
Within the stool samples the *Bacteroides* genus was most abundant when 16S classification was used (Huttenhower *et al.*, 2012).

Both the MetaHIT and HMP studies used stool samples as a proxy for intestinal microbial diversity. Stool sampling is a robust and non-invasive method to retrieve microbial populations from the distal gut, however these samples are not representative of the length of the tract, which is likely to harbour a variety of ecological niches and therefore differing communities along its length. As the procedure for sample harvest on internal sites is invasive, studies observing population diversity across the length of the G.I. tract have been undertaken in animal models. Mice display an increase in phylogenetic diversity in the early and late G.I. tract (gastric, duodenal; cecum, colon and faeces) compared to the mid tract (jejunum and ileum) where total diversity dropped (Gu *et al.*, 2013). The authors noted that despite large between subject variability the *Bacilli* class, *Lactobacillaceae* family and *Lactobaccilus* genus were enriched in gastric and small intestine samples, and in contrast the *Clostridia* class and the *Lachnospiraceae*, *Ruminococcaceae, Prevotellaceae*, *Rikenellaceae* and *Bacteroidaceae* families were enriched in the large intestinal and faecal samples. The authors speculate that this observation is due to changes in the availability of oxygen along the intestinal tract, based on the facultative and obligate natures of observed bacteria (Gu *et al.*, 2013). Similar changes in microbial diversity have also been observed along the pig G.I. tract, where a shift in the dominant phyla (Firmicute rich to Bacteroidetes rich) occurs from the early to late tract (Kim & Isaacson, 2015).

1.1.4. The Infant Microbiota

The infant is colonised by microbes rapidly, with some evidence to suggest colonisation initially takes place *in utero* (Aagaard *et al.*, 2014; Funkhouser & Bordenstein, 2013). After delivery, environmental microbes are key colonisers (Penders *et al.*, 2006; Bäckhed *et al.*, 2015). The microbial composition of infants occurs in successive stages, initially Firmicutes and Proteobacteria are dominant, with a steady increase of Actinobacteria over time, by six months of age Bacteroidetes begin to dominate whilst Proteobacteria and Actinobacteria gradually decline (Koenig *et al.*, 2011; Vaishampayan *et al.*, 2010). Over the first two years of life, the microbiota continues to experience compositional changes, before becoming adult-like around the age of three (Yatsunenko *et al.*, 2012).

During early development Actinobacteria, in particular the *Bifidobacterium* genera, are prominent (Bäckhed *et al.*, 2015; Koenig *et al.*, 2011). Human Milk Oligosaccharides (HMOs) present in mothers' milk have been shown to support *Bifidobacterium spp.*, likely accounting for the presence of this genus (Sela *et al.*, 2008; Garrido *et al.*, 2011; Marcobal & Sonnenburg, 2012).

A number of factors may influence the formation of the early microbiota, including delivery mode (Biasucci *et al.*, 2008; Bäckhed *et al.*, 2015) and feeding (Guaraldi & Salvatori, 2012). The developing microbiota are vulnerable; disruptive antibiotic treatment at early stages has been shown to exacerbate obesity in mice, demonstrating that perturbations during early development could impact health later in life (Cox *et al.*, 2014). Additionally, a correlation between low microbial diversity during infancy and development of allergic disease has been observed (Bisgaard *et al.*, 2011). In infants with severe or moderate acute malnutrition the development of the microbiota is retarded (Subramanian *et al.*, 2014).

1.1.5. The Microbiota and Disease

Given the central role of the microbiota toward the maintenance of digestive health, it is unsurprising that during disease states the microbial community is affected. Microbial dysbiosis is not a standardised medical term, but refers to the concept of "imbalance" within the microbial ecology (Hawrelak & Myers, 2004). Dysbiosis may refer to any state in which the microbial community is divergent from what is expected within healthy individuals, but the concept of dysbiosis is suggestive of either the loss/reduction of key members of the community, or the pronounced over representation of others. A clear, established example of microbial dysbiosis is the destruction of the normal microbiota during antibiotic treatment resulting in reduced species diversity (Dethlefsen *et al.,* 2011). Such loss of species diversity and community robustness may lead to antibiotic associated diarrhoea (AAD) via the harmful overgrowth of otherwise normal members of the community such as *C. difficile* (Bien *et al*., 2013; McFarland, 1998).

Acute dysbiosis due to antibiotic treatment or pathogen invasion is usually a transient state which can be reversed. Chronic dysbiosis, however, is a restructuring of entire communities over the course of a long term illness, which may span an individual's entire life. In many of these cases, the underlying cause of dysbiosis is obscure and can be a consequence of faulty environmental and/or internal host factors. Loss of diversity within the microbiota is a consistent indicator of dysbiosis and occurs across a variety of disease states including old age, where this phenomenon is correlated with frailty and dietary intake (Claesson *et al.*, 2012); increased adiposity, insulin resistance, dyslipidaemia and inflammation (Chatelier *et al.*, 2013); Crohn's disease, where there is a marked overall decrease in species richness coupled with aberrant shifts at the family level (Gevers *et al.*, 2014) and in type two diabetes where although only a small dysbiosis effect was observed, there was an increase in abundance of opportunistic pathogens coupled with loss of normal butyrate producing microbes (Qin *et al.*, 2012).

It is clear that the interplay between endogenous microbes and the host is incredibly complex and during disease, pathogenesis may occur through a multitude of pathways and mechanisms.

It seems that the gut microbiota plays an integral role in disease onset and may exacerbate existing conditions, therefore current and future medical interventions which can increase species diversity, and/or stimulate selected populations of beneficial microbes will be of value for treating a number of digestive diseases, and as a prophylactic treatment for maintaining host health.

1.2. Dietary Fibres, MACs and the Human Microbiota

1.2.1. Dietary Fibre and Microbiota Accessible Carbohydrates (MACs)

Monosaccharides such as glucose and fructose, and starch or starch derived oligosaccharides are digested and absorbed in the small intestine; however, all other oligo- and polysaccharides cannot be accessed by human enzymes and pass, partially digested or undigested, to the colon. These glycans, collectively referred to as dietary fibre, are derived from dietary sources such as plant storage polysaccharides, components of the plant cell wall, animal cartilage and tissue (Flint *et al.*, 2008). Not all complex dietary carbohydrates are accessible to the microbiota, indeed the greater portion of ingested cellulose remains inaccessible to endogenous microbes (Chassard *et al*., 2010). Furthermore glycans from non-dietary sources such as host mucins are accessible to some members of the microbiota. In addition, the metabolic capacity of each individual's microbiota to access glycan varieties will be personalised to some extent. Therefore a more appropriate term to describe those carbohydrates which supply energy to the microbiota was coined: Microbiota Accessible Carbohydrates or MACs (Sonnenburg & Sonnenburg, 2014). MACs provide the necessary carbon and energy for the microbiota to subsist and consists of a portion of the dietary fibre as mentioned, with the addition of host glycans and carbohydrates from other gut micro-organisms, such as mannan from yeast cell walls or other capsular glycans (**Figure 1.7;** Koropatkin *et al*., 2012).

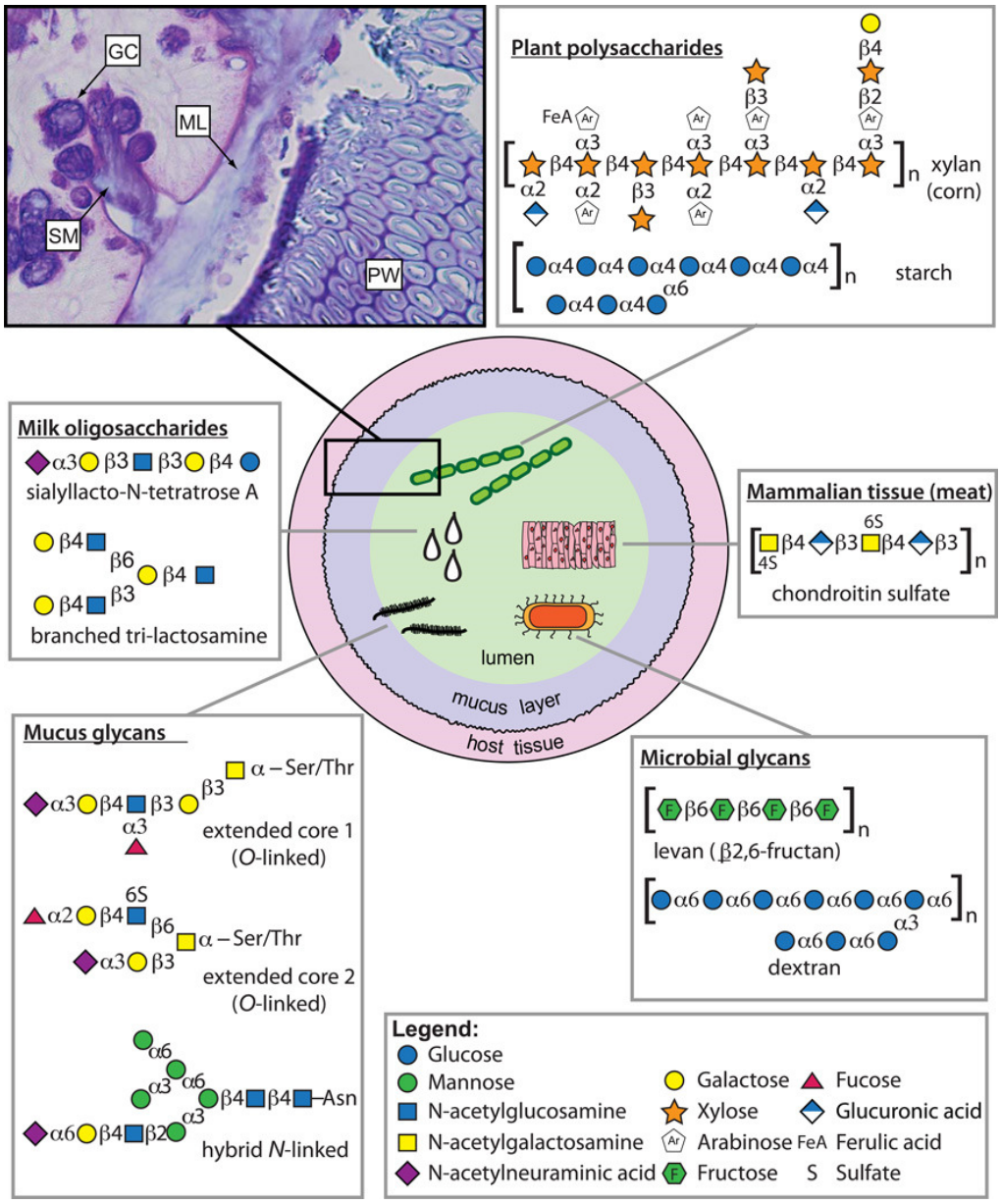


Figure 1.7. Examples of glycan sources and structures available to the gut microbiota (Koropatkin *et al.*, 2012). Cross section of the intestine (centre) demonstrating the availability and structures of five sources of microbiota accessible carbohydrates; dietary plants, dietary animal tissue, endogenous microorganisms, the mucus layer and breast milk. Representative structures are shown for each source, however true structural diversity is incredibly vast. A section of germfree mouse colon (inset) is stained to reveal host mucus-secreting goblet cells (GC), secreted mucus (SM), the mucus layer (ML) and a fragment of plant cell wall (PW); a demonstration of several glycan sources (Koropatkin *et al.*, 2012).

Whilst the human genome changes slowly across generations the microbiome is plastic. Prolonged selective pressure through the diet may sculpt the metabolic power of the microbiota; *Bacteroides plebius* isolated from Japanese individuals contains genes encoding for porphyranases and agarases able to catalyse the depolymerisation of porphyran and agar, the authors provide evidence that these enzymes are acquired from marine bacteria and likely reflect the niche present in Japanese populations where seaweed is commonly consumed, compared to north American populations which do not contain these porphyranase genes as part of the microbiome (Hehemann *et al.*, 2010).

New technologies in agriculture and the food industry have revolutionised the human diet in the developed countries. Western diets now typically contain more refined sugar, more meat and less fibre than ancestral diets (Cordain *et al.*, 2005; Jew *et al.*, 2009). Traditional diet groups have been the focus of several large metagenomic studies; traditional diet groups consist of human populations which follow a pre-urbanization diet and lifestyle, such as hunter-gatherers or diets similar to those experienced by early human settlers during the birth of agriculture (Segata, 2015).

The Hadza contained microbes with a higher diversity of carbohydrate active enzymes compared with urban Italians (Rampelli *et al.*, 2015) which may reflect more varied MAC content of the Hadza diet compared with their Italian counterparts and the need for functional flexibility across seasonal change in diet. Changes in microbial composition were observed in the Hadza vs. the Italian control group, most notably an increase in the *Prevotella* genus which is also seen in other traditional diet groups. A traditional diet group from Burkina Faso, consisting of 14 children were compared with 15 children of EU origin (De Fillippo *et al.,* 2010). The microbial composition of each group is consistently divergent; the Burkina Faso microbiota is, like the Hadza, dominated by the *Prevotella* genus (De Fillippo *et al.*, 2010; Schnorr *et al.*, 2014).

Interestingly, the youngest children surveyed from both groups had enriched populations of *Actinobacteria* consistent with the hypothesis that these microbes are supported by breast feeding (De Fillippo *et al.,* 2010). The Yanomami tribe, a previously uncontacted tribe isolated for >11,000 years demonstrated the highest bacterial and functional diversity of any human group (Clemente *et al.*, 2015).

Several traditional diet groups from rural populations in South America and Africa were examined by Yatsunenko and shown to have microbial populations more similar to each other than to their western counterparts, despite distant geological locations. This suggests that dietary intake is a key factor in determining microbial composition (Yatsunenko *et al.*, 2012).

Dietary intake of fibre has reduced considerably in the western diet over recent decades. Many foods, such as dairy, refined sugars and oils, contain little or no complex carbohydrates and these make up an increased proportion of our caloric intake compared with typical diets from previous centuries or from populations following more traditional diets (Cordain *et al.*, 2005; Jew *et al.*, 2009). In the U.S.A. the average adult dietary fibre intake was reported at 15.8g per day between 1999 and 2008, with little change over the course of the study; far from the recommended daily intake of 25-38g per day. The Bukina Faso children consumed more fibre than their EU counterparts (1-2 year old: 10g/day compared with 5.6g/day; 2-6 year old: 14.2g/day compared with 8.4g/day), the fibre intake of other traditional diet groups was not monitored as closely. However, foods available to the Hadza and the groups examined by Yatsunenko and colleagues also suggested fibre enriched diets (De Fillippo *et al.*, 2010; Schnorr *et al.*, 2014; Yatsunenko *et al.*, 2012).

This relatively recent reduction of dietary fibre in the western diet may account for the divergence of the western microbiota compared to the traditional diet groups and subsequent reduction in diversity (Segata, 2015). Loss of diversity through failing to provide enough dietary MAC to ensure diverse microbial growth may leave individuals vulnerable to digestive diseases (King *et al.*, 2012; O'Keefe *et al.*, 2015; Sonnenburg & Sonnenburg, 2014).

1.2.2. The Concept of Prebiotics

The concept of prebiotics was first introduced in 1995 by Gibson and Roberfroid, who stated that a prebiotic was:

"A non-digestible food ingredient that beneficially affects the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon, and thus improves host health" (Gibson & Roberfroid, 1995).

However the authors updated this description in 2004 for clarity, describing the characteristics the food ingredient should exhibit to be classified as a prebiotic. This updated description allows better distinction between prebiotic compounds, added dietary fibre or other food ingredients. The new classification contained three major distinctions for a food ingredient to be classified as a prebiotic:

"(1) resists gastric acidity, hydrolysis by mammalian enzymes and gastrointestinal absorption;
(2) is fermented by the intestinal microbiota;
(3) stimulates selectively the growth and/or activity of intestinal bacteria associated with health and wellbeing." (Gibson *et al*, 2004)

It is imperative that a prebiotic survives gastric acidity to a good extent, as hydrolysis within the early G.I. tract may reduce the prebiotic to monosaccharide components accessible to the host in the small intestine, preventing the prebiotic from coming into contact with the target microbial populations. Having passed through the early digestive system, the prebiotic must be fermented by the microbiota. Many recalcitrant glycans, such as crystalline cellulose, cannot be classed as prebiotics if the bulk of these pass through the G.I. tract without being fermented by the microbial community (Chassard *et al*., 2010). Finally, a prebiotic should act upon a targeted species or community (Gibson *et al.*, 2004); this final point is crucial for distinguishing a prebiotic from MACs.

Prebiotics should stimulate a defined species or range of species to promote host health via an informed manner, rather than broad spectrum effect which may be attributed to mechanisms outside of, or in addition to, microbial fermentation. In this manner, all carbohydrate prebiotics are MACs, but not all MACs may be defined as prebiotics. Additionally, glycans described as MACs tend to be ingested passively through the diet, and may include glycans from non-dietary sources; in contrast prebiotics are administered deliberately with the purpose of enhancing host health. Fructo-oligosaccharides (FOS) and galacto-oligosaccharides (GOS) are established prebiotics, but the benefits of more novel non-digestible glycans such as xylo-oligosaccharides and isomaltooligosaccharides are being explored with promising results (Patel & Goyal, 2012; Singh *et al*., 2015).

As discussed in the previous section (Chapter 1.2.1) dietary MACs have been dramatically reduced within the western diet, potentially underpinning the divergent microbial profiles observed. The reintroduction of this dietary MAC as a prebiotic may help in a broad spectrum manner to nourish the microbiota and encourage greater species diversity.

The industrial production of HMOs is currently infeasible due to the cost of purification from breast milk and the technical complications of *in vitro* production, however with the addition of FOS and GOS to infant formula the disparity between the microbial profiles of formula fed and breast fed infants was reduced (Knol *et al.*, 2005). The addition of inulin negated the detrimental effects of antibiotic intervention on *in vitro* cultures, suggesting that prebiotics may be valuable as a prophylactic therapy during antibiotic treatments in humans (Johnson *et al.*, 2015). Subjects fed 16 g/day FOS experienced higher levels of GLP-1 and PYY, hormones involved in appetite control, and a reduced caloric intake suggesting that prebiotic treatment may assist in weight loss therapies. The authors speculate that this effect could be due to increased production of SCFAs within the gut (Verhoef *et al.*, 2011). Prebiotic compounds have also demonstrated therapeutic effects across a number of disease states including inflammation and diabetes (Roberfroid *et al.*, 2010; Rastall & Gibson, 2015).

The classical prebiotic targets are members of the *Bifidobacterium* genus, which are thought to be health promoting (Gibson & Roberfroid, 1995; Koenig *et al.*, 2011; Fanning *et al*., 2012a; Fanning *et al*., 2012b). *Bifidobacterium spp.* are able to ferment colonic glycans to produce butyrate and modulate host health through the exclusion of pathogens, priming of the immune system and reduction of inflammation (Picard *et al.,* 2005; Fukuda *et al.*, 2011).

1.3. Fructans

1.3.1. Structure and Function

Fructans are β-linked fructose polymers which, like many complex carbohydrates, are inaccessible to human enzymes and pass through the early digestive tract intact (Gibson *et al.*, 2004). Inulin and levan are β-2,1 and β-2,6 linked homopolymers respectively (**Figure 1.8**). Many levans, particularly bacterial levans, are highly branched (Blake *et al.*, 1982).

More complex non-linear fructans comprising both linkage types may consist of an inulin backbone with levan branches such as found in garlic bulbs and agave stems (Baumgartner *et al*,. 2000; Velazquez-Martinex *et al*., 2014; Mancilla-Margalli & Lopez, 2006) or may consist of a levan backbone with inulin branch-points, such as found in winter wheat (Kawakami & Yoshida, 2012). Other fructan structures may include bifurcated inulins, such as those created by the onion 6G-fructosyltransferase (Vijn *et al.*, 1997). Whilst non-linear fructans may be relatively common, there is a scarcity of data regarding fructan structures in common dietary sources, consequently the extent of branching in fructans accessible to the microbiota is difficult to ascertain.

Fructans of any structure with a DP less than around 10 units may be referred to as FOS. Several branched fructan structures are illustrated (**Figure 1.9 A**).

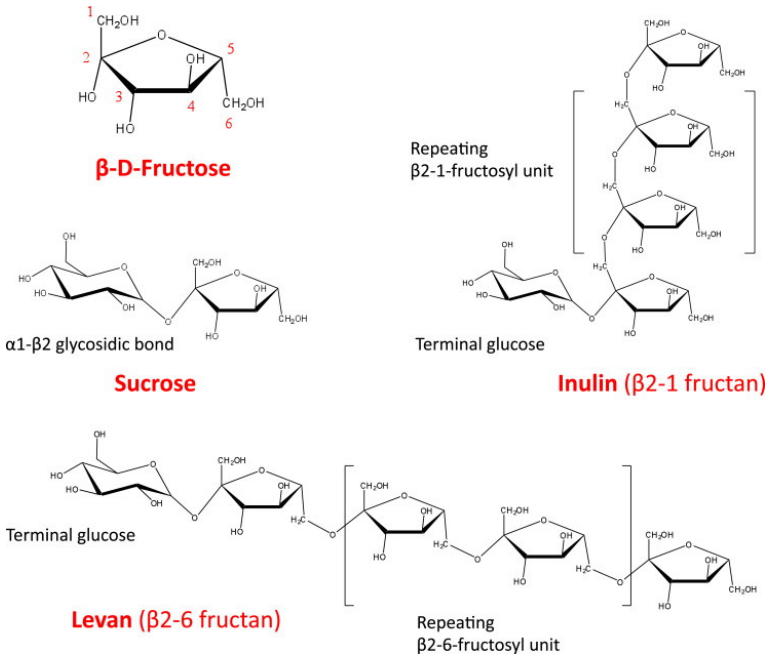


Figure 1.8. The molecular structure of fructose, sucrose, linear inulin and levan (Sonnenburg *et al.*, 2010). This figure shows the structure of polymeric linear fructans (β2-1 linked inulin, β2-6 linked levan), sucrose and fructose (Sonnenburg *et al.*, 2010).

Fructan is synthesised through the action of an enzyme, such as those belonging to CAZy family GH68 (further described in Chapter 1.4.2). Fructan synthesising enzymes characterised to date construct fructan polymer using sucrose as the substrate, as fructose is appended to a starting sucrose moiety, unmodified fructans are synthesised with one terminal sucrose unit, at what would otherwise be the reducing end of the sugar.

Inulin is a more flexible polysaccharide with the furanose rings protruding from the backbone, whereas the furanose rings are part of the backbone of levan. Levan is less flexible as hydrogen bonding holds the helical structure in place (**Figure 1.9. B**). These secondary structural features likely contribute towards the properties and functionality of these glycans within fructan synthesising organisms (Valluru & Van den Ende, 2008; Vereyken *et al.*, 2003).

Linear inulin, low molecular weight levan and non-linear fructans are plant-derived and generally serve as an energy store in plants (Vijn & Smeekens, 1999) and can be found in a variety of plants including cereals such as wheat and barley, grasses, and vegetables including garlic, onion and leek (Vijn & Smeekens, 1999; Muir *et al.,* 2007; Campbell *et al.*, 1997; Van Loo *et al.,* 1995). In addition to acting as reserve carbohydrate, fructans in plants may serve to regulate the sucrose levels within the vacuole, preventing sugar-induced feedback inhibition of photosynthesis (Vijn & Smeekens, 1999; Wagner & Wiemken, 1983) and/or protect against stressors such as drought (Valluru & Van den Ende, 2008). Dietary sources of fructans are illustrated (**Figure 1.10**). In contrast, high molecular weight levan, often highly branched, is found as a component of the extracellular polysaccharide matrix surrounding biofilms or as a bacterial capsule. Bacterial levan has a high molecular weight and a high degree of branching. *E. herbicola*, *Z. mobilis* and *B. subtilis* make levan with similar aqueous properties; these solutions are surprisingly elastic and likely assist in biofilm cohesion (Blake *et al.*, 1982; Benigar *et al.*, 2014).

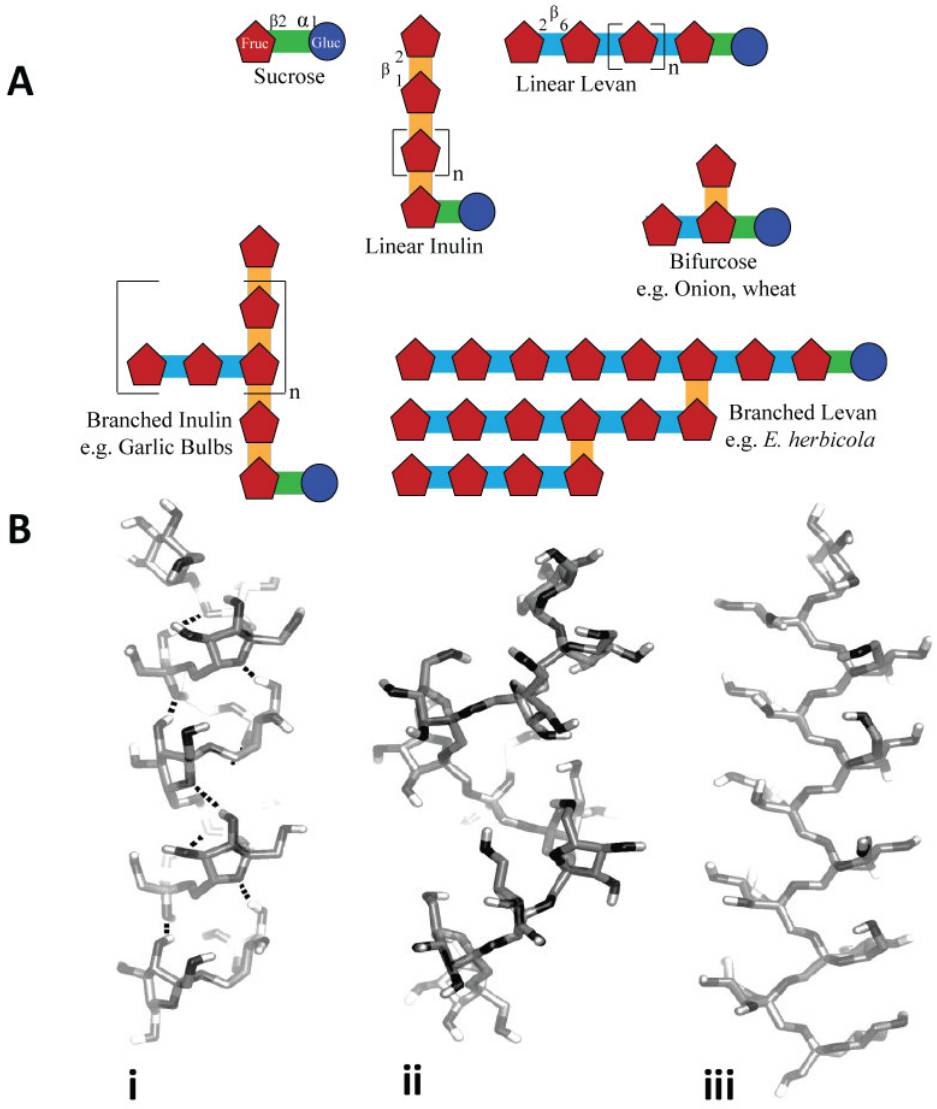


Figure 1.9. An overview of fructan secondary structure (Panel B from Vereyken *et al*., 2003). (**A**) Fructans are depicted in cartoon form with the inclusion of three non-linear examples (Baumgartner *et al.*, 2000; Blake *et al.,* 1982; Vijn & Smeekens, 1999, Cimini *et al.*, 2015). (**B**) Predicted conformations of decamers by Vereyken *et al.*, 2003; (i) Levan decamer adopts a rigid helix supported by hydrogen bonding. (ii) and (iii) two preferred helical conformations of the more flexible inulin decamer (Vereyken *et al*., 2003)

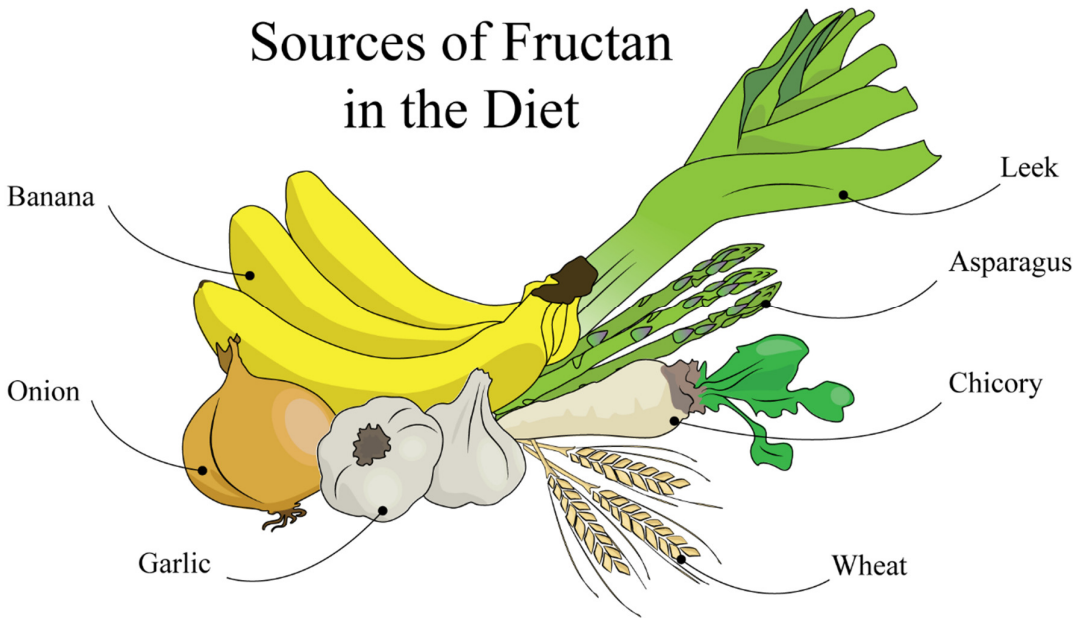


Figure 1.10. Sources of fructan in the diet. Humans consume fructans from a wide variety of fruits and vegetables, including but not limited to onion, garlic, leek, chicory, asparagus, wheat and banana (Muir *et al.,* 2007; Campbell *et al*., 1997; Van Loo *et al.,* 1995).

1.3.2. Dietary Prevalence

Americans, on average, consumed 2.6g of inulin and 2.5g of FOS per day between the years 1994 and 1996 (Moshfegh *et al.*, 1999). In Europe, the daily intake of inulin was approximated at 3-11g (Van Loo *et al.*, 1995). In contrast, levan intake is difficult to estimate, as there is little information regarding levan content of edible plants. Moreover, it is likely that significant microbiota accessible levan may arise from other members of the microbiota through expressed glycan capsules, or environmental bacteria, and not through dietary routes; making it incredibly difficult to predict the amount of levan available to the microbiota. One estimate places prehistoric fructan intake near 135g per day for an average male residing around the Chihuahuan Desert area; this estimate was made based upon well preserved coprolites (Leach & Sobolik, 2010). Regardless of the true quantity of fructan consumed by our ancestors, fructan intake has almost certainly reduced in recent years, as the total MAC content of our diets have decreased (Sonnenburg & Sonnenburg, 2014).

1.3.3. Fructans and Industry

Whilst levan has many potential applications, especially in the cosmetics industry, these are relatively under-explored when compared with inulin (Srikanth *et al.*, 2015). In contrast inulin has become increasingly industrially significant. Sinistrin, an inulin-type fructan with short β2-6 branches, is more soluble than linear inulin. It can be used to determine glomerular filtration rate as part of a diagnostics tool for kidney disorders as it is not broken down within the blood, and sinistrin clearance can be directly measured. Sinistrin is manufactured in large quantities from red squill for these medical assays (Spies *et al.*, 1992; Zitta *et al.*, 2013).

Inulin and FOS have been considered prebiotics since the inception of prebiotic treatments (Gibson & Roberfroid, 1995; Roberfroid *et al.*, 1998). Both inulin and FOS can support *Bifidobacterium spp. in vitro* (Watson *et al.*, 2013) and ingestion of inulin or FOS can promote these species *in vivo* (Gibson *et al,*. 1995; Slavin, 2013). Moreover, inulin and FOS can stimulate mineral absorption (Griffin *et al.*, 2002) and increase the production of beneficial short chain fatty acids (Van de Wiele *et al.*, 2004). Inulin has also been reported to have a beneficial effect on bone mineral content, blood lipid profiles, the immune system and energy homeostasis (Schaafsma & Slavin, 2015). Consumption of 5-8 g/day of inulin is thought to be sufficient to elicit a prebiotic effect (Kolida & Gibson, 2007).

Inulin is an ideal functional food; it is cheap and easy to extract, soluble, neutral tasting and has a mouthfeel and texture similar to fat (Franck, 2002).

Inulin is particularly useful in low-calorie dairy drinks, desserts and ice-cream as fat content may be reduced in these foods without compromising taste or texture in the final product when supplemented with inulin (Buriti & Saad, 2013; Akalin & Erisir, 2008). These properties and the versatile nature of inulin have no doubt contributed to its success within the food industry, in addition to the role it plays as a prebiotic.

1.4. Carbohydrate and Fructan Active Enzymes

1.4.1. Carbohydrate Active Enzymes (CAZymes)

Protein-carbohydrate interactions underpin an enormous number of biological processes; from central metabolism, to sensing and manipulating the environment. Carbohydrates are able to perform such a variety of functions in part, due to the structural complexity which can be achieved by these compounds. Complexity of monosaccharides arises through variable hydroxyl group positioning, chirality and stereochemical conformations (e.g. furanose and pyranose). Monosaccharides may also be modified in a number of ways, for example the incorporation of sulphate, N-acetyl or carboxyl groups. Glycosidic bonds between monosaccharides may be either α or β conformation depending on the configuration of the anomeric carbon in relation to the configuration of the chiral centre furthest away, typically C5 (Koshland *et al.*, 1954). From this wide variety of monosaccharide building-blocks and linkages, incredibly complex and diverse polysaccharide chains can arise. In order to facilitate such carbohydrate complexity, huge numbers of carbohydrate active enzymes are expressed across all domains of life which construct, deconstruct or modify these compounds.

Study of Carbohydrate Active enZymes, or CAZymes, is aided by organisation and classification of proteins into sequence related families (Lombard *et al.*, 2014; Cantarel *et al.*, 2009). The CAZy database (www.cazy.org) aims to catalogue CAZymes, and by doing so enable deeper understanding and prediction of unknown CAZymes. Enzymes are split into functional classes, Glycoside Hydrolases (GH), Polysaccharide Lyases (PL), Carbohydrate Esterases (CE) and Auxiliary Activities (AA).

GH enzymes facilitate the cleavage of glycoside bonds with the inclusion of water (hydrolysis). This large class of enzymes contains diverse members with a variety of folds and operate through acid-base catalysis with either retention or inversion of the stereochemical configuration of the glycosidic bond (**Figure 1.11**).
The retaining mechanism has two stages and a covalent glycosyl-enzyme intermediate is formed, stereochemistry at the anomeric carbon (C1) is retained.

The reaction requires a negatively charged nucleophile and a protonated acid/base, normally situated 5.5 Å apart; typically acidic residues glutamate and/or aspartate. In the first stage, the catalytic nucleophile attacks the C1 and at the same time the acid/base donates a proton to the leaving group. Catalysis proceeds through an oxocarbenium ion-like transition state and the covalent intermediate is formed. The second stage is carried out by the acid/base residue, which now acts as a general base, a proton abstracted from water forms a hydroxyl ion which attacks C1 to cleave the covalent intermediate. The inverting mechanism consists of a single step without an intermediate, the stereochemistry of C1 is inverted. Two residues, a protonated general acid and a general base are required, situated between 6 – 11 Å apart. A proton from water is abstracted by the general base, forming a hydroxyl ion which attacks C1. At the same time, a proton is donated from the general acid which assists departure of the leaving group. Catalysis proceeds through an oxocarbenium ion-like transition state. (McCarter & Withers, 1994; Davies *et al.*, 1995).

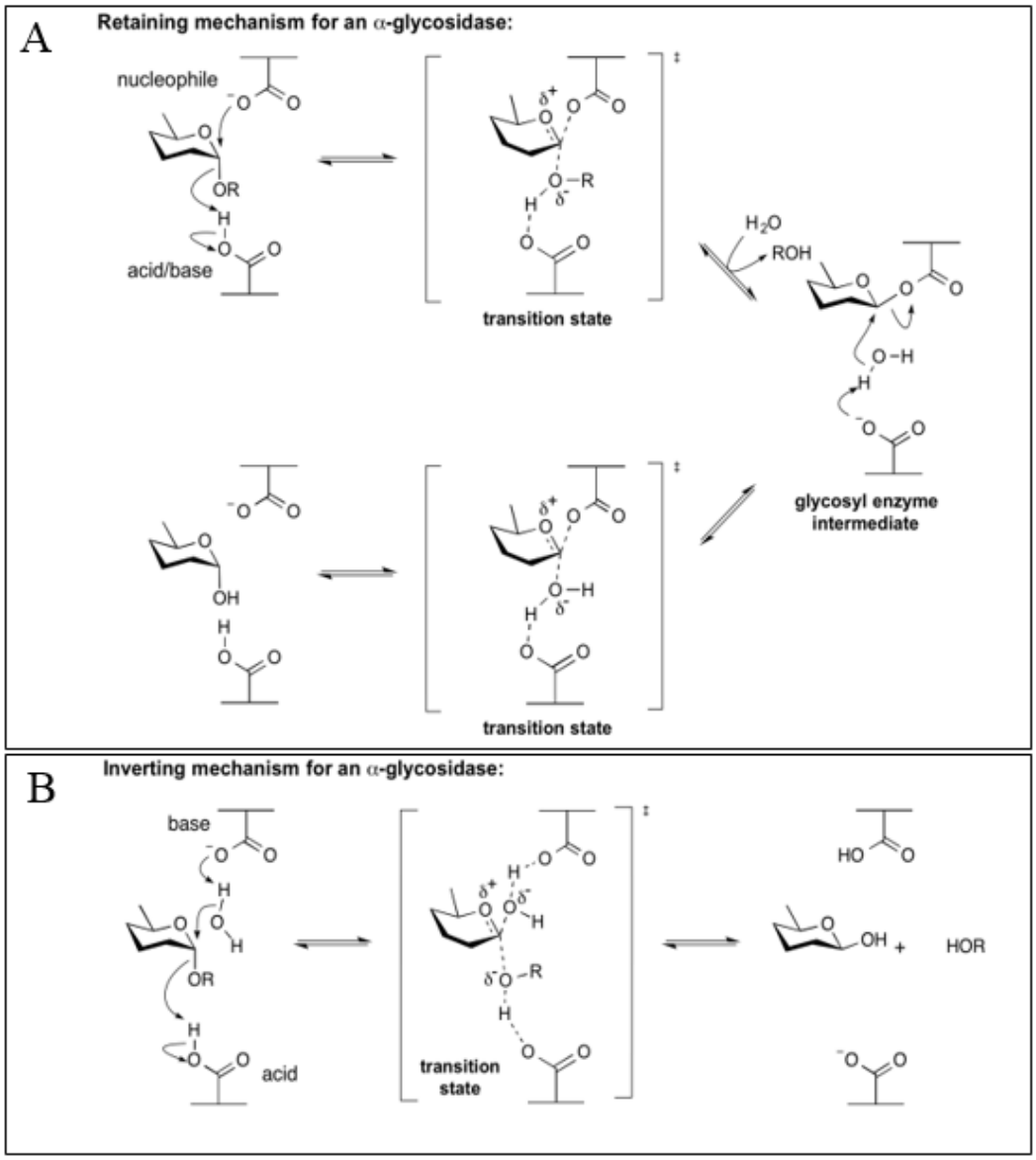


Figure 1.11. The two mechanisms employed by glycoside hydrolases. These diagrams adapted from Withers & Williams, "Glycoside Hydrolases" in CAZypedia, available at URL http://www.cazypedia.org/. (**A**) The retaining mechanism, whereby a glycosyl enzyme intermediate is formed between two successive steps. Two carboxylate residues, an acid/base and nucleophile are required for catalysis. The stereochemistry of the anomeric carbon is retained. (**B**) The inverting mechanism is undertaken in one single displacement enzymatic step. Two amino acid residues, a general acid and base, are required for catalysis. The stereochemistry of the anomeric carbon is inverted.

PL enzymes cleave polysaccharide chains containing uronic acid moieties through a β-elimination type mechanism with the formation of an unsaturated product.

In the CAZy database, enzymes within each class are split into sequenced-based families. Families are represented numerically, and new families are added incrementally. Within families, characterised examples are annotated and any observed activities and mechanisms are listed. The database provides links to structural studies which provide insight into the enzyme fold and mechanism. All of this information allows predictions to be made about unknown members of the family, and allow new members to be more easily uncovered in newly sequenced genomes. Of course, grouping families by sequence can become misleading for predictive purposes in some cases, especially where the family already exhibits a diverse range of activities. The predictive power of the CAZy database relies on robust characterisation of numerous family member examples, and in some enzyme families this predictive power is very weak, with relatively few members characterised, however the target linkage type (α or β) and chirality (L or D) are almost always consistent within families, furthermore several families target a single group of polysaccharides, such as the fructan-active GH32 family, and identification of a GH32 encoding gene is highly predictive of a fructan associated role.

1.4.2. Fructan Active Enzymes: Glycoside Hydrolase Families 32, 68 and 91

Three CAZy families are associated with fructans; GH32, GH68 and GH91. Family GH32 are the largest β-fructosidase family, with 3,365 members at the time of writing. This family is the main hydrolase family associated with fructans and enzymes within this group employ the retaining mechanism to cleave both β2-1 and β2-6 fructans in either an exo- or endo- fashion. So far, only fructanase activity is associated with GH32 family enzymes, therefore the presence of a GH32 encoding ORF strongly predicts a function associated with fructan degradation.

GH32 family enzymes have been extensively studied. The catalytic mechanism was first identified as retaining by Koshland and Stein in 1954 by using labelled oxygen within water (H_2O^{18}) and examining where this oxygen was incorporated during the reaction (Koshland and Stein, 1954). A catalytic aspartic acid was later identified in yeast invertase using conduritol B epoxide, a β-glucosidase inhibitor which irreversibly alters the catalytic residue serving as the nucleophile (Reddy & Maley, 1990). Two further residues were also elucidated as essential to catalysis, a glutamic acid which serves as the acid/base catalyst and a second aspartic acid which most likely stabilises a transition state complex (Reddy & Maley, 1990; Meng & Fütterer, 2003). The catalytic residues are present within three conserved motif regions: WMN**D**SWP, R**D**P and **E**C, where the residues in bold are the nucleophile, transition state stabiliser and catalytic acid/base respectively (Lammens *et al*, 2009). The conservation of surrounding residues suggests that these play a role in forming a functional active site. Both exo- and endo-acting GH32 enzymes retain the same mechanism of action, however the nucleophilic aspartate may be replaced with a glutamate in the endo-acting enzyme (Pouyez *et al*., 2012; Vandamme *et al*., 2013).

The family GH68 contains a few members which exhibit β-fructosidase activity, however the majority of characterised members are levan- or inulinsucrase enzymes. These enzymes build nascent fructan chains by breaking sucrose into the constituent monosaccharides and forming new glycoside bonds between fructose monomers in a manner consistent with glycoside transferase activity. The catalytic apparatus of GH68 enzymes is very similar to that of family GH32 and operates in reverse; the two families are likely to have diverged from a shared origin, and contain the same three catalytic residues within similarly structured conserved motifs (Lammens *et al*, 2009).

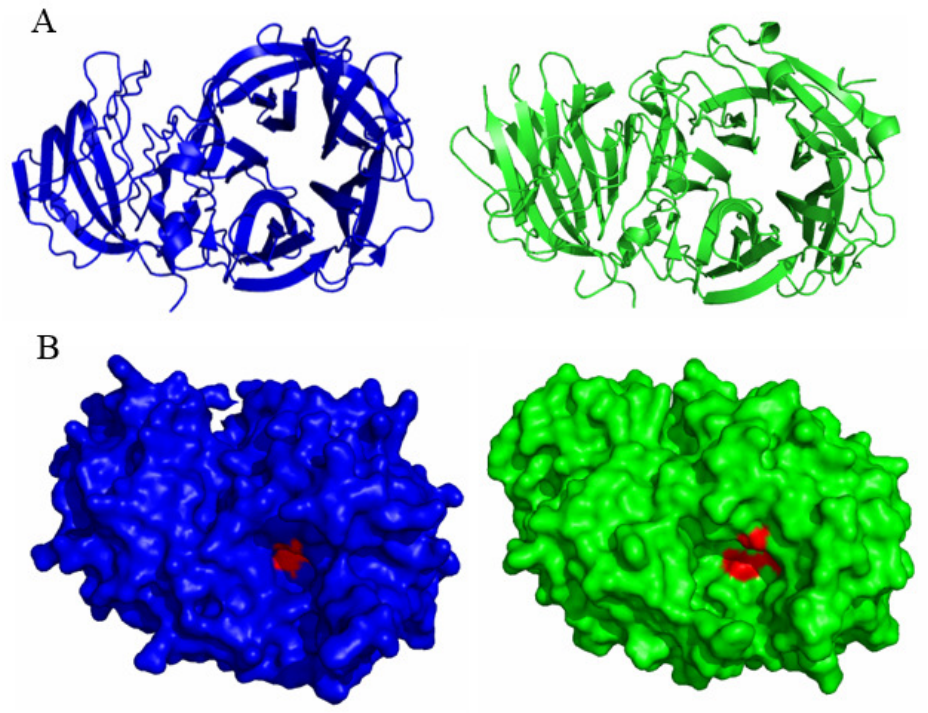


Figure 1.12. Comparison of endo- and exo- acting GH32 family enzyme structures. A comparison of the structures of endo-acting INU2 (PDB ID: 3SC7) shown in green and exo-acting invertase from *T. maritima* (PDB ID: 1UYP) shown in blue; two GH32 family enzymes which act on inulin. (**A**) A cartoon representation demonstrates the overall GH32 fold; a five bladed β propeller and C-terminal β-sandwich. The overall fold is consistent between exo- (left) and endo-acting (right) enzymes. (**B**) A surface representation with the catalytic residues in red, the exo-enzyme forms a pocket (left) whereas the endo-enzyme (right) has a larger cleft (Pouyez *et al.,* 2012; Alberto *et al.*, 2004).

The GH91 family is very small, containing only 55 members at the time of writing. Only a few reported activities have been observed which can be roughly divided into two groups; the degradation of inulin through the production of difructose anhydride (transferase-like) or the degradation of DFA to inulinbiose (hydrolase).

At present, only a few members of this family have been characterised, however most fall into the former category and are not technically hydrolases, rather employing a fructotransferase mechanism.

One member, BsIFTase (derived from *Bacillus subtilis* Inulin-fructotransferase), has been characterised and x-ray crystal structures in the presence and absence of fructobiose have been determined (Jung *et al.*, 2007). BsIFTase is a homotrimeric complex, with an active site formed at the interface between each β-helix monomer; three functional active sites are formed. Two catalytic residues were identified as part of this study, D233 and E244, these residues appear to be conserved in most sequenced GH91 members. The function of D233 is unclear, but it is suggested that this residue orientates the substrate within the catalytic site. The mechanism is proposed as a single displacement mechanism whereby a general base, E244, allows nucleophilic attack. E244 propagates this nucleophilic attack, which deprotonates a fructose at the 3' position. This negatively charged hydroxyl group attacks the C1 carbon, forming a second glycosidic bond, the attack results in inversion of the stereochemistry of C1. The reaction results in the creation of a difructose anhydride, DFA-III (**Figure 1.13**; Jung *et al.*, 2007).

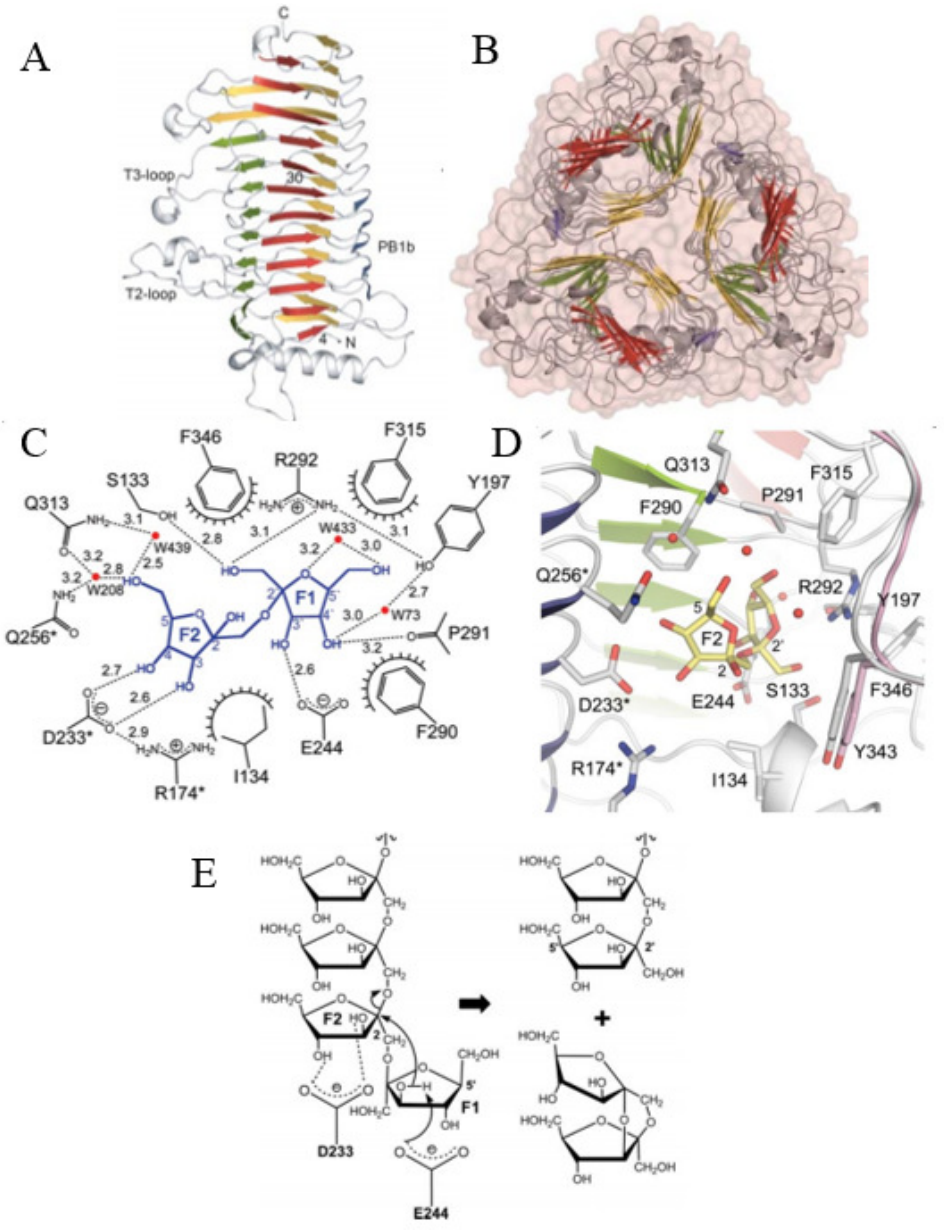


Figure 1.13. The catalytic mechanism of BsIFTase, a GH91 family enzyme (Adapted from Jung *et al*., 2007). The structure of BsIFTase was solved by Jung and colleagues and remains the only characterised GH91 family enzyme with the inclusion of a crystal structure. (**A**) Side view and (**B**) top-down view of BsIFTase crystal structure, showing homotrimeric conformation. (**C**) 2D and (**D**) 3D representation of the catalytic site including fructobiose ligand bound. (**E**) Proposed catalytic mechanism for BsIFTase, as described in the text (Jung *et al*., 2007).

1.5. Fructan Binding Proteins

Carbohydrate Binding Modules (CBMs) are discrete modules appended to enzymes, most commonly, glycoside hydrolases. CBMs are ordered into sequenced based families (www.cazy.org) and enhance enzyme activity, for example by bringing insoluble substrate into proximity to the active site thereby allowing catalysis, through targeting of the enzyme to specific regions of the substrate and/or, through the straightforward increase in the local concentration of substrate around the active site (Bolam *et al.*, 1998; Cécile *et al.*, 2010; Gilbert *et al.*, 2013).

CBM family 66 are generally appended to GH32 enzymes and recognise the non-reducing fructose termini of fructans. The founding member of this CBM family, BsCBM66 targets a non-specific β-fructosidase, SacC, to the chain termini of branched levans, enhancing the activity of this enzyme on levan by ~100 fold. The CBM is not required for active site formation, as a truncated version of the protein, lacking the BsCBM66 retained maximal affinity for non-targeted substrates including sucrose and FOS (Cuskin *et al.*, 2012).

A master's thesis published prior to this doctoral thesis examined a novel domain appended to an apparently inactive GH91 enzyme, BACOVA_04502. This domain, referred to as BACOVA_04502CTD (derived from C-Terminal Domain) bound to inulin with high affinity (K_d = 17 μM) during Isothermal Titration Calorimetry (ITC) analysis (**Figure 1.14**). A crystal structure of BACOVA_04502CTD in complex with kestopentaose (a sucrose containing pentasaccharide, $G_{\alpha 1\text{-}2\beta}F_{\beta 2\text{-}1}F_{\beta 2\text{-}1}F_{\beta 2\text{-}1}F$) was obtained and the binding site was elucidated using targeted alanine mutagenesis and isothermal titration calorimetry (**Figure 1.14**). Despite being appended to an enzyme, the BACOVA_04502CTD was not considered a CBM as BACOVA_04502 was thought to be an inactive duplicate of BACOVA_04503 (Shapiro, MRes thesis, 2012).

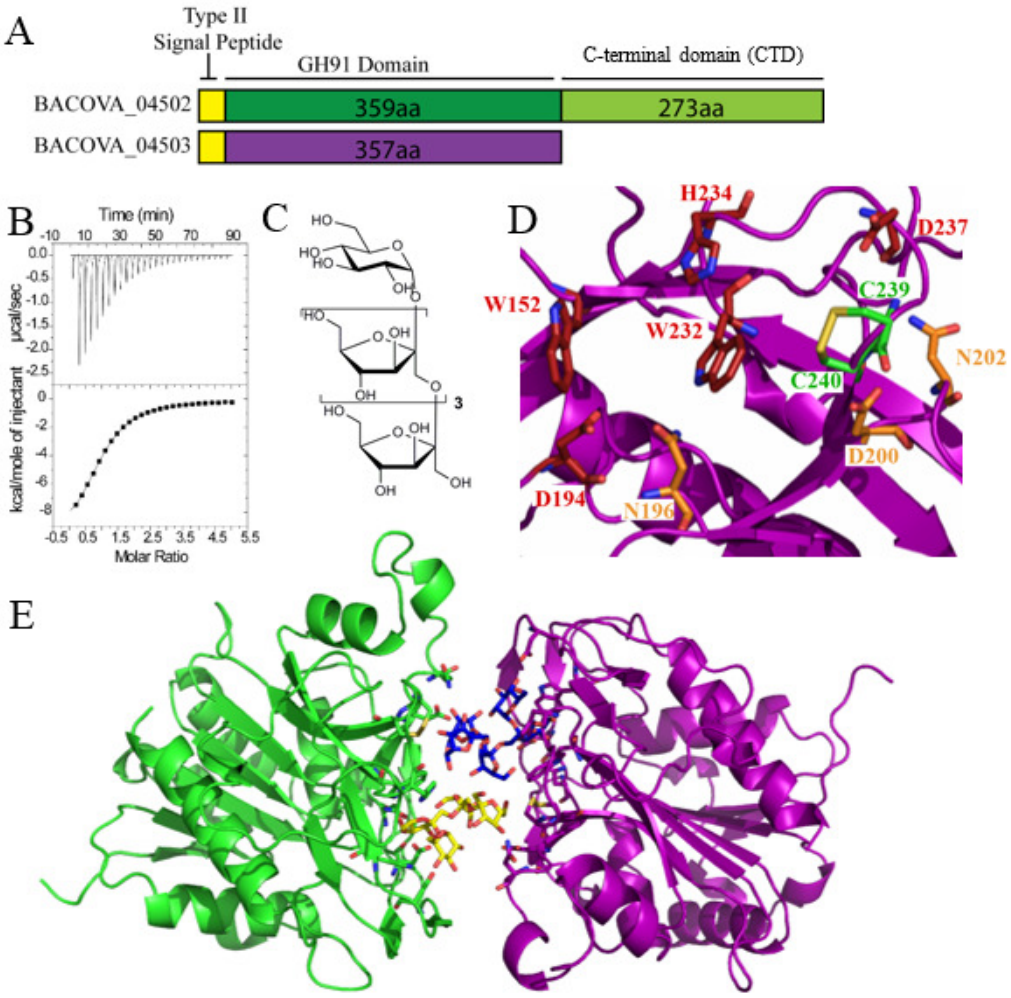


Figure 1.14. Recognition of inulin by a discrete binding domain (Shapiro, MRes, 2012). (**A**) Two consecutive genes encoding GH91 family enzymes were identified in *B. ovatus*. BACOVA_04502 contained a C-terminal domain (BACOVA_04502 CTD). (**B**) BACOVA_04502 CTD bound inulin (Kd = 17 µM) during ITC analysis, the top frame shows the raw heats and the bottom shows the integrated peak areas fitted to a onesite model using Microcal origin software. The BACOVA_04502 CTD was crystallised in complex with kestopentaose, the structure of which is shown in panel (**C**). (**D**) 3D representation of the binding cleft, residues are highlighted (red: alanine mutant cannot bind, orange: alanine mutant binds weakly, green: a di-sulphide bridge critical to binding). (**E**) Cartoon representation of BACOVA_04502 CTD structure. Two K5 molecules bound across two protein monomers, mediating a dimer only observed in the crystal structure. K5 does not bind well during ITC analysis and the binding of two ligand molecules in this manner seems to mimic the preferred substrate, which has a DP greater than 8 (Shapiro, MRes thesis, 2012).

It was speculated that the full length BACOVA_04502 protein performed an analogous function to the SusE of the Starch Utilisation System in *Bacteroides thetaiotaomicron*, this SusE protein binds to polymeric starch at the cell surface (Cho & Salyers*.,* 2001; Chapter 1.7.1). This hypothesis is challenged within Chapter 4, where evidence is provided for BACOVA_04502CTD being the founding member of a novel CBM family.

Fructan binding has also been documented in a SusD-homologue, BT1762, which is presented on the surface of *B. thetaiotaomicron* where it assists the import of levan through the SusC-homologue outer-membrane channel, a critical function during growth on levan (Sonnenburg *et al*., 2010).

Extracellular Substrate Binding Proteins (ESBP) have been shown to operate in tandem with ABC permeases to import small molecules and short oligosaccharides across the bacterial cell membrane (Bertsson *et al.*, 2010; Ejby *et al.*, 2013). Putative extracellular solute binding proteins have been identified within several inulin utilisation gene clusters, providing a tantalising glimpse of novel inulin recognition mechanisms, but little characterisation of these proteins has been undertaken (Scott *et al*., 2011; Garrido *et al*., 2011).

There is a paucity of literature on fructan binding proteins, therefore the mechanisms by which fructans are recognised by proteins remains largely unclear.

1.6. Carbohydrate Capture and Utilisation by the Gut Microbiota

1.6.1. The Distribution of CAZymes Within the Human Microbiota

Carbohydrate metabolism is an essential function to all life. Given the unique environment the intestine provides – a continuously moving, dynamic environment furnished with complex carbohydrates indigestible to human enzymes which represent the main source of carbon – it is not surprising that the resident microbiota have become extremely adept at carbohydrate harvest. An impressive array of CAZymes are encoded by the microbiota, many of which are highly specialised to the available glycans within the human gut (Kaoutari *et al*., 2013).

CAZy enzymes have been identified from most major gut residents and analysis of these enzymes demonstrates, to some extent, the ecological niches occupied by each group (Kaoutari *et al.*, 2013). Kaoutari and colleagues used bioinformatics techniques and the data available via the CAZy database to analyse the wealth of CAZymes encoded by sequenced members of the microbiota (**Figure 1.15**). The team used 177 genomes in proportions equivalent to within healthy humans (104 Firmicutes, 29 Bacteroidetes, 22 Proteobacteria, 12 Actinobacteria, 2 Fusobacteria and one each from the phyla Verrucomicrobia, Lentisphaerae, Spirochaetes, Thermotogae, Synergistetes, Tenericutes, Elusimicrobia and Cyanobacteria) to construct a "mini-microbiome" with which they conducted their analysis. The authors show that this mini-microbiome contains 9,120 (57 %) glycoside hydrolases, 5,520 (35 %) glycosyltransferases, 950 (6 %) carbohydrate esterases and 292 (2 %) polysaccharide lyases.

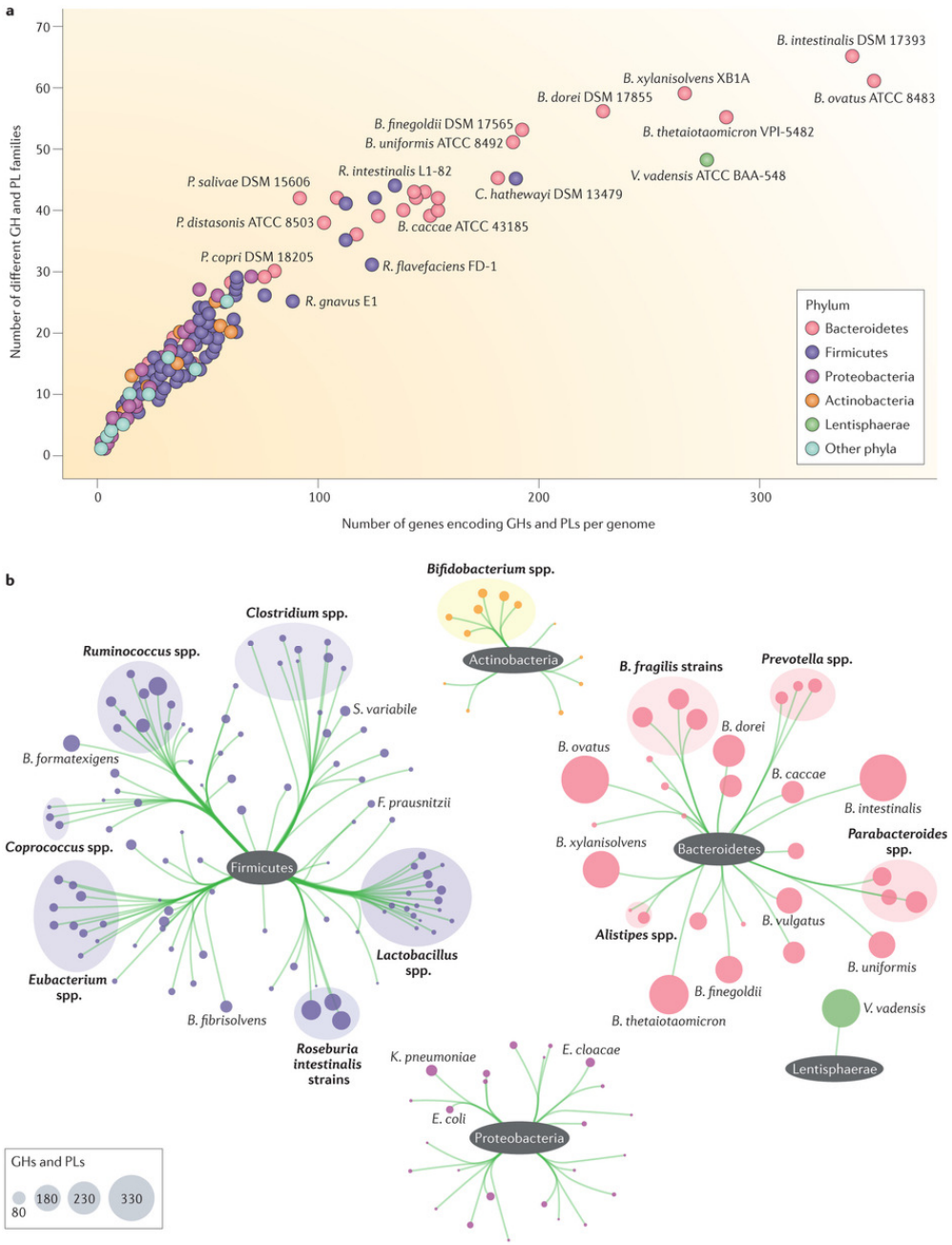


Figure 1.15. Diversity of Glycoside Hydrolyses and Polysaccharide Lyases within a Constructed "mini-microbiome" (Kaoutari *et al.*, 2013). In this figure, the authors demonstrate the diversity of GH and PL enzymes within the constructed "mini-microbiome" used for analysis. (**A**) The number of enzymes mapped against the number of represented CAZy families (diversity). (**B**) The distribution of CAZymes within related groups. This graph demonstrates that members of the Bacteroidetes phylum contain the highest numbers of CAZymes including high numbers of diverse families. Members with high numbers of CAZymes may be able to occupy multiple gut niches, whereas members with a narrow set of CAZymes may occupy a more specific niche (Kaoutari *et al.*, 2013).

Most commonly, these enzymes target plant and animal glycans. In comparison, the human genome is known to encode 91 glycoside hydrolases, only 8 of which have a known digestive function. Furthermore, the authors demonstrate the diversity and number of encoded enzymes within each genome, allowing comparisons to be made between species and phyla – *Bacteroides* contained by far the most enzymes per genome, including a large number of diverse families. The data suggest that *Bacteroides* genus can target a wider range of glycans, whereas other genomes are more specific and likely occupy tight niches within the intestinal environment (Kaoutari *et al.*, 2013).

1.6.2. Carbohydrate Harvest by Members of the Bacteroidetes phylum

Currently, glycan utilisation within the intestine has been most widely studied in the *Bacteroides* genus. The Bacteroidetes phyla is dominated by the *Bacteroides* class (Arumugam *et al*., 2011; Huttenhower *et al.*, 2012), a group noted for the astonishing ability to recognise, import and degrade a wide variety of glycan structures; for example *Bacteroides thetaiotaomicron* and *Bacteroides ovatus* have between them the capacity to use almost all of the classes of glycan known (Martens *et al*., 2011). *Prevotella spp.* are also prominent members of Bacteroidetes, particularly within the gut of individuals following a higher-fibre hunter-gatherer or agrarian diet (Yatsunenko *et al.,* 2012; De Fillippo *et al.,* 2010; Rampelli *et al.*, 2015). *Prevotella spp.* have also been shown to utilise non-cellulose plant cell wall glycans (Flint *et al.*, 2012), though carbohydrate utilisation in this genus is relatively under explored.

Polysaccharide Utilisation Loci (PULs) were first identified in *Bacteroides thetaiotaomicron* in the 20th century (Anderson & Salyers, 1989), however genome sequencing has revealed that all sequenced Bacteroidetes contain these closely linked, tightly regulated, PULs, each specific for a distinct glycan (Martens *et al.*, 2011; Rosewarne *et al*., 2014). *Bacteroides* species have PUL-rich genomes, *B. thetaiotaomicron* contains 88 PULs, which occupy roughly 18% of the genome (Martens *et al.*, 2009).

Recently, advances in prediction software have been made allowing automated PUL identification from genomic data (Terrapon *et al.*, 2015).

Bacteroides spp. have become models for the study of glycan utilisation within the Bacteroidetes phyla as they are prominent members of the gut microbiota, are culturable and encode a multitude of PULs (Arumugam *et al*., 2011; Huttenhower *et al.*, 2012; Martens *et al.*, 2011). The first PUL, encoding for a starch utilisation system (SUS) was described by Anderson and Salyers in 1989. Subsequently, detailed biochemical studies have provided insight into how the starch utilisation system works on a molecular level. The starch utilisation system is a suite of periplasmic and outer membrane localised proteins, SusABCDEFGR, which bind to, import and degrade starch (Martens *et al.*, 2009). A typical Sus-like system is often identified by the presence of a SusC homologue, a SusD homologue, a sensor/regulator and a number of CAZymes.

Outer membrane SusC homologues are TonB coupled oligosaccharide importers (Reeves *et al.,* 1996) and SusD homologues are predicted surface located oligosaccharide binding lipoproteins (Shipman *et al.,* 2000). Initial research on *susD* knock-out strains suggested that the binding role of SusD is essential to the function of the starch utilisation system (Koropatkin *et al.*, 2008).

As the SusD protein has previously been found to interact with the SusC import protein (Shipman *et al.*, 2000) it is likely that SusD orients or channels the polysaccharide for import, however the binding role of SusD appears to be superfluous when the system is activated using maltose, though the presence of the protein is still required; therefore the role of this protein during import and regulation remains unclear (Cameron *et al.*, 2014). In addition to the SusC/SusD complex other outer membrane carbohydrate binding apparatus is present (Cameron *et al.*, 2014). Although a minority of SusC/D homologue pairs appear to stand alone within *Bacteroides* genomes, all currently characterised PUL contain CAZymes which act to deconstruct polysaccharide chains. PUL encoded enzymes are highly diverse and play a multitude of roles within glycan degradation systems.

Many PUL, including the recently elucidated xyloglucan and xylan utilisation systems (Rogowski *et al.*, 2015; Larsbrink *et al.*, 2014) and the canonical starch utilisation system (Koropatkin & Smith, 2010; Shipman *et al.*, 1999) contain an endo-acting enzyme localised to the cell surface. This vanguard enzyme breaks up large polymers into shorter oligosaccharides for import mediated by the SusC/D pair through the outer membrane and is critical for growth on the polysaccharide (Larsbrink *et al.*, 2014; Reeves *et al.*, 1997). It is worth noting that endo-acting enzymes have also been found to localise to the periplasm, such as in the case of mannan degradation in *Bacteroides thetaiotaomicron* and xylan degradation in *Bacteroides ovatus* (Rogowski *et al.*, 2015; Cuskin *et al.*, 2015). The localisation of these enzymes may depend on the utilisation strategy employed by the organism, external endo-acting enzymes will release short oligosaccharides to the extracellular environment, which could lead to loss of substrate to competitors. Internalised endo-acting enzymes may be part of a "selfish" mechanism, where glycans are kept as intact as possible prior to import, preventing this carbohydrate loss (Cuskin *et al.*, 2015).

1.6.3. Carbohydrate Harvest by Members of the Firmicutes and Other Phyla

The SUS paradigm is far from the only mechanism by which intestinal microbiota degrade glycans (Flint *et al.*, 2012). Most members of the gut microbiota contain significant numbers of glycoside hydrolases (Kaoutari *et al.*, 2013), in this section a few selected examples of glycan harvest by gut microbiota through non-SUS like mechanisms are illustrated, to demonstrate the diversity and complexity of these systems.

ABC transporters are near-ubiquitous and can mediated the import of extracellular oligosaccharides (Yan, 2013). ABC transporters are coupled with substrate binding domains in some systems which, upon binding, propagate a conformational change through the transmembrane domain allowing substrate to pass through the channel.

Substrate binding domains may be part of a complex with the transporter, periplasmic or lipid-anchored to the membrane (Berntsson *et al.*, 2010).

Eubacterium rectale is a member of the Firmicutes phylum and utilises soluble starch through the expression of multiple ABC permeases linked with extracellular located substrate binding proteins (ESBP) which co-ordinate ABC permease import through solute recognition (**Figure 1.16;** Cockburn *et al.*, 2014).

The *Bifidobacterium* genus are Gram positive members of Actinobacteria phylum able to utilise a range of short oligosaccharides. As discussed previously, members of the *Bifidobacterium* genus are commonly abundant during infancy and are the classical target for prebiotic therapy (Chapter 1.1.4, Chapter 1.2.2). *Bifidobacterium longum* dedicates ~8% of its genome toward glycan catabolism, typically encoding ABC transporter systems regulated by LacI-class repressors (Schell *et al.*, 2002). Members of this genera may utilise human milk oligosaccharides, HMOs, present within mother's milk. HMOs are complex oligosaccharides with a range of secondary structures, *Bifidobacterium longum subsp. infantis* encodes a number of ABC permeases and HMO binding proteins, enabling it to grow efficiently within the infant gut (Garrido *et al.*, 2011)

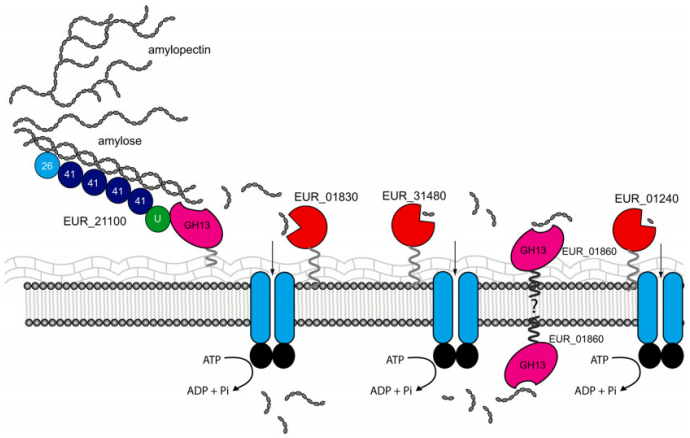


Figure 1.16. *Eubacterium rectale* employs three ABC transporter and ESBP pairs during starch harvest (Cockburn *et al.*, 2014). Tethered to the peptidoglycan layer is the multidomain EUR_21100 structure, which binds to starch and releases malto-oligosaccharides. Released oligosaccharides may be targeted by the ESBPs (red) EUR_01830, EUR_31480 and EUR_01240 for import into the cell. The location of the second GH13 enzyme, EUR_01860 is unknown (Cockburn *et al.*, 2014).

1.7. Fructan Degradation by the Microbiota

1.7.1. Fructan Utilisation Loci

As discussed previously, saccharolytic members of the Bacteroidetes phylum express PULs to degrade target polysaccharides (Martens *et al.*, 2011, Chapter 1.6.2). To date, however, only one fructan PUL has been characterised, the levan utilisation system of *Bacteroides thetaiotaomicron* (**Figure 1.17**).

Levan is broken up by an extracellular endo-acting GH32 enzyme, BT1760. Levan and levanoligosaccharides are bound at the cell surface by a SusE-like protein, BT1761 and a SusD-homologue, BT1762. Once imported through the SusC-homologue, BT1763, levan is broken down to fructose by two exo-acting GH32 enzymes; BT1759 and BT3082.

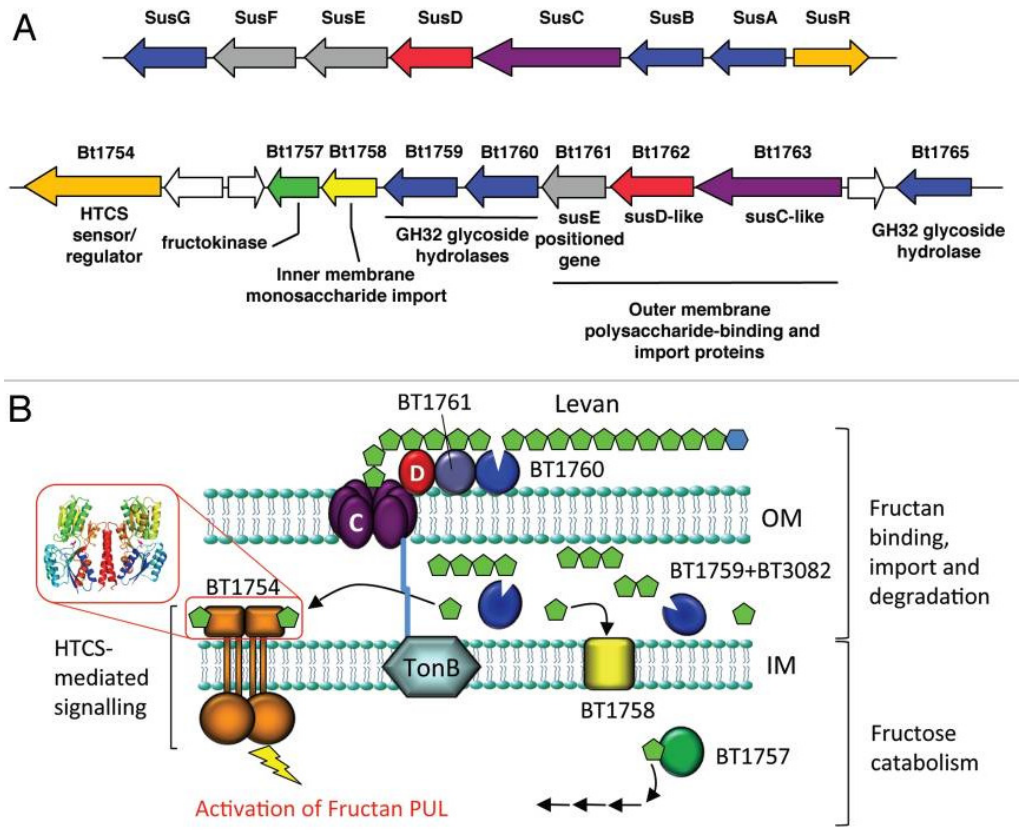


Figure 1.17. The levan utilisation system from *Bacteroides thetaiotaomicron*. The levan utilisation system from *B. thetaiotaomicron* was characterised (Bolam and Sonnenburg, 2011). (**A**) The canonical starch PUL compared with the levan PUL in *B. thetaiotaomicron*. (**B**) The predicted model for levan utilisation in *B. thetaiotaomicron* (Bolam and Sonnenburg, 2011).

The sensor domain of the Hybrid Two Component System (HTCS) BT1754 recognises periplasmic fructose as a proxy for levan polysaccharide and is able to up-regulate the PUL through a DNA binding domain (Sonnenburg *et al.*, 2010; Bolam & Sonnenburg, 2011).

The authors also identified putative fructan PUL from *B. ovatus*, *B. caccae*, *B. fragilis*, *B. vulgatus* and *B. uniformis* during the study, by searching for homologues of the HTCS, BT1754. When cultured on minimal media supplemented with fructose, all species grew. When cultured upon levan, only *B. thetaiotaomicron* was able to grow. When cultured on inulin *B. ovatus*, *B. caccae, B. fragilis*, and *B. uniformis* grew. *B. vulgatus* growth was not supported by either polysaccharide (**Figure 1.18**).

B. ovatus and *B. caccae* likely degrade inulin in the same manner, and contain two GH91 encoding genes in addition to a further two GH32 encoding genes. *B. uniformis* contains four GH32 enzymes which are most divergent from those characterised in *B. thetaiotaomicron* (Bolam & Sonnenburg, 2011), this divergence may indicate a divergent strategy for inulin degradation. *B. fragilis* is also unusual in that it displays two distinct loci, each containing a HTCS gene with high sequence similarity to BT1754, the reason for this gene organisation is unclear. *B. vulgatus* was able to utilise fructose, but not polymeric substrates of either linkage, this result was consistent with the lack of a SusC/SusD homologue pair and no other candidate genes for glycan import across the outer membrane. The presence of one GH32, BV1663, suggests that small fructans may be broken down in some manner, perhaps outside of the cell as there is no obvious fructan import apparatus such as a SusC/D-homologue pair (Sonnenburg *et al*., 2010; Bolam & Sonnenburg, 2011).

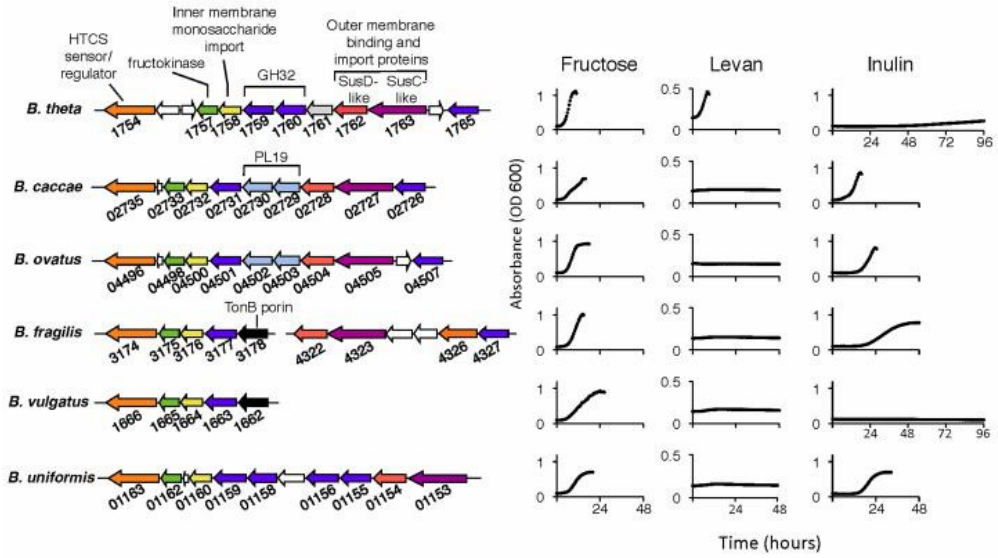


Figure 1.18. Six putative fructan PUL were identified in Bacteroides species (Sonnenburg *et al*., 2010). Sonnenburg and colleagues predicted and compared six putative fructan PUL from six Bacteroides species (left). All species were cultured on minimal media supplemented with fructose, levan or inulin to identify whether the strains could utilise these substrates (right). *B. vulgatus* could utilise only fructose, whereas the remaining five utilised either levan or inulin in addition to fructose. *B. thetaiotaomicron* was the only levan user, with *B. ovatus, B. caccae, B. fragilis* and *B. uniformis* able to utilise inulin (Sonnenburg *et al*., 2010).

1.7.2. Other Fructan Utilisation Systems

Fructan utilisation has been observed in several non-*Bacteroides* members of the microbiota. A wide scale study by Watson and colleagues shows a number of *Bifidobacterium* and *Lactobacilli* species, desirable as probiotic species, are able to grow up FOS and inulin (Watson *et al.*, 2013). Whilst there is little mechanistic insight into how fructans are imported and degraded within probiotic species, several GH32 family CAZymes have been identified within the available genomes and are catalogued within the CAZy database. A GH32 enzyme isolated from *Bifidobacterium longum* KN29.1 was characterised with the inclusion of a crystal structure by Bujacz and colleagues in 2011.

This enzyme is capable of mediating the depolymerisation of short and mid length β2-1 linked FOS and to some extent is active upon polymeric inulin and sucrose. The authors did not include data regarding β2-6 linked FOS and levan (Bujacz *et al.*, 2011).

Roseburia inulinovorans is a motile member of the Firmicutes phylum able to harvest inulin. A cluster of genes, including a GH32 family enzyme and ABC transporter is upregulated in the presence of inulin. Co-cultures of *R. inulinovorans* and *B. longum* on inulin show very little growth of *B. longum*, which is only able to utilise short chain FOS or fructose. These data suggest that inulin is imported prior to degradation, such that no breakdown products are accessible to *B. longum* (Scott *et al.*, 2011). This is in contrast to other systems, such as inulin utilising strains of *Bacteroides* which break up inulin prior to import, enabling the growth of non-inulin utilising species of *Bifidobacterium* in co-culture (Falony *et al.*, 2009; Rakoff-Nahoum *et al.*, 2014).

1.8. Research Objectives

- To provide mechanistic insight into inulin recognition and utilisation by prominent glycan degrading members of the human gut microbiota.
    - In a Gram-negative species, *Bacteroides ovatus.*
    - In a Gram-positive species, *Bifidobacterium adolescentis.*

Chapter 2. Materials and Methods

2.1. Chemicals, Commercial Kits and Water

A comprehensive list of chemicals and commercial kits used in this project can be found in Appendix II.

Water used generally throughout this project was double distilled and purified to 18.2 Ω with a Millipore Milli-RO 10 plus Water Purification System. Water was used as the solvent for most solutions in this project and this can be assumed unless stated otherwise.

Where stated PCR-Grade 0.1 µm filter sterilised DNase, RNase and Protease free water was used (Sigma-Aldrich).

2.2. Sterilisation

Media and glassware were sterilised by autoclaving at 121 °C, 15 1b in^{-2}(psi) for 20 minutes. Solutions under 1 l were filter sterilised using an 0.25 µm pore Millipore filter disc (Supor® Acrodisc® 3.2, Gelman Sciences) and a pre-sterilised syringe (Plastipak®, Becton Dickinson). Solutions of a larger volume were filter sterilised by passing through a 0.2 µm filter (Supor® 200-S PES membrane, Pall Sciences) under a vacuum.

2.3. Storage Practices

2.3.1. DNA

DNA was generally stored in PCR grade water or 10 mM Tris-Cl pH 8.5 at -20 °C. DNA was stored at 4 °C for a maximum of 6 weeks if used regularly to minimise freezing and thawing.

2.3.2. Protein

Protein was stored at 4 °C for a maximum of 2 weeks, with regular positive control tests to ensure no loss of function had occurred. Longer term storage was not possible with all protein constructs due to degradation however if possible proteins were snap-frozen in liquid nitrogen and stored immediately at -80 °C for a maximum of 1 year. Experiments involving purified proteins were performed from fresh purifications in at least one instance so that any degradation of the protein during storage could be monitored.

2.3.3. Bacterial Cultures

Bacterial cell cultures were stored on plates at 4 °C for a maximum of 5 weeks. Liquid cultures containing 25 % (v/v) glycerol were stored at -80 °C in cryovials.

2.3.4. Carbohydrates

Carbohydrate solutions were frozen at -20 °C for long term storage or kept at 4 °C over a short period (< 6 weeks). If a large yield of carbohydrate was obtained, this was freeze-dried and kept at room temperature.

2.4. Vectors

Several vectors were needed for the work carried out during this project, these are listed below (**Table 2.1**).

Table 2.1. An index of vectors used during this Study

Plasmid	Size (Kbp)	Supplier/Reference	Features
pET28a-b	5.4	Novagen	Kan^r, T7, *lac*, *laciq* Cloning region includes His tag.
pET32b	5.9	Novagen	Amp^r, T7, *lac*, *laciq* Cloning region includes Trx tag, His tag.
pET22a	5.5	Novagen	Amp^r, T7, *lac*, *laciq* Cloning region includes *PelB,* His tag.
pET21a-b	5.4	Novagen	Amp^r, T7, *lac*, *laciq* Cloning region includes His tag.
pGEX_6P_1	5.0	GE Healthcare	Amp^r, tac, *lac*, *laciq*. Cloning region includes cleaveable GST tag.
Zero Blunt®	3.5	Invitrogen	Kan^r *, lac*, *laciq*
pExchange-*tdk*	4.2	Provided by Nicole Koropatkin (Koropatkin *et al*, 2008)	Amp^r

2.5. Bacterial Strains

A variety of bacterial strains were used during the project, optimised *E. coli* strains were used for the purposes of cloning and protein expression. Wild-type and genetically modified (*Bacteroides ovatus* only) strains of *Bifidobacterium* and *Bacteroides* species were used for direct observation/characterisation during experimentation. These strains, together with a description of their use are displayed below (**Table 2.2.**)

Table 2.2. An index of bacterial strains used during this project.

Name	Genotype	Description
BL21 (DE3)	F– *ompT gal dcm lon hsdSB*(rB-mB-) λ(DE3 [*lacI lacUV5*-T7 gene 1 *ind1 sam7 nin5*])	*E. coli* strain optimised for protein expression using a T7 promoter. Routinely used to over-express recombinant proteins. (Studier & Moffatt, 1986)
BL21 (DE3) pLysS	*F- ompT gal dcm lon hsdSB*(rB-mB-) λ(DE3) pLysS(*cmR*)	As BL21, with additional pLysS plasmid (T7 lysozyme) to lower background expression of the target gene prior to induction if leaky expression of the recombinant protein interfered with the growth of *E.coli* cells. (Moffatt & Studier, 1987)
Tuner (DE3)	F- *ompT hsdS$_B$*(r_B^- m_B^-) *gal dcm lacY1*(DE3)	As BL21 with the addition of a lac permease mutation to allow uniform diffusion of IPTG across cells, establishing a linear relationship between IPTG concentration and expression levels. Tuner was used when control over expression levels was desired e.g. to improve protein solubility. (Novagen)
Top10	F-*mcrA*Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ ΔM15 ΔlacX74nupGrecA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1* λ-	Routinely used for plasmid propagation and cloning. (Invitrogen)
CC118 λ-pir	*Δ(ara-leu) araD ΔlacX74 galE galK phoA20 thi-1 rpsE rpoB argE (Am) recA1* λ pir	Used for plasmid propagation and cloning (pExchange *tdk* only) (Herrero *et al*., 1990)
S17.1 λ-pir	*hsdR recA pro* RP4-2 (Tc::Mu;KmTn7)	Conjugation of pExchange *tkd* plasmids from this strain to *Bacteroides ovatus* (Skorupski & Taylor, 1996)
B. ovatus	Wild-Type ATCC 8483	DSM-1896, Type strain (HMP Consortium, 2010).
B. ovatus Δtdk	*Δtdk*	Used to generate genomic mutants or gene deletions in *Bacteroides ovatus* through FuDR selection (Chapter 2.16). Provided by Eric Martens.
B. ovatus Δ04502	*Δtdk Δ04502*	*bacova_04502* deletion strain, created as part of this project
B. ovatus 04503 M	*Δtdk 04503M*	*bacova_04503* E196Q genomic mutant created as part of this project

Table 2.2. An index of bacterial strains used during this project (continued).

Name	Genotype	Description
B. ovatus Δ04504	*Δtdk Δ04504*	*bacova_04504* deletion strain created as part of this project
B. ovatus Δ04505	*Δtdk Δ04505*	*bacova_04505* deletion strain created as part of this project
B. ovatus 04507M	*Δtdk 04507M*	*bacova_04501* D265A genomic mutant created as part of this project
B. caccae	Wild-Type ATCC 43185	DSM-19024, Type strain (HMP Consortium, 2010).
B. theta -iotaomicron	Wild-Type ATCC 29148	DSM-2079, Type strain (Xu *et al*., 2003).
B. xylansolvens	Wild-Type CCUG 53782	DSM-18836, Type strain (HMP Consortium, 2010).
B. fragilis	Wild-Type ATCC 25285	DSM-2151, Type strain (Cerdeño-Tárraga *et al.*, 2005).
B. uniformis	Wild-Type ATCC 8492	DSM-6597, Type strain (HMP Consortium, 2010).
B. intestinalis	Wild-Type JCM 13265	DSM-17393, Type strain (HMP Consortium, 2010).
B. theta -iotaomicron 8764	Wild-Type	Provided by Eric Martens
B. adolescentis	Wild-Type	DSM-20083, Type strain (Suzuki *et al.,* 2006)
B. longum	Wild-Type ATCC 15707	DSM-20219, Type strain (HMP Consortium, 2010).

2.6. Growth Media

Antibiotic selection was used within media to ensure only cells transformed with a desired plasmid could grow as outlined (**Table 2.3.**).

Media were used as plates with the addition of 2 % (w/v) of agar or alone as broth.

Rich media was used when a high cell density was required. Minimal media were used in conjunction with 0.5 % final w/v of the desired carbon source to monitor bacterial utilisation of target carbohydrates (minimal medium lacking this carbohydrate component was unable to support any growth). Selenomethionine medium was used for the production of proteins with the inclusion of selenomethionine for use during x-ray crystallography analysis in order to solve the phase problem. All media, together with a description of use is listed (**Table 2.4.**)

Table 2.3. Antibiotics

Antibiotic	Working Concentration	For use with:
Ampicillin	100 µg/ml	pET-32b, pET-22b, pET-21a, PCR Blunt®, pGEX_6P_1, pExchange *tdk*$^-$
Kanamycin	50 µg/ml	pET-28b
Chloramphenicol	30 µg/ml	pLysS
Gentamycin	200 µg/ml	pExchange *tdk*$^-$
Erythromycin	25 µg/ml	pExchange *tdk*$^-$

Table 2.4. Media

Medium	Composition	Quantity (per litre)	Description/Method
Luria-Bertani (LB)	LB powder, as supplied (Sigma-Aldrich)	25 g	Rich culture medium for *E. coli* Components dissolved in water and pH adjusted to 7.2 with NaOH.
Tryptone Yeast Extract Glucose (TYG)	Tryptone	10 g	Rich culture medium for Bacteroides
	Yeast Extract	5 g	
	Glucose	2 g	Components were dissolved in water and adjusted to 1 l. This medium is self pHing to 7.2. His-heme* was added prior to innoculation at a rate of 1 µl per 1 ml.
	Cysteine, free base	0.5 g	
	1 M KPO_4 pH 7.2	100 ml	
	0.4 mg/ml $FeSO_4$	1 ml	
	1 mg/ml Vitamin K	1 ml	
	0.8 % $CaCl_2$	1 ml	
	0.25 mg/ml Resazurin	4 ml	
	TYG Salt Solution ($MgSO_4$ 0.5 g/l, $NaHCO_3$ 10 g/l, NaCl 2 g/l)	40 ml	
Minimal Media Bacteroides (mm-target glycan)	NH_4SO_4	1 g	Minimal Medium optimised for growth of Bacteroides with the addition of a target glycan
	Na_2CO_3	1 g	
	Cysteine, free base	0.5 g	
	1 M KPO_4 pH 7.2	100 ml	
	0.4 mg/ml $FeSO_4$	10 ml	Prepared as with TYG. The target glycan (250 µl of a 10 % (w/v) solution), was added prior to inoculation. His-heme* was added prior to innoculation at a rate of 1 µl per 1 ml.
	1 mg/ml Vitamin K	1 ml	
	0.01 mg/ml Vitamin B_{12}	0.5 ml	
	0.25 mg/ml Resazurin	4 ml	
	MM Salt Solution (NaCl 18 g/l,$CaCl_2$ 0.53 g/l, $MgCl_2$ 0.4 g/l, $MnCl_2$ 0.2 g/l, $CoCl_2$ 0.2 g/l)	50 ml	
Brain-Heart Infusion (BHI)	BHI powder, as supplied (Sigma-Aldrich)	37.5 g	Rich culture medium for *Bacteroides* and *E.coli.* His-heme* was added prior to innoculation at a rate of 1 µl per 1 ml.

Table 2.4. Media (Continued)

Medium	Composition	Quantity (per litre)	Description/Method
Minimal Media Bifidobacterium (mm-target glycan)	Peptone	6.5 g	Minimal medium optimised for growth of Bifidobacterium species with the addition of a target glycan. His-heme* was added prior to innoculation at a rate of 1 µl per 1 ml. (Heavily adapted from Van der Meulen *et al*., 2006)
	Tryptone	2.5 g	
	KCL	2 g	
	$NaCO_3$	0.2 g	
	NaCl	4.5 g	
	$MgSO_4$	0.5 g	
	$CaCl_2$	0.45 g	
	$MnSO_4$	0.2 g	
	$FeSO_4$	0.005 g	
	$ZnSO_4$	0.005 g	
	Cysteine, free base	0.5 g	
	1 mg/ml Vitamin K	0.005 ml	
Clostridial Media (CM)	CM powder, as supplied (Sigma-Aldrich)	33 g	Rich culture medium for the growth of Bifidobacterium. His-heme* was added prior to innoculation at a rate of 1 µl per 1 ml.
His-Heme*	Hematin	1.2 g	Added to several of the above media as specified, for enhanced growth of *Bifidobacterium* and *Bacteroides* species.
	0.42 g/l Histidine HCL pH 8	1 l	

* His-Heme was added to most media after sterilisation, prior to inoculation to improve the growth of *Bifidobacterium* and *Bacteroides* species.

2.7. Routine Equipment Usage

2.7.1. Centrifugation

Large volumes were centrifuged using a Beckman J2-21 centrifuge with a JA-10 fixed angle rotor in 500 ml centrifuge tubes (Nalgene) at a maximum speed of 10,000 rpm (17,650 *x g*). A JA25.5 rotor with 50 ml centrifuge tubes (Nalgene) was used for smaller volumes at a maximum speed of 25,000 rpm (75,398 *x g*). A Harrier 18/80R centrifuge with swing out rotor was used with vivaspin® 20 concentrating centrifuge tubes (Sartorius Stedim Biotech) with a molecular weight cutoff of 10, 30 or 50 kDa to concentrate solutions.

Volumes between 2 – 25 ml were centrifuged using 30 ml universal tubes (Sterilin) in a Hettich Mikro 220R benchtop centrifuge with fixed angle rotor at a maximum of 14,000 rpm (4,025 *x g*).

Volumes 2ml or under were centrifuged in a Haraeus Instruments Biofuge *pico* using 2 ml or 1.5 ml microfuge tubes (Eppendorf).

2.7.2. Incubators, Heat Blocks and Water Baths

Agar plates were inverted and incubated using a static incubator at 37 °C (LEEC Ltd). *E. coli* cultures were incubated, shaking (180 rpm) at 37 °C during growth or 16 °C during protein expression using an orbital incubator (Sanyo Biomedical).

A heat block was used to incubate volumes under 2 ml using microfuge tubes (Eppendorf), most commonly for *E. coli* transformations at 42 °C. A water bath was used to incubate non cell culture material to temperatures between 30-80 °C.

Bacteroides and *Bifidobacterium* cultures were incubated (static) at 37 °C in an anaerobic chamber (Don Whitely Scientific).

2.8. Transformation of *E.coli*

Competent *E.coli* were made by Mr. Carl Morland and stored at -80 °C in 100 µl aliquots for up to a year. Each aliquot was incubated on ice until thawed and approximately 200 ng of the desired plasmid was added. Cells were mixed by gentle flicking and incubated on ice for 1 hour. Cells were then subjected to 42 °C for 1 minute before being transferred back to ice for a further 2 minutes. LB broth (400 µl) was added to the cell suspension and the culture was allowed to grow at 37 °C for 2 hours. Cells were centrifuged gently (1,000 *x g*) for 5 minutes and 400 µl of the supernatant removed. The remaining 100 µl of gently suspended cell culture was plated onto plasmid appropriate selective medium and grown at 37 °C overnight.

2.9. DNA Manipulation

2.9.1. Polymerase Chain Reaction (PCR)

The Polymerase Chain Reaction (PCR) originally developed by Mullis and Faloona in 1987 was used throughout this project for the amplification of targeted regions of DNA for use in downstream applications. PCR requires a thermostable polymerase protein which extends a region of DNA to which a primer, a small DNA fragment, has annealed. Cycles of different temperatures are carried out to first melt double stranded DNA, then allow the oligonucleotide primers to anneal and finally allow the polymerase to extend the primer through the addition of free nucleic acids to the growing DNA chain. Each reaction was made to a final volume of 50 µl using PCR grade water in a 200 µl sterile PCR tube (Eppendorf). Reactions were constructed as shown (**Table 2.5.**)

Table 2.5. Polymerase Chain Reaction Composition

Reagent	Concentration in 50µl Reaction
KOD Hot Start DNA Polymerase*	0.25 U
10X Buffer for KOD Hot Start DNA polymerase*	0.12 M Tris-HCL, 10 mM KCl, 6 mM $(NH_4)_2SO_4$, 0.1 % Triton X-100, 0.001 % BSA, pH 8.0
dATP*	0.2 mM
dTTP*	0.2 mM
dGTP*	0.2 mM
dCTP*	0.2 mM
$MgCl_2$*	0.125 mM
DMSO	5 % (v/v)
Forward Primer	50 pM
Reverse Primer	50 pM
Template DNA	1 µg/50 µl)

* Component from Novagen® Hot Start KOD polymerase kit.

Routine PCR reactions were carried out with the following program as standard; however annealing temperature was sometimes optimised according to the melting temperature of the primer set (3-5 °C below T_m). Additionally, a final storage step of 10 °C could be set indefinitely if required.

Table 2.6. Standard Program for Polymerase Chain Reaction.

Initial Denaturation		95 °C	2 minutes
35 cycles	Denaturation	95 °C	1 minute
	Annealing	55 °C	0.5 minutes
	Elongation	68 °C	0.5 minutes per 1kb to be amplified
Final Elongation		68 °C	10 minutes

Sewing PCR required an additional step before running a routine amplification. A program of 10 cycles of denaturation, annealing and elongation (lacking the initial denaturation and final elongation steps) were carried out.

QuikChange® mutagenesis was carried out in order to alter DNA sequences within a defined plasmid. PCR reactions and program were constructed and performed as stated in the protocol (QuikChange®, Stratagene).

2.9.2. Primer design and usage

Polymerase Chain Reaction (PCR) requires two primers, each complementary to one strand of the DNA and placed, 5'→ 3', at the terminals of the region to be amplified. Primers in this project were designed differently to perform one of four functions:

1) Amplification Primers: To amplify specific regions of DNA for further use (e.g. cloning).
2) Probe Primers: To amplify a small region of DNA within a target gene for detection purposes.
3) Sewing Primers: To "stitch" two regions of DNA together. These primers were used in conjunction with amplification primers.
4) Mutatgenesis Primers: To mutate a small region of DNA which has been cloned into a plasmid.

Amplification primers were the most widely used primer type in this study. Each primer was designed to be roughly 20 base pairs in length and, ideally, have a melting temperature (T_m) above 55 °C. These primers were prefixed with a single codon and a nuclease restriction site. The initial codon was present to allow the nuclease protein more space on the DNA fragment to cleave, as efficiency of these proteins may drop when the site is present at a fragment terminus. Target DNA sequences were checked for naturally occurring restriction sites using the online tool, NEBcutter (Vincze *et al*, 2003). Where possible, primer ends were G-C rich, as this forms an additional hydrogen bond and can increase annealing efficiency.

Probe primers were required for quantitative PCR as discussed later (Chapter 2.15). Each primer amplified a small region of DNA (100-300 base pairs) and was designed without any additional prefix. Primer3Plus (Untergasser *et al.*, 2007) was used to generate potential probes, and these were checked for specificity using routine PCR to ensure only one band of the correct size could be detected.

Probes must be specific to within the target gene, and must not amplify any other regions.

Sewing PCR was used to prepare plasmids for gene deletions. This technique is discussed separately as part of a larger protocol (Chapter 2.16.).

A set of two mutagenesis primers were created by taking a 30 base pair stretch of DNA and making the necessary alterations to the sequence, then creating a complementary primer at the same location, with the same mutation. Only one codon was altered per round of mutagenesis. Mutagenesis reactions were carried out as per the instruction manual (QuikChange ®, Stratagene).

Melting temperatures of primers were checked using the online web tool OligoCalc (Kibbe, 2007). This tool uses the following formula to calculate T_m;

$$\mathrm{Tm} = \frac{64.9 + 41(\mathrm{yG} + \mathrm{zC} - 16.4)}{(\mathrm{wA} + \mathrm{xT} + \mathrm{yG} + \mathrm{zC})}$$

Where w,x,y and z are the number of bases A, T, G and C in the sequence, respectively.

All primers were manufactured by MWG-Biotech and supplied lyophilised. Primers were dissolved in PCR grade water to 100 pmol/µl.

A list of key primers used within this project can be found in Appendix I (Table I.7.).

2.9.3. Agarose Gel Electrophoresis

Linearised DNA fragments were separated by size and visualised by electrophoresis in submerged horizontal agarose gels (Brody & Kern, 2004) regularly during manipulation to ensure the DNA was the correct size and concentration.

Agarose gels (1 % w/v) were made by combining 0.5 g of agarose in 50 ml of TBE buffer (89 mM Tris-borate and 2 mM EDTA, pH 8.3). The agarose suspension was dissolved by boiling and allowed to cool (<50 °C) before addition of ethidium bromide (1 µg/ml). Ethidium bromide is a dye which can intercalate into the DNA structure and be visualized by UV light.

Gels were cast in mini-gel trays (Applied Biosystems) and once set submerged in 50 ml TBE buffer. Loading buffer (10x concentration: 0.25 % w/v Bromophenol Blue, 50 % v/v Glycerol, 10x TBE) was added (1 µl) to the DNA sample (9 µl). DNA samples were loaded into the gel wells adjacent to the appropriately sized DNA standards (Hyperladder I or IV, Bioline) and gels were electrophoresed at 70 volts for 1 hour.

Gels were visualised using a gel documentation system (Bio-Rad Gel Doc 1000, Molecular Analyst^{TM}/PC windows Software). Photographs were printed with linked Mitubishi Video Copy Processor P68B with Mitsubishi thermal paper. Gel images were archived digitally for the duration of this project.

The size of linear double stranded DNA is determined by comparison to the known DNA standards (Hyperladder^{TM} I or Hyperladder^{TM} IV, Bioline), these standards are shown (**Figure 2.1**). DNA fragments migrate through the gel matrix at a rate inversely proportional to the $\log_{10}$ of the size of the nucleic acid. An example of an agarose gel from this project is displayed (**Figure 2.2**).

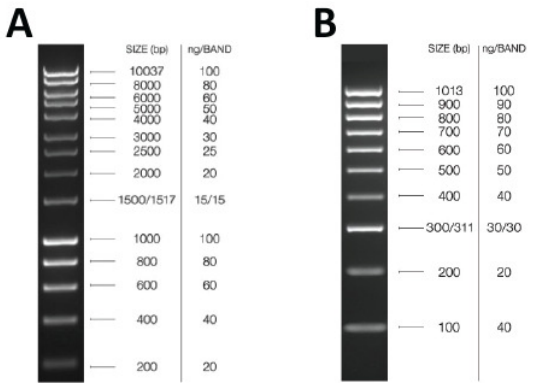


Figure 2.1. DNA ladders. This figure is adapted from the Bioline online catalogue. Two DNA ladders were used for comparison during this project, both from Biolone. (**A**) Hyperladder^{TM} I was used for viewing fragments between 200 and 10,000 bp and (**B**) Hyperladder^{TM} IV was used for viewing smaller fragments, between 100 and 1,000 bp.

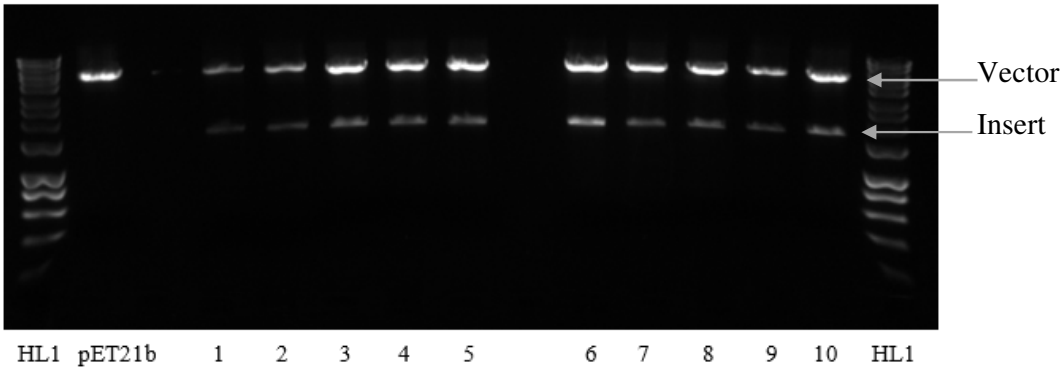


Figure 2.2. Agarose Gel Visualisation. This agarose gel shows a typical step in the cloning procedure. Newly cloned plasmids are digested with the two endonuclease enzymes used during plasmid construction, resulting in a band which corresponds to the vector, and a band which corresponds to the insert when visualised via agarose gel electrophoresis. HyperladderTM I (HL1) and the empty linearised vector (Pet21b) are run for comparison. Multiple clones are screened in this manner in case ligation efficiency of the desired fragment is poor, in this case all 10 clones screened contained an insert of the correct size (SusD-homologue, 1.7 kb).

2.9.4. DNA purification

PCR products and other DNA fragments were purified using a QIAquick PCR Purification Kit as per the manufacturer's instructions (Qiagen). In cases where the desired DNA fragment was impure a Gel Extraction Kit was used as per the manufacturer's instructions (Qiagen). Purified fragments are eluted in 10 mM Tris-Cl pH 8.5.

2.9.5. DNA Digestion and Ligation

Endonuclease restriction sites were chosen during primer design to match those required by the chosen vector. Digestion reactions were made to 40 µl total volume with PCR grade water and comprised of 1 µg DNA and the appropriate concentration of each enzyme in the buffer recommended by the manufacturer (DoubleDigest Caculator, Thermo Scientific). Restriction digests were incubated for 1 hour at 37 °C in sterile microfuge tubes (Eppendorf). Digested DNA fragments were purified after digestion to remove endonuclease enzymes and buffer.

Digested DNA fragments and vectors containing matched cohesive ends were used for ligations. 20 µl final volume made up with PCR grade water was used for ligation reactions. 4 µl of 5x T4 ligase buffer (250 mM Tris/HCL pH 7.6, 50 mM $MgCl_2$, 25 µM ATP, 25 mM dithiothreitol, 25 % (w/v) polyethylene glycol 8,000) and 1 µl (5U) T4 ligase were added per reaction. DNA was added so that insert was in at least 3 times molar excess of vector. This 3x excess was calculated using the formula:

$$\text{Mass of insert (ng)} = 3\left(\frac{\text{Size of Insert (bp)}}{\text{Size of Vector (bp)}}\right) * \text{Mass of Vector (ng)}$$

T4 ligase and T4 ligase buffer were produced by Invitrogen.

2.9.6. Determination of DNA and RNA concentration

DNA concentration was determined by comparison to known DNA standards on an agarose gel. DNA and RNA was quantified using a NanoDrop 2000 UV-Vis spectrophotometer (ThermoScientific). For DNA quantification both methods were used and respective concentrations were compared to ensure fidelity.

2.9.7. Preparation of Plasmids for Protein Expression

Plasmids designed for protein expression were compatible with BL21 or Tuner *E. coli* expression strains. Each plasmid contained an appropriate antibiotic resistance gene. All plasmid systems contained a lac operator and *lacI* repressor from the lac operon system (Jacob and Monod, 1961), which allows induction of protein when IPTG or lactose is present in the media.

PlysS was used in conjunction with an expression vector in order to reduce expression during exponential phase before induction. This could be done to increase yield of target protein or reduce the effect of protein toxicity during over expression. All constructs used for protein expression contained a His- or GST-tag for purification.

Purification of protein using these constructs is further described in Chapter 2.12.

2.9.8. Preparation of Plasmids for Genomic Disruption

A plasmid carrying a region homologous to part of the *Bacteroides* genome, but lacking the target gene, was required for genomic disruption in *Bacteroides ovatus*. Genomic disruption in using these plasmids was undertaken as described in Chapter 2.16. The CC118 λ pir strain of *E.coli* was used for cloning of these fragments into the pExchange *tdk* plasmid. Restriction sites BamHI and XbaI were routinely used, and SalI or SpeI used if necessary.

To remove a gene, a fragment was required which possessed homology to a region of at least 1,000 bases upstream and 1,000 base pairs downstream of the gene, but lacking the gene ORF entirely. These regions are referred to as the upstream and downstream flanks.

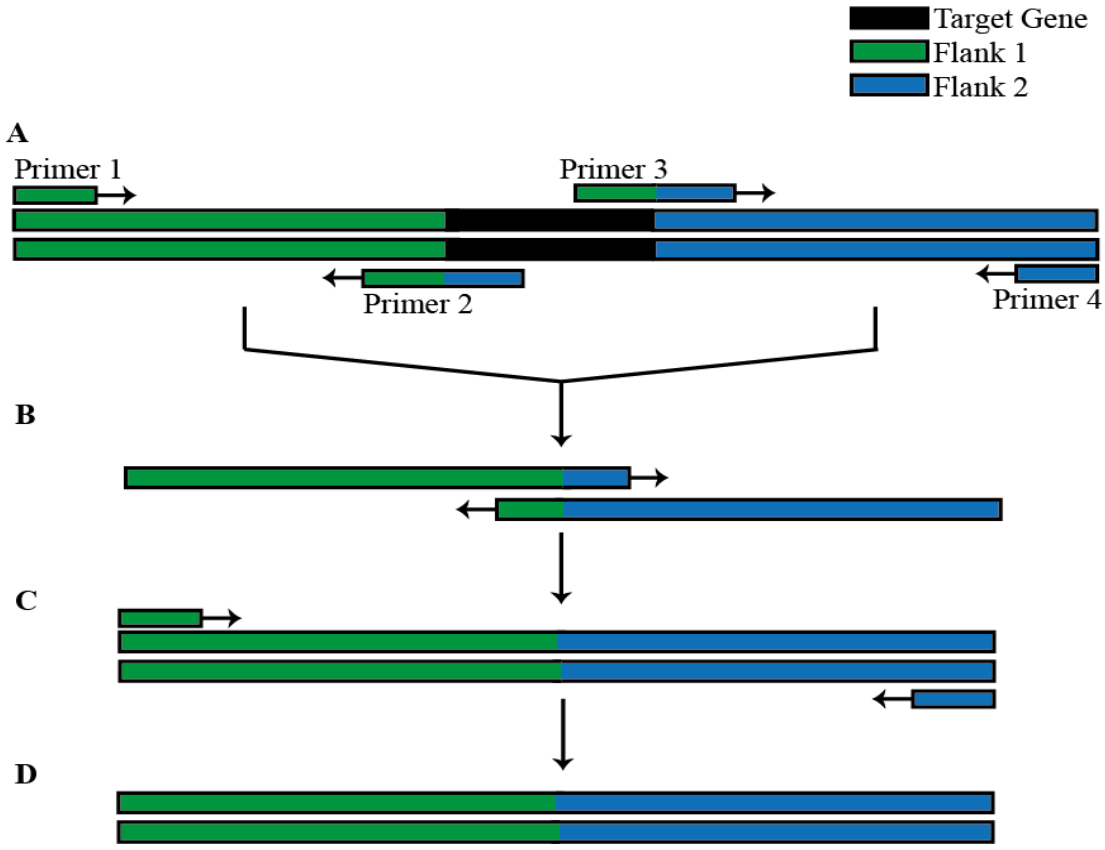


Figure 2.3. Sewing PCR was used to Generate Gene Knockout Fragments. Sewing PCR was used to remove a gene from a fragment of DNA. (**A**) Two 1,000 bp regions flanking the gene of interest were amplified during two distinct reactions using an amplification primer (primer 1, primer 4) and a sewing primer containing homology to the terminus of the flank to be joined (primer 2, primer 3). This step creates two fragments with a region of complementarity. (**B**) The first step of the two-step sewing PCR cycle contains only the two gene flanks, which act as both primer and template DNA to each other. (**C**) The second step of the two-step sewing PCR cycle is undertaken with the addition of the two amplification primers (primer 1, primer 4). (**D**) A hybrid fragment is created which lacks the target gene.

Four sets of primers were required for construction of this fragment, two amplification primers and two internal "sewing" primers which allowed the two flanks to be joined (**Figure 2.3**). The upstream primer (Primer 1) and the downstream primer (Primer 4) were designed as for routine amplification primers with appropriate restriction sites.

Primer 3 and 4 were designed by selecting the 20 bases pairs immediately upstream from the start codon and the 20 base pairs directly after the stop codon and adding these together to form a 40 base pair sequence. This sequence was used in antisense (Primer 3) and sense (Primer 4) to target each DNA strand.

The first step (**Figure 2.3 A**) is to create the two flanks, with homologous region. This is done by using primer set 1 + 2 and primer set 3 + 4 in two distinct routine PCR amplifications.

The second step (**Figure 2.3 B**) is to join these flanks together. This is done using a two-step PCR protocol as outlined (Chapter 2.9.1). The first cycle is performed with no primers or template DNA, but both flanks in equal concentration around 1 µg. As the flanks contain complementarity they will act as both primer and template to each other. After 10 cycles the amplification primers are added (Primer set 1 + 4) in order to efficiently amplify the newly constructed 2,000 base pair fragment (**Figure 2.3 C**).

It was essential to begin with both flanks at equal concentration; otherwise the reaction would favour the amplification of one flank above the other and result in an undesirable 1,000 base pair product. Once obtained, the 2 kb fragment was then purified and ligated into pExchange *tdk* vector.

If obtaining the sewing fragment was unusually difficult, PCR Blunt vector was used as an intermediate vector after the second PCR step. PCR Blunt accepts undigested PCR products directly and may be transformed. The sewing fragment could then be digested from this intermediate vector and re-ligated into the desired conjugation vector pExchange *tdk*.

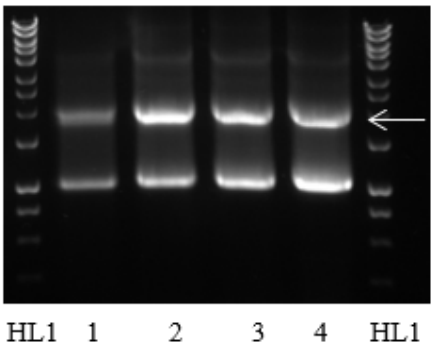


Figure 2.4. An Example Sewing Fragment. This 2 kb fragment was obtained during preparation of a Δ*bacova_04507* knockout construct. Hyperladder™ I (HL1) was used for size comparison. The sewing fragment (arrowed) is 2 kb in size, however there are contaminating flank regions which will lower ligation efficiency. The sewing reaction was done in quadruplicate so that successful fragments could be pooled.

2.9.9. DNA Sequencing

Automated DNA sequencing was performed using the MWG value read service (MWG- Biotech). 100 ng of DNA was dried by vacuum lyophilisation at room temperature in a 1.5 ml microfuge tube (Eppendorf) with a pre-ordered label provided by the company applied. This was sent to the company by postal service.

Sequencing primers were synthesised, stored and used by MWG-Biotech during automated DNA sequencing. Most plasmids contained regions targeted by standard sequencing primers, such as T7 (TAATACGACTCACTATAGGG) and T7term (CTAGTTATTGCTCAGCGGT). If no standard primers were available, custom primers were designed to amplify the region of interest, these were designed as for normal amplification primers (Chapter 2.9.2.) however no prefix was required.

Sequences returned by the company were aligned to the expected sequence using online alignment tools such as Clustal Omega and Multalign (Sievers *et al*., 2011; Corpet 1988) to ensure the sequence is as expected.

2.10. Routine Protein Analysis

2.10.1. Determination of Protein Concentration

Protein concentration could be roughly estimated by comparison to known markers using SDS-PAGE. This was performed prior to further analysis to ensure the protein sample was pure, and the target protein was of the expected size.

The absorbance of the appropriately diluted pure protein was measured using a NanoDrop 2000 UV-Vis spectrophotometer (ThermoScientific) and the concentration of protein determined using Beer-Lambert's law, below:

$$A = \varepsilon cl$$

Where **A** = Absorbance at 280-320 nm, **ε** = Extinction co-efficient (M) of the target protein calculated online using the EXPASY ProtParam program (Gasteiger, 2005), **c** = concentration (M), and **l** = path length (cm).

2.10.2. Adjustment of Protein Concentration and Buffer Exchange

Filtered protein solutions were concentrated using 20 ml or 2 ml Vivaspin™ centrifugal concentrators (VivaScience) by centrifugation. Buffer exchange could be achieved through concentration of the protein solution followed by dilution with the required buffer. At least two rounds of concentration were undertaken to ensure complete buffer exchange.

For larger volumes buffers were exchanged by dialysis. Protein samples were contained with a section of dialysis tubing with a MW cut-off of 13.5 kDa and both ends sealed with clips. This was submerged in 4 l of the appropriate buffer and 16-24 hours was allowed at 4 °C, stirring, for buffer exchange to take place.

2.10.3. Sodium Dodecylsulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein was visualised by SDS-PAGE as described by Laemmli (Laemmli, 1970) to determine the size, relative purity and relative quantity of the protein. 12.5 % polyacrylamide gels (Acrylogel 3; BDH Electran®) were routinely used for protein visualisation in conjunction with the AE-6450 apparatus from ATTO Corporation (Genetic Research Instruments) which utilises 12 cm x 10 cm glass plates sealed with a rubber gasket. All reagents for the creation of these buffers and gels are listed (**Table 2.7, Table 2.8**). The resolving gel was poured into the plates, covered with water and allowed to polymerise. The water was then removed and the stacking gel poured on top of the resolving gel. A comb is added and polymerisation of this second layer is allowed to take place. Before use, the comb and rubber seal is removed and the gel affixed within the gel tank, which is filled with running buffer. Loading dye is added to samples at a ratio of 1:2 and samples are boiled for 2 minutes to denature the proteins. Samples were loaded, alongside standards for comparison (**Figure 2.5**), into the gel wells and a current of 30 A (per gel) was applied.

After electrophoresis, the gel was soaked in InstantBlue™ stain (Expedeon) for 15 minutes to reveal protein bands, after which they were washed in distilled water. Gels were then photographed and catalogued (Bio-Rad Gel Doc 1000, Molecular Analyst™/PC windows Software). An example gel is shown (**Figure 2.6**).

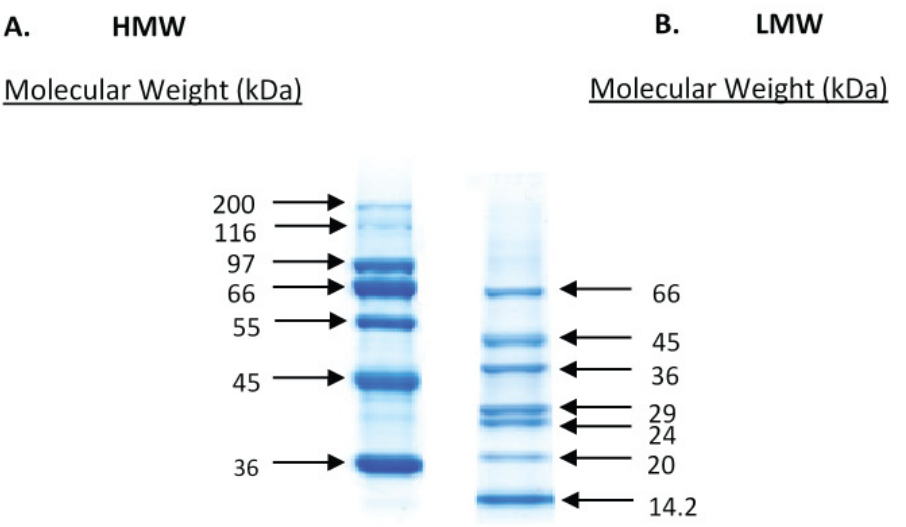


Figure 2.5. SDS-PAGE Standards. This figure is adapted from Sigma-Aldrich online catalogue. A high molecular weight (**A**) and a low molecular weight (**B**) standard (Sigma-Aldrich) were run on each SDS-PAGE gel for size comparison.

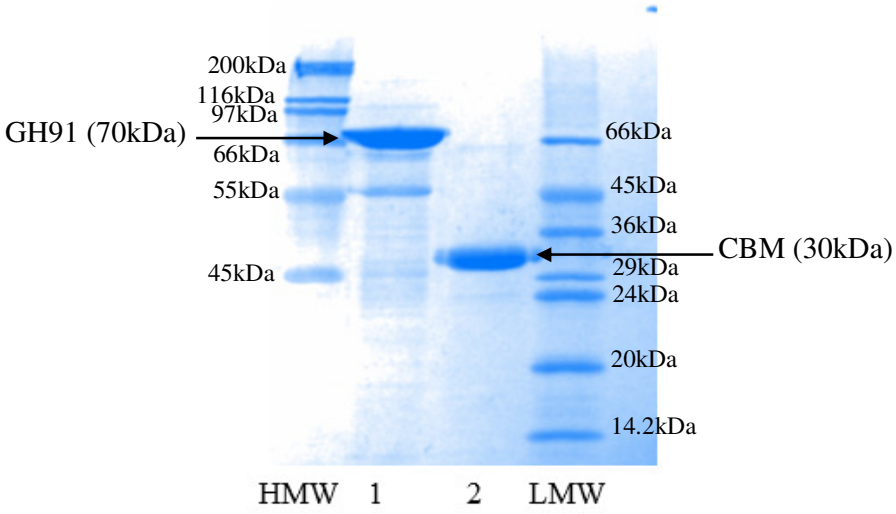


Figure 2.6. An Example SDS-PAGE Gel. The markers (HMW and LMW) are run so that sample sizes can be determined. Well 1 contains a protein of ~70 kDa (GH91 Enzyme) and well 2 contains a protein domain of ~33 kDa (Inulin Binding Domain).

Table 2.7. Preparation of SDS-PAGE Buffers

Component	Reagent	Volume or Concentration
Running Buffer	32 mM Tris/190mM glycine, pH 8.3	350 ml
(For 1 l*)	SDS	0.1 % (w/v)
Loading Buffer	SDS	10 % (w/v)
(For 10 ml*)	0.25M Tris/HCl, pH 8.8	5 ml
	Glycerol	25 % (w/v)
	β-mercaptoethanol	2.5 ml
	Bromophenol blue dye	0.1 % (v/v)

* Made up to this volume with d.d. H_2O.

Table 2.8. Preparation of SDS-PAGE Gels

Component	Reagent	Volume per gel
Resolving Gel	0.75 M Tris/HCl, pH 8.8 with 0.2 % SDS	2.35 ml
	40 % Acrylamide (BDH Electran acrylamide, 3 % (w/v) bisacrylamide)	1.45 ml
	d.d. H_2O	0.875 ml
	10 % (w/v) Ammonium persulphate	22.5 µl
	TEMED	7.5 µl
Stacking Gel	0.25 M Tris/HCl, pH 8.8 with 0.2 % SDS	0.938 ml
	40 % Acrylamide (BDH Electran acrylamide, 3 % (w/v) bisacrylamide)	0.188 ml
	d.d. H_2O	0.75 ml
	10 % (w/v) Ammonium persulphate	15 µl
	TEMED	5 µl

* Total volume listed is for one gel; gels were routinely made in batches of four.

2.11.Protein Expression

2.11.1. Protein Overexpression in *E.coli*

Plasmids were designed as described (Chapter 2.9.7) and transformed (Chapter 2.8). A single colony was picked from the resulting transformation plate and grown in 5ml of LB broth containing the appropriate concentration of antibiotic for 16 hours.

The entirety of this culture was used to inoculate 1 l of LB broth with antibiotic and this was grown for roughly 4 hours until the culture reached an optical density of 0.4-0.8 at 600 nm. Flasks were then cooled to 16 °C by incubation for up to 1 hour before induction with IPTG (100 mM, final concentration). Induced flasks were incubated overnight at 16 °C.

Cells were harvested by centrifugation (10 minutes at 4412.2 *x g*, 4 °C) and re-suspended in TALON^{TM} (20 mM Tris/HCL pH 8.0, 300 mM NaCl) or GST binding buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) at a rate of 5 ml buffer per 1 l of cell culture. This suspension could be stored at -20 °C for up to 3 months before crude protein extracts were prepared.

2.11.2. Preparation of Cell Free Extracts

Cell suspensions containing the target protein were thawed by incubation at room temperature. This cell suspension was sonicated for 2 minutes using a B. Braun Labsonic U sonicator set at low intensity (~42 watts and 0.5 second cycling) to lyse the cells. Insoluble cell debris was compacted to a pellet by centrifugation (30 minutes at 27143.1 *x g*, 4 °C) and the soluble supernatant, the cell free extract, was retained for further use. The insoluble fraction was re-suspended in the appropriate buffer at a rate of 10 ml per 1 l cell culture and a sample was visualised by SDS-PAGE alongside a sample of cell free extract to ensure the target protein was present in the soluble fraction.

2.12.Protein Purification

2.12.1. Immobilised Metal Affinity Chromatography (IMAC)

His tagged proteins were separated using Immobilised Metal Affinity Chromatography (IMAC). Histidine residues interact with electropositive transition metals such as cobalt and nickel. Metals are immobilised in a column and the His-tag, which is generally six histidine residues long, will stick to this column through interaction with the immobilised metal.

Imidazole disrupts this interaction by competitively binding to the column; in this manner bound proteins may be released and eluted with an imidazole gradient. In this project, TALON™ resins containing cobalt ions were used (Clontech Laboratories Inc.).

Columns were prepared by allowing 3 cm^3 of TALON™ resin settle in a gravity flow column. Column volumes were increased up to 5 cm^3 if a high protein yield was expected. A new column was used for every 2 l of starting culture. Columns were equilibrated with 3x column volumes of TALON™ buffer and eluent discarded. Cell free extracts (5-20 ml) were applied to the column and the eluent collected and retained for SDS-PAGE analysis.

Columns were washed with TALON™ buffer to remove unbound proteins; the eluent was collected and retained for SDS-PAGE analysis.
Columns were then subjected to two concentrations of imadazole (10 mM or 100 mM imidazole in TALON™ buffer) to elute weakly bound proteins prior to tightly bound proteins in 10 ml fractions. This gradient was important for protein purity and could be adjusted for each target protein in subsequent purifications.

A sample of each fraction (initial eluent, TALON™ buffer wash, and imidazole eluent fractions) was retained for SDS-PAGE analysis, often alongside samples of soluble cell free extract and insoluble cell debris. An example of this routine analysis is displayed (**Figure 2.7**).

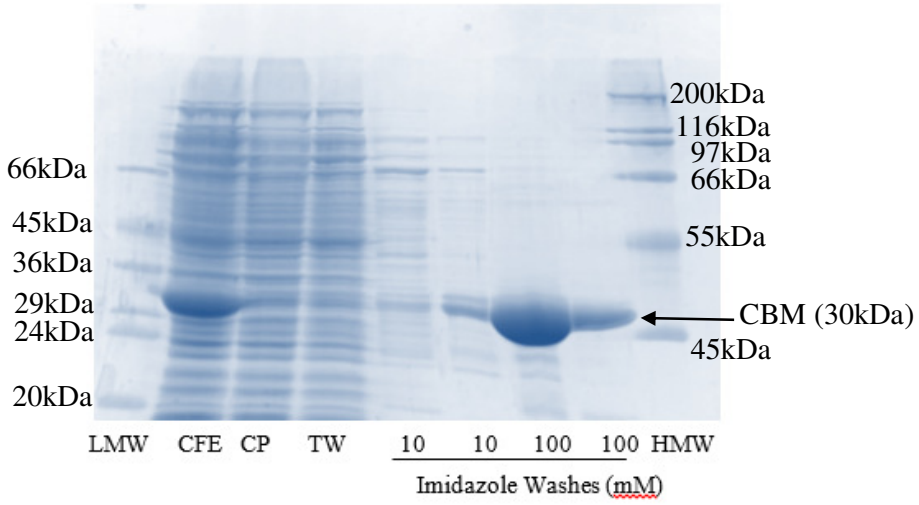


Figure 2.7. Protein Purification using a His-tag. An example of a protein purified using a His-tag. Cell free extract (CFE), Cell Pellet (CP) and the initial TALON^{TM} wash (TW) contain an impure mix of cell proteins. The target protein is then purified using two 5 ml washes of 10 mM imidazole and two 5 ml washes of 100 mM imidazole. The protein, the C-terminal region of BACOVA_04502 was ~30kDa.

The TALON^{TM} column could also be used to trap cleaved his-tag after protease digestion, allowing pure untagged protein to be collected in the eluent.

2.12.2. GST Tag purification using Glutathione Sepharose 4B

Glutathione S Transferase (GST) is a small protein which binds with high affinity to reduced glutathione. GST fusion proteins can be purified based upon this interaction and may also benefit from increased solubility. In this study, GST fusion proteins were selectively purified using a glutathione sepharose 4B resin (GE Healthcare). GST-fusion proteins bind to immobilised glutathione present in the column and are released in fractions by the repeated application of an excess of free reduced glutathione.

1 ml of this resin was allowed to settle in a gravity flow column. This was equilibrated with 10 ml GST binding buffer. A stopper was applied to the column and 5-10 mL of cell free extract was added and incubated for 10 minutes.

The stopper was then removed and the flow through collected. A further 10 ml of binding buffer was applied to the column to wash off excess unbound proteins. 1 ml of an elution buffer (50 mM Tris-HCl, 10 mM reduced glutathione, pH 8.0) was then applied to elute GST-fusion proteins, this step was repeated up to 5 times, and fractions containing target protein were pooled. SDS gel visualisation was routinely performed during GST purification to ensure the target protein was bound and eluted completely. An example of this routine analysis is displayed (**Figure 2.8**).

The glutathione sepharose 4B column could also be used to trap cleaved GST tag after protease digestion, allowing pure untagged protein to be collected in the eluent.

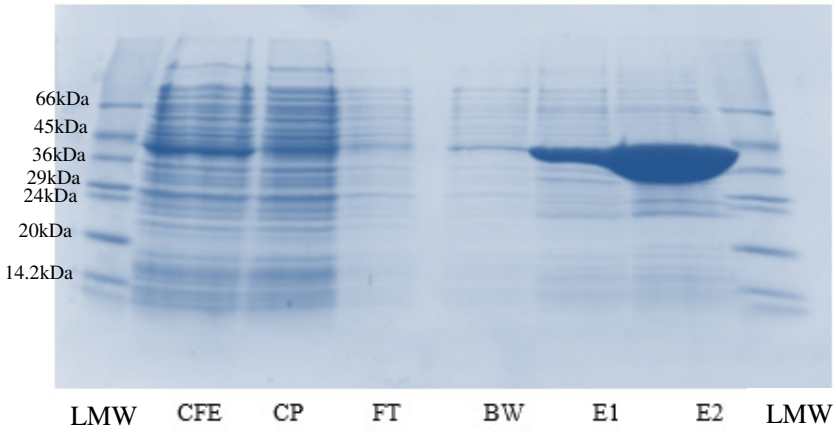


Figure 2.8. Protein Purification using a GST tag. LMW ladder was run for comparison. Cell Free Extract (CFE) and Cell Pellet (CP) contain a crude protein mix. The initial Flow Through (FT) from the column and the Buffer Wash (BW) shows little target protein, suggesting it has bound to the column. The protein is eluted, in this example, using two washes (E1, E2) of elution buffer resulting in purified protein. This protein is the 3C protease and is ~43 kDa.

2.12.3. Gel Filtration Chromatography

Proteins were purified by gel filtration chromatography using Fast Protein Liquid Chromatography (FPLC) with a HiLoad™ 16/60 Superdex™ 200 Prep grade column (GE Healthcare).
The column was used in conjunction with an ÄKTA pure 25 chromatography system (GE Healthcare). The column was equilibrated with two column volumes of buffer A (50 mM Tris, 150 mM NaCl, pH 7.5) at a rate of 1 ml/minute. Crude protein was concentrated to a volume under 1 ml or 5 ml and loaded with a clean 1 ml or 5 ml loop respectively.

Proteins were separated on the column using Buffer A at a flow rate of 1 ml/minute and collected using a fraction collector. Fractions were visualised using SDS gels and those containing high yields of pure protein were pooled for further use.

2.12.4. Protease Treatment

Protein constructs were tagged for purification. In some cases this tag became undesirable after this purification step and was removed through protease treatment; most commonly this was done in preparation for crystallisation trials where a flexible tag region could hamper crystal lattice formation.

During cloning it was ensured a protease cleavage site was present between the protein of interest and its purification tag. Thrombin or Enterokinase was used in conjunction with a pET vector His-tag; 3C protease was used in conjunction with P-GEX 6p-1 GST-tag.

Due to differing exposure of the cleavage site within different protein constructs, pilot assays (**Figure 2.9**) were carried out to identify the correct concentration of protease to use. Freshly prepared protease (10-50 µM) was diluted and added to the protein sample (50+ µM) and incubated for three hours. Based on the results,

an appropriate concentration of protease for the sample was chosen and used during for full scale reaction.

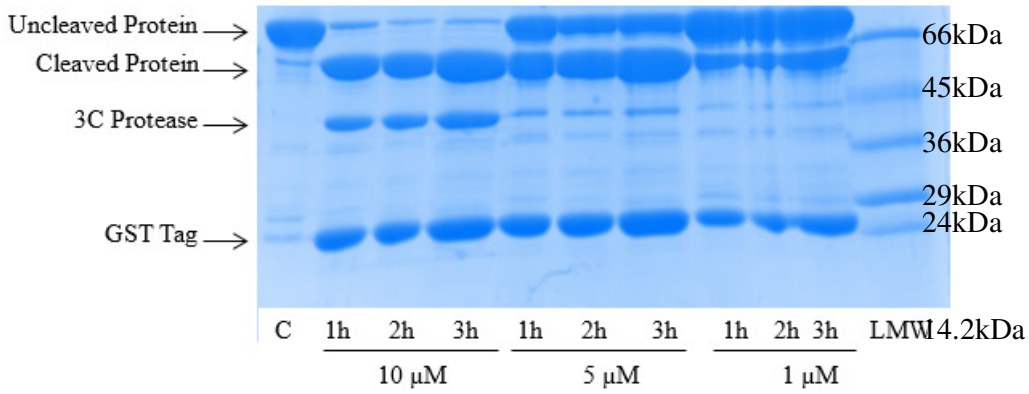


Figure 2.9. A Pilot 3C Protease Assay. This assay shows the cleavage of a GST tag from a protein by the 3C protease. In this assay, 3C protease was used at 10 µM, 5 µM and 1 µM and samples were taken at 1, 2 and 3 hours. Uncleaved protein (C) and a LMW marker were run as references. Most cleavage appeared to take place within the initial two hours, and no further cleavage was seen after overnight incubation. In this instance, 3C protease was used at the high concentration during the scaled up experiment.

2.13.Protein Characterisation

2.13.1. Fructose Detection Linked Enzyme Assays

Fructose was detected using the Megazyme fructose detection kit (Megazyme) in order to monitor β-fructosidase activity. This kit relies on three linked enzymatic steps to generate nicotinamide-adinine dinucleotide phosphate (NADPH); free fructose is phosphorylated by hexokinase, the resulting fructose-6-phosphate is then converted to glucose-6-phosphate by phosphoglucose isomerase. In the presence of glucose-6-phosphate dehydrogenase, glucose-6-phosphate is oxidised by nicotinamide-adenine dinucleotide phosphate (NADP+) to gluconate-6-phosphate, generating the reduced NADPH (**Figure 2.10**).

One molecule of NADPH is representative of one liberated fructose unit, however as the kit also detects glucose two NADPH molecules are generated upon the

breakdown of one sucrose molecule (this was accounted for by dividing the final values by two when sucrose was the substrate).
NADPH, thus liberated fructose, can be measured by the increase in absorbance at 340 nm using the extinction coefficient (6300 M^{-1} cm^{-1}).

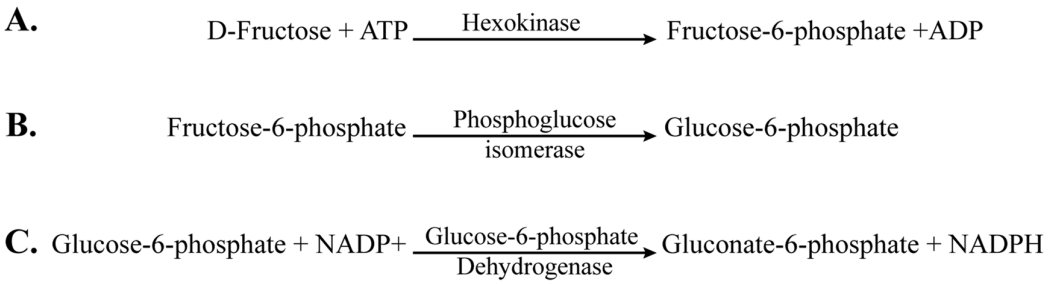


Figure 2.10. The Megazyme Linked Fructose Detection Kit. (**A**) Liberated fructose is phosphorylated and (**B**) converted to glucose-6-phosphate. (**C**) Glucose-6-phosphate is oxidised by nicotinamide-adenine dinucleotide phosphate (NADP+), resulting in the formation of gluconate-6-phosphate and NADPH. NADPH can be measured at 340 nm and has an extinction coefficient of 6300 M^{-1} cm^{-1}.

A spectrophotometer (Pharmacia Ultrospec 4000) was used in conjunction with quartz cuvettes to continuously measure absorbance at 340 nm. Cuvettes and reaction components were pre-incubated to 37 °C. Reactions were carried out as per the manufacturer's instructions in triplicate with protein from at least two fresh preparations using a range of substrate concentrations so that kinetic parameters could be elucidated. Data was plotted using Prism 6.0 (GraphPad).

2.13.2. Isothermal Titration Calorimetry (ITC)

ITC was used to characterise protein – carbohydrate interactions. A MicroCal™ VP-Isothermal Titration Calorimeter was used at 25 °C following standard procedures. Proteins were dialysed extensively in buffer (20 mM HEPES, pH 7) and ligands were dissolved in this dialysis buffer to minimise any heats from dilution. Degassed protein solution (1.4331 ml) at high concentration (45-100 µM) was equilibrated in the reaction cell. Ligand was titrated into this cell in 28 aliquots of 10 µl with rapid stirring (307 rpm) at 200 second intervals.

Ligand concentration was optimised to elucidate binding parameters, but typically concentrations around 10 mM for oligosaccharides, and 10 mg/ml for polysaccharides were used.

Upon binding, heat is commonly released (exothermic reaction) or absorbed (endothermic reaction); during titration the difference in electrical power required to maintain the temperature of the reaction cell vs. the reference cell was recorded and from these differences the heat change on binding was calculated.

The molar concentration of binding sites present in polysaccharide ligands could be calculated following the method used by Szabo and colleagues (Szabo *et al.*, 2001). If the protein contained a known number of binding sites (e.g. from structural studies or previous titrations against oligosaccharides of known molecular weight) the concentration of polysaccharide could be estimated; concentration was fitted in an iterative fashion until the n-value was as close as possible to one. This method was used to calculate the concentration of inulin polysaccharide, a previously characterised exo-binding inactive enzyme (BT3082 D147A) with one binding site was used (**Figure 2.11**; Szabo *et al.*, 2001; Cuskin, Doctoral Thesis, 2012). The levan used in this thesis was estimated at 1 % = 0.25 mM in the manner described above by Cuskin, 2012 and this estimation was used in this thesis (Cuskin, Doctoral Thesis, 2012).

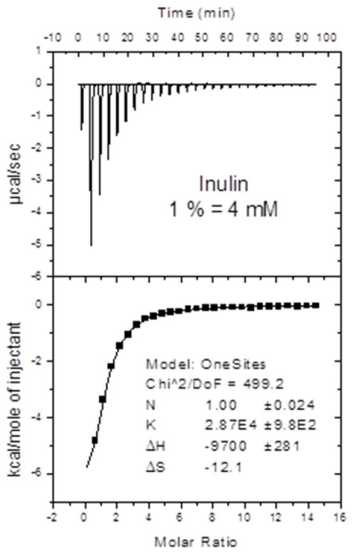


Figure 2.11. Estimation of inulin concentration. ITC titration for characterised inactive enzyme (BT3082 D147A) predicted the concentration of 1 % (w/v) inulin to be approximately 4 mM using the method outlined in the text (Szabo *et al*., 2001, Cuskin, Doctoral Thesis, 2012; Cuskin *et al.*, 2012).

Integrated heat effects, after correction for heats of dilution where necessary, were analysed by nonlinear regression using a single site-binding model (Microcal Origin, version 7.0.). The fitted data yield the association constant (K_a) and the enthalpy of binding (ΔH). Further parameters could be elucidated using the following equation:

$$-RTlnKa = \Delta G = \Delta H - T\Delta S$$

Where R is the gas constant (1.99 cal.K^{-1}.mol^{-1}), T is the absolute temperature (298.15 K), ΔG is the change in free enthalpy and ΔS is the entropy of binding.

Multiple titrations were undertaken, and average values were calculated for each ligand. As the number of datasets collected was variable, these are stated alongside binding parameters. Where only one dataset was collected, values were estimated from the fitted model.

2.13.3. Protein Crystallisation

Proteins solutions were prepared for crystallisation by IMAC or Glutathione Sepharose 4B purification, followed by a further purification step by gel filtration. Proteins were concentrated to between 10 mg/ml and 50 mg/ml in water.

Initial protein crystallisation screens (**Figure 2.12 A**) were performed using commercially available screening kits: JCSG+, PACT, Morpheus and Structure Screen I (Molecular Dimensions). Screens were conducted on MRC 96 well crystallisation plates (Molecular Dimensions) using two sitting drops containing 200 nl protein solution + 100 nl crystallisation condition and 100 nl protein solution + 100 nl crystallisation condition. Screens were set up using a Mosquito^{TM} (TTP Labtech) and sealed. Co-crystallisation was performed by incubating protein solutions with 20-200 mM required ligand prior to crystallisation. Plates were examined using a Leica MZ-6 crystallisation microscope (Leica MicroSystems).

Generally, crystals were harvested from screen conditions, however hanging-drop 24 well Linbro plates (Molecular Dimensions) were set up in some cases to achieve larger crystals (**Figure 2.12 B**). These plates also contained two drops (1 µl protein solution + 1 µl crystallisation condition and 2 µl protein solution + 1 µl crystallisation condition) and were assembled manually.

Three crystal structures were obtained during this thesis, the crystallisation conditions are shown in Table 2.9 below:

Table 2.9. Crystallisation conditions

Protein	[Protein] (mg/ml)	Ligand: [Ligand] (mM)	Condition	Cryoprotectant
SusD-homologue BACOVA_04504 Chapter 4.7.4	25	Sucrose: 200	20% (w/v) PEG 3350 0.2 M Sodium nitrate	20% PEG 400
ESBP Open Form BAD_1330 Chapter 5.7.2	10	N/A	10% (w/v) PEG 20 000, 20% (v/v) PEG MME 550, 0.02 M of each monosaccharide, 0.1 M bicine/Trizma base pH 8.5 As described (Gorrec, 2009)	ethylene glycol included in screen condition
ESBP Closed Form BAD_1330 Chapter 5.7.2	10	Kesto-tetraose: 200	12.5 % (w/v) PEG 1000, 12.5 % (w/v) PEG 3350, 12.5 % MPD, 0.03 M ethylene glycol, 0.1 M bicine/Trizma base pH 8.5 As described (Gorrec, 2009)	ethylene glycol included in screen condition

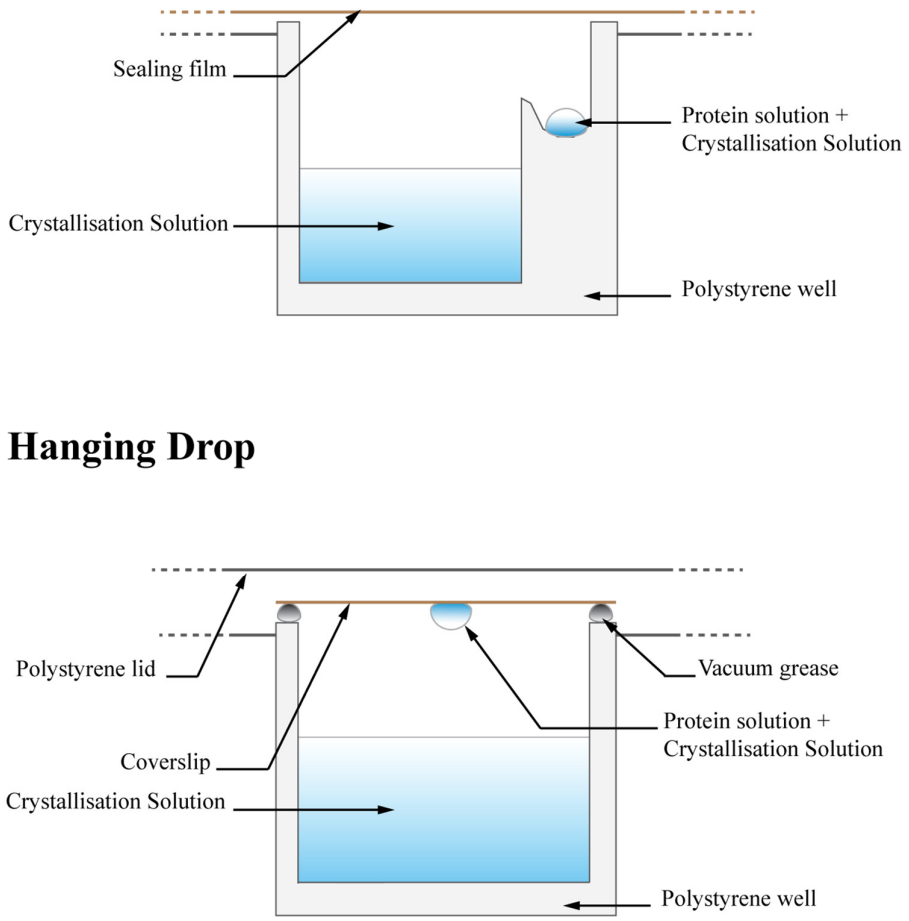


Figure 2.12. Sitting Drop and Hanging Drop Crystallisation. Crystal screens were conducted using (**A**) sitting drop crystallisation plates, consisting of a protein + crystallisation solution drop which sits upon a shelf above the reservoir of crystallisation solution. (**B**) Hanging drop crystallisation plates were used to optimise conditions for larger crystals. These were constructed using vacuum grease to seal an inverted coverslip containing the protein + crystallisation solution drop above the reservoir.

2.14.Bacterial Cultures

2.14.1. Culture Preparation and Monitoring

Bacterial cultures were set up by inoculation of the desired medium using a rate of 10 µl active bacterial culture or 20-50 µl glycerol stock per 1ml of growth medium. All experimental cultures were inoculated from actively growing cultures (these were seeded using glycerol stocks).

Bacterial cultures could be monitored directly throughout growth using a 96 well or 24 well corning® costar ® culture plate (Sigma-Aldrich) in conjunction with an Epoch microplate spectrometer (Biotek Instruments Ltd.) inside of an anaerobic chamber (Don Whitely Scientific), data was manipulated in Gen5 2.05 software and later plotted using Prism 6.0 (GraphPad). 96 well plates allowed for culture volumes of 250 µl, and 24 well plates of 1 ml. The plate reader measured and recorded the optical density (at 600 nm) of each well at 15 minute intervals. Each well was prepared in triplicate and the data averaged. Media without bacterial inoculum was always run as a control to ensure no contamination has occurred throughout the growth period.

For larger cultures glass test tubes were used to hold 5 ml aliquots of media. These were plugged with cotton wool to prevent contamination prior to sterilisation. Tubes were inoculated and incubated in an anaerobic chamber (Don Whitely Scientific). OD (at 600 nm) was measured using a CO 7500 spectrophotometer (Biochrom).

2.14.2. DNA Extraction

DNA extraction from 5 ml cultures was undertaken using the GenElute™ Bacterial Genomic DNA Kit (Sigma Aldrich) according to the manufacturer's instructions.

2.14.3. RNA Extraction and cDNA generation

RNA extraction from 5 ml cultures was undertaken by first stabilising the culture using 10 ml RNAprotect Bacteria Reagent (Qiagen) as per manufacturer's instructions. RNA was extracted using the RNeasy Mini Kit as per manufacturer's instructions. RNA was kept on ice for the duration of the extraction and all equipment washed and sterilised prior to the extraction to minimise contamination.

RNA was quantified after extraction using the NanoDrop 2000 UV-Vis spectrophotometer (ThermoScientific). As RNA is unstable, it was converted to cDNA immediately after extraction using the QuantiTect Reverse Transcriptase Kit (Qiagen) as per manufacturer's instructions.

2.14.4. Inactive Whole Cell Assays

5 ml cell cultures were gently harvested by centrifugation (5,000 *x g*), the supernatant removed and the pellet washed with Phosophate Buffered Saline Buffer (PBS). This step was repeated twice to ensure thorough washing before the pellet was finally re-suspended using 1 ml PBS. In the presence of oxygen, the cells are metabolically inactive but retain structural integrity. Thus, proteins which do not require ATP (such as CAZymes) presented at the cell surface remain active and can be observed.

Whole cells preparations were used as the catalytic agent during assays to detect CAZyme activity. 1 ml total reaction volume was used (500 µl whole cell assays, 500 µl 1 % glycan solution in PBS). Cells were boiled for 10 minutes and this matter used as a control reaction. To stop this reaction, samples were centrifuged (to remove cells) and the supernatant boiled. These samples were analysed by TLC.

2.14.5. Supernatant Analysis

5 ml cell cultures were pelleted by centrifugation (5,000 *x g*) and the supernatant removed. The supernatant was filter sterilised and visualised via TLC. Enzyme activity could be tested by adding glycan to this supernatant and monitoring any degradation by TLC visualisation.

2.15. Real-time Quantitative Polymerase Chain Reaction (qPCR)

cDNA or gDNA harvested from monocultures or batch cultures could be quantified using qPCR (in this thesis, all qPCR was real time qPCR). qPCR relies on a dye which releases measurable light during amplification. This project used SYBR Green I (Roche), an intercalating dye which absorbs light at 497 nm and emits light at 520 nm when intercalated into double stranded DNA. During each amplification step the samples are illuminated at 497 nm and emitted light at 520 nm is measured. The more amplification which occurs during the PCR reaction, the greater the intensity of light released, as the dye is incorporated into new double strands. A CQ (quantification cycle) value is obtained at the cycle where fluorescence from the sample exceeds background florescence, this is the point at which it is clear that a fragment is being amplified. A low CQ value means fewer cycles were required to detect amplification, whilst a higher CQ value shows that more cycles were required. Every experiment included a control with the probe set but no template DNA to ensure that background florescence was not classed as amplification throughout the experiment.

10 µl Reactions were set up using 5 µl of the SYBR Green I Master Mix (Roche), 1 µl of each probe (5 µM forward primer, 5 µM reverse primer), 2 µl of template gDNA or cDNA and 1 µl of PCR grade water. qPCR was carried out using a LightCycler® 480 (Roche) or using a Roche LightCycler® 96.

The qPCR program used is shown below (**Table 2.10**).

Table 2.10. qPCR Program and Parameters used throughout this Project

Initial Denaturation		95 °C	600 seconds
45 cycles	Denaturation	95 °C	10 seconds
	Annealing	57 °C	10 seconds
	Elongation	72 °C	10 seconds
	Measurement*	72 °C	-

* Measurement taken by exciting samples at 497 nm and measuring emitted light at 520 nm.

2.16.Genomic Disruption of Genes in *Bacteroides ovatus*

pExchange *tdk* plasmids containing a knockout fragment were prepared (Chapter 2.9.8). These were transformed into S17 λ *pir E.coli* cells, referred to as the "donor" strain. *Bacteroides ovatus tdk-* is the "recipient" strain. The donor and recipient strains were cultured (5 ml) to roughly equivalent cell densities in LB broth and TYG media respectively (**Figure 2.13 A**).

Cells were harvested by centrifugation and washed in TYG medium. Equal sized cell pellets were then re-suspended in 1 mL TYG medium and spread evenly on the surface of BHI plates with no antibiotic. These plates were incubated agar side down and grown for 16-24 hours until a thick lawn has formed; *E.coli* should grow first, creating an anaerobic environment underneath this growth in which *Bacteroides ovatus* can thrive, providing the necessary conditions for plasmid conjugation from the donor to the recipient strains (**Figure 2.13 B**). This biomass was scraped from the plate and re-suspended in 5 ml TYG medium. 100 µl of this solution, along with three serial dilutions (1:10, 1:100, 1:1000) were plated onto BHI + gentamycin (200 µg/ml) + erythromycin (25 µg/ml) plates. These antibiotics select for the recipient strain and the pExchange *tdk* plasmid, thus colonies represent single recombinant where the pExchange *tdk* has recombined with the genomic DNA via one of the flanks. These plates were incubated anaerobically for up to 2 days or until colonies formed, then 10 colonies were picked and re-streaked onto fresh BHI + gentamycin + erythromycin plates to minimise wild type contamination (**Figure 2.13 C**). 10 colonies were cultured overnight in TYG medium, 1 ml of each culture was taken and a pooled stock created**.** A glycerol stock can be made at this stage for safekeeping (**Figure 2.13 D**).

The pooled stock alongside three serial dilutions (1:10, 1:100, 1:1000) was plated upon BHI + FUdR (200 µg/ml) and allowed to grow anaerobically for 2 days or until colonies appeared. FUdR is toxic to strains able to synthesise thymidine. The recipient strain lacks the tdk gene, but this has been complemented within the pExchange *tdk* plasmid, in this manner FUdR selects for the second recombination event, whereby the second flank incorporates into the genome and the pExchange *tdk* sequence is eliminated.

10 FUdR resistant colonies were re-streaked onto fresh BHI + FUdR plates to minimise wild-type contamination (**Figure 2.13 E**). 10 resistant colonies were picked and cultured in 5 ml of TYG so that genomic DNA could be extracted and glycerol stocks could be made (**Figure 2.13 F**).

Isolated DNA was screened for successful knockout mutations using PCR. The downstream and upstream primers used to create the plasmid (primer 1 & primer 4, **Figure 2.3**) were used to amplify the clones, using wild-type *Bacteroides ovatus* as a control; the wild-type strain will produce a fragment which is the length of the target gene (500-2000 bp), plus the length of both flanks (1,000 bp each). Any successful knockouts will lack the target gene, yielding a fragment of 2,000 bp. Clones which appeared successful after screening were then sequenced to ensure the correct mutation had taken place.

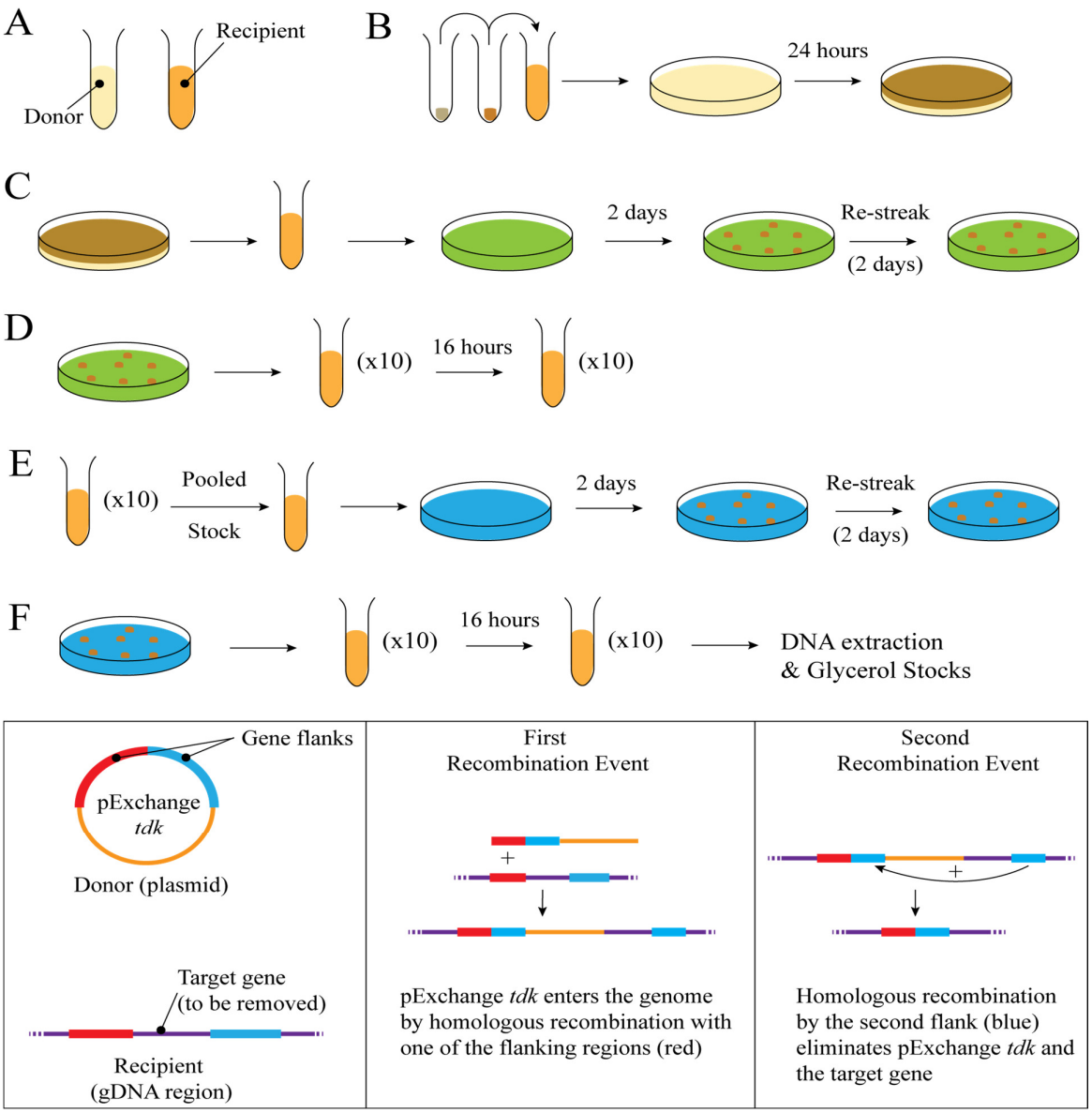


Figure 2.13. Generating Knockout Strains of *Bacteroides ovatus*. (**A**) The donor and recipient strains were cultured in 5 ml of LB and TYG media respectively. (**B**) Equal size cell pellets were harvested by centrifugation, washed in TYG, combined and re-suspended in 5 ml TYG and plated onto BHI plates containing no antibiotics (yellow). These plates were not inverted during growth. (**C**) The plates were scraped and the biomass re-suspended in 5 ml TYG. This was plated onto BHI plates containing gentamycin (200 µg/ml) and erthyromycin (25 µg/ml) (green). Resistant colonies were re-streaked onto fresh plates to minimise wild-type contamination. (**D**) 10 colonies (these represent the first recombination event) were picked and cultured overnight in TYG. (**E**) The cultures were pooled into one stock, which was plated onto BHI containing FUdR (200 µg/ml) (blue) to select for the second recombination event, as before these are re-streaked. (**F**) 10 resistant colonies are cultured overnight in TYG. Glycerol stocks are prepared and DNA extracted for analysis. Inset: A visualisation of the donor and recipient DNA, and the first and second recombination events.

2.17.Analysis of Carbohydrates

2.17.1. Purification of Fructans from Plant Material

Fructans were purified successfully from garlic bulbs using this method, but it can be applied to other fructan containing plant material. This method was adapted from Baumgartner *et al.*, 2000.

200 g of garlic (raw, broken into segments with the papery skin removed) was chopped roughly and transferred to a mortar and pestle and crushed into a rough paste. This garlic paste was added to 200 ml hot water at 80 °C. Some of this water was used to rinse the mortar in order to retain as much plant matter as possible. The mixture was incubated at 80 °C for 1 hour before boiling for 5 minutes. The mixture was then cooled and strained through cheese cloth to remove the solid material. The cloudy filtrate was then centrifuged at 14,000 *x g* for 20 minutes to pellet any additional insoluble material. The supernatant (which is now clear) should contain the soluble fructan fraction and can be freeze dried or stored at -80 °C.

Analysis by TLC or HPAEC-PAD was required to ensure fructan was present. If too much low molecular weight FOS or fructose was contaminating the sample further purification could be done by dialysing the solution in water using dialysis tubing with an appropriate molecular weight cut off.

2.17.2. Acid Hydrolysis Analysis of Polysaccharides

Polysaccharides could be hydrolysed to their constituent parts or partially hydrolysed into oligosaccharides using acid hydrolysis.

Complete acid hydrolysis was undertaken by incubating the glycan solution (1 % w/v) at room temperature (25 °C) for 1 hour with 1 M HCl. Samples were neutralised carefully with NaOH and the dilution factor noted. Samples were visualised using TLC.

Partial acid hydrolysis was done as with complete hydrolysis, but an initial pilot assay was required; samples were taken at 10 minute intervals and visualised by TLC. The time taken for the desired partial hydrolysis was noted and used as the stop-point for the scaled-up hydrolysis reaction.

2.17.3. Visualisation of Sugars by Thin Layer Chromatography (TLC)

Thin layer chromatography (TLC) uses a solvent system to separate non-volatile mixtures, in this project sugars, across a foil backed silica plate (Silicagel 60, 20 x 20, Merck). The plate is cut to an appropriate size with a height of at least 10 cm and a line drawn 1 cm from the bottom edge across the plate. A small cross was drawn every 1 cm along this line to serve as a guide when loading the plate. The adsorbent silica layer is loaded with a small quantity of glycan solution (3-5 µl) by spotting this onto a cross and allowing the spot to dry. If the solution is too low a concentration to accurately detect, additional spots can be added on top of the first spot allowing each application to dry. The solvent (comprised of 1-butanol/acetic acid/water at 2:1:1) was added to a glass chromatography tank (50 ml per tank of dimensions 23 cm x 23 cm x 7.5 cm) and the tank covered tightly. Vapours were allowed to equilibrate for at least 2 hours prior to use. The dry TLC plate was then placed into the tank, with the solvent reaching just below the pencil line. Different sugars will migrate with the solvent at different rates; this is due to differences in solubilisation and attraction to the adsorbent layer. Generally, smaller compounds will migrate more rapidly than larger compounds, but this is not always the case. The TLC was left until the solvent line reached within 1 cm of the top edge of the plate, it was then removed, dried and replaced and this mobile phase repeated, this results in clearer separation of sugars with similar sizes and solubility within the solvent system. Plates were then dried a second time and immersed for 20 seconds in orcinol sulphuric acid (comprised of sulphuric acid/ethanol/water 3:70:20 v/v, 1 % orcinol), this reagent stains sugars and allows for visualisation. Finally, these plates were dried and baked at 60-120 °C until sugar bands were visible.
TLC standards consisting of relevant sugars were always run for comparison. An example TLC plate is shown (**Figure 2.14**)

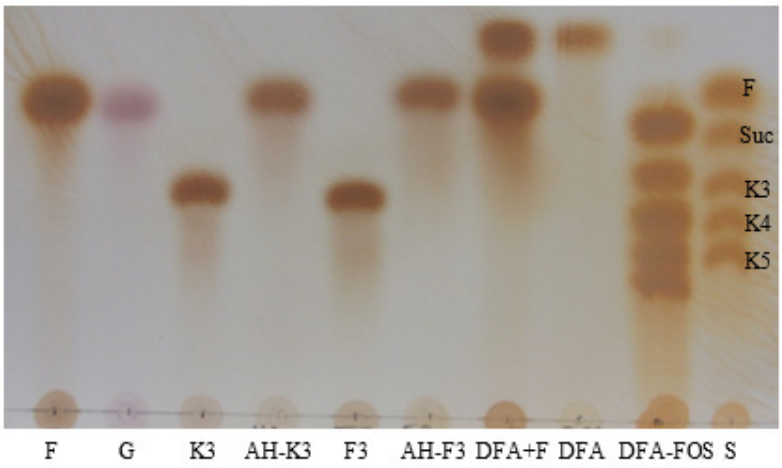


Figure 2.14. An Example TLC plate. This TLC plate shows several different sugars at 10 mg/ml. Standards (S) consisting of 2 mg/ml (each) of fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) are run for comparison. Samples run were fructose (F), glucose (G), kestose (K3), acid-hydrolysed kestose (AH-K3), fructotriose (F3), acid-hydrolysed fructotriose (AH-F3), Difructose anhydride + Fructose mixture (DFA + F), DFA, and DFA-terminated FOS (DFA-FOS).

2.17.4. Separation of Sugars by High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD)

Oligo- and polysaccharides were also analysed by HPAEC-PAD using an analytical CARBOPAC™ PA-100 anion exchange column (Dionex) equipped with a CARBOPAC™ PA-100 guard column.

The fully automated system (ICS-3000 gradient pump, detector compartment, electrochemical detector, auto sampler) had a loop size of 100 µl, flow rate of 1.0 ml/min, pressure of ≈ 2000 psi and was kept at room temperature; sugars were detected by pulsed amerpometric detection (PAD), the PAD settings were E_1 = +0.05, E_2 = +0.6 and E_3 = -0.6. The column was equilibrated with sodium hydroxide before a gradient of sodium acetate was applied. 100% sodium acetate was then applied to elute any material still adhered to the column after the gradient step.

The column was washed with 500 mM sodium hydroxide and equilibrated with 66 mM sodium hydroxide for the next run (**Table 2.11**). Appropriate standards were run for comparison. Sugars were loaded onto the column at concentrations of around 10 µg/ml. A glucose standard was run before and after data collection to control for any loss of detection or variance in elution time over the course of the collection period. Data were collected and manipulated using Chromeleon™ Chromatography Management System V.6.8 (Dionex) via a Chromeleon™ Server (Dionex). Final graphs were drawn with Prism 6.0 (GraphPad).

Table 2.11. Typical HPAEC-PAD Buffers and Program.

Step	Solution	Gradient (mM)	Time (min)
1	66mM sodium hydroxide	N/A	10
2	500mM sodium acetate in 66mM sodium hydroxide	0mM – 200mM 0mM – 500mM	50 90
3	500mM sodium acetate in 66mM sodium hydroxide	N/A	10
4	500mM sodium hydroxide	N/A	10
5	66mM sodium hydroxide	N/A	10

* Step two consisted of a gradient whereby sodium acetate concentration was increased linearly over the allotted time, polysaccharide samples were analysed with a longer gradient step than mono- or oligo- saccharides.

2.17.5. Identification of small sugars by Matrix Assisted Laser Desorption/Ionisation (MALDI) – Time of Flight (TOF) Mass Spectrometry

MALDI-TOF mass spectrometry was undertaken using an ABI Voyager-DE ™ STR Biospectrometry™ Workstation (Applied Biosystems) through the Pinnacle Facility at Newcastle University. Sugars were mixed with a matrix solution (2,5-dihydroxybenzoic acid) and applied to the metal plate. Multiple datasets (3x) were collected to ensure peaks were consistent.
An adduct calculator was used to determine peak identity, this calculator was developed by the Fien Lab at UC Davis and is available online (http://fiehnlab.ucdavis.edu/staff/kind/Metabolomics/MS-Adduct-Calculator).

2.18. pH Controlled Batch Cultures using Human Faecal Inoculum

Batch cultures were performed in collaboration with Ms. Roberta Grimaldi, Dr. Adele Costabile and Prof. Glenn Gibson at Reading University, with use of their facilities and equipment. Batch cultures were performed as described by Brück *et al.*, 2002:

Each culture was set up inside a temperature and pH controlled glass 100 ml vessel consisting of an N_2 inlet, gas outlet, glass stopper, pH probe, vessel clamp, vessel, lid, rubber ring seal, magnetic flea and an acid/base feeder. The magnetic flea was placed inside the vessel and the lid sealed on using the rubber seal, a liberal application of vacuum grease and the vessel clamp. The vessel has five ports within the lid allowing for the placement of the pH probe, the acid/base feeder, the gas outlet, the N_2 inlet and the stopper, and two ports on the vessel jacket which connect to the water bath; all ports were fitted with vacuum grease to form a tight seal, the vessels were then sterilised using an autoclave.

The vessels were connected to a water bath through two ports on either side of the vessel which allow water flow through a glass jacket around the vessel, keeping the contents at 37 °C.

All the ports were connected to the necessary apparatus, the pH controller was connected to the acid/base feeder and set to correct (by feeding small quantities of NaOH or HCl into the vessel) for fluctuations in pH to keep this between 6.7 and 6.9 during growth.

Autoclaved basal media (35 ml), 5 % (w/v) carbohydrate (10 ml) and freshly prepared homogenised faecal slurry in PBS (5 ml) were added to each vessel (**Table 2.12**). Growth occurred over 48 hours; two 1 ml samples were removed from each vessel at 0 hours, 4 hours, 8 hours, 24 hours and 48 hours post inoculation. These samples were split into a cell pellet fraction and a supernatant fraction by gentle centrifugation (5,000 *x g*) and frozen.

Samples were transported to Newcastle University on dry ice where supernatant fluid fractions were visualised by TLC and DNA extracted from the cell pellet fraction was examined using qPCR.

Table 2.12. Reagents used for Batch Culture Experiments.

Reagent	Composition	Description
Basal Medium (1 litre)	1 g peptone 1 g yeast extract 0.05 g NaCl 0.02 g K_2HPO_4 0.005 g MgSO4·7H2O 0.005 g CaCl2·6H2O 1 g NaHCO3 1 ml Tween 80 0.0025 g haemin 5 µl vitamin K_1 0.25 g cysteine HCl 0.25 g bile salts	Media was dissolved in 1 l deionised water and adjusted to pH 7.0 before sterilisation by autoclaving.
Carbohydrate Solution (10 ml)	0.5 g Target Carbohydrate	5 % (w/v) carbohydrate solution. This was 1 % (w/v) final in the vessels.
Faecal Slurry	10 % (w/v) faecal sample in PBS	The fresh faecal sample was collected and stored in an anaerobic jar at the beginning of the experiment and processed immediately. Faecal matter was weighed and the appropriate amount of PBS buffer added for a 10 % (w/v) solution (e.g. 10 g per 100 ml). This was homogenised and particulate matter removed prior to use.

2.19.Bioinformatics Tools

2.19.1. Signal Peptide Prediction

- **LipoP 1.0 Server**, sequence based prediction of lipid anchored (Type II) signal peptides (Bagos *et al*., 2008).
- **SignalP 4.1 Server**, sequence based prediction of periplasmic (Type I) signal peptides (Petersen *et al*., 2011).
- **PRED-Lipo,** sequence based prediction of lipid anchored (Type II) signal peptides (Juncker *et al*., 2003).

2.19.2. Crystallography

- **PyMol**, visualisation and manipulation of protein structures (www.pymol.org).
- **DynDom V2.0**, analysis of dynamic domains (http://fizz.cmp.uea.ac.uk/dyndom/).
- **DALI Server**, identification of structural homologues (Holm & Rosenström, 2010).
- **PDBSum Generate**, for analysis of protein secondary structure (de Beer *et al.*, 2014).

2.19.3. Multiple Sequence Alignments

- **Multalin**, multiple sequence alignment (Corpet, 1988).
- **Clustal Omega** and **Clustal W2**, multiple sequence alignment (Sievers *et al.*, 2011; Larkin *et al*., 2007).

2.19.4. Protein Parameters

- **ProtParam**, calculates protein parameters based on sequence (Gasteiger *et al*., 2005).

2.19.5. Primer Design

- **Oligocalc**, calculates oligosaccharide parameters based on sequence (Kibbe, 2007).
- **Primer3Plus**, assists probe primer design within a given sequence (Untergasser *et al.*, 2007).
- **NEBcutter V2.0.**, checks a DNA sequence for native restriction sites (Vincze *et al.*, 2003).

2.19.6. Genetics Tools

- **BLAST** (Basic Local Alignment Search Tool), Suite of tools used to compare sequence alignments against each other or against a reference database (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul, *et al.*, 1990)
- **IMG** (Integrated Microbial Genomes), database and toolset for the searching and comparison of microbial genome and metagenome datasets (Markowitz *et al.*, 2014).

Chapter 3 – Analysis of Non-Linear Fructan Structures

3.1. Chapter Overview

There is a great deal of research into the modulation of the microbiota by linear inulin and FOS compared with non-linear fructan structures, including levan and plant derived branched inulins; this is likely due to the relevance of linear inulin as a prebiotic, however non-linear fructans are found within commonly consumed plants and environmental bacteria and therefore presumably make up a significant portion of total fructan accessible to the microbiota, although there are too few data available at present to provide an accurate estimation of non-linear fructan intake. Furthermore, the main body of literature regarding non-linear fructan remains within the sphere of plant physiology; with little known regarding their role in human nutrition or as MACs. What can be asserted, however, is the presence of non-linear fructans from common dietary sources (**Table 3.1**); wheat, onion, garlic and agave have all been shown to contain non-linear fructans (Baumgartner *et al.*, 2000; Arrizon *et al.*, 2010; Velazquez-Martinez *et al.*, 2014; Verspreet *et al.*, 2007; Bancal *et al.*, 1991; Vijn *et al.*, 1997). Fructan extracts from these plants were either made or obtained for use during this project, an overview of the literature on these fructan structures is shown (**Table 3.1**).

Table 3.1. Common Non-Linear Fructan Structures

Fructan Source	Description*	Reference
Wheat	Low molecular weight (LMW) branched structures containing both β2-1 and β2-6 linkages, including bifurcose.	(Verspreet *et al.*, 2013; Bancal *et al.*, 1991; Cimini *et al*., 2015)
Garlic	High molecular weight (HMW) β2-1 backbone with regular, short β2-6 branches.	(Baumgartner *et al.*, 2000)
Agave	HMW fraction with β2-1 backbone with regular, short β2-6 branches; low molecular weight fraction is highly branched.	(Arrizon *et al.*, 2010; Velazquez-Martinez *et al*., 2014; Praznik *et al.*, 2013; Lopez *et al.*, 2003)
Onion	LMW fraction includes bifurcose.	(Vihn *et al.*, 1997)
Bacterial Levan	Bacterial levan, found in species such as *E. herbicola, Z. mobilis* and *B. subtilis* is highly branched but contains predominantly β2-6 linkages.	(Benigar *et al.,* 2014)

* A visualisation of the structures of bifurcose, branched inulin from garlic, bacterial levan and linear fructan homopolymers are shown in Figure 1.9, Chapter 1.3.1

Previous studies show that *Bacteroides ovatus* and *Bacteroides thetaiotaomicron* display a mutual exclusivity for linkage type and will only grow on inulin or levan polysaccharide respectively (Sonnenburg *et al.*, 2010). The levan utilisation system from *B. thetaiotaomicron* was characterised previously and the inulin utilisation system from *B. ovatus* has been characterised as part of this thesis (Sonnenburg *et al.*, 2010; Chapter 4). In this chapter the ability of these two species to use mixed-linkage fructans are investigated.

3.2. Objectives

- To visualise extracts from common crop plants using TLC and HPAEC-PAD, in order to provide evidence for mixed linkage or branched structures.
- To investigate whether *Bacteroides ovatus* and *Bacteroides thetaiotaomicron* are able to degrade mixed-linkage or branched fructans.

3.3. Acquisition and Analysis of Common Fructan Structures

3.3.1. Fructan Sources

Fructan extracts from plants purported to contain non-linear fructans (**Table 3.1**) were obtained in order to further investigate these structures with regards to the microbiota. Crude fructan extract was obtained by hot water extraction performed as outlined (Chapter 2.17.1) for garlic. Other fructan extracts, also obtained through hot water extraction, were a gift from Megazyme (Wheat, Onion and Agave). Linear inulin and branched bacterial levan were obtained from Sigma-Aldrich (Inulin: Chicory and Dahlia; Levan: *Z. mobilis* and *E. herbicola*).

Inulin and levan substrates used in this study have been characterised and are comprised of exclusively linear β2-1 linkages in the case of the two inulin substrates used, Chicory and Dahlia, and mostly β2-6 linkages in the case of both levan substrates from *E. herbicola* and *Z. mobilis* (Niness, 1999; Blake *et al.*, 1982; Benigar *et al.*, 2014). Levan samples were of a higher molecular weight than inulin samples, and were insoluble above 5% w/v. All other fructan samples were soluble at concentrations used within the study, up to 10% w/v.

3.3.2. Extraction of Garlic Inulin from Raw Garlic Cloves.

Garlic fructan was extracted from garlic cloves. To ensure that this garlic extract obtained contained soluble carbohydrate polysaccharide, this was visualised via TLC (**Figure 3.1 A**). The fructose content of the extracted carbohydrate was then examined by acid hydrolysis of the garlic extract to component monosaccharides. Acid hydrolysis revealed that the extracted carbohydrates were comprised of fructose and therefore fructans (**Figure 3.1 B**).

The absence of contaminating monosaccharides after hydrolysis indicates that the glycan component of the extract is exclusively fructan. The presence of glucose was expected, as fructans generally contain a sucrose terminus and thus glucose is a minor component of most fructan chains, however the glucose component of this glycan may be too low in concentration to detect, or masked by the presence of fructose. Once it was confirmed that the extract was comprised of fructan it was freeze-dried until a fine dry powder was obtained.

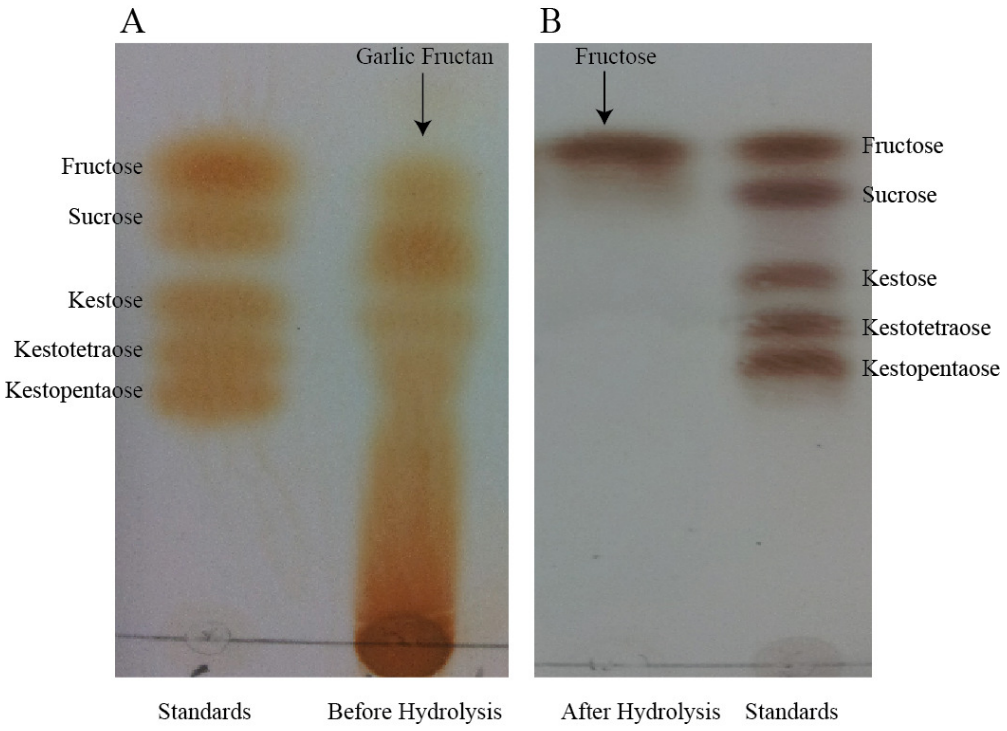


Figure 3.1. Acid hydrolysed garlic extract. Two TLC plates are shown before (**A**) and after (**B**) acid hydrolysis of extracted garlic fructan. Hydrolysis products were visualised after diluting with water 1 in 20 for clarity. Standards (0.2 mg/ml of each: fructose, sucrose, kestose, kestotetraose, kestopentaose) were visualised for comparison. The presence of large quantities of fructose and no obvious contaminants indicate that extracted glycans are fructans.

3.3.3. Visualisation of Fructan Structures

3.3.4. Thin Layer Chromatography (TLC) Analysis of Fructan Sizes

Fructans were separated based upon size and visualised using TLC as described (Chapter 2.17.3). Onion, wheat, agave, chicory, dahlia and garlic hot water plant extracts, neosugar® and raftilose® FOS mixtures and bacterial levan from *Z. mobilis* were visualised (**Figure 3.2**). Dark circles at the bottom of the plate indicate high molecular weight (HMW) structures which do not migrate (DP > ~8). Low molecular weight (LMW) structures could be separated (DP < ~8). Defined bands represent structures which have very similar molecular weights and similar solubility within the solvent system used; when smears are seen multiple different structures are present which do not separate well into distinct bands.

Defined bands usually indicate a single molecular weight product which can be separated, increasing DP can be observed in the kestose-series standards used which produce a clear series of bands.

This pilot study shows that onion, wheat, agave and garlic fructans contain a multitude of relatively low molecular weight fructan structures (**Figure 3.2**). These did not separate well into defined bands and are thought to represent multiple branched or decorated isomers with similar charges. Agave, chicory, dahlia and garlic fructans contained a large proportion of HMW structures which did not migrate. Levan (*E. herbicola*) did not migrate. Onion, wheat and agave contained LMW FOS, fructose and sucrose to varying extents. Onion and wheat displayed distinct bands, indicating relatively little structural diversity, compared with garlic and agave, which migrated in smears. The two FOS fractions run in this experiment contained defined oligosaccharides which migrated with the kestose series standards used (**Figure 3.2**).

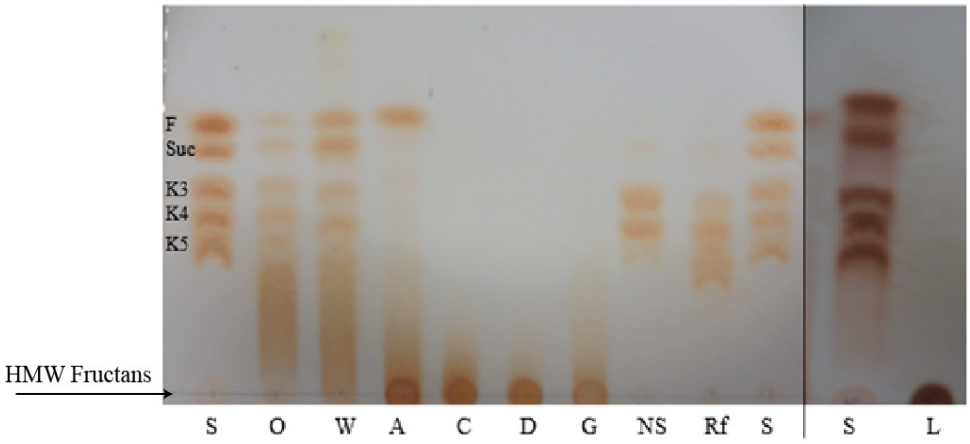


Figure 3.2. TLC visualisation of the different fructans used in this study. Standards (S) comprising of 2 mg/ml each of fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentase (K5) were loaded for reference. Onion (O) and wheat (W) fructans contain distinct bands with relatively low DP which do not co-migrate with each other. Agave (A) fructan is contaminated with fructose monosaccharide but it, chicory (C), dahlia (D) and garlic (G) do not migrate indicating that the bulk of this substrate is HMW. Neosugar® (NS) and Raftilose® (Rf) are two FOS mixtures obtained from Megazyme which contain DP 3-4 and DP 3-6 respectively. Levan (L) from *E. herbicola* was loaded separately with the same standards for reference, and did not migrate.

3.3.5. HPAEC-PAD Analysis of Fructans

Fructans were visualised by HPAEC-PAD (**Figure 3.3.**). Onion, wheat, garlic and agave fructan display a profile of several irregularly spaced peaks, which are not well defined, whilst chicory and dahlia displayed distinct, regular peaks characteristic of linear inulins with various chain lengths, despite the bulk of the glycan eluting during the wash. Linear kestose series standard with DP up to five were used to show increasing DP. The less distinct profiles of onion, wheat, garlic and agave suggest that multiple, LMW fructan structures with similar molecular weights (i.e. isomers, rather than increasing DP) exist within these extracts (**Figure 3.3.**). These data are consistent with the structures identified in the literature (**Table 3.1.**) and the TLC data (**Figure 3.2**).

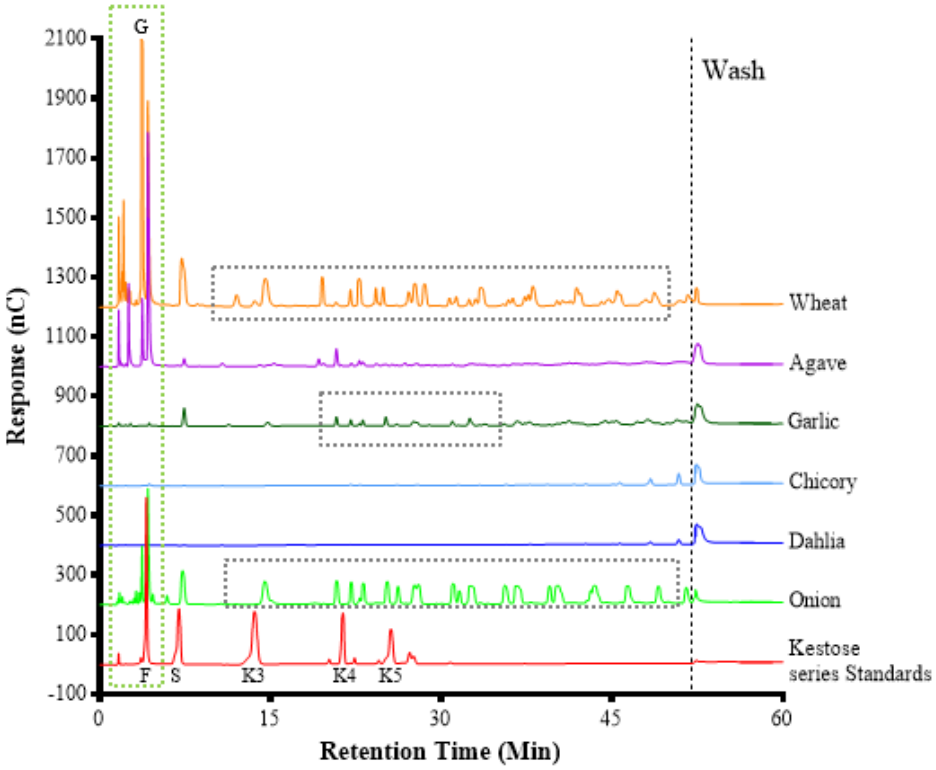


Figure 3.3. HPAEC-PAD Visualisation of Fructans. HPAEC-PAD visualisation of wheat, agave, garlic, chicory, dahlia and onion fructan extractions. Kestose series standards, fructose (F), sucrose (S), kestose (K3), kestotetraose (K4) and kestopentaose (K5) were loaded for comparison. A glucose standard was also run but is not shown on the graph, instead, a glucose peak (G) from wheat is labelled. All fructans were eluted on a sodium acetate gradient (0-200mM). A wash (500mM sodium acetate) was undertaken after the gradient was completed. For each trace above the standards, the Y values have been staggered to prevent overlapping. Wheat, Agave and Onion extracts contain a substantial diversity of small sugars (green box) in addition to glucose (G) and fructose (F). Agave, Garlic, Chicory and dahlia contained HMW fructans which did not elute until the wash step (dotted line). Chicory and dahlia show three regularly spaced peaks prior to the wash step, suggestive of linear inulin of increasing DP. Onion, wheat, and to some extent, garlic fractions contained small to mid-length FOS which did not align well with the standards and appear to represent multiple structures for each DP, indicating these are non-linear fructan structures (grey boxes).

Fructans with HMW from dahlia, agave, chicory and garlic were further analysed using a steeper sodium acetate gradient (0-500mM) during HPAEC-PAD visualisation, as these could not be satisfactorily visualised using the lower gradient shown above (**Figure 3.4.**).

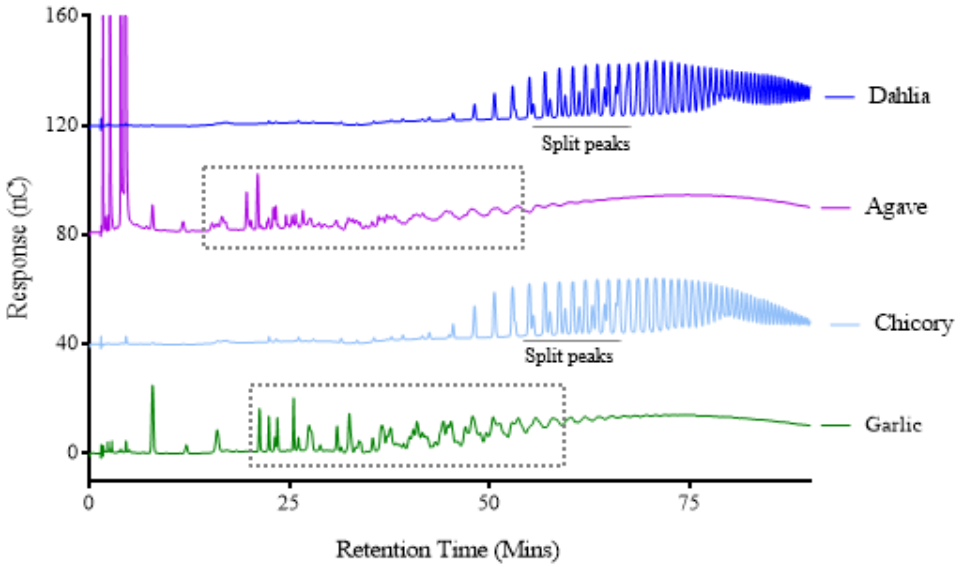


Figure 3.4. HPAEC-PAD analysis of HMW Fructans. HMW fructans were eluted using a 0-500mM sodium acetate gradient. Agave and Garlic fructans elute earlier from the column, without defined peaks (grey boxes). The lack of distinct, regular peaks indicate multiple different fructan structures with similar molecular weights, these isoforms likely represent variations in the structure, which cannot occur in linear linkage homopolymers, therefore it is likely these structures are non-linear. In contrast, dahlia and chicory fructan has a defined peak for each additional fructose unit. Early peaks in these two substrates seem split (labelled), it may be that these peaks represent fructans with and without sucrose terminations.

Fructan from chicory has been previously characterised as linear β2-1 linked inulin (Niness, 1999), dahlia inulin appears to be highly similar to that of chicory (**Figure 3.4.**). Garlic and Agave fructan both display a multitude of overlapping peaks highly suggestive of a variety of fructan secondary structures with similar molecular weights (**Figure 3.4.**).

3.4. *B. ovatus* and *B. thetaiotaomicron* Cultures Provide Insight into Non-Linear Fructan Utilisation in the Gut

Most of the avaliable fructan substrates were used as the sole carbon source during the growth (**Figure 3.5 A**) of a levan specific *Bacteroides* species, *B. thetaiotaomicron*, and an inulin utilising species, *B. ovatus* (Sonnenburg *et al.*, 2010) to determine if the non-linear fructans were levan, inulin or a mixture of linkages.

Spent culture supernatant fluid was visualised after 35 hours via TLC to show any unused fructan components (**Figure 3.5 B**). Control cultures without bacterial inoculum were visualised for comparison. Both species were cultured in minimal media supplemented with glucose as a control as glucose can be utilised very effectively by both species (Appendix I – **Figure I.5**).

As previously observed by Sonnenburg and colleagues, *B. ovatus* grew on inulin but not levan, and *B. thetaiotaomicron* grew on levan, but not inulin (Sonnenburg *et al.*, 2010). The TLC visualisation (**Figure 3.5 B**) shows a very slight release of fructose from inulin by *B. thetaiotaomicron* despite lack of growth; levan is completely depleted from the medium. *B. ovatus* was unable to degrade levan, which was expected as it did not grow (**Figure 3.5**). *B. ovatus* grew on inulin and depleted the polysaccharide from the media with the production of a waste product (later identified as di-fructose anhydride or DFA, Chapter 4.6.7). A defined β2-1 linked tetrasaccharide (Kestotetraose) was also tested, which demonstrated that *B. thetaiotaomicron* is unable to use even small inulin oligosaccharides.

Garlic fructan supported the growth of both species, with *B. ovatus* able to utilise this substrate more rapidly. The visualised spent supernatant fluid samples demonstrate that *B. ovatus* is able to utilise the entire fructan chain without creating DFA, whereas *B. thetaiotaomicron* is able to utilise the LMW component of garlic fructan only (**Figure 3.5**).

Wheat fructan supported both species. *B. thetaiotaomicron* was able to clear the substrate entirely from the culture medium and *B. ovatus* depleted much of, but not all substrate. No DFA was produced by *B. ovatus* during growth on wheat fructan (**Figure 3.5**).

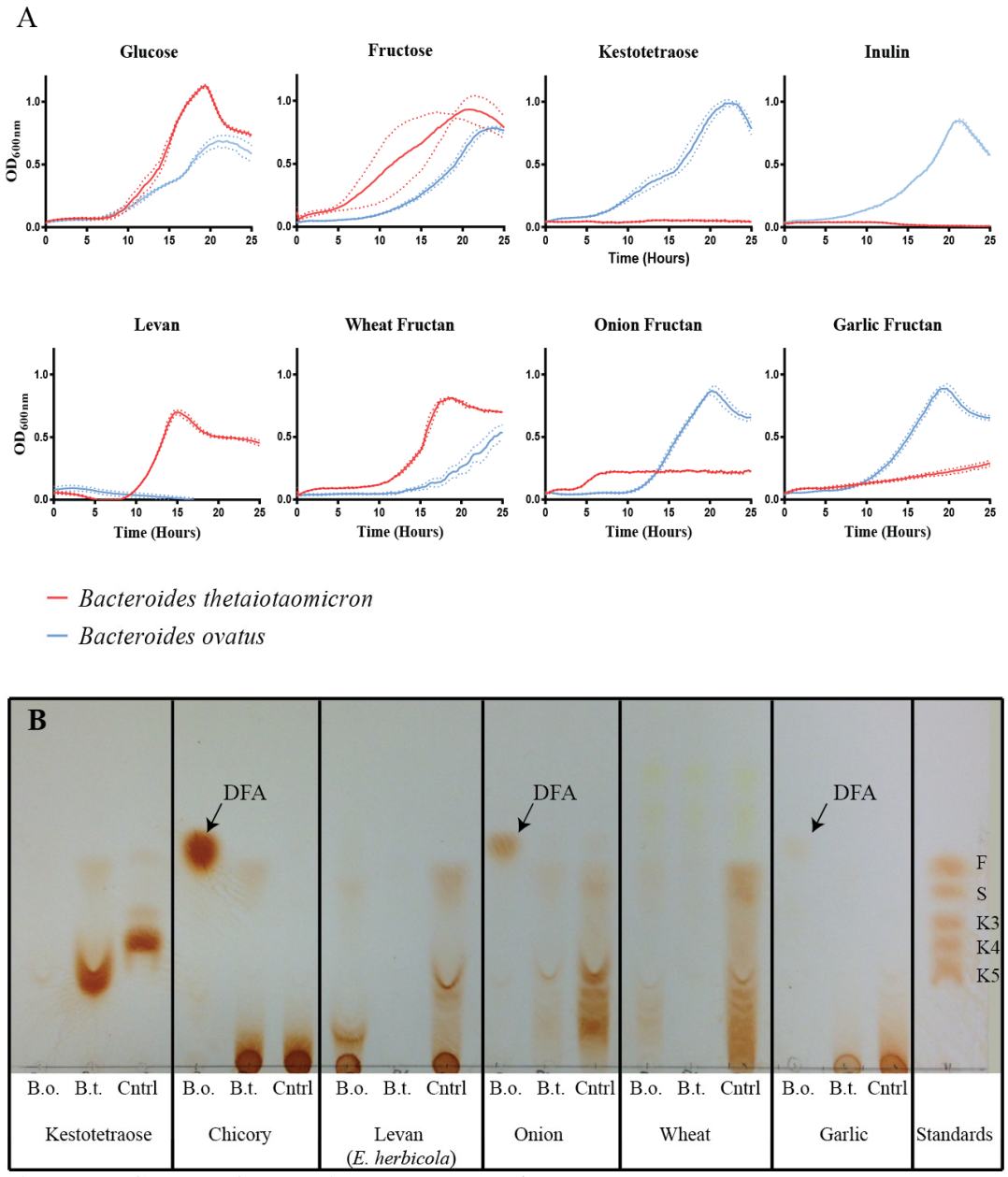


Figure 3.5. Growth of an Inulin User, *Bacteroides ovatus*, and a Levan User, *Bacteroides thetaiotaomicron* on various Fructan Extracts. Both species were cultured in minimal media supplemented with 0.5 % of each fructan extract, which contain non-linear fructans. (**A**) *B. ovatus* (blue) grew upon kestotetraose, chicory, onion, wheat and garlic extracts but did not grow on levan. *B. thetaiotaomicron* (red) grew robustly on levan and wheat extract, but weakly on onion and garlic extracts and did not grow at all on chicory inulin or kestotetraose. Triplicate technical repeats were performed and averaged as shown here, several datasets were collected to ensure consistency, and these data are representative. (**B**) Spent supernatant taken after 35 hours post growth reveals the extent of degradation of each fructan structure. A standard containing 2 mg/ml each of fructose (F), Sucrose (S), Kestose (K3), Kestotetraose (K4) and Kestopentaose (K5), was run for comparison.

Onion fructan supported the growth of *B. ovatus* well and resulted in the accumulation of DFA suggesting that a component of this glycan is linear inulin accessible to the GH91 enzyme. *B. thetaiotaomicron* was also supported, though to a lesser extent (**Figure 3.5**).

3.5. Discussion

The findings obtained for each fructan extract were in agreement with previous literature exploring the structure of these compounds.

Wheat fructan contained LMW FOS which did not map well to the known, linear inulin standards (**Figure 3.2** & **Figure 3.3**). This was consistent with the structures expected from the literature, where wheat fructan was comprised of both β2-1 and β2-6 linkages, with branching occurring from a "bifurcose" moiety (Cimini *et al*., 2015; Bancal *et al.*, 1991). It remains unknown whether *B. ovatus* can utilise short levan oligosaccharides, however *B. thetaiotaomicron* cannot use LMW β2-1 linked FOS; because *B. thetaiotaomicron* can fully clear wheat fructan from the culture supernatant during growth (**Figure 3.5 B**) it is unlikely that wheat fructan is comprised of β2-1 linked FOS. DFA may only be produced from *B. ovatus* growth on linear inulin (Chapter 4.6), no DFA is observed which would be expected if wheat fructan is non-linear. Wheat fructan does not display a regular pattern of peaks associated with increasing chain length and is unlikely to be solely linear levan; therefore branched levan or mixed linkage fructans are likely to be present (**Figure 3.3**). *B. ovatus* is able to clear most of the fructans in the wheat extract mixture (**Figure 3.5 B**) and therefore must be able to utilise LMW levans, or mixed-linkage fructans such as bifurcose. *B. ovatus* grows on wheat more slowly than *B. thetaiotaomicron*, though it reaches a similar final optical density (**Figure 3.5 A**) suggesting that it is not as efficient in degrading the substrate but may utilise most or all of what is available.

Onion fructan was expected to contain bifurcose (Vijn *et al.*, 1997) but indeed must contain a portion of linear inulin as DFA is made during growth on onion fructan by *B. ovatus* (**Figure 3.5 B**).

B. thetaiotaomicron may use some of the available substrate as it rapidly grows but to a much lower maximal optical density than *B. ovatus*, these data indicate that a portion of the fructan mixture, presumably the linear inulin, cannot be utilised by *B. thetaiotaomicron* (**Figure 3.5 A**). Therefore we suggest that a portion of onion fructan is bifurcose but this substrate is mainly inulin. Interestingly, *B. thetaiotaomicron* is able to clear much of the larger substrate from the media, potentially indicating the presence of large bifurcose type sugars which could be utilised or branch points which may be targeted.

Garlic fructan was shown to have a β2-1 linked inulin backbone, with regular β2-6 linked branch points of 2 or 3 fructose units each (Baumgartner *et al.*, 2000). Consistent with this, *B. ovatus* is able to utilise garlic fructan well, and *B. thetaiotaomicron* may only grow slowly, and to a much lower final OD (**Figure 3.5**).

The data suggest that whilst a general preference for either inulin or levan is observed in *B. ovatus* and *B. thetaiotaomicron* respectively, they may still have somewhat overlapping niches regarding non-linear fructan within the human intestine.

Chapter 4. Inulin Utilisation by *Bacteroides ovatus*

4.1. Chapter Overview

The *Bacteroides* genus is highly abundant in the healthy human intestine. This genus is the dominant genus within the prominent Gram negative Bacteroidetes phylum (Arumugam *et al*., 2011; Huttenhower *et al*., 2012; Eckburg *et al.*, 2005; Walter & Ley, 2011).

All saccharolytic members of the Bacteroidetes phylum have been found to encode polysaccharide utilisation loci or PULs (Marten *et al.*, 2011). PULs are co-regulated gene clusters which encode SUS-like systems; SUS systems, named for the canonical Starch Utilisation System, are gene products which act collectively to target, import and degrade specific glycans (Koropatkin *et al*., 2012; Martens *et al*., 2009; Flint *et al.*, 2012). Members of the *Bacteroides* genus encode staggering numbers of these systems in order to utilise a wider variety of glycans (Martens *et al.*, 2011). The success of this genus in the gut may indeed be due, at least in part, to the efficiency and wide variety of these glycan utilisation systems. *Bacteroides ovatus* is a prominent member of the microbiota, and has been shown previously to utilise inulin (Martens *et al.*, 2011; Sonnenburg *et al.*, 2010). As such, *Bacteroides ovatus* represents a major potential pathway for inulin degradation within the human bowel.

Bacteroides species, particularly *B. thetaiotaomicron*, have served as models for studying PUL. *B. thetaiotaomicron* was found to utilise levan, but not inulin, through a levan PUL, which was characterised using biochemical and genetic techniques (Sonnenburg *et al.*, 2010). Sonnenburg and colleagues identified a putative fructan utilisation locus within *B. ovatus* and showed that this species utilised inulin, rather than levan (Sonnenburg *et al.*, 2010). However, no insight has yet been obtained regarding the mechanism of this putative PUL.
This chapter describes the biochemical and genetic characterisation of the *B. ovatus* inulin utilisation locus.

4.2. Chapter Objectives

- To experimentally confirm the predicted *B. ovatus* inulin utilisation apparatus.
- To biochemically and genetically characterise the *B. ovatus* inulin utilisation apparatus.
- To use these data to construct a molecular model for inulin recognition, import and degradation by *Bacteroides ovatus*.

4.3. *Bacteroides ovatus* Releases a Range of Glycan Breakdown Products during Growth on Inulin

B. ovatus had previously been shown to utilise inulin (Sonnenburg *et al.*, 2010). *B. ovatus* was cultured using minimal media supplemented with inulin or with glucose as a control (**Figure 4.1**). Culture samples were taken over the course of growth and cells removed by centrifugation. The supernatant fluid was visualised via TLC to monitor degradation of inulin within the culture media. Extracellular oligosaccharides were produced from very early stages of growth; these accumulated during the exponential growth phase and were only depleted after stationary phase had been reached. A waste product, later identified as DFA (Chapter 4.6.7) accumulated in the media.

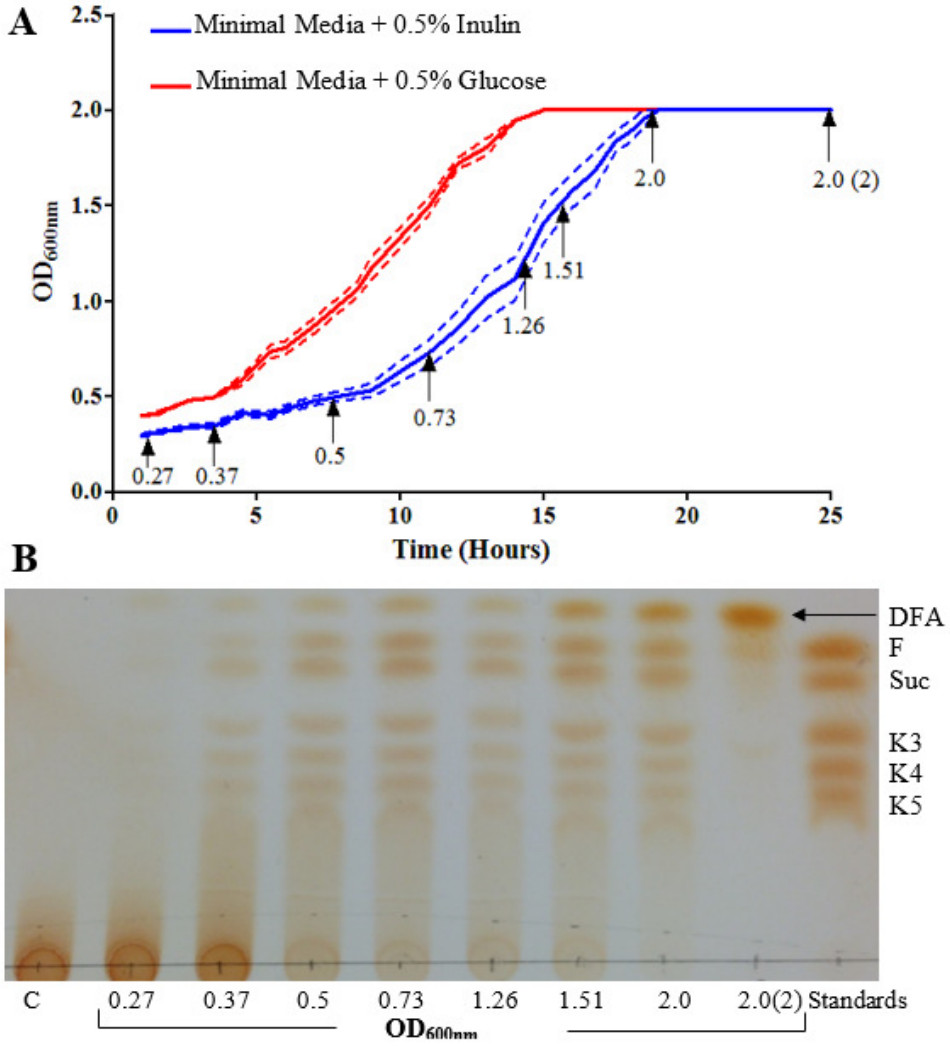
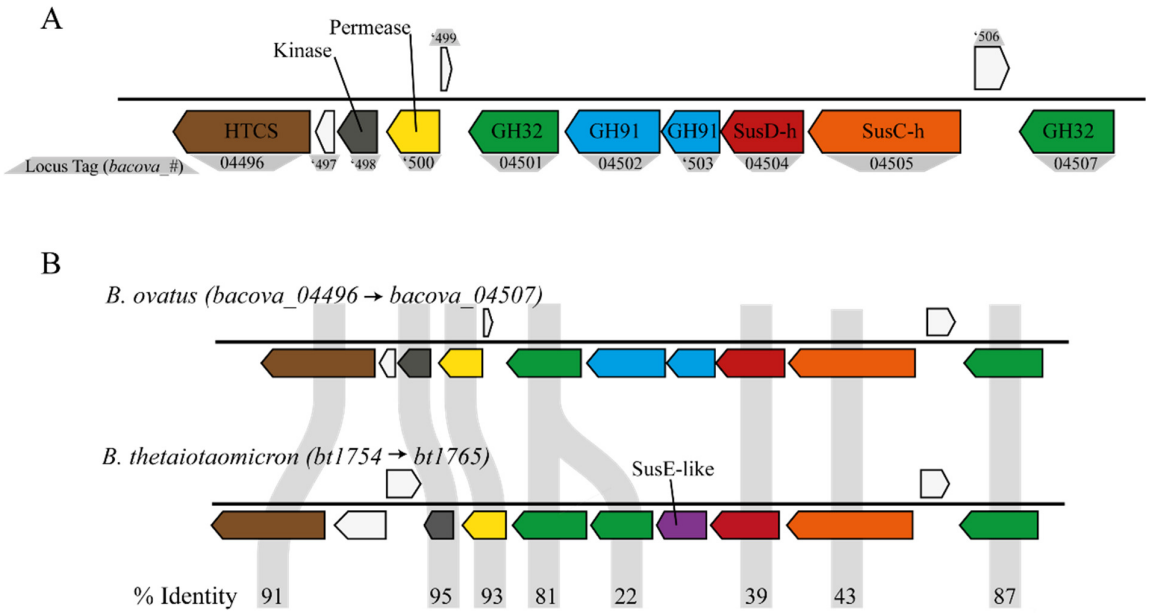


Figure 4.1. Glycan breakdown products are accumulated in the media over the course of *B. ovatus* growth on inulin. (**A**) Growth was monitored by measuring optical density (OD) at 600 nm at 1 hour intervals. Samples were taken over the course of growth, sampling points and OD are arrowed (maximum measurable OD_{600nm} = 2). Cultures were conducted in duplicate, continuous error bars are displayed using a dotted line. (**B**) Sample supernatant fluid was visualised using TLC. A media only control, C, was run alongside the *B. ovatus* cultures, this was harvested with the last sample to check that inulin did not degrade over time in a *B. ovatus* independent manner. Standards (2 mg/ml each, stock) consisting of fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) were run for comparison. A waste product later identified as Difructose anhydride (DFA, arrowed, Chapter 4.6.7) accumulated and did not appear to be degraded.

4.4. The Putative *Bacteroides ovatus* inulin PUL

The *B. ovatus* ATCC 8483 genome has been sequenced as part of the HMP Reference Genomes project and is avaliable on the Integrated Microbial Genomes Database (HMP Consortium, 2010). This genome was used throughout the project. The putative PUL identified by Sonnenburg *et al.*, consisted of twelve annotated ORFs (*bacova_04496* → *bacova_04507*) as illustrated (**Figure 4.2. A**; Sonnenburg *et al.*, 2010). Several genes including a hybrid two component system (HTCS) sensor-regulator, fructose kinase and permease and two GH32 family enzymes bore high sequence identity to genes from the previously characterised *B. thetaiotaomicron* levan PUL (Sonnenburg *et al.*, 2010; **Figure 4.2 B**); however two GH91 family enzymes are present in *B. ovatus* but not in *B. thetaiotaomicron*



and the SusD- and SusC-homologues are more divergent. *B. thetaiotomicron* has a SusE-like gene, however *B. ovatus* does not contain an obvious SusE-like candidate (**Figure 4.2,** Sonnenburg *et al.,* 2010).

Figure 4.2. The putative inulin PUL from *B. ovatus* compared with the levan PUL from *B. thetaiotaomicron*. (**A**) The putative inulin PUL from *B. ovatus*, predicted gene product functionality is annotated. (**B**) The *B. ovatus* PUL contains genes which share high sequence identity with genes from the previously characterised *B. thetaiotaomicron* levan PUL (Sonnenburg *et al.*, 2010), however the SusC/D pair is divergent and two GH91 family enzymes are present in *B. ovatus* but not *B. thetaiotaomicron*.

BACOVA_04498 is a predicted fructokinase with high sequence identity to a previously characterised fructokinase, FruK, from *Prevotella intermedia* (Fuse *et al.*, 2012) and BACOVA_04500 is a member of the major facilitator superfamily (MFS), this family of proteins mediate the passage of small solutes across lipid membranes (Yan, 2013). BACOVA_04500 shares 22% sequence identity with FucP, a fucose permease from *E. coli* which has been found to also import fructose (Gunn *et al.*, 1994; Kornberg & Lourenco, 2006); it is likely that BACOVA_04498 phosphorylates fructose within the cytoplasm of the cell and BACOVA_04500 mediates transport of fructose monosaccharide across the inner membrane.

4.5. The putative inulin PUL is up-regulated in the presence of inulin and fructose

Template cDNA was prepared from cultures grown with inulin or fructose as the sole carbohydrate source. Quantitative PCR (qPCR) was performed as outlined (Chapter 2.15) with probe sets targeting each predicted gene from the locus, excluding *bacova_04497* and *bacova_04499* which were very small. Cultures grown on minimal media with fructose and minimal media with inulin were compared with cultures grown on minimal media with glucose, which should not elicit up-regulation of genes involved in complex carbohydrate processing. The fold change of the genes grown on the target glycans was obtained compared to the genes grown on glucose.

A pilot experiment focused on the SusC-homologue encoding gene *bacova_04505* for which a probe set was designed. SusC-homologues have previously been used successfully as proxies for whole PUL activation (Martens *et al.,* 2011; Rogowski *et al.*, 2015). The *sus*C-homologue, *bacova_04393*, from the *B. ovatus* xylan PUL was used as a control (Martens *et al.*, 2011; Rogowski *et al.*, 2015). When *B. ovatus* was grown on xylan only *bacova_04393* was upregulated, in accordance with the results obtained by Martens and colleagues (**Figure 4.3**). When grown on inulin and fructose, *bacova_04505* was strongly up-regulated and *bacova_04393* was not. These data support that the putative inulin utilisation locus was activated in the presence of inulin and fructose.

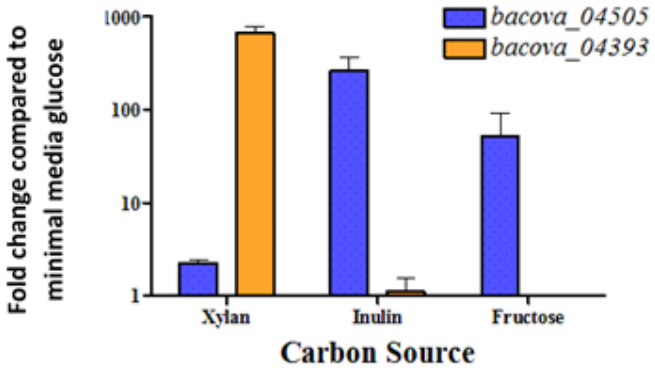


Figure 4.3. Expression of *sus*C-homologues on xylan, inulin and fructose compared to glucose. The abundance of transcripts encoding two *sus*C-homologues from *B. ovatus* cultures grown on minimal media supplemented with a target glycan was compared to that of cultures grown on minimal media with glucose to obtain a fold change. The *sus*C-homologue from the xylan PUL, *bacova_04393* was up-regulated on xylan but not on inulin or fructose. The *sus*C-homologue from the putative inulin PUL, *bacova_04505*, was up-regulated on inulin and fructose, but not xylan.

Probes were designed to target the remaining nine genes and fold upregulation of each gene was examined in the same manner (**Figure 4.4**). Inulin consistently elicits a higher fold up-regulation than fructose in most cases, with the exception of *bacova_04500*, the putative fructose permease.

Up-regulation of genes in the presence of the monosaccharide fructose in addition to that of the polysaccharide strongly indicates that fructose is recognised as a signalling molecule for the presence of polymeric inulin, these data match the findings for the HTCS present in the *B. thetaiotaomicron* levan PUL which binds to fructose within the periplasm (Sonnenburg *et al.*, 2010).

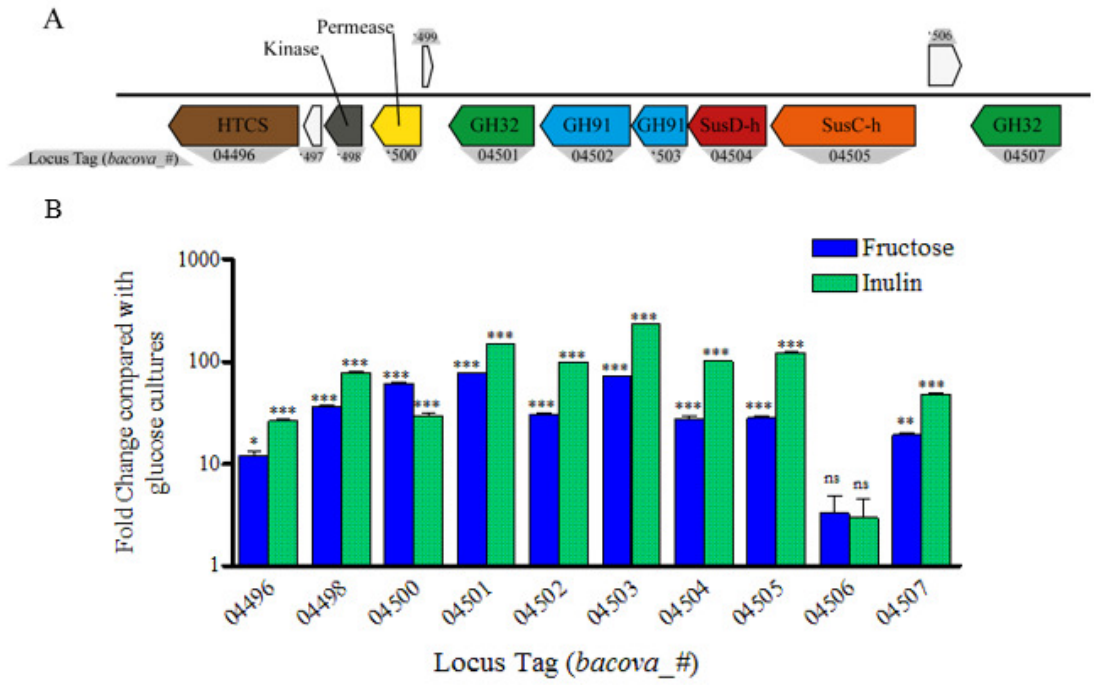


Figure 4.4. The putative *B. ovatus* fructan locus is upregulated in the presence of inulin and fructose. (**A**) PUL diagram illustrating the genes probed and their annotated function. (**B**) The fold up-regulation of each gene in the presence of fructose or inulin compared with glucose is shown, including standard error of the mean (SEM). All genes, with the exception of *bacova_04506* were up-regulated on inulin compared with glucose. Statistical significance (ns = $P > 0.05$, * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$) of the experimental value compared with the value obtained from the same probe set on glucose was determined using a one way ANOVA followed by Tukey's multiple comparison test. Full data is shown in Appendix I Table I.2.

4.6. Inulin is processed at the Cell Surface by a Heteromeric GH91 Family Enzyme

4.6.1. Identification of Endo-Inulinase Activity

To explore the initial step in inulin breakdown by *B. ovatus* supernatant fluid from cultures grown on inulin was analysed by TLC (**Figure 4.1**). The data show that the supernatant contained FOS (DP ~2-6), indicating that inulin is broken down externally prior to import - either at the cell surface or within the media. To determine if this activity was localised to the cell surface, *B. ovatus* was grown on inulin and glucose until mid-exponential phase was reached (OD_{600nm} = 0.8-1) and inactive whole cells were prepared as outlined to probe surface activity (Chapter 2.14.4). Whole cells prepared in this manner are metabolically inactive due to aerobic conditions, and are unable to actively transport glycans. Endo-inulinase activity was observed (**Figure 4.5**) which matched the activity seen in actively growing cultures, this shows that activity is localised to the cell surface. Activity was greatly reduced in cells grown on glucose, demonstrating that the enzyme(s) responsible are up-regulated in response to inulin.

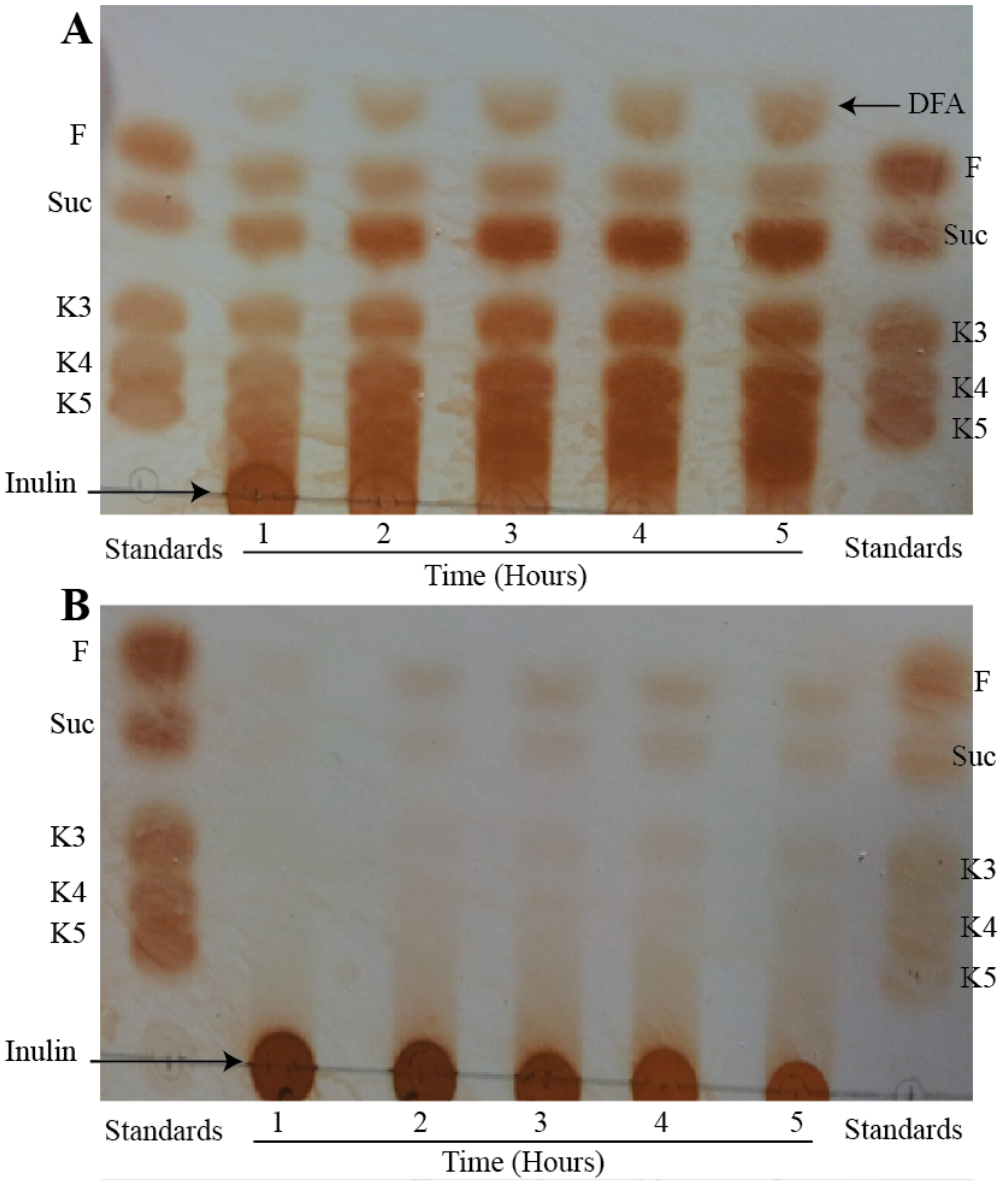


Figure 4.5. Whole cell assays show surface endo-inulinase activity. TLC was used to visualise the degradation of inulin by surface associated enzymes in *B. ovatus*. Assays were performed in the presence of inulin with cells harvested from either minimal media with 0.5 % inulin (**A**) or minimal media with 0.5 % glucose (**B**). Reactions were incubated for 5 hours, with sampling every hour. There is clear endo-inulinase activity from the cells initially grown on inulin, with FOS, fructose and an unknown product, later identified as DFA (Chapter 4.6.7) accumulating over time. The glucose grown cells also display some endo-inulinase activity, but this is much less than from cells grown on inulin. Standards (2 mg/ml each, stock) consisting of fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) were loaded for reference.

4.6.2. The Two GH91 Enzymes (BACOVA_04502 & BACOVA_04503) are Candidates for Endo-activity

The surface endo-activity detected in growing cultures was initially attributed to one of the two GH91 family enzymes present within the putative inulin PUL; BACOVA_04502 or BACOVA_04503.

Both genes products contained a predicted type II signal peptide sequence (LipoP 1.0 Server) indicating that both localised to the cell membrane by attachment to a lipid anchor (**Figure 4.6**). Previously characterised glycan utilisation systems from *Bacteroides* often contain proteins localised to the cell surface through the presence of these Type II signal sequences; making these enzymes candidates for mediating the endo-activity seen during whole cell assays (Rogowski *et al.*, 2015, Cuskin *et al.*, 2015; Cameron *et al.*, 2014; Larsbrink *et al.*, 2014).

The proteins are members of the GH91 family. BACOVA_04502 is significantly larger than BACOVA_04503, containing a region of 273 amino acids at the C-terminus which does not share homology with other GH91 family enzymes and forms a novel carbohydrate binding module which has been previously characterised (Shapiro, MRes thesis, 2012; Chapter 1.5). When aligned together without this region, BACOVA_04502 and BACOVA_04503 are 31.51 % identical (**Figure 4.7**). Two residues, an aspartate and a glutamate, are thought to be directly involved in catalysis in GH91s, as discussed in more detail in the introduction (Chapter 1.4.2). These residues were mapped on to the alignment by comparing each enzyme with other GH91 family members. Unexpectedly, neither protein has both residues; BACOVA_04502 has the catalytic aspartate (D185) and BACOVA_04503 has the catalytic glutamate (E196). This information suggested that these proteins operated through a different mechanism than the previously characterised BsIFTase (Jung *et al.*, 2007).

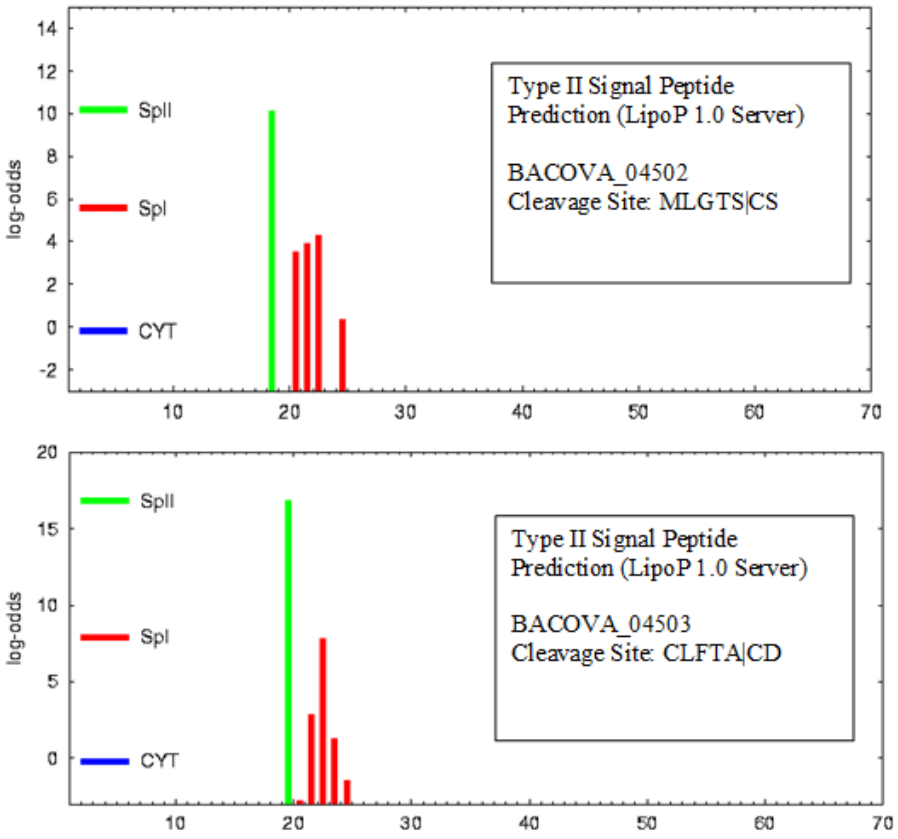


Figure 4.6. Prediction of Type II signal sequences in BACOVA_04502 and BACOVA_04503 using LipoP 1.0 Server. These data are the output files from the LipoP1.0 Server, which show strong type II signal sequences predictions for both BACOVA_04502 and BACOVA_04503. In *Bacteroides* species, this suggests that the proteins may localise to the cell surface (Sonnenburg *et al*., 2010; Rogowski *et al*., 2015; Larsbrink *et al.*, 2014).

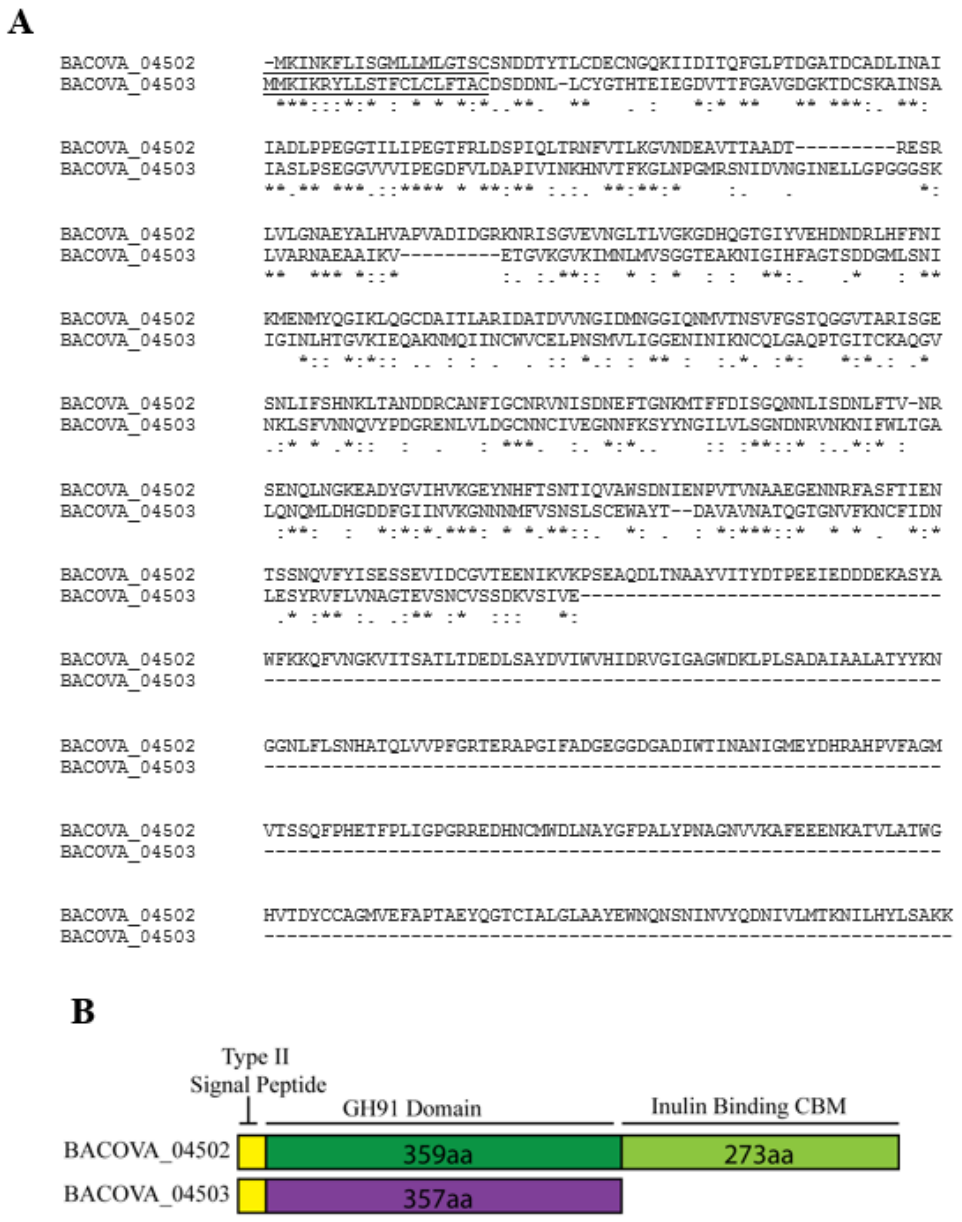


Figure 4.7. An alignment of the two GH91 enzymes, BACOVA_04502 and BACOVA_04503. (**A**) Sequence alignment between the two GH91 family enzymes (BACOVA_04502 0-359 aa; BACOVA_04503 0-357 aa). The predicted type II signal peptides are underlined and the two predicted catalytic residues are highlighted; note that BACOVA_04502 has the catalytic aspartic acid at 185, but has lost the glutamic acid at 196; BACOVA_04503 has the glutamic acid at 196, but has lost the aspartic acid at 185 (**B**) A visual representation of the gene domains, including the 273 aa inulin binding domain from BACOVA_04502.

4.6.3. The GH91 Proteins Interact to form an Active Endo-Inulinase

Each GH91 enzyme was cloned and expressed using recombinant *E. coli* as described (Chapters 2.9.7 and 2.11). Proteins were purified using IMAC (Chapter 2.12.1). Further construct details are listed (Appendix I.8).

Both GH91 proteins were incubated individually with inulin, however no activity could be detected with either enzyme. Reaction conditions were optimised by altering pH (5-8), salt (0 mM or 150 mM), buffer (tris or sodium phosphate) and calcium content (0 mM – 5 mM) to no effect.

However when the two GH91 proteins were added to the same reaction, enzyme activity was detected (**Figure 4.8**). The largest product formed appears to be around DP ~6, therefore this enzyme cannot be active upon inulin series oligosaccharides with DP < ~6 (**Figure 4.8**).

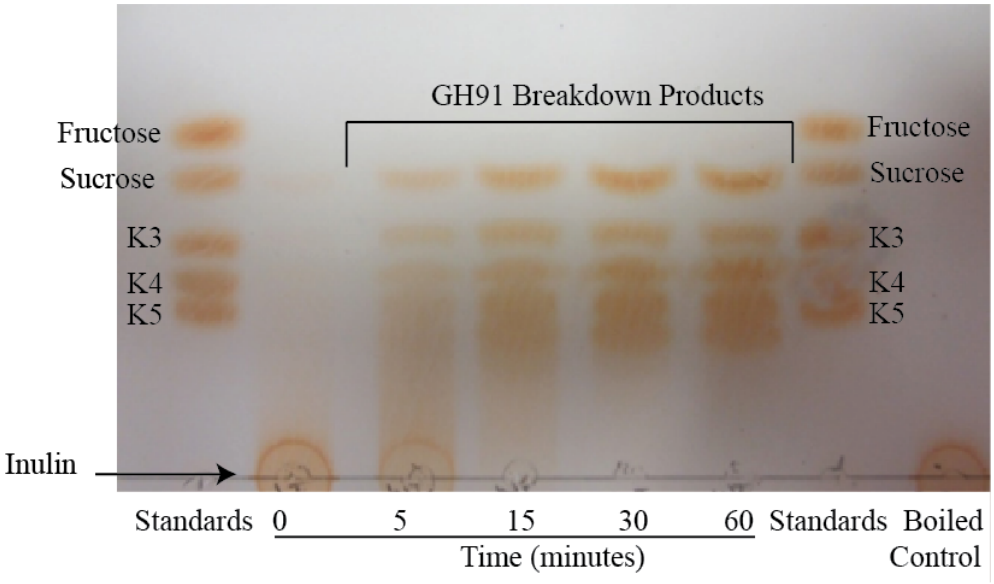


Figure 4.8. Endo-inulinase activity requires both GH91 enzymes. To observe the progression of the GH91 reaction, both recombinant enzymes (1µM each) were added to a 1ml reaction containing 0.25% inulin in 20 mM Tris, 150 mM NaCl, pH 7 at 37 °C. Samples were taken at 0, 5, 15, 30 and 60 minutes. A control reaction containing boiled enzyme (Boiled Control) was also run for the full 60 minutes before visualisation. Under these conditions, the reaction is complete by 30 minutes. All samples and relevant standards (stock. 2 mg/ml each) consisting of fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) were visualised via TLC.

To see whether inulinase activity was due to sequential action of each protein or the formation of an active complex, reactions were set up where each protein was incubated with inulin and then boiled to stop the reaction. Reactions were returned to 37 °C and incubated with the other GH91 family protein. If both enzymes formed an active complex, no activity would be seen under these conditions, however if one enzyme modified inulin for cleavage by its functional partner activity would be seen in one of these conditions. To test whether an active complex was formed, both proteins were also incubated together (**Figure 4.9 A-E**). Activity was only seen when both proteins were mixed together, supporting that activity is due to an active complex rather than sequential activity.

To explore the importance of the predicted catalytic residues in the GH91 enzyme (BACOVA_04502/3) the candidate catalytic residues were mutated to alanine, this experiment also allows us to speculate on the formation of the catalytic site between the two protein monomers. BsIFTase D233 aligned to BACOVA_04502 D185 and BsIFTase D244 aligned to BACOVA_04503 E196. Neither of the mutant proteins (BACOVA_04502 D185A, BACOVA_04503 E196A) were active when incubated with its wild-type partner (**Figure 4.9. F & G**).

Finally, the enzymatic function of the C-terminal CBM of BACOVA_04502 (377 aa-650 aa) was explored by expressing this region independently (BACOVA_04502 CTD) and by expressing a truncation (0 aa-376 aa) which lacks this region (BACOVA_04502 NTD). These constructs were probed for activity with the wild type BACOVA_04503 partner. Activity was observed with BACOVA_04502 NTD but not with BACOVA_04502 CTD indicating that this C-terminal region was not crucial for enzyme function (**Figure 4.9. H & I**).

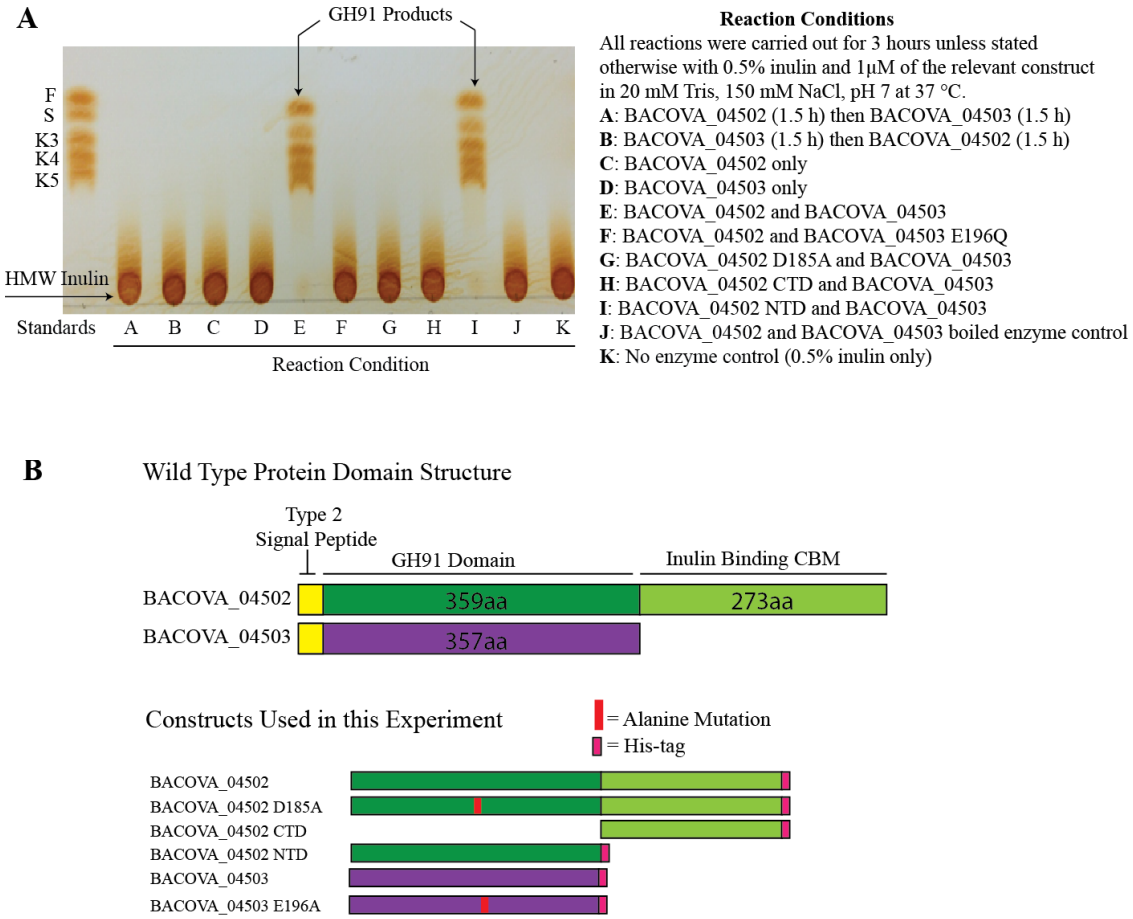


Figure 4.9. BACOVA_04502 and BACOVA_04503 form an active complex. (**A**) A TLC visualisation of several reaction conditions used to determine the relationship between the two GH91 enzymes, BACOVA_04502 and BACOVA_04503. Standards (2 mg/ml each, stock) of fructose (F), sucrose (S), kestose (K3), kestotetraose (K4) and kestopentaose (K5) were visualised for reference. A positive control showing enzyme activity (E) and two negative controls (J and K) show no activity. Reaction products are arrowed. All reactions (outlined below) contained 1 µM of the relevant construct(s), 0.5 % inulin in 20 mM Tris, 150 mM NaCl, pH 7 at 37 °C and were incubated for 3 hours unless otherwise stated. (**B**) A visual guide to the constructs used during this experiment

Reaction Condition:

A: BACOVA_04502 incubated for 1.5 hours, inactivated. BACOVA_04503 added and incubated a further 1.5 hours.
B: As 1 with BACOVA_04503 incubated prior to addition of BACOVA_04502.
C: BACOVA_04502 only.
D: BACOVA_04503 only.
E: BACOVA_04502 and BACOVA_04503.
F: BACOVA_04502 and BACOVA_04503 E196A.
G: BACOVA_04502 D185A and BACOVA_04503.
H: BACOVA_04502 CTD and BACOVA_04503.
I: BACOVA_04502 NTD and BACOVA_04503.
J: BACOVA_04502 and BACOVA_04503 boiled enzymes control.
K: No enzyme control.

An attempt was made to restore enzyme activity within homogenous enzyme reactions by reinstating the lost catalytic residue in each monomer (BACOVA_04502 D196E and BACOVA_04503 K185D) however neither of these constructs was active, suggesting that active complex formation requires specific interaction between BACOVA_04502 and BACOVA_04503 (**Figure 4.10. D & E**). We also investigated whether the mutant BACOVA_04502 D185N was also able to form an active complex, given that the equivalent mutation was not active in BsIFTase (Jung *et al.*, 2007). Asparagine is similar in structure to the aspartate it replaces, containing a nitrogen ($-NH_2$) in place of an oxygen (=O) at C_4. This substitution should form a hydrogen bond at the same place as the oxygen within the active site, and it was surprising in BsIFTase that activity was lost, though substrate recognition was retained when this mutation was made. Activity is retained in BACOVA_04502 D185N (**Figure 4.10. C**) showing that the sugar can still be co-ordinated within the active site by the asparagine in a way that activity remains possible - it remains unclear why this is not the case in BsIFTase (Jung *et al.*, 2007).

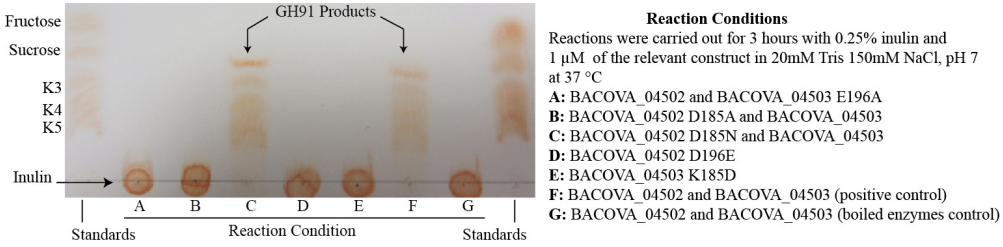


Figure 4.10. Mutant GH91 constructs reveal putative active site residues. TLC visualisation of the end-point samples from enzyme assays with different GH91 mutant constructs. All assays were performed using 1 µM of each enzyme and 0.25 % inulin in 20 mM Tris, 150 mM NaCl, pH 7 and were incubated at 37 °C for 3 hours. Standards (stock, 2mg/ml each) consisting of fructose, sucrose, kestose (K3), kestotetraose (K4) and kestopentaose (K5) were run for reference. Positive (F) and negative (G) controls were run to visualise reaction products and no reaction respectively.

Reaction Condition:
(**A**) BACOVA_04502 + BACOVA_04503 E196A.
(**B**) BACOVA_04502 D185A + BACOVA_04503.
(**C**) BACOVA_04502 D185N + BACOVA_04503.
(**D**) BACOVA_04502 D196E.
(**E**) BACOVA_04503 K185D.
(**F**) BACOVA_04502 + BACOVA_04503 positive control.
(**G**) BACOVA_04502 + BACOVA_04503 boiled enzymes control.

4.6.4. The C-Terminal CBM of BACOVA_04502 GH91 Contributes to the Rate of Endo-Inulinase Activity

The C-terminal truncation construct, BACOVA_04502 NTD, lacks the inulin binding CBM. Incubation of BACOVA_04502 NTD and BACOVA_04503 on inulin did not result in altered end-point products compared with wild-type BACOVA_04502. To determine if the CBM region contributed to enzyme activity rate the BACOVA_04502 NTD and BACOVA_04503 construct mix was assayed over a time course of 1 hour and compared to wild type (BACOVA_04502 + BACOVA_04503) activity (**Figure 4.11**).

Both wild type GH91 enzymes resulted in complete formation of end products between 15 and 30 minutes, however when BACOVA_04502 NTD was used, the reaction took longer to reach completion, between 30 minutes and 60 minutes, suggesting that the CBM may, unusually, play a role during catalysis (**Figure 4.11**). This may be explained as the CBM may bring inulin closer to the active site, or loss of the CBM region may disrupt protein-protein interactions formed by

the two monomers. The accepted paradigm for CBM activity suggests that CBM do not enhance activity rate on soluble material, however confer a distinct advantage on insoluble matter (Boraston *et al.,* 2004).

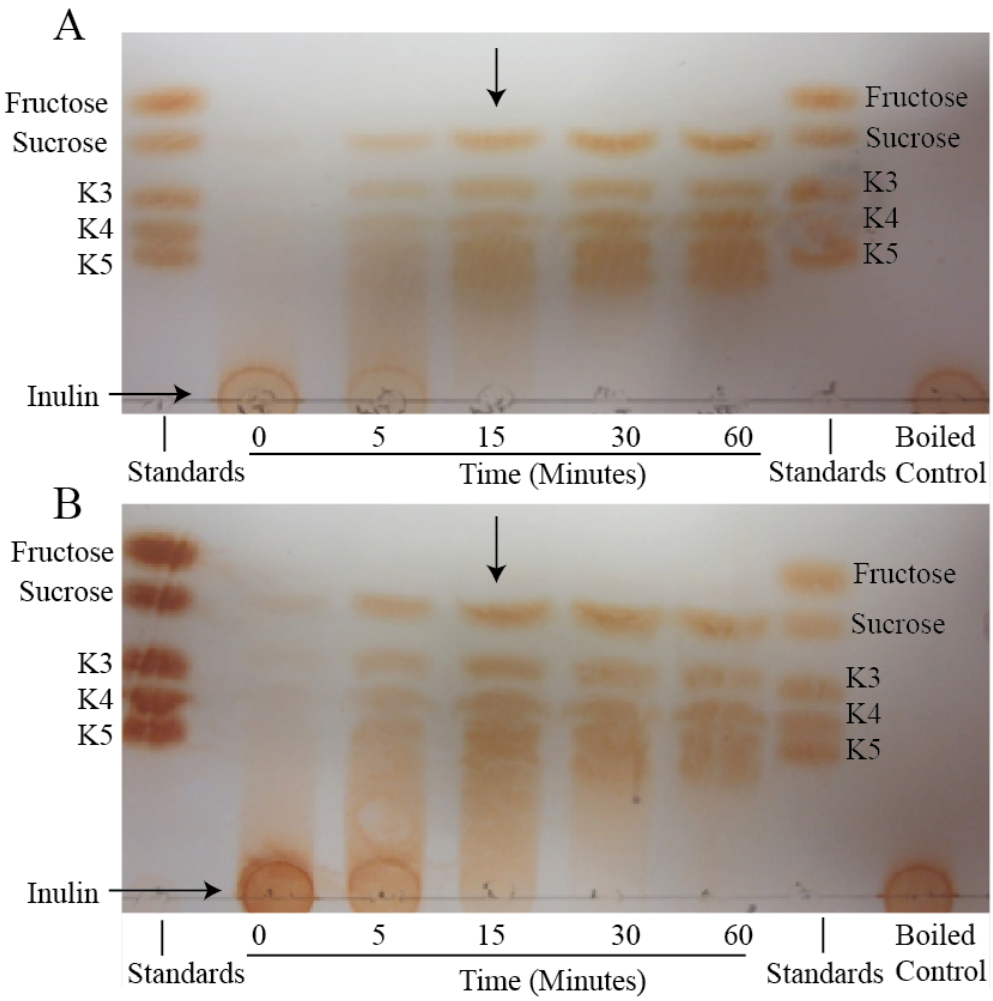


Figure 4.11. The C-terminal CBM appended to the GH91 has an effect on enzyme activity rate. Two identical enzyme reactions were set up using either 1 µM BACOVA_04502 (**A**) or BACOVA_04502 NTD (**B**) and wild-type BACOVA_04503, 0.25 % inulin in 20 mM Tris, 150 mM NaCl, pH 7. Samples were taken at 0, 5, 15, 30 and 60 minutes and visualised via TLC with the appropriate standards (stock, 2 mg/ml each) of fructose, sucrose, kestose (K3), kestotetraose (K4), kestopentaose (K5). The end time point of a boiled enzyme control was also run in each case as a negative control. At 15 minutes (arrowed) there is a clear difference between the progresses of the two reactions.

4.6.5. HPAEC-PAD Visualisation of GH91 Inulin Breakdown Products

The GH91 (BACOVA_04502/3) inulin breakdown products were visualised via HPAEC-PAD along with kestose series standards. Previous TLC data suggest these products co-migrate fairly well with the standards (see **Figure 4.8**) however when visualised with HPAEC-PAD the inulin breakdown products appear larger (**Figure 4.12**).

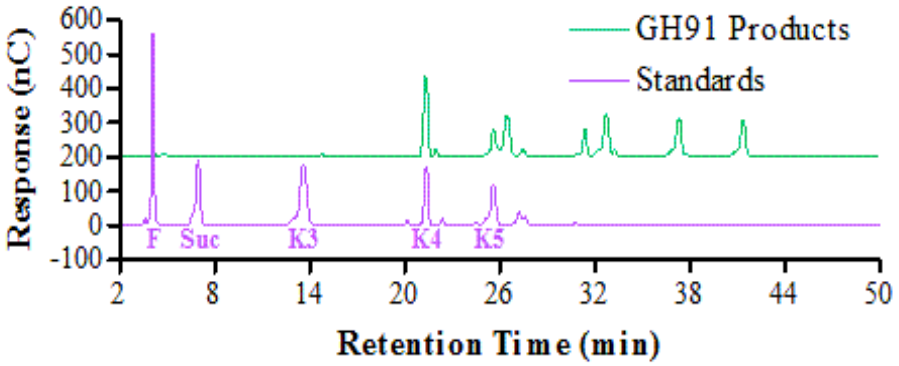


Figure 4.12. HPAEC-PAD visualisation of the GH91 products. HPAEC-PAD was used to visualise GH91 products (0.1 % w/v). Kestose-series standard consisting of 0.1 mg/ml each of fructose (F), sucrose (S), kestose (K3), kestotetroase (K4), kestopentaose (K5) were run for comparison.

4.6.6. Enzymatic Hydrolysis of GH91 Products Reveal an Unknown Compound

BsIFTase creates a difructose anhydride (DFA) product (Jung *et al.*, 2007) and we wished to investigate the possibility that the BACOVA_04502/3 GH91 enzyme created an anhydrous terminus during internal cleavage of inulin. This could also explain the discrepancy observed between the co-migration of the GH91 inulin breakdown products with the kestose series standards on TLC but not HPAEC-PAD (**Figure 4.8; Figure 4.12**).

GH91 inulin breakdown products were hydrolysed using BACOVA_04501, a GH32 enzyme isolated from the *B. ovatus* PUL (Chapter 4.8). GH91 products (from 2.5 % inulin) were incubated with 1 µM BACOVA_04501 for 3 hours. HPAEC-PAD visualisation of the hydrolysed products revealed three products; fructose and glucose and an unknown sugar (**Figure 4.13**). This unknown sugar was suspected to be DFA.

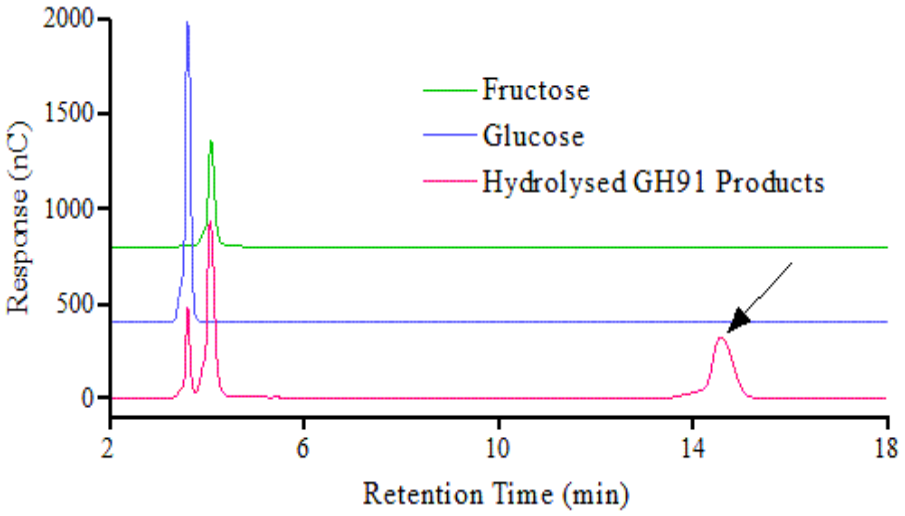


Figure 4.13. Hydrolysis of GH91 inulin breakdown products by an exo-fructosidase. HPAEC-PAD was used to visualise the three end-products created by digesting GH91 products using am exo-fructosidase from the PUL (BACOVA_04501). Two of the three products co-migrate well with known monosaccharide standards (0.125 mM each of fructose and glucose). The third product (arrowed) did not co-migrate with either of the standards used.

4.6.7. DFA-FOS is a Product of the GH91 Enzyme BACOVA_04502/3

4.6.7.1. MALDI-TOF Mass Spectrometry

The GH32 hydrolysed GH91 products were analysed using MALDI-TOF mass spectrometry (Chapter 2.17.5), in which the samples were calibrated using peaks of known mass present in the matrix. Several datasets were collected to ensure peaks detected were reproducible and not machine artefacts (**Figure 4.14**).

The hydrolysed GH91 products sample (containing glucose, fructose and unknown sugar) detected a compound at 347.1 m/z (**Figure 4.14 A**). A control sample was prepared in the same manner by hydrolysing inulin directly with the GH32, BACOVA_04501 (i.e. without GH91 treatment, thus containing only fructose and glucose) and in this instance the peak at 347.1 m/z was not detected (**Figure 4.14 B**). Sugars often form lithium adducts, and in the presence of lithium a new peak was detected at 331.1 m/z, which was not present in the control sample (**Figure 4.14 C**). The shift in mass in the presence of lithium was consistent with a sugar forming lithium adducts. Lithium has a known mass of 6.941, therefore the mass of the unknown sugar could be calculated as 324.2, and the peak detected at 347.1 m/z in the absence of lithium is congruent with a sodium adduct of the same size (324.2 + 23 = 347.2). As the GH91 enzyme is an endo-inulinase (Chapter 4.6.3), glucose was ruled out as a component of this compound. Fructose has a molecular weight (MW) of 180.16, upon glycosidic bond formation a water molecule (MW = 18.015) is lost. Therefore a fructose disaccharide with one glycosidic bond (180.16 x 2 – 18.015) would produce a peak at 342.3 m/z plus adduct mass and a fructose disaccharide with two glycosidic bonds (180.16 x 2 – 18.015 x 2) would induce a peak at 324.3 m/z plus adduct mass. An error of 0.1 m/z is well within normal detection error of this technique and thus the unknown sugar was concluded to be a difructose-anhydride with a retained β2-1 linkage and a secondary glycosidic bond of unknown linkage. This is consistent with products made by other members of the GH91 family.

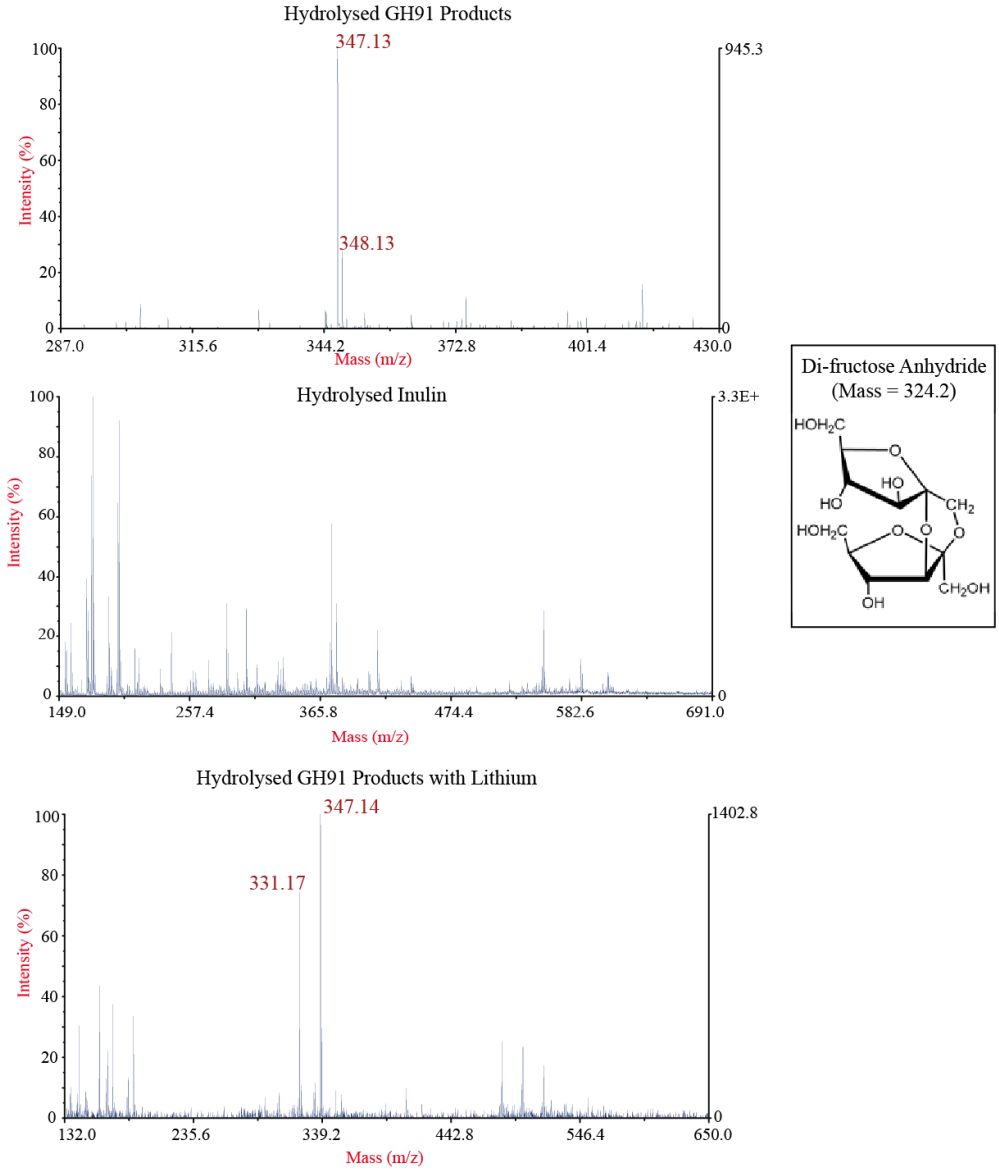


Figure 4.14. MS of GH91 inulin breakdown products after hydrolysis by BACOVA_04501 GH32 exo-fructosidase. (**A**) Hydrolysed GH91 inulin breakdown products (glucose, fructose, unknown) yield a peak at 347.1 m/z which is not seen in (**B**), the hydrolysed inulin (glucose, fructose) control. (**C**) Hydrolysed GH91 products in the presence of lithium result in a lithium adduct of the unknown sugar with a mass of 331.17 m/z. Given that lithium has a mass of 6.94 m/z, the unknown sugar has a mass of 324.2 m/z. The peak at 347.1 m/z is an adduct of the unknown sugar and sodium (23 m/z). Based upon the predicted mechanism of GH91 family enzymes the unknown sugar therefore was identified as difructose-anhydride, the structure of DFA is shown (inset).

4.6.7.2. DFA Accumulates in Cultures of *Bacteroides ovatus* Grown on Inulin

It was suspected that the sugar which accumulates over time in the culture medium during growth of *B. ovatus* on inulin (**Figure 4.1**) was the terminal DFA moiety produced by the GH91 enzyme BACOVA_04502/3 during inulin breakdown. To demonstrate this, hydrolysed GH91 products containing DFA and spent culture supernatant fluid were visualised by HPAEC-PAD (**Figure 4.15**). The peak identified as DFA during MS (**Figure 4.14**) co-migrates with the peak detected in the culture supernatant fluid, which is strongly supportive of the presence of DFA. The GH91 enzyme (BACOVA_04502/3) produces a range of DFA-terminated oligosaccharides, these data demonstrate that whilst the majority of the breakdown products can be utilised, this DFA moiety is not fermented and is released into the media where it accumulates.

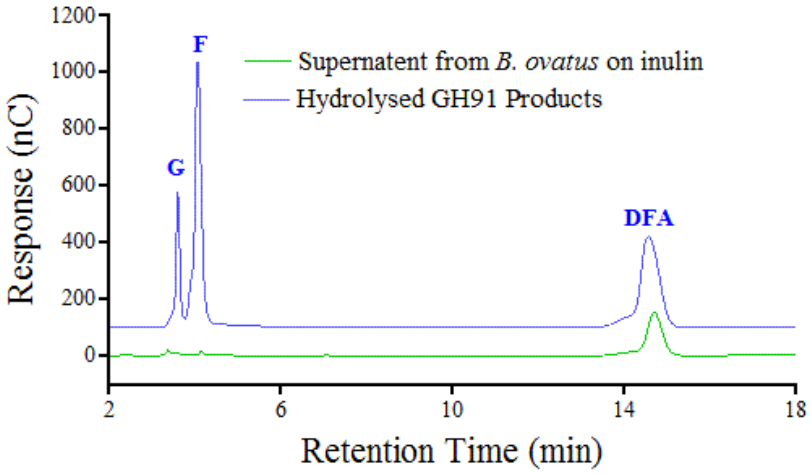


Figure 4.15. DFA produced by the GH91 enzyme is released into the culture medium. Hydrolysed GH91 products contain glucose (G), fructose (F) and difructose anhydride (DFA). The known DFA from this sample co-migrates during HPAEC-PAD visualisation with the unknown sugar from supernatant fluid harvested from late stationary phase *B. ovatus* cultures (minimal media + 0.5 % inulin, e.g. such as shown in Figure 4.1 at approx. 24 hours). This demonstrates that the unknown sugar within the culture media is DFA produced by the GH91 enzyme during inulin breakdown.

4.6.8. Analysing the Interaction between the BACOVA_04502 and BACOVA_04503 GH91 Monomers

BsIFTase is a homotrimer with a functional catalytic site at the interface between each monomer, forming three catalytic sites (Jung *et al.*, 2007). It is unclear whether BACOVA_04502/3 is also a trimer. As each subunit contributes a catalytic residue (**Figure 4.10**) only one functional catalytic site would be created in either a heterodimer or heterotrimer conformation, therefore, it seems likely that BACOVA_04502/3 would form a dimer.

To determine both the ratio of subunits within the functional protein, and to characterise the interaction between the two proteins a range of techniques were used including affinity gel electrophoresis (AGE), gel filtration and ITC (**Figure 4.16**). Where possible (AGE, ITC) experiments were conducted in the presence and absence of inulin, in case binding is mediated by substrate; additionally, experiments were conducted with wild-type protein or with one of the two catalytic mutants (BACOVA_04502 D185A, BACOVA_04503 E196A) to prevent catalytic activity interfering with binding kinetics. In all cases no interaction between the two protein monomers could be detected.

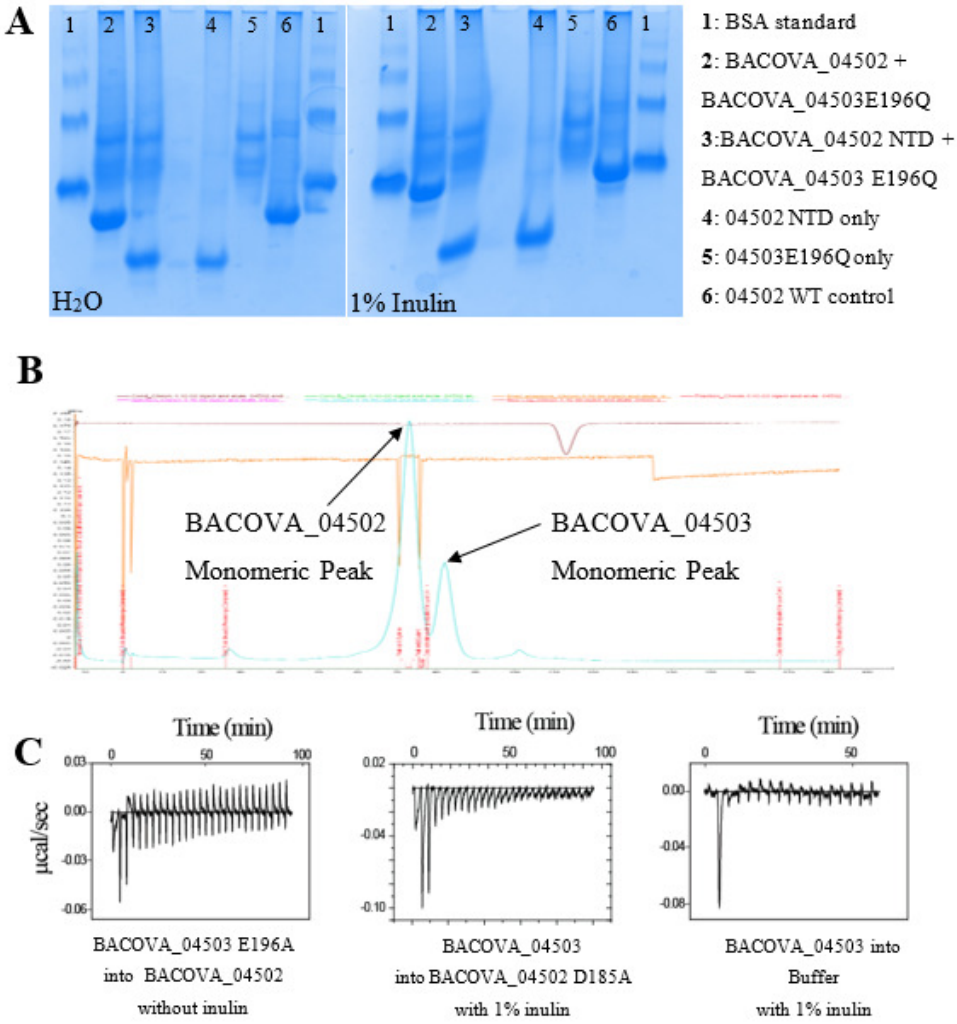


Figure 4.16. Analysis of the interaction between BACOVA_04502 and BACOVA_04503 GH91 proteins. (**A**) When BACOVA_04502 (6 and 2) is loaded the band is retarded on inulin (right) compared with water (left) due to the interaction of the inulin binding CBM with inulin. To monitor other interactions the CTD truncation, BACOVA_04502 NTD was used in conjunction with the BACOVA_04503 catalytic mutant, BACOVA_04503 E196Q (3) so that no inulin breakdown could occur. The proteins were run individually (4, 5) and BSA (1) was also run as a control. AGE analysis showed no difference in band pattern in any of the conditions tried, therefore there was no evidence for oligomerisation. (**B**) Gel filtration chromatograms showed only two peaks corresponding to BACOVA_04502 and BACOVA_04503 in monomeric state. Full gel filtration data is available in Appendix I (Figure I.2 and I.3). (**C**) ITC analysis. The control (BACOVA_04503 into buffer) demonstrates a dilution event is occurring, and no evidence for binding was seen.

4.6.9. Phyre2 Modelling Allows Insight into GH91 Structures

Despite extensive screening, no crystals were obtained for either GH91 protein, both individually and together. To provide insight into the possible structure adopted by the GH91 oligomer (BACOVA_04502/3) we modelled the structure of each monomer using BsIFTase as a template using Phyre2, a program to predict the 3D structures of proteins with structural homologues (Kelly *et al.*, 2015). BACOVA_04502 NTD and BACOVA_04503 models were aligned with the subunits from BsIFTase and the ligand bound with BsIFTase, fructobiose, was retained (**Figure 4.17**). A long dimerization interface joins the two subunits together, and there is a wide cleft containing the two catalytic residues and the ligand modelled with BsIFTase. BsIFTase is exo-acting and the catalytic site is within a pocket which forms interactions around the chain terminus of inulin (Jung *et al*., 2007; **Figure 4.17 A**). As BACOVA_04502/3 is endo-acting and binds to inulin internally on the polysaccharide chain, the active site is predicted to adopt a more cleft-like conformation, with space for the polysaccharide to extend in both directions; the BACOVA_04502/3 model shows a space around the catalytic site which is not apparent in BsIFTase, supporting the requirements for endo-activity. In the model the two known catalytic residues are close to each other (~11Å), which also fits what is expected from the *B. ovatus* GH91 enzyme.

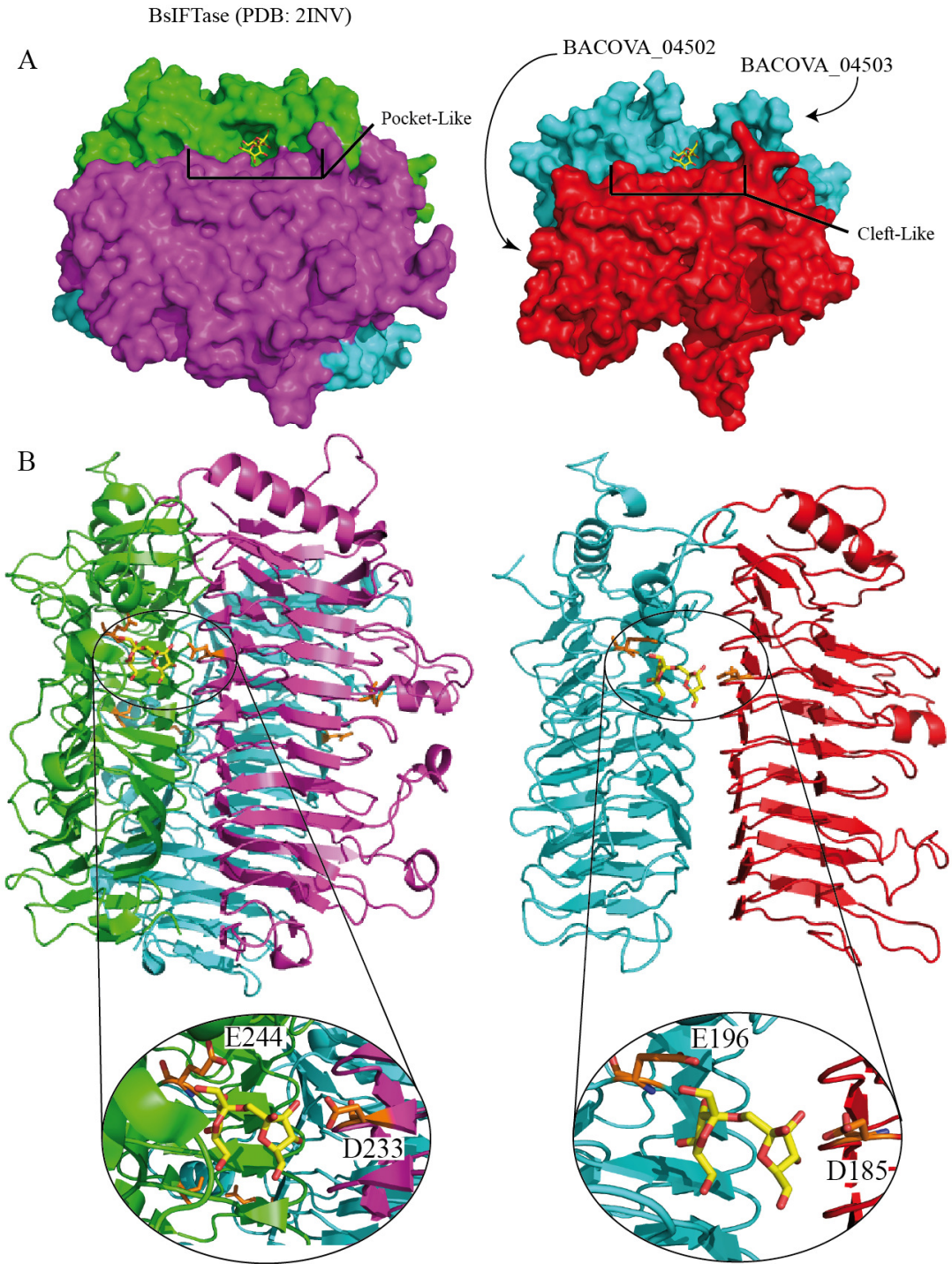


Figure 4.17. Phyre2 predictions of BACOVA_04502/3 structure, using BsIFTase (2INV) as a model. The structure of BsIFTase (2INV) compared with the Phyre2 predictions for the structure of BACOVA_04502/3 aligned into a putative dimer. Fructobiose (yellow sticks) from 2INV has been modelled in to the Phyre2 predictions. (**A**) Surface representation highlighting the binding pocket of 2INV vs. the cleft predicted in the Phyre2 model. (**B**) The proteins in cartoon representation, with a close up of fructobiose and the two catalytic residues (orange) within the active site.

4.6.10. The GH91 Enzyme is not Required for Growth on Inulin or FOS

To explore the role of the GH91 heteromeric enzymes (BACOVA_04502/3), two strains were made to disrupt catalysis; *Δbacova_04502* and *Δbacova_04503 E196Q*. These were made as described in (Chapter 2.16) and are a complete knockout of *bacova_04502* (one of the GH91 monomers) and a catalytic mutant of the full enzyme, respectively. Neither of these mutants displayed surface endo-inulinase activity, confirming that the two gene products are required for surface endo-activity and spent supernatants from neither mutant contains DFA (**Figure 4.27; Figure 4.18**).

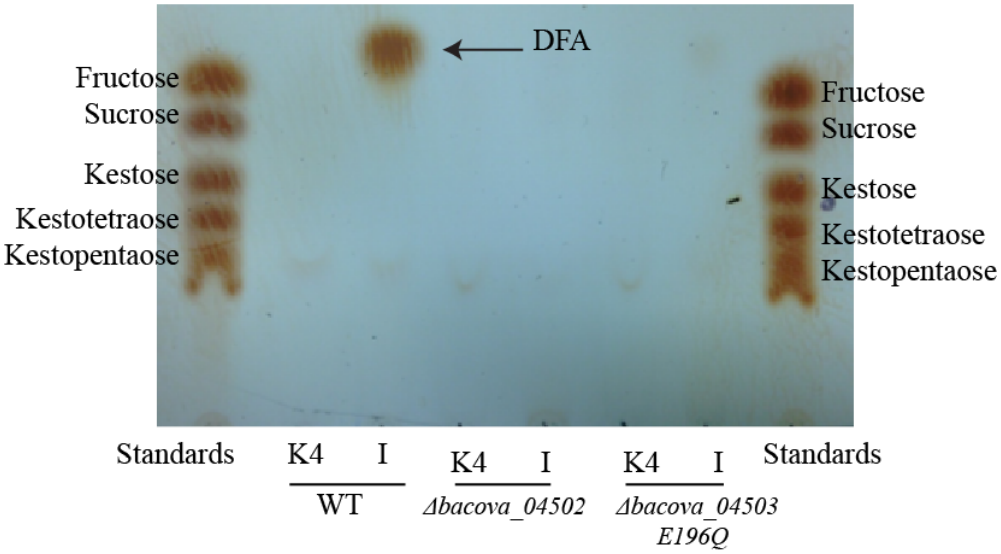


Figure 4.18. Spent culture supernatant fluid analysis of GH91 mutants. *B. ovatus* wild type (WT), *Δbacova_04502* and *Δbacova_04503 E196Q* strains were grown on kestotetraose (K4) or inulin for 24 hours until late stationary phase (OD_{600nm} > 2). Supernatant fluid from each culture was harvested by centrifugation. Wild type *B. ovatus* produces DFA (arrowed) as a waste product from inulin however DFA is not produced during growth on K4 or by the GH91 genetic mutant strains *Δbacova_04502* and *Δbacova_04503 E196Q*. These data show that the GH91 is inactive in *Δbacova_04502* and *Δbacova_04503 E196Q* and on short FOS such as kestotetose.

Both GH91 mutant strains grew on all substrates similar to wild type (**Figure 4.19**). These data show that the GH91 is not essential for growth on any of the fructan substrates tested.

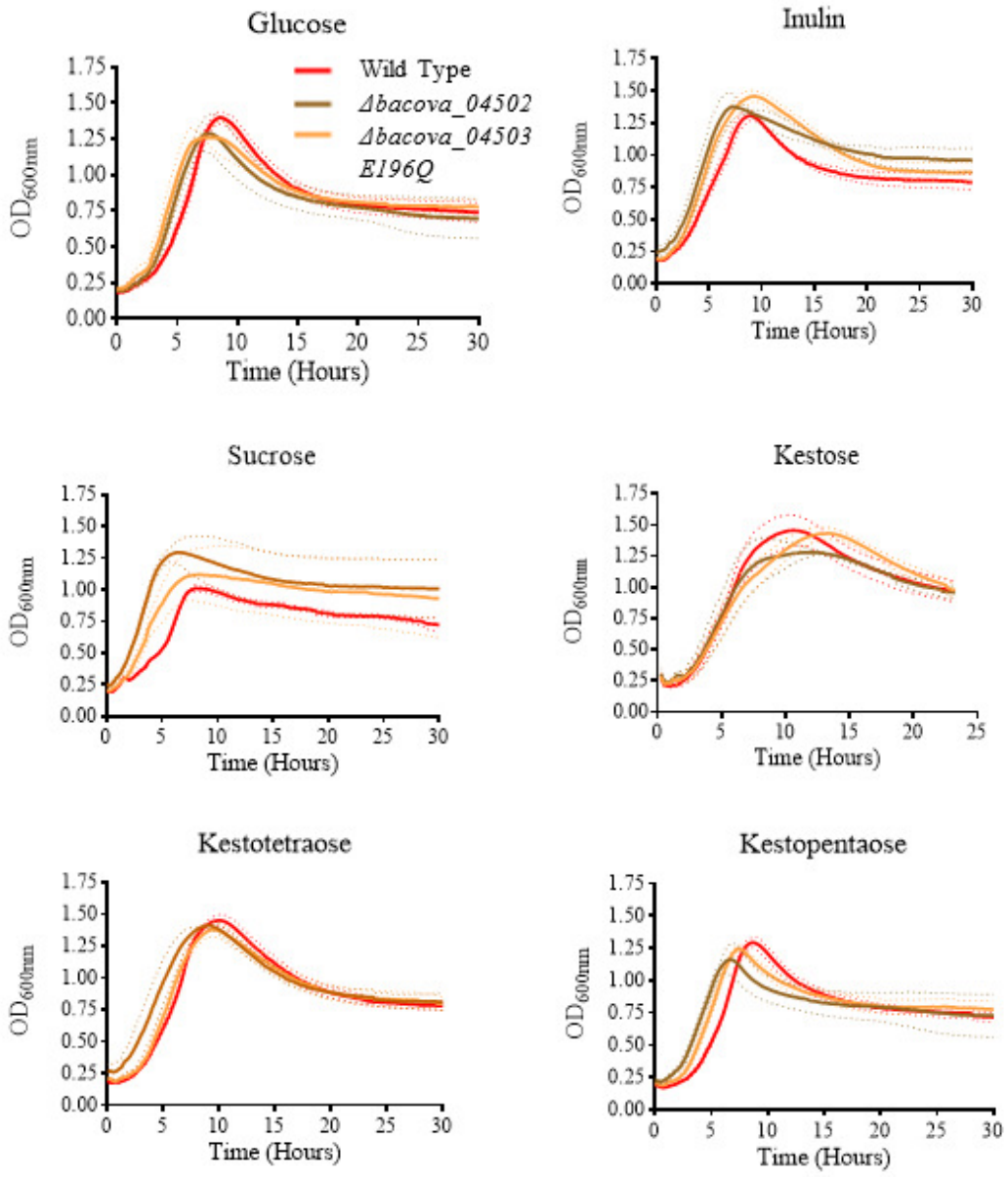


Figure 4.19. Disruption of the GH91 enzyme does not result in a growth defect on any of the substrates tested. The growth of *Δbacova_04502* and *Δbacova_04503 E196Q* was monitored in comparison to wild type, these two strains lack active GH91 surface endoactivity. All strains were grown on minimal media supplemented with 0.5 % of either inulin, glucose, sucrose, kestose, kestetraose or kestopentaose. No growth defects were observed in either of these mutants. Multiple datasets were obtained for each condition (2-4) and representative data performed in triplicate are shown with continuous error bars.

4.7. Unprocessed Inulin and FOS are Targeted for Import by a SusD-homologue

4.7.1. The *Bacteroides ovatus* PUL Contains a SusD-homologue, BACOVA_04504

Other characterised glycan utilisation systems, including the canonical SUS system import glycans into the periplasmic space through a SusC/D-homologue pair. The SusD-homologue fulfils a binding role and typically targets the breakdown products produced by extracellular enzymes (Sonnenburg *et al*., 2010; Rogowski *et al*., 2015; Cuskin *et al*., 2015; Larsbrink *et al*., 2014; Koropatkin *et al*., 2008).

A SusD-homologue, BACOVA_04504, is encoded within the *B. ovatus* inulin PUL. This protein has sequence identity to other SusD family proteins, including the SusD, with which it retains 24 % sequence identity.

The SusD is known to localise to the cell surface (Shipman *et al*., 2000). All characterised SusD-homologues contain a conserved type II signal peptide and cysteine which is anchored to a lipid within the membrane after signal sequence cleavage. All characterised SusD-homologues are thought to localise to the cell surface (Sonnenburg *et al*., 2010; Rogowski *et al*., 2015; Cuskin *et al*., 2015; Larsbrink *et al*., 2014; Koropatkin *et al.,* 2008); as BACOVA_04504 retains the signal sequence and cysteine it is also thought to be a surface localised lipoprotein.

BACOVA_04504 was cloned and expressed using recombinant *E. coli* as described (Chapters 2.9.7 and 2.11). Proteins were purified using IMAC or GST-purification (Chapter 2.12). Further construct details are listed (Appendix I.8).

4.7.2. The SusD-homologue, BACOVA_04504, Binds to Inulin, FOS and Sucrose

ITC analysis revealed that the SusD-homologue, BACOVA_04504, bound to inulin, FOS and sucrose with an affinity (K_d ~0.05 – 0.5 mM) similar to reported affinities of other SusD-homologues to their target substrates (Koropatkin *et al.*, 2008; Cuskin *et al.*, 2015; Sonnenburg *et al.*, 2010). As no significant increase in affinity was associated with increasing chain length, BACOVA_04504 was predicted to bind to a chain terminus. BACOVA_04504 did not bind to free fructose or glucose. Full binding parameters of this protein with the ligands described are displayed (**Table 4.1.**) and representative ITC traces for binding to each ligand are shown (**Figure 4.20**).

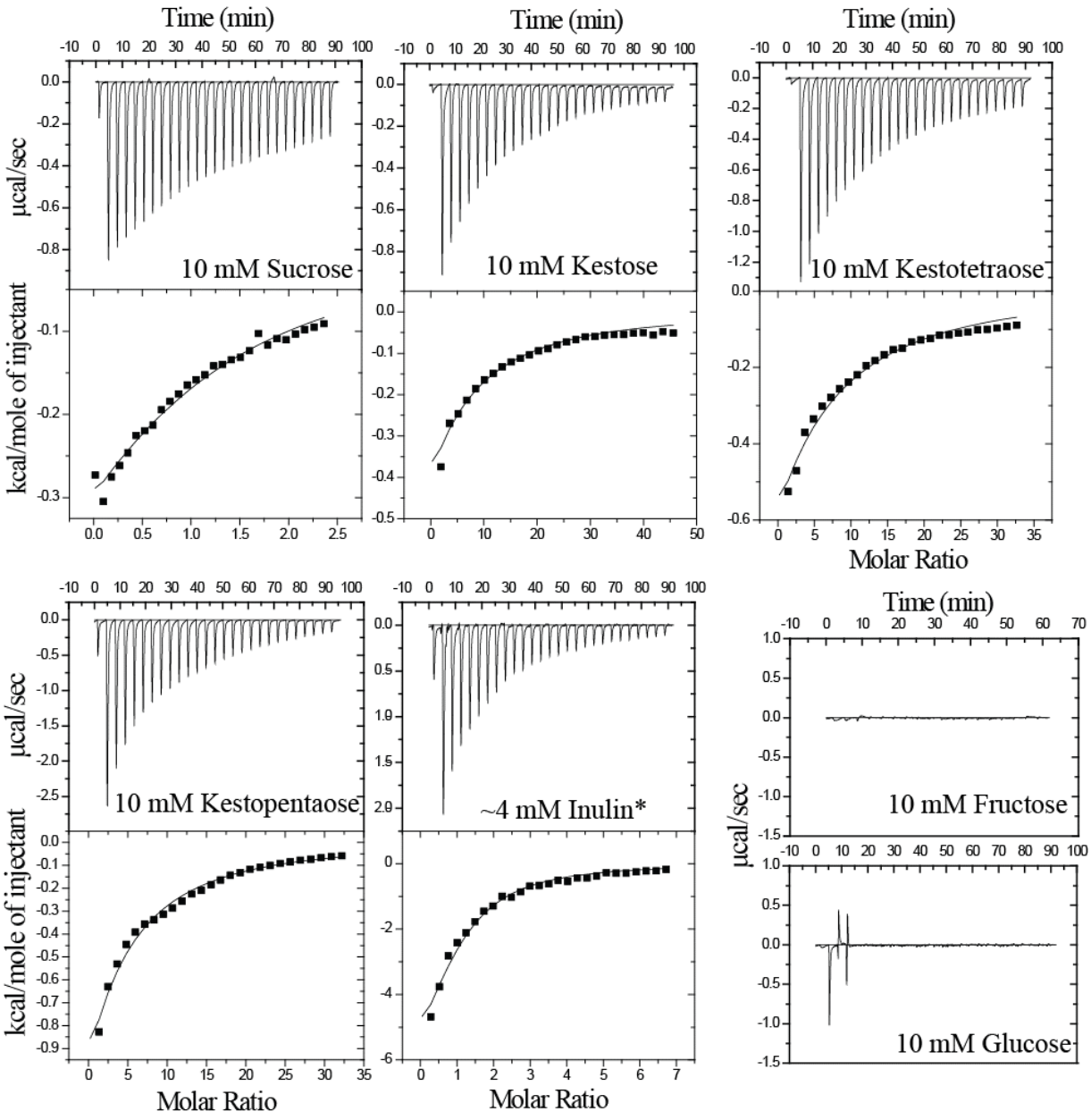


Figure 4.20. ITC traces showing the SusD-homologue, BACOVA_04504, binding to fructan derived ligands. Representative ITC binding traces of the SusD-homologue, BACOVA_04504, with β2-1 linked fructan oligosaccharides and inulin. Binding was not observed with fructose or glucose. Where binding was observed the data were fitted to a single site model shown in the lower panels. Multiple runs for each condition were undertaken and each graph shown is representative within this set of repeats. 50-60 µM of protein was used during each assay. The buffer was 20 mM HEPES, pH 7.0.
*Concentration of inulin was estimated as outlined in Chapter 2.13.2

Table 4.1. ITC data for the SusD-homologue, BACOVA_04504, binding to inulin, FOS and sucrose.

Ligand	N	K_a x 10^3 (M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	TΔS (kcal mol^{-1})
Sucrose	1.0 ± 0.1	0.8 ± 0.3	-3.9	-9 ± 8.2	-5.1
Kestose	1.0 ± 0.1	0.9 ± 0.3	-4.0	-7.1 ± 1.6	-3.1
Kesto-tetraose	1.1 ± 0.1	1.2 ± 0.4	-4.2	-10.7 ±7.6	-6.5
Kestopentaose	1.0 ± 0.3	1.2 0.4	-4.2	-20 ±9.4	-15.8
Inulin*	1.0 ± 0.1	4.3 ± 0.6	-5.0	-7.8 ± 1.1	-2.8

Multiple (2-6) datasets were used to determine binding parameters for each ligand, average values ± standard deviation (SD) are shown.
*The concentration of inulin was estimated as described (Chapter 2.13.2)

4.7.3. The SusD-homologue, BACOVA_04504, Recognises a Sucrose Chain Terminus

Experiments were designed to determine which end of the fructan chain the SusD-homologue, BACOVA_04504, recognised. Unprocessed inulin has two non-reducing termini, as it is capped by a sucrose moiety at what would otherwise be the reducing end of the polysaccharide. Breakdown of the inulin chain through acid hydrolysis will result in the presence of reducing fructose ends as the sucrose moiety is lost. Additionally, the GH91 enzyme (BACOVA_04502/3) creates a DFA terminus during breakdown (Chapter 4.6.7). Therefore, four candidate chain end types for inulin exist; reducing fructose and non-reducing fructose, sucrose and DFA termini.

To increase the molar concentration of both reducing and non-reducing fructose ends inulin was treated with acid (partial hydrolysis).

As other SusD-homologues recognise extracellular breakdown products (Rogowski *et al.*, 2015; Cuskin *et al.*, 2015; Koropatkin *et al.*, 2008) we hypothesised that the SusD-homologue (BACOVA_04504) may recognise GH91 inulin breakdown products which contain an increase in the molar ratio of both DFA and non-reducing fructose ends.

If BACOVA_04504 recognises any of these three end types (reducing fructose, non-reducing fructose or DFA), saturation of the assay will occur more rapidly in one or both of the treated inulin conditions as more binding targets per weight of substrate are available. We demonstrate this effect using a characterised GH32 mutant (BT3082) which binds to non-reducing fructose ends. BT3082 saturates far more quickly on both GH91 products and acid hydrolysed inulin, as the number of non-reducing fructose ends available has substantially increased (**Figure 4.21 A**). BACOVA_04504 did not saturate more quickly under any of these conditions compared to untreated inulin (**Figure 4.21 B**) leading to the conclusion that neither the non-reducing fructose, reducing fructose and DFA terminus was recognised by BACOVA_04504.

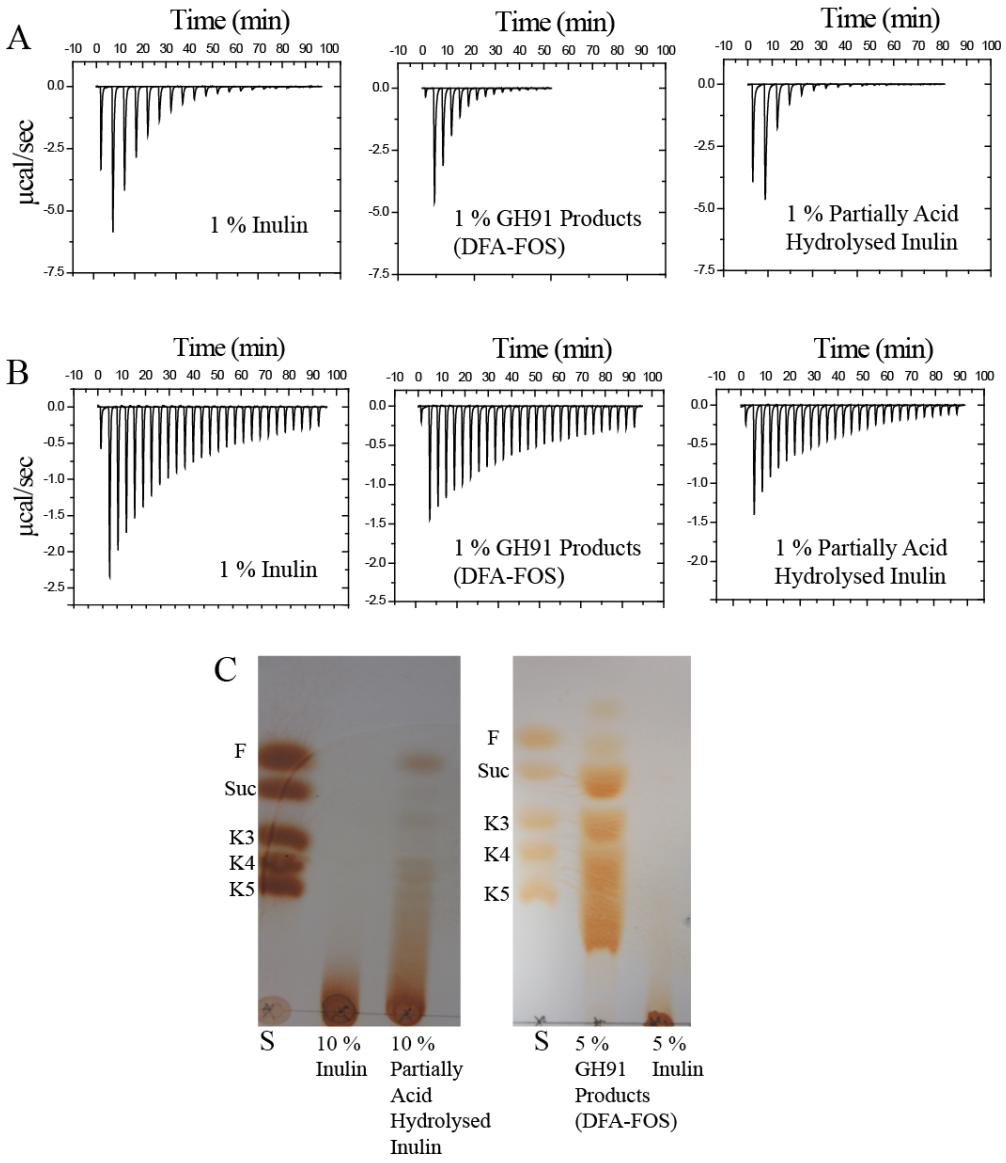


Figure 4.21. ITC reveals the SusD-homologue, BACOVA_04504, has a similar affinity for untreated and treated inulin. (**A**) BT3082 recognises non-reducing fructose ends and the titration of this protein into either GH91 products (DFA-FOS) or partially acid hydrolysed inulin saturates far more quickly than on unprocessed inulin. (**B**) The SusD-homologue bound in a similar manner to untreated inulin polysaccharide, partially acid hydrolysed inulin and inulin treated with the GH91 enzyme, though reduced heats indicates less available substrate. Saturation was not achieved more rapidly in either treatment condition suggesting that BACOVA_04504 did not recognise DFA or fructose ends. (**C**) Two TLC plates show inulin and partially acid hydrolysed inulin (left) and GH91 products and inulin (right). Both TLC plates contain standards of fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) for comparison. ITC experiments were conducted in 20 mM HEPES, pH 7 with 50-60 µM of protein. Representative ITC binding traces for each condition are shown.

To confirm that the SusD-homologue recognised the sucrose we compared binding of BACOVA_04504 to Kestose and Fructotriose, which contain and lack the sucrose moiety respectively (**Figure 4.22 A**). BACOVA_04504 bound kestose (K_d ~1.2 mM), and bound to fructotriose with a significant drop in affinity, the binding was too weak to fit to a model (**Figure 4.22 B**). These data allow us to conclude that BACOVA_04504 shows a strong preference for the sucrose terminal present on unprocessed inulin.

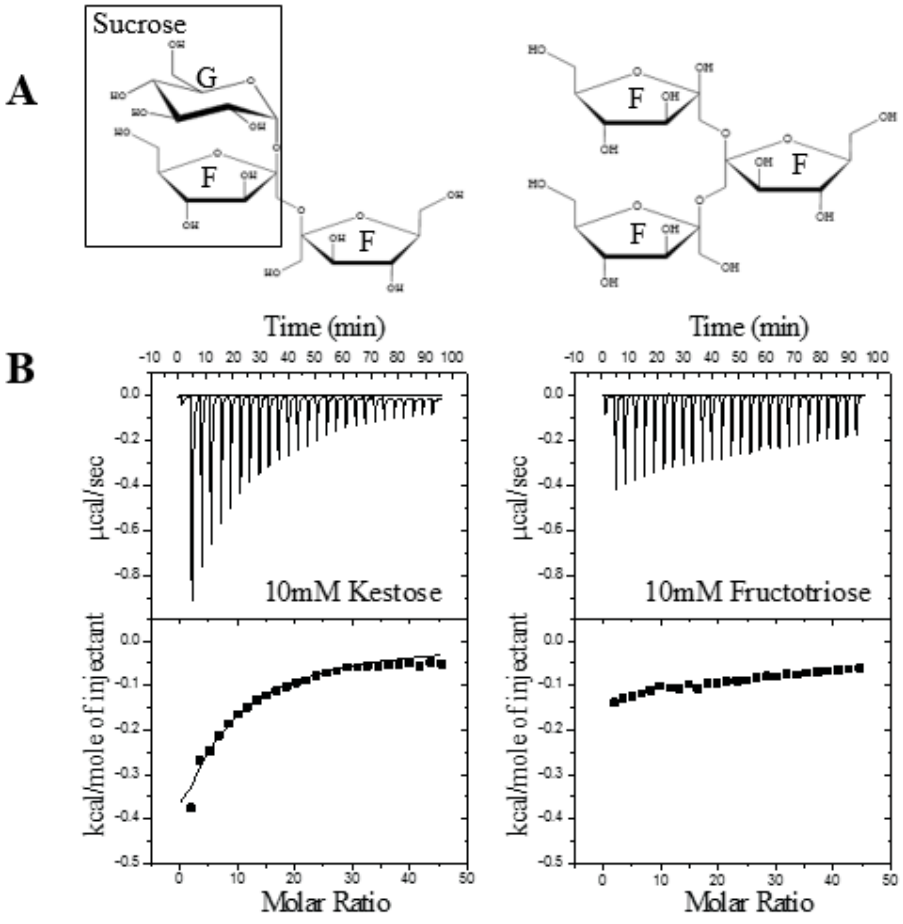


Figure 4.22. The SusD-homologue, BACOVA_04504, binds to kestose but not fructotriose. (**A**) The chemical structures of kestose (left) and fructotriose (right) are pictured. (**B**) ITC analysis reveals that BACOVA_04504 binds to kestose, but not fructotriose, indicating that a terminal sucrose motif is recognised. Binding was too weak to fit for fructotriose. Multiple datasets (3-5) were obtained, shown are representative datasets. Experiments were conducted in 20 mM HEPES, pH 7.0, using 50 µM protein.

4.7.4. X-Ray Crystallography Reveals the Structure of BACOVA_04504

BACOVA_04504 was expressed without the type II signal peptide in *E. coli* using PGEX 6P-1 and the GST tag removed for crystal trials. Initial screening was unsuccessful. The N-terminus of the crystal structure of the SusD is not visible and therefore may be disordered (Koropatkin *et al.,* 2008). Screening with various N-terminal BACOVA_04504 truncations (**Figure 4.23**) was therefore undertaken to remove any disordered region which may be preventing crystal formation.

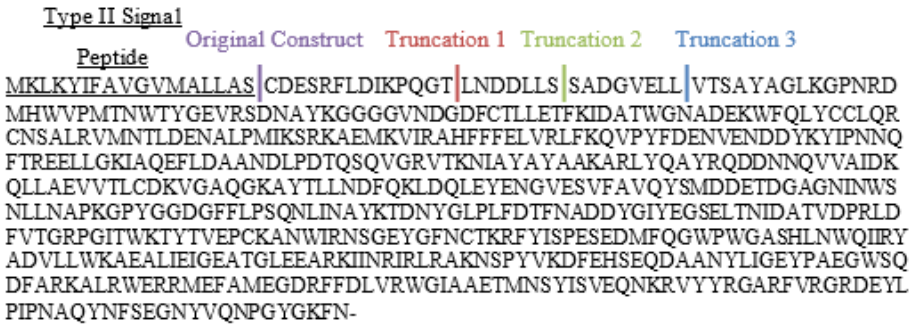


Figure 4.23. BACOVA_04504 truncations made for crystal screening. Screening with the original construct (purple) was unsuccessful. The structure of SusD-homologues, including SusD (PDB: 3CKB) did not have visible N-terminal regions, suggesting this area may be flexible. Flexible regions may hamper crystal lattice formation, therefore three new constructs with truncated N-terminal regions were made (red, green and blue) to reduce this potential flexible region and improve crystal formation.

Truncation 1 (31-585 aa) and truncation 2 (39-585 aa) did not yield crystals, but truncation 3 yielded crystals in several conditions and the structure was solved as outlined (Chapter 2.13.3).

Good candidate crystals were passed to Dr Arnaud Baslé who harvested crystals, collected datasets, solved the phase problem and fit the structural model. Full crystal model data from Dr Arnaud Baslé is shown (Appendix I, **Table I.4** and **I.5**).

The BACOVA_04504 structure (**Figure 4.24**) is highly similar to other SusD-homologue structures, including SusD itself. BACOVA_04504 is a monomer with an α-helical fold. It contains 21 α-helices and 3 pairs of anti-parallel β-strands.

The SusD-homologue retains the four tetratricopeptide repeat (TPR) domains found in SusD and other SusD-homologues (Koropatkin *et al.*, 2008, Koropatkin *et al.*, 2009). These are formed from eight of the helices in pairs: α1 (43-53 aa), α3 (104-130 aa), α4 (138-162 aa), α5 (185-204 aa), α6 (215-231 aa), α7 (244-258 aa), α15 (434-449 aa) and α16 (450-470 aa). These units form a right handed super helix along one side of the structure, another helix loop helix α17 (498-515 aa) and α18 (519-528 aa) is packed against this super helix, connecting it to the rest of the structure.

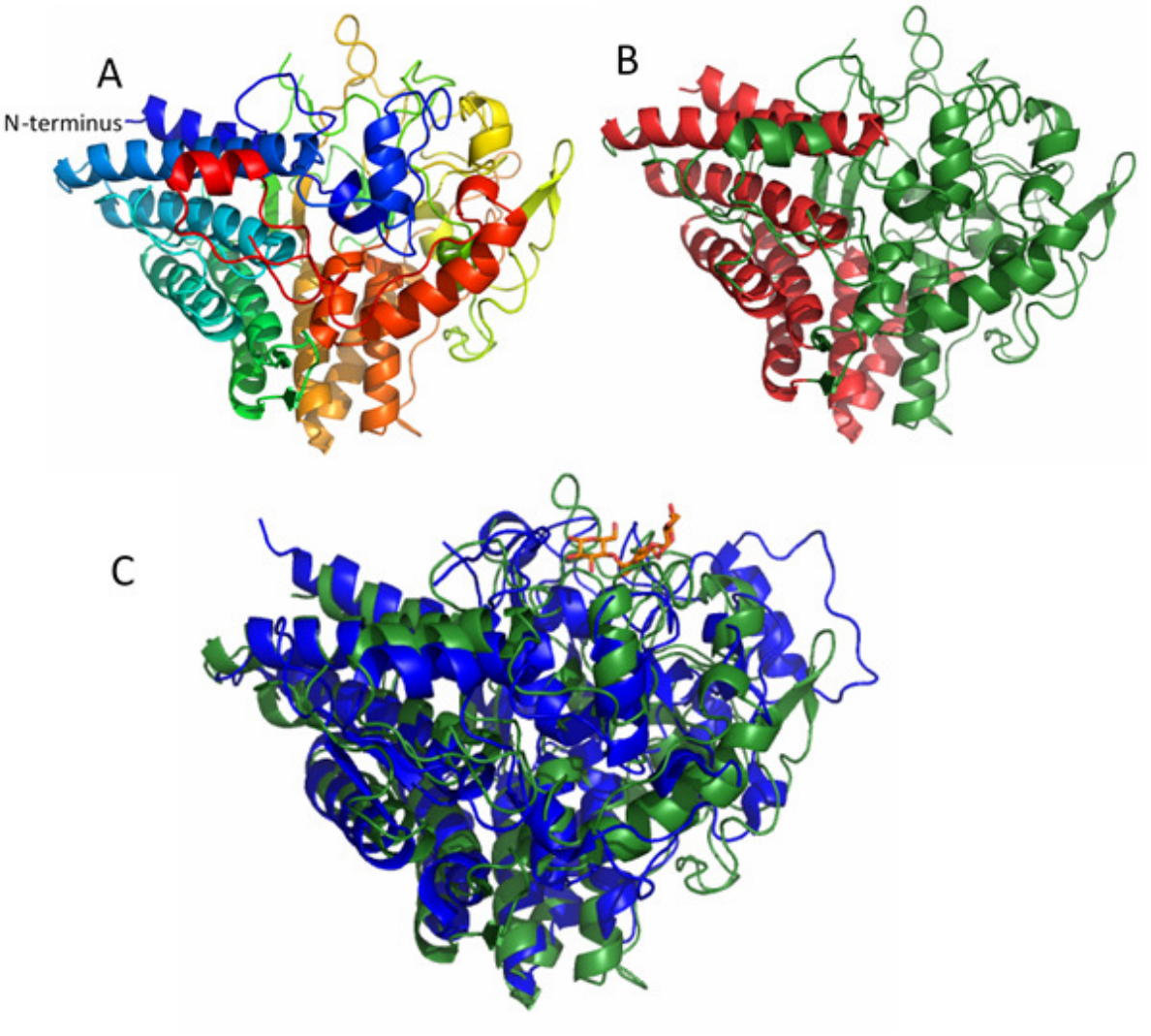


Figure 4.24. The structure of the SusD-homologue, BACOVA_04504. The structure of BACOVA_04504 was solved through x-ray crystallography. (**A**) Cartoon representation of BACOVA_04504 coloured from N-terminus (Labelled, Blue) to C-terminus (Red). (**B**) BACOVA_04504 is shown as a cartoon in green, with TPR helices coloured in red (**C**) The structure of BACOVA_04504 in green is overlaid with SusD, BT3701, (PDB ID 3CKB) shown in blue. The ligand from 3CKB, maltotriose, was retained in the overlay and is shown as orange sticks. We predicted the binding site in BACOVA_04504 is in a similar location as SusD.

4.7.5. Targeted Mutagenesis Reveals the Binding Site of BACOVA_04504 SusD-homologue

Despite several crystallisation screens in the presence of sucrose or kestose no crystals could be obtained with ligand. Therefore, to determine the location of the binding site putative ligand binding residues were mutated to alanine and the capacity of each mutant to bind to inulin was assessed.

Structural alignment of the SusD (BT3701; PDB ID 3CK9) to the inulin binding SusD-homologue, BACOVA_04504, revealed a similar topology (**Figure 4.25**). A putative binding site, similar in overall structure to SusD, was present. Several candidate aromatic residues and residues which may form hydrogen bonds with the sucrose terminal of inulin were identified in the BACOVA_04504 structure. These residues are highlighted (**Figure 4.26 A**) and mutant constructs were made whereby each annotated residue was changed to an alanine. Of the BACOVA_04504 mutants created D88A and R394A did not express; F321A, W420A and W22A bound inulin with a wild-type like affinity (K_d = ~0.4 mM), and three mutants, F321A, W420A and W422A resulted in protein which no longer bound inulin (**Figure 4.26**).

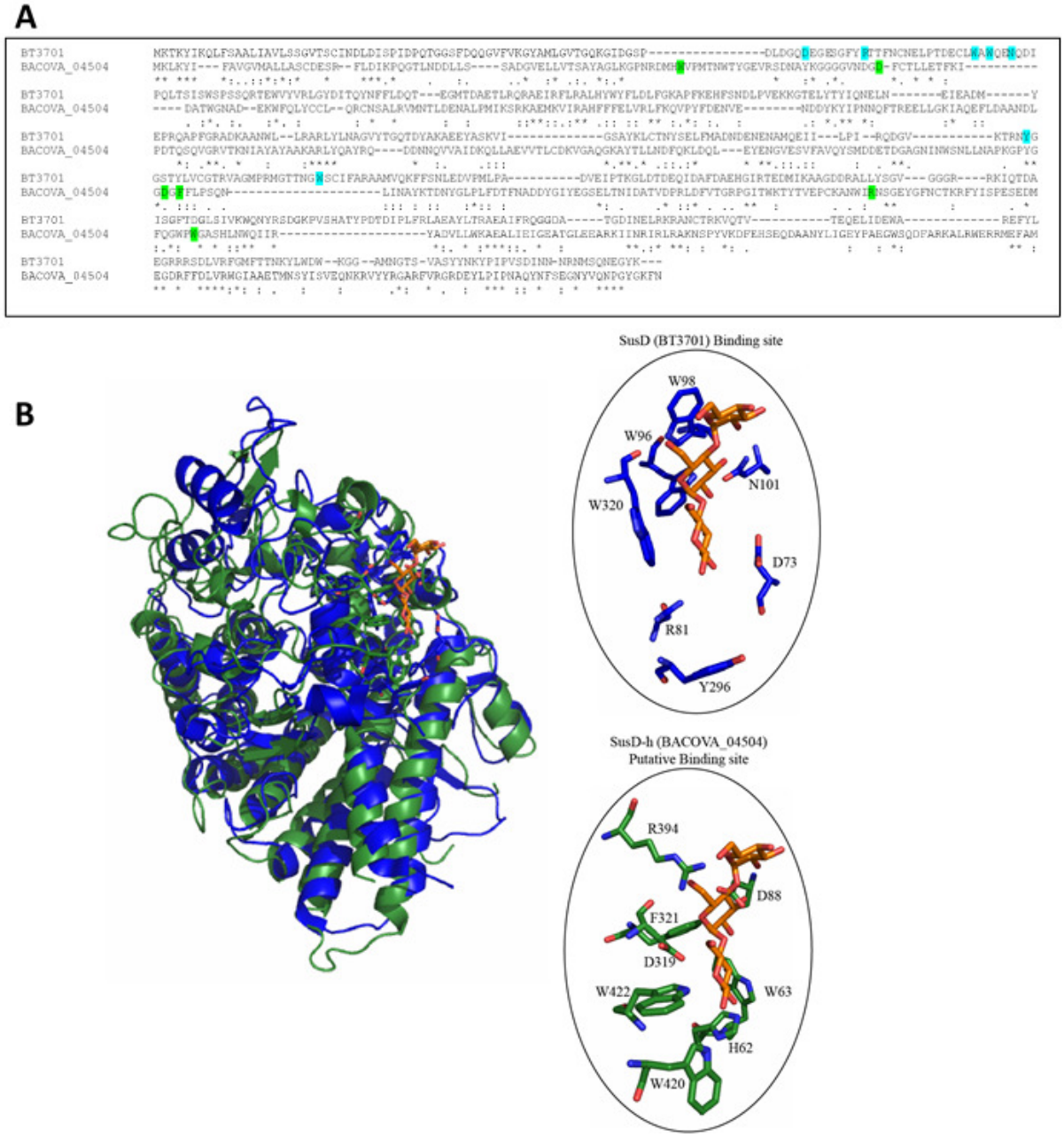


Figure 4.25. The SusD-homologue, BACOVA_04504 was compared with the canonical SusD. (**A**) Shows the sequence alignment of BACOVA_04504 SusD-homologue (SusD-h) and BT3701 (SusD). Putative SusD-h binding residues (green) and known SusD binding residues (blue) are shown (**B**) The overlaid SusD-h (green) and SusD (blue) structures with binding residues from both proteins shown as sticks. Maltotriose from the SusD structure is shown (orange sticks). Inset: The SusD binding site with maltotriose (top) and the SusD-h putative binding site with maltotriose (orange sticks) retained from the overlay.

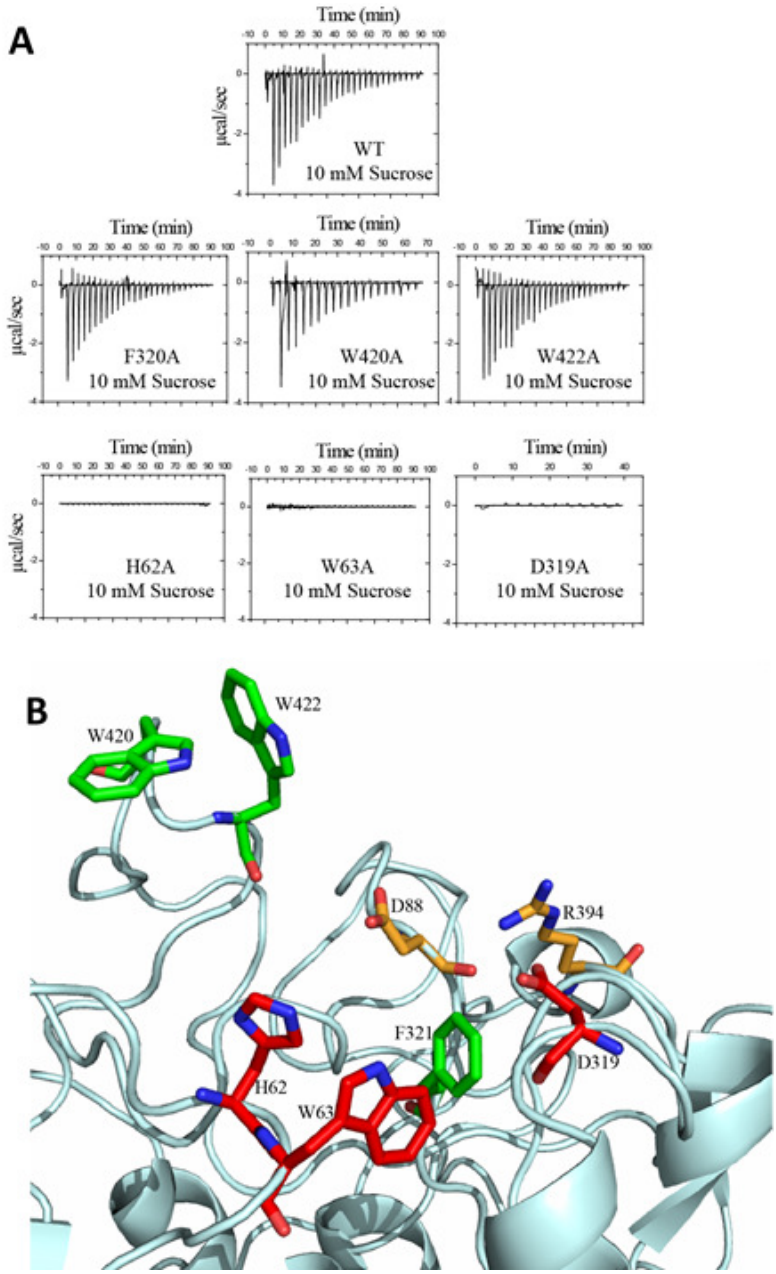


Figure 4.26. Identification of the ligand binding site of BACOVA_04504 SusD-homologue. (**A**) ITC analysis of alanine mutants of the putative binding residues from BACOVA_04504 SusD-h. F321A , W420A, and W422A bound sucrose with wild type affinity whilst H62A, W63A, and D319A were no longer able to bind sucrose. Representative datasets from multiple titrations (2-4) are shown. These were conducted in 20 mM HEPES, pH 7. 50-60 µM protein was used. (**B**) A visualisation of the putative binding site showing the residues mutated to alanine, residues which retained binding (green), were not tested (orange) and residues which could not bind (red) to sucrose are shown. Resides which (upon mutation to alanine) were no longer able to bind sucrose are likely to be involved in ligand recognition.

4.7.6. Importance of the SusD-homologue, BACOVA_04504, during Fructan Utilisation

The role of the SusD-homologue within glycan utilisation systems has been previously shown to be critical to PUL function on polymeric substrates (Sonnenburg *et al.*, 2010; Cuskin *et al*., 2015). SusD (BT3701) from the SUS system is required for growth on polymeric starch (Koropatkin *et al*., 2008) however whilst the presence of SusD protein is absolutely required, a SusD mutant which does not bind starch is able to grow on starch in the presence of small quantities of maltose despite being unable to grow on starch alone (Cameron *et al*., 2014). Maltose is able to induce up-regulation of the SUS, however the exact role of the SusD remains unclear; SusD may be fulfilling both a key structural role through mediating protein-protein interactions, and a less critical binding role which may be circumvented through the addition of small quantities of maltose (Cameron *et al*., 2014). The levan specific SusD-homologue from *B. thetaiotaomicron* (BT1762) was necessary for rapid growth on levan, but removal of this gene from the genome did not result in complete lack of growth (Sonnenburg *et al.*, 2010). It was expected that the *Δbacova_04504* strain would display a significant growth defect when grown on polymeric inulin, similar to the levan SusD-homologue BT1762. The strain, *Δbacova_04504*, was made as described in (Chapter 2.16) and is a complete deletion of the *bacova_04504* gene.

Δbacova_04504 did grow when cultured with inulin as the sole carbon source but had a reproducible lag phase of about 5 hours compared with wild-type *B. ovatus* (**Figure 4.27**). This lag phase was also seen in kestose series fructo-oligosaccharides, but not with sucrose or the glucose control. These data suggest that BACOVA_04504 confers an advantage during growth on inulin and FOS, but is not essential.

The growth defect of *Δbacova_04504* was not as pronounced on inulin as the defect of *Δbt1762* on levan as described by Sonnenburg and colleagues (Sonnenburg *et al.*, 2010).

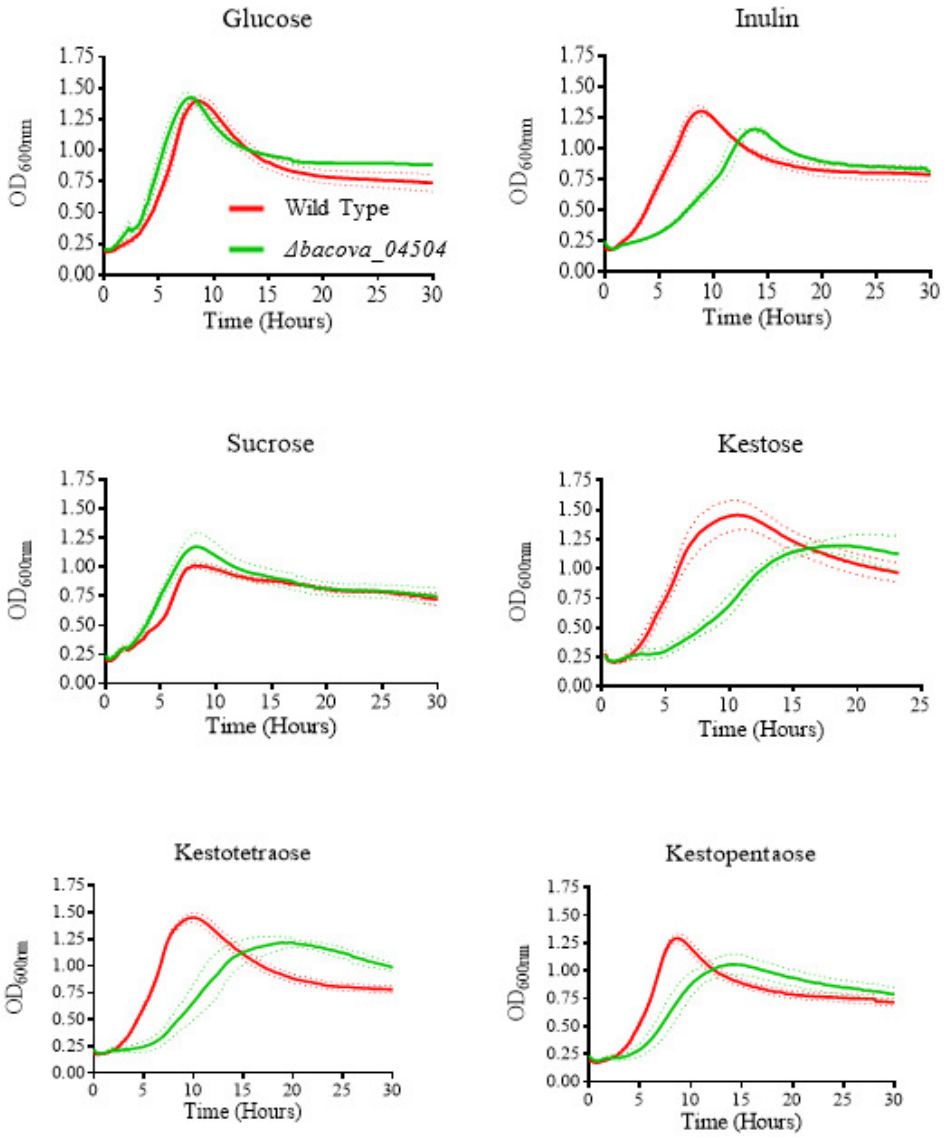


Figure 4.27. The growth of the SusD-homologue knockout *Δbacova_04505* is retarded on inulin and FOS. The growth of *Δbacova_04504* compared with wild-type *B. ovatus* was monitored when grown on minimal media supplemented with 0.5 % of either inulin, glucose, sucrose, kestose, kestetraose or kestopentaose. Growth of *Δbacova_04504* on inulin, kestose, kestotetraose and kestopentaose is retarded, but growth on glucose and sucrose are comparable to wild-type *B. ovatus*. Multiple datasets were obtained for each condition (2-4) and representative data performed in triplicate are shown with continuous error bars.

4.8. Characterisation of the Two PUL-Encoded GH32 Enzymes

4.8.1. BACOVA_04501 and BACOVA_04507 are GH32 family enzymes Which Localise to the Cell Envelope

Two GH32 family enzymes are encoded by the *B. ovatus* inulin PUL. BACOVA_04501 contains a putative type I signal peptide, type I signal peptides typically contain a hydrophobic region followed by a cleavage site (**Figure 4.28 A**); enzymes containing a type I signal peptide have been previously shown to localise to the periplasmic space in *Bacteroides* PUL systems (Cuskin *et al.,* 2015; Rogowski *et al.*, 2015). This type I signal peptide was predicted by Signal P4.1. BACOVA_04507 contains a predicted type II signal peptide (**Figure 4.28 B**). This was predicted using LipoP1.0. Lipoproteins are generally cleaved at a position around 18 to 20 amino acids from the N-terminus after a hydrophobic region, a cysteine present at this position is covalently attached to a lipid post-cleavage and anchored into the membrane. In *Bacteroides spp.* these lipoproteins are often on the outside face of the cell membrane (Shipman *et al*., 2000; Cuskin *et al*., 2015; Rogowski *et al.*, 2015).

Whilst predictive localisation tools are useful, localisation of the protein should be experimentally confirmed. If either GH32 is present at the cell surface, exo-β-fructosidase activity will take place; we predict that BACOVA_04507 will be present at the cell surface due to the N-terminal type II signal sequence, whilst BACOVA_04501, with a Type I signal sequence, will localise to the periplasm.

A GH91 mutant *Δbacova_04503 E196Q*, was used to examine surface activity of *B. ovatus* inactive whole cells, without the known activity of the GH91 endo-inulinase (chapter 4.6.2). Indeed, fructosidase activity was detected at the cell surface (**Figure 4.29**).

It is suspected that this activity is due to BACOVA_04507 as this enzyme appears to be a lipoprotein, and is likely to localise to the cell surface (Shipman *et al*., 2000; Cuskin *et al*., 2015; Rogowski *et al*., 2015). As fructose is accumulated, we expect this enzyme to be exo-acting.

A

MNVLFSFKQVQTFLTILVMMIFSFLDASADTSSLLIKDLGEGHCLVQINTNQRYLLLPVEEVMPDVRVSMIVNNKEVNVADV
RLAVNRVDYFVPLDLSAYAGKNILLKFKLGSNDPIRGKLSAVCCKEMKLSDTFDTGNREKFRPTYHFSPLYGWMNDPNGMVY
KDGEYHLFYQYNPYGSKWGNMSWGHAISQDLVNWKHLPVAIAPDALGTIFSGSAVVDFDNTAGFGAGAIIAIYTQNSDRQVQ
SIAYSTDNGRTFTKYENNPVLTSEARDFRDPKVFWYENTKRWIMVLAVGQEMQIFSSPNLKDWTFESRFGEGQGAHGGVWEC
PDLFELPVEGTNDKKWVLLCNLNPGGPFGGSATQYFVGSFNGKEFVNEFPSKTKWMDWGKDHYATVTWSDAPDNRRIAIAWM
SNWQYANDVPTSQYRSPNSVPRDLSLFTVDGETYLQSAPSPELLALRDASKKRSFKVNGTRTIKEMIPGNDGAYEIELTIEN
QRADVIGFRLYNDKGEEVDMQYDMKEKKFSMDRRKSGEVSFNENFPMLTWTAIEQGGNEEQGGKEALKLRLFVDKSSVEAFG
NGGRFVMTNQVFPSEPYNHINFYSKGGAYKVDSFVVHKLKP-

B

MKLCKGAILALAVSYGLTYCHTTKSKLTLEQKGDSLTVIHITNPTNYILLPIEEEAAESQVLLDTGEAADTDMGIRLAQTQV
DYFVPFALPVGAKTATVRIRNKSKEALCWKEIKLSDTFDTTNTEKFRPVYHHTPLYGWMNDANGLVYKDGEYHLYFQYNPYG
SKWGNMHWGHSVSKDLVHWEHLAPAIARDTLGHIFSGSSIVDQENIAGYGAGSILAFYTSASDKNGQIQCLAFSKDNGRTFT
KYEKNPILCPADGLKDFRDPKVFRYEPEDKWVMIVSADKEMRFYDSKNLKDWNYMSSFGEGYGVQPCQFECPDMVELPVDGD
LNRKKWALIVNVNPGCYFGGSATQYFTGDFDGKKFSCDNQPNVTKWLDWGKDHYATVCFSNTGERTIAVPWMSNWQYCNIVP
TKQFRSANALPRELSLYTQDGEIYLSAAPVLEIKTLRKEKKETPAFTVANDYHIDSLLADNDGAYELALEITAGEAEIMGFS
LFNDKGEKVDIYFNLPEKRLVMDRTKSGIVDFGKKSVPHEIEVHDRRKTTSINYIDDFALATWAPIKKENKYTLDVFVDKCS
VEIFLNGGKVAMTNLIFPSAPYNRMCFYSKGGSFQVDSFNAYRLGL-

Figure 4.28. Predicting Localisation of the two GH32 enzymes: BACOVA_04501 and BACOVA_04507. (**A**) BACOVA_04501 contains a putative type I signal peptide highlighted in yellow (**B**) BACOVA_04507 contains a putative type II signal peptide highlighted in yellow, with the candidate cysteine highlighted in cyan.

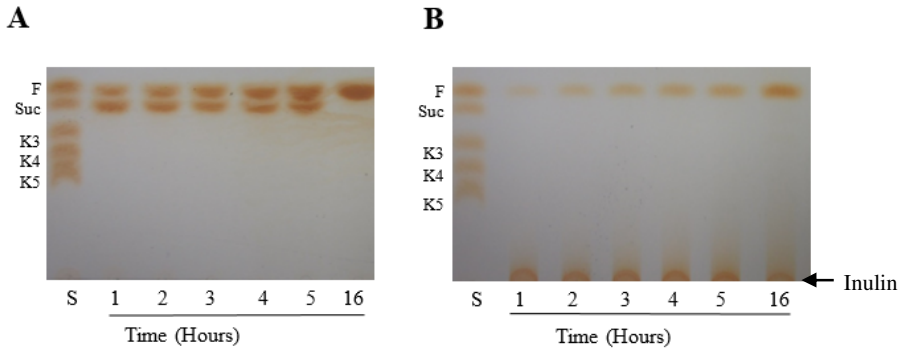


Figure 4.29. Analysis of inactive whole cells for exo-fructosidase activity. The GH91 mutant stain, *Δbacova_04503 E196Q* initially grown on minimal media with inulin was used to prepare inactive whole cells for assays to detect surface exo-β-fructosidase activity, these were visualised via TLC. Whole cells were incubated with (**A**) 0.5 % sucrose or (**B**) 0.5 % inulin. Standards (S) containing 2 mg/ml each of fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) were run for reference. Fructose release indicative of exo-fructosidase activity was detected on both substrates, indicating the presence of one of the GH32 family enzymes (BACOVA_04501 or BACOVA_04507) at the cell surface. We suspect this activity is attributed to BACOVA_04507 as signal peptide analysis suggests that this enzyme has a type II signal peptide, which are often extracellular in *Bacteroides* PUL systems (Shipman *et al*., 2000; Cuskin *et al*., 2015; Rogowski *et al.,* 2015).

4.8.2. BACOVA_04501 and BACOVA_04507 are Both Exo-acting β-fructosidases

Both GH32 enzymes (BACOVA_04501 and BACOVA_04507) were cloned and expressed in recombinant *E.coli* as described (Chapters 2.9.7 and 2.11). Proteins were purified using IMAC (Chapter 2.12.1). Further construct details are listed (Appendix I.8).

Pilot enzyme assays were undertaken and visualised via TLC to examine the product profiles of each GH32 enzyme (**Figure 4.30**). Both enzymes release fructose from inulin over time indicating that both of these enzymes are exo-acting (**Figure 4.30**).

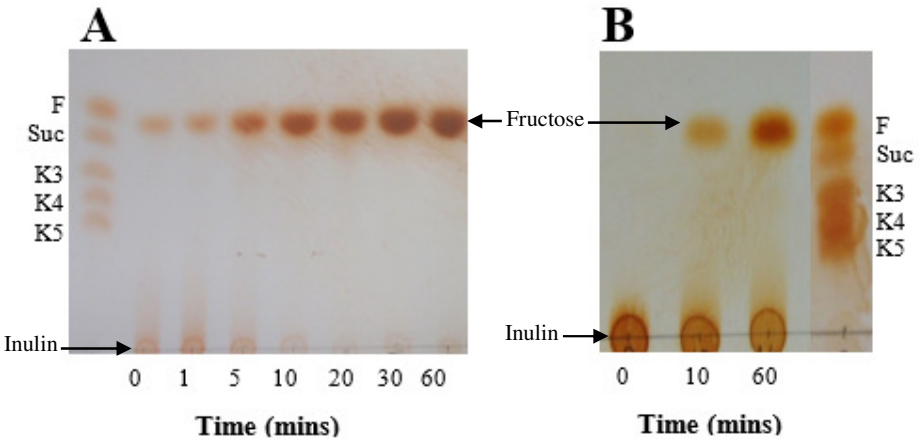


Figure 4.30. Examining the product profiles of both GH32 enzymes on inulin. (A) BACOVA_04501 and (**B**) BACOVA_04507 were in incubated with inulin for one hour, both enzymes released fructose, which is suggestive of exo-activity. Standards (2 mg/ml each, stock) consisting of fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) were run for comparison. Both GH32 reactions were carried out in 20 mM phosphate buffer pH 7.0 with 0.5 % inulin and 1 µM enzyme.

As fructose is released, the kinetic parameters for both GH32 enzymes (BACOVA_04501 and BACOVA_04507) were elucidated through fructose detection assays (**Table 4.2**). BACOVA_04501 displayed a preference for β2-1 linkages, but would degrade levan. BACOVA_04507 appears to be a sucrase, with decreasing affinity for longer substrates. Our data are similar to those elucidated for the two GH32 homologues from the characterised levan PUL

(Sonnenburg *et al.*, 2010). Key Michaelis-Menton curves from both enzymes are shown (**Figure 4.31**).

Table 4.2. Kinetic parameters for the two PUL encoded GH32 exo-fructosidases on sucrose, FOS, Inulin and Levan

	BACOVA_04501			BACOVA_04507		
	K_{cat} (Min-1)	**K_m (mM)**	**K_{cat}/K_m (min^{-1} /mM^{-1})**	**K_{cat} (Min-1)**	**K_m (mM)**	**K_{cat}/K_m (min^{-1}/ mM^{-1})**
Sucrose	5367 ± 416	3 ± 0.8	1789	12087 ± 555	1.7 ± 0.3	7110
Kestose	1669 ± 115	1.7 ± 0.4	982	1506 ± 182	12.2 ± 3.2	123
Kestotetraose	2461 ± 272	1.4 ± 0.5	1758	1295 ± 45.5	2.6 ± 0.26	498
Kestopentaose	-	-	-	3040 ± 655	12 ± 4.3	253
Fructotriose	-	-	-	2065 ± 280	5.6 ± 1.9	369
Inulin*	2353 ± 208	0.8 ± 0.2	2941	-	-	~ 322
Levan*	-	-	~2900	-	-	~ 220

* Concentration in mM for polysaccharides was calculated as described in Chapter 2.13.2.

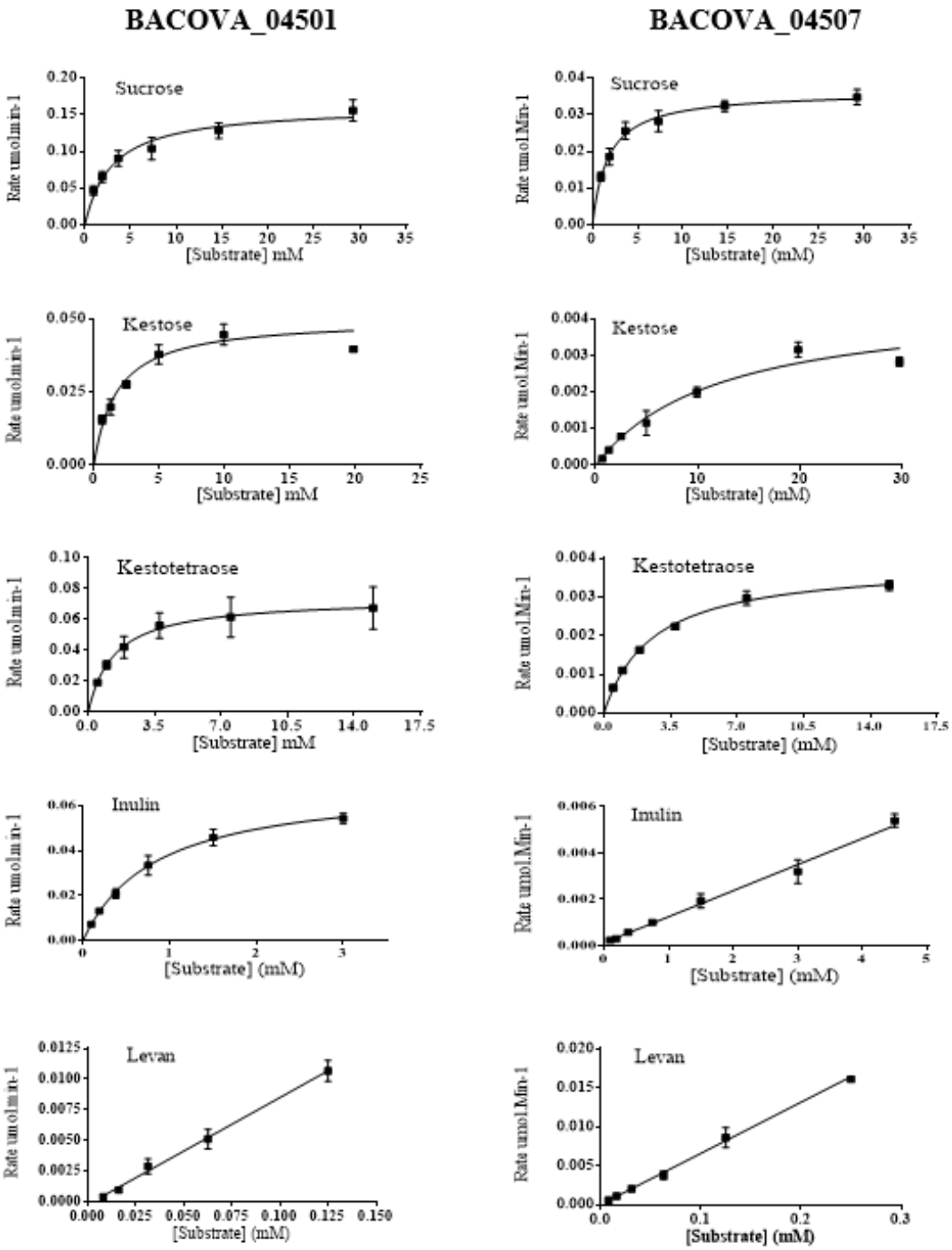


Figure 4.31. Michaelis-Menton plots from PUL encoded GH32 enzymes on defined substrates. Michaelis-mention curves from BACOVA_04501 and BACOVA_04507 on sucrose, kestose series FOS, inulin and levan. Experiments were undertaken using the megazyme fructose detection kit in 20 mM Tris, 150 mM NaCl buffer, pH 7.0.

4.8.3. The GH32, BACOVA_04507, Does not have a Clear Role During Inulin Harvest

The role of the GH32 exo-fructosidase, BACOVA_04507, was probed through the creation of a catalytically inactive mutant strain, *Δbacova_04507* D265A. BACOVA_04507 is a predicted lipoprotein (**Figure 4.28**) and appears to target sucrose and short chain FOS.

No growth defect was observed when *Δbacova_04507 D265A* was grown on minimal media supplemented with a range of defined fructan substrates when compared with the growth of wild type *B. ovatus* (**Figure 4.32**). These data suggest that this enzyme is not critical for growth on any of the substrates tested.

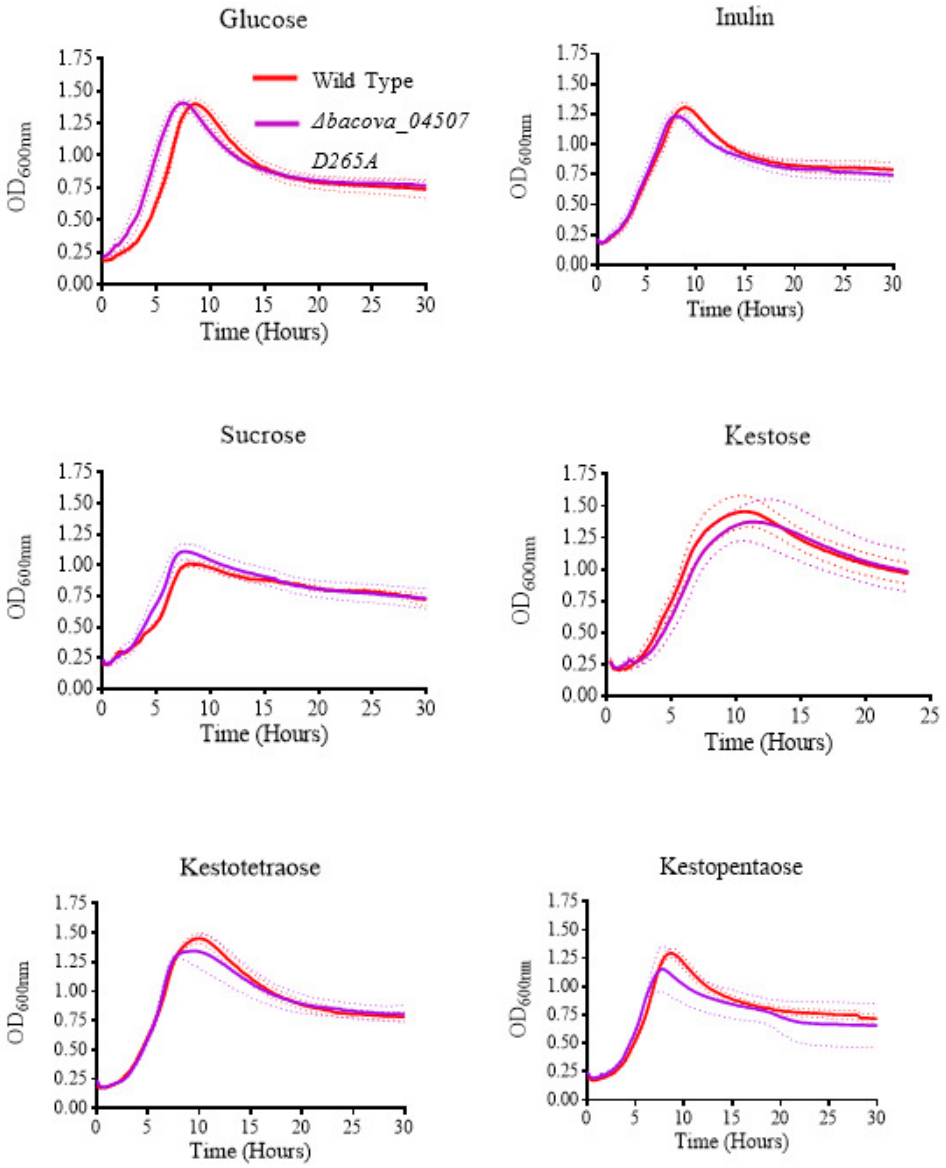


Figure 4.32. The growth of the GH32 catalytic mutant, *Δbacova_04507*, is not different to growth of wild type *B. ovatus* on the fructan substrates tested. The growth of *Δbacova_04507 D265A* compared with wild-type *B. ovatus* was monitored when grown on minimal media supplemented with 0.5 % of either inulin, glucose, sucrose, kestose, kestetraose or kestopentaose. All mutant growth conditions were comparable to wild-type *B. ovatus*. Multiple datasets were obtained for each condition (2-4) and representative data performed in triplicate are shown with continuous error bars.

4.9. Exploring fructan utilisation by *Bacteroides ovatus* Through Generation of Knock-out Strains

4.9.1. Importance of the PUL encoded SusC-homologue, BACOVA_04505, during Fructan Utilisation

The SusC-homologue, BACOVA_04505, was hypothesised to be critical to PUL function. SusC-homologues are outer membrane transport proteins which partner with SusD-homologues to form a complex (Shipman *et al*., 2000; Koropatkin *et al*., 2008). SusC-homologues are TonB-dependent porins, outer membrane spanning transporters which are able to transfer macromolecules through the outer member to the periplasmic space using energy derived from the TonB-ExbBD complex (Martens *et al*., 2009). A knockout strain of *B. thetaiotaomicron* lacking the prototypic SusC is unable to utilise starch (Cameron *et al*., 2014).

The entire gene region was removed from the genome following the protocol outlined (Chapter 2.16). The growth of *Δbacova_04505* was monitored and compared with that of wild type on minimal media supplemented with target fructans (**Figure 4.33**). The data showed a significant growth defect compared with wild type on inulin and FOS however there is no noticeable defect between wild type *B. ovatus* and *Δbacova_04505* growth on sucrose (**Figure 4.33**).

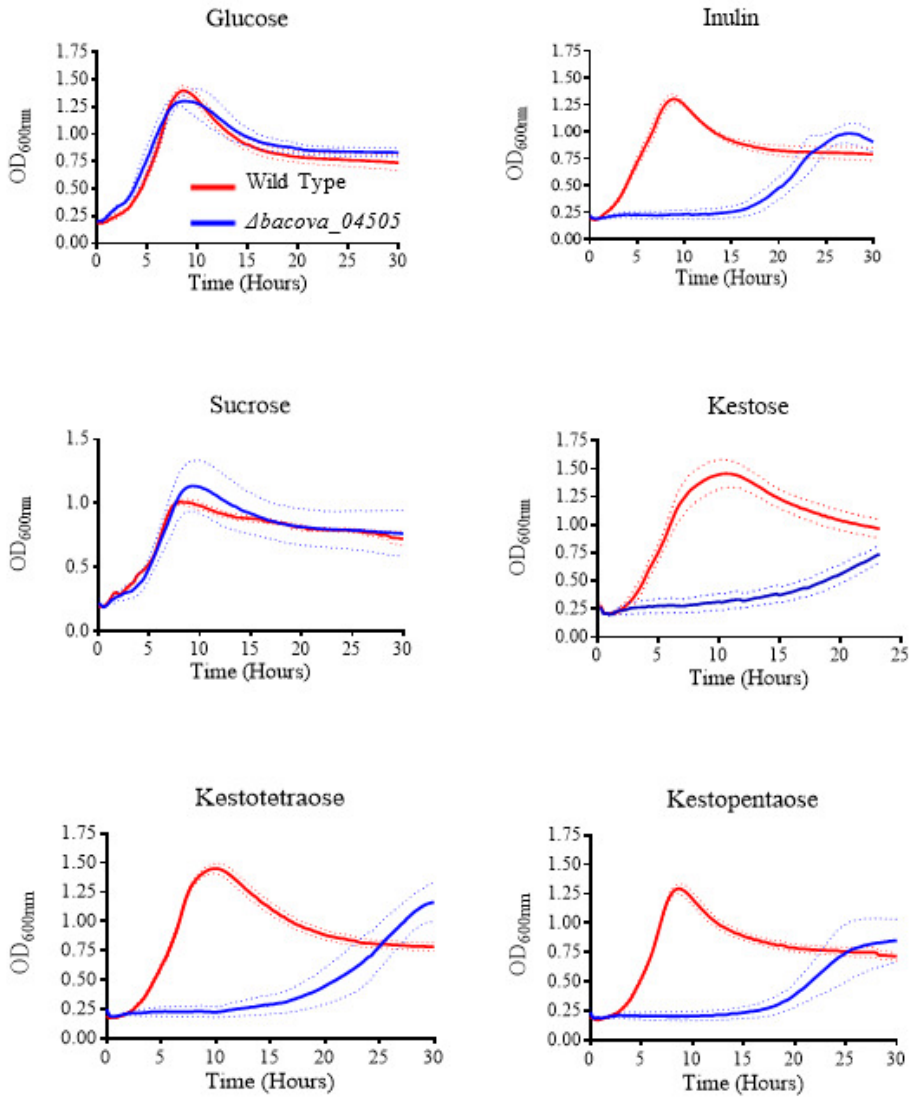


Figure 4.33. The SusC-homologue knockout *Δbacova_04505* has a growth defect on inulin and FOS. The growth of the SusC-knockout strain *Δbacova_04505* was monitored and compared with wild-type *B. ovatus* when grown on minimal media supplemented with 0.5% of either inulin, glucose, sucrose, kestose, kestetraose or kestopentaose. The SusC-knockout has an extremely long lag before growth on inulin and FOS, but grows at a similar rate and to a final density comparable to WT on glucose and sucrose. Multiple datasets were obtained for each condition (2-4) and representative data performed in triplicate are shown with continuous error bars.

4.9.2. Growth of the Inulin PUL Mutant Strains on a Range of Fructans

Each mutant plus wild type *B. ovatus* was grown with different fructan structures as the sole carbon source under the same conditions (**Figure 4.34**). These assays allow insight into both the structures of the fructans used and of the limitations and functionality of the *B. ovatus* inulin utilisation system. Linear inulin (chicory) and glucose were used as controls with wheat, onion and garlic extracted fructans tested.

Wheat fructan supported the growth of all mutants and wild type, but to a lesser extent than glucose or inulin control (**Figure 4.34**). The SusC-homologue knockout, *bacova_04505,* grows to a lower OD than the other strains, which suggests that some structures within this fructan mixture require import through the SusC-h transporter. Both the SusD-h knockout, *Δbacova_04504* and the putative surface exo-fructosidase mutant, *Δbacova_04507 D265A,* seem to have a slight growth defect. The defect from *Δbacova_04504* suggests that a portion of this fructan extract is comprised of short chain FOS. The *Δbacova_04507 D265A* defect is intriguing as it is possible that wheat fructan contains the preferred substrate for this enzyme.

Fructan extracted from onion produces similar growth profiles to the chicory inulin control for each strain (**Figure 4.34**). It is therefore likely that this substrate is comprised mostly of linear inulin, or has a secondary structure very similar to linear inulin. DFA is produced during growth demonstrating that the GH91 endo-inulinase (BACOVA_04502/3) is active, this enzyme is not active on short FOS, therefore linear inulin with DP > ~6 must be present within this fructan extract.

Fructan extracted from garlic has a similar growth profile to the inulin control (**Figure 4.34**), however no DFA is produced in the supernatant (Chapter 3. **Figure 3.5 B**).

These data support the characterisation of garlic inulin as a β2-1 inulin polymer with short β2-6 linkages, as described by Baumgartner *et al.*, 2011. Short levan decorations may occlude the action of the GH91 enzyme, preventing DFA production.

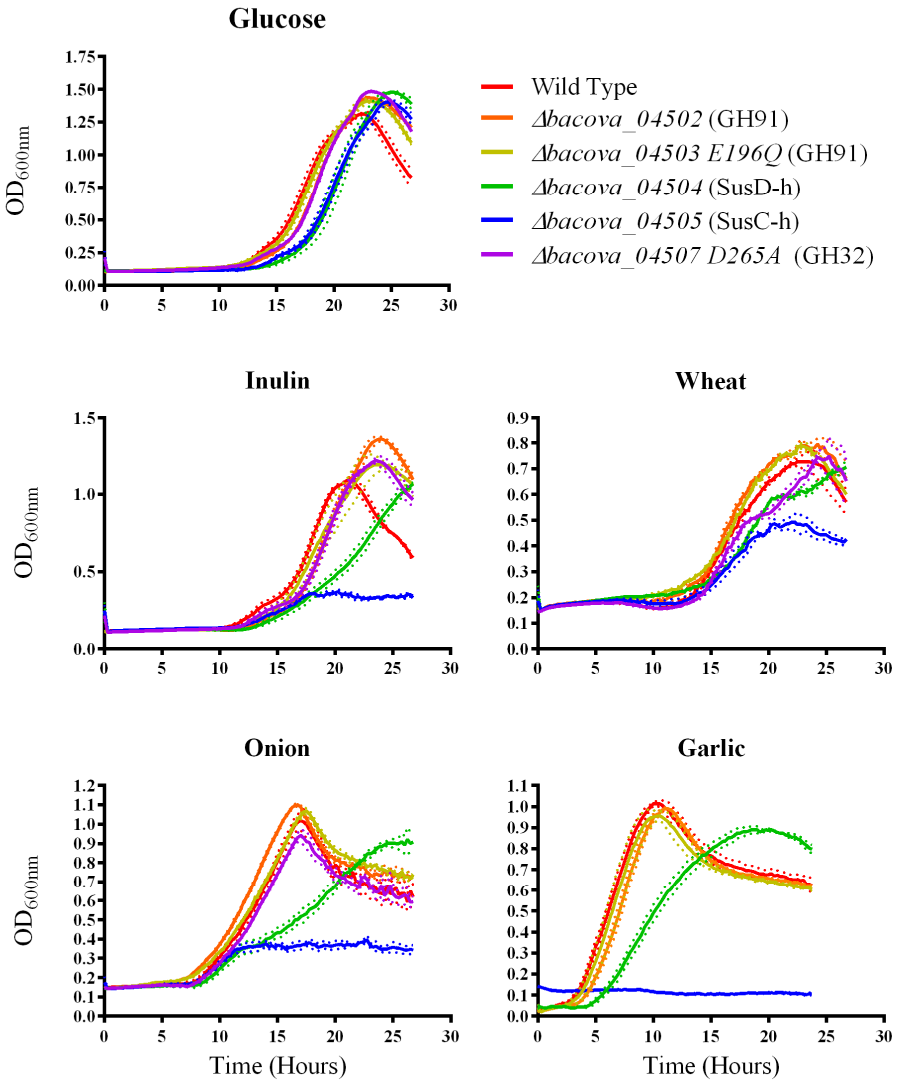


Figure 4.34. *B. ovatus* mutants were grown on a variety of fructan structures.
Mutants were grown on various fructan substrates to determine any growth defects. Minimal media supplemented with 0.5 % of either linear inulin, glucose, wheat, onion or garlic fructans was monitored. Growth of *bacova_04507 D265A* on garlic was not reproducible and has not been shown. Multiple datasets were obtained for each condition (2-4) and representative data performed in triplicate are shown.

4.10. Exploring the Creation and Degradation of DFA and DFA-FOS within the Gut using Faecal Samples.

4.10.1. Detection of DFA Within Cultures Supplemented with Inulin

Bacterial cultures obtained from fresh human faecal samples can shed light upon how glycans are degraded *in vivo* without requiring extensive or invasive procedures; faecal slurry from three healthy human donors was used to inoculate pH controlled batch cultures as described (Chapter 2.18) to investigate if DFA is produced from inulin under these conditions. Basal media was supplemented with Inulin, FOS (Orafti ® 95) and a negative control without sugar for comparison. Samples from each vessel were taken at 0, 4, 8, 24 and 48 hours. Supernatant fluid from each sample was visualised using TLC so that any DFA created could be observed (**Figure 4.35**).

No DFA was created from vessels containing FOS, which is consistent with the characterisation of the *B. ovatus* GH91 which is not active upon inulin with a DP < ~6. DFA was detected within vessels containing inulin (**Figure 4.35**).

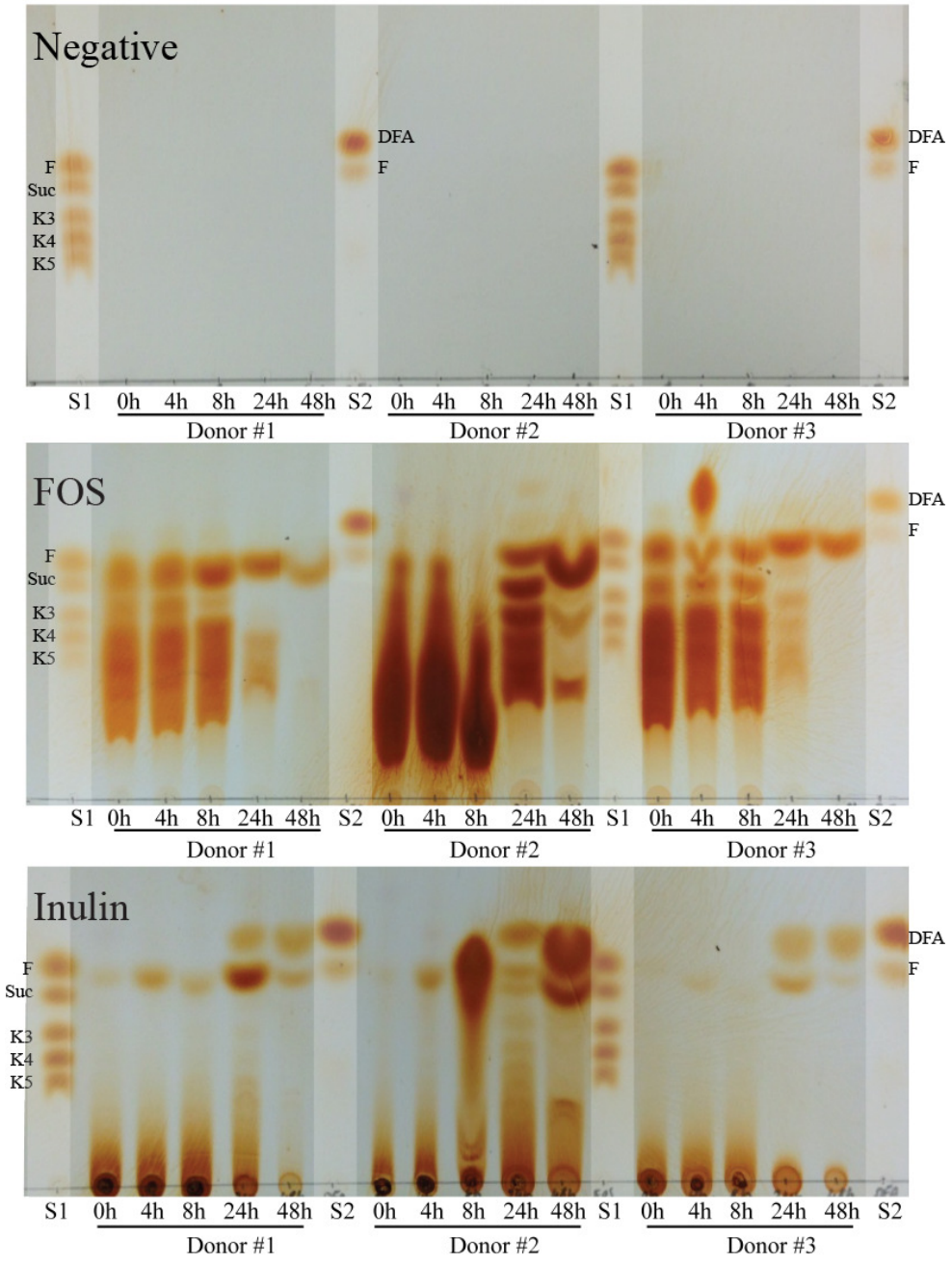


Figure 4.35. TLC visualisation of DFA production from inulin by the faecal cultures. TLC visualisation of culture supernatant fluid removed from the fermentation vessels at defined time points (0 – 48 hours) allowed visualisation of inulin and FOS breakdown products by complex faecal microbiota. All three donors are shown for each glycan, also shown are 2 mg/ml each of standards (highlighted for clarity): S1 contains fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentase (K5), S2 contains DFA and fructose. DFA was produced from inulin cultures but not FOS cultures.

4.11. Examining the Degradation of DFA-FOS and DFA by the Faecal Microbiota

Using batch cultures set up at the same time and in the same manner as the previous experiment, the degradation of DFA by the faecal microbiota was probed. Two additional batch cultures, containing DFA and DFA-FOS were set up, and samples taken at 0, 4, 8, 24 and 48 hours (**Figure 4.36**). DFA-FOS was utilised over time, however DFA was released and accumulated in the media. This implies that the DFA-FOS produced extracellularly by *B. ovatus* (or other species) can be utilised either directly by *B. ovatus* or by other species, however the DFA portion of the chain cannot be utilised. Vessels supplemented with DFA did not show any significant reduction in DFA content over the course of the experiment (**Figure 4.36**). These data, taken with those from the previous experiment (**Figure 4.35**) suggest that DFA is produced during inulin degradation by the normal gut microbiota, but that this DFA is not utilised.

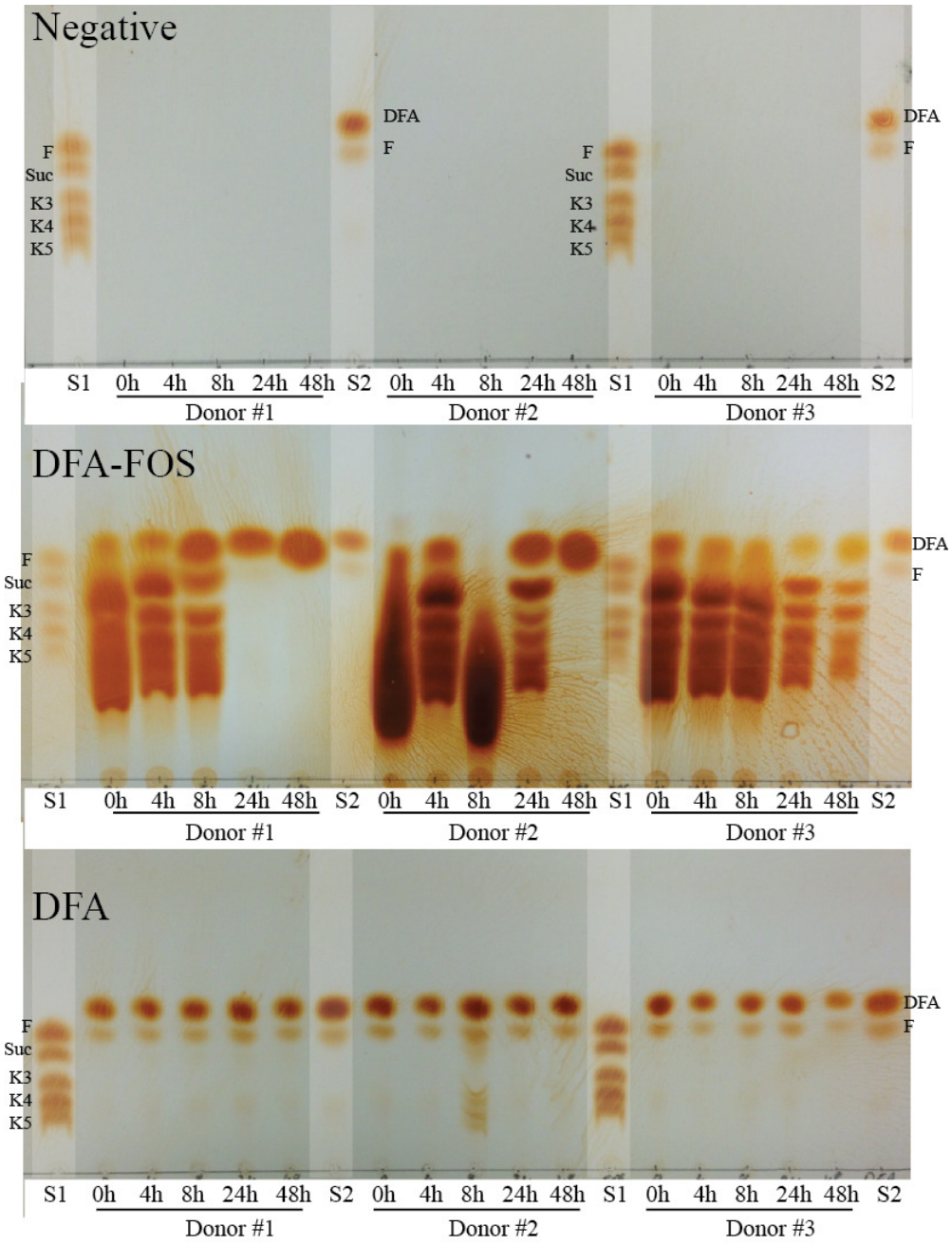


Figure 4.36. TLC visualisation shows that DFA is not degraded by the faecal microbiota. This experiment was conducted at the same time and in the same manner as the previous experiment and shows culture supernatant fluid removed from the fermentation vessels at defined time points (0h-48h). All three donors are shown for each glycan, also shown are 2 mg/ml each of standards (highlighted for clarity): S1 contains fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentase (K5), S2 contains DFA and fructose. DFA samples were diluted 1 in 10 for clarity. DFA was not degraded in either condition.

4.12. Discussion

4.12.1. Fructan Utilisation in other *Bacteroides spp.*

In this chapter, we demonstrate how inulin is utilised by the *B. ovatus* inulin PUL, to put this system in context here we explore the fructan utilisation apparatus of other *Bacteroides species*.

Linear inulin with a DP > ~6 is broken up at the *B. ovatus* cell surface by a heteromeric GH91 endo-inulinase (Chapter 4.6.2). Further inulin processing requires the action of one or both exo-fructosidase enzymes (BACOVA_04501 and BACOVA_04507) which do not display a linkage preference (Chapter 4.8). This system is similar to that of *B. thetaiotaomicron* levan PUL, where specificity for levan is observed in the surface endo-levanase (BT1760) but not in the exo-fructosidase enzymes (BT1759, BT1765 and BT3082). The exo-fructosidase enzymes from *B. ovatus* and *B. thetaiotaomicron* have high sequence identity (**Figure 4.2**) and are biochemically similar (Chapter 4.8; Sonnenburg *et al.*, 2010); therefore these enzymes perform a core role in fructan utilisation that is not-linkage specific. Furthermore, as the SusC/D-homologue pair is divergent between the fructan PUL of *B. ovatus* and *B. thetaiotaomicron* we suspect that this import apparatus must be specialised for the target fructan; we demonstrate that the *B. ovatus* SusD-homologue (BACOVA_04504) recognises a sucrose moiety in inulin (Chapter 4.6) and the *B. thetaiotaomicron* SusD-homologue recognises levan but not inulin (Sonnenburg *et al.*, 2010) supporting that the import apparatus is specific for the target glycan.

To explore this further we examined the GH32 content from various *Bacteroides species*, looked for other GH91 encoding *Bacteroides species* and examined the divergence of SusD-homologues from other fructan PUL. The aim of this was to understand whether there was a connection between the PUL encoded apparatus and linkage specificity. Seven *Bacteroides spp.* were grown on inulin and levan (Appendix I. **Figure I.6.**) to confirm fructan specificity.

Two FOS users (*B. fragilis*, *B. vulgatus*), two levan users (*B. thetaiotaomicron, B. intestinalis*) and three inulin users (*B. ovatus, B. thetaiotaomicron 8763, B. uniformis*) were examined.

The putative fructan PUL from each species was identified by searching for homologues to the BT1754 HTCS as described by Sonnenburg *et al*., 2010. Within each PUL up to four GH32 encoding genes were present, these were aligned and a neighbour joining tree was generated (**Figure 4.37**). BT1760 is an endo-levanase (Sonnenburg *et al*., 2010) and it is likely that BACINT_03199 is also an endo-levanase as *B. intestinalis* is a levan user and the two enzymes are closely related. *B. uniformis* is an inulin user; BACUNI_01155 contains reversed catalytic residues as seen in a GH32 endo-inulinase enzyme (Puyez *et al*., 2012), suggesting that this enzyme is likely to be an endo-inulinase, however it is predicted to localise to the periplasm, this enzyme is more similar to the endo-levanases BT1760 and BACINT_03199 (**Figure 4.37**). All polymeric fructan users contain a BACOVA_04501 homologue, whereas the two FOS users, *B. fragilis* and *B. vulgatus*, contain a more divergent predicted lipoprotein more similar to BACOVA_04507 exo-fructosidase (**Figure 4.37**). The inulin using *B. thetaiotaomicron 8736* contains homologues of the *B. thetaiotaomicron* exo-fructosidases but lacks the endo-levanase (**Figure 4.37**); this is interesting as it supports the hypothesis that an endo-inulinase is not necessary for growth on inulin (Chapter 4.9.3) and shows that the exo-fructosidase enzymes are not linkage specific.

These data assist predictions regarding how *Bacteroides spp.* utilise fructans, and demonstrate that several fructan acquisition strategies are apparent within the *Bacteroides* genus.

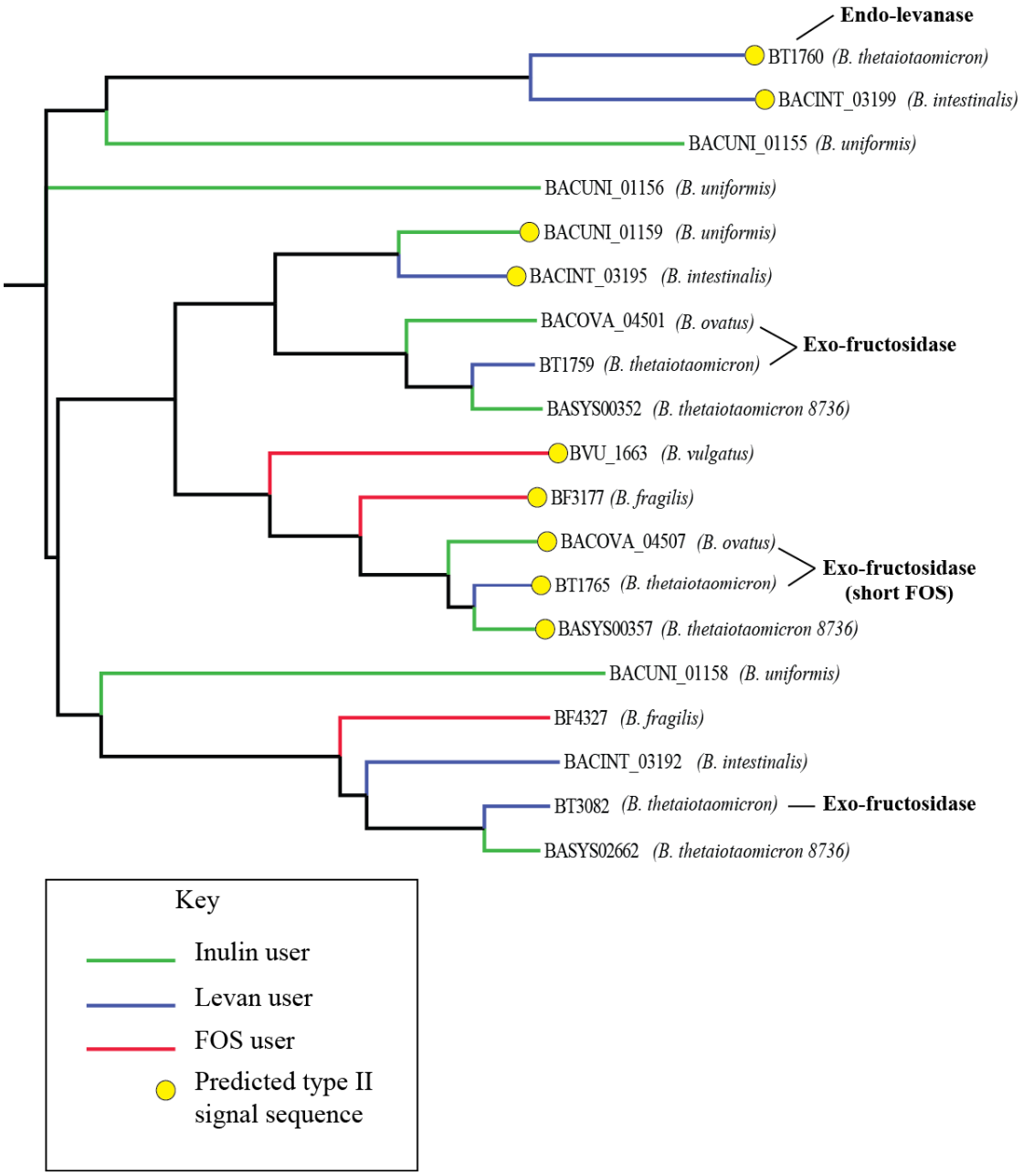


Figure 4.37. The GH32 enzymes from seven *Bacteroides* fructan PUL. A tree was generated using FigTree (http://tree.bio.ed.ac.uk/software/figtree/) from an alignment of 19 GH32s from seven *Bacteroides spp*. colour coded by fructan preference of the originating species; FOS (red), inulin (green) or levan (blue). Sequences with a putative type II signal sequence (predicted extracellular lipoproteins) are labelled with a yellow node terminus, all others have a putative Type I signal sequence (predicted periplasmic). Characterised enzymes from *B. ovatus* and *B. thetaiotaomicron* are annotated.

The distribution of GH91 encoding genes within sequenced *Bacteroides* genomes are very narrow. *Bacova_04502/3* homologues are found within *B. caccae* and *B. xylanisolvens* where the genomic context of the inulin PUL is conserved (**Figure 4.38**). No other GH91 enzymes have been identified from *Bacteroides spp* at this time.

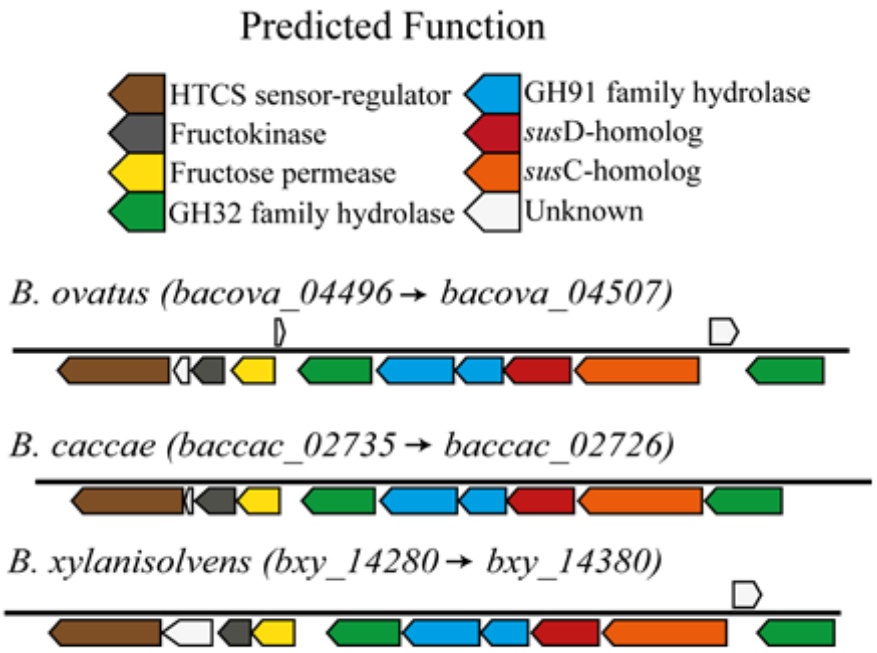


Figure 4.38. Exploring the distribution of GH91 encoding genes within *Bacteroides spp.* Three species of *Bacteroides*, including the characterised *B. ovatus* contain a PUL encoded GH91. *B. xylanisolvens* and *B. caccae* contain homologues of the BACOVA_04502/3 enzyme which is embedded within a homologous inulin utilisation PUL.

The distribution of GH91 enzymes within the CAZy database demonstrate that GH91 encoding genes are more commonly associated with species isolated from soil, such as from the Bacillus, Arthrobacter and Rhizobium genera (www.cazy.org). GH91 enzymes have been isolated from *Klebsiella spp.,* this genus has been isolated from a wide range of environments including human samples (Brisse *et al.*, 2006).

As the distribution of GH91 encoding genes is narrow and *Bacteroides spp*. are prominent resident microbiota, it is likely that much or all of the DFA produced by faecal cultures is generated by the action of the inulin utilisation PULs from *Bacteroides spp*. (**Figure 4.35**; **Figure 4.38**).

The SusD-homologues from each fructan PUL were aligned and a tree generated (**Figure 4.39**), *B. vulgatus* was not included as this species does not have a candidate fructan binding SusD-homologue. The SusD-homologues clustered by preference for inulin or levan, with *B. fragilis* FOS SusD-homologue the most divergent (**Figure 4.39**).

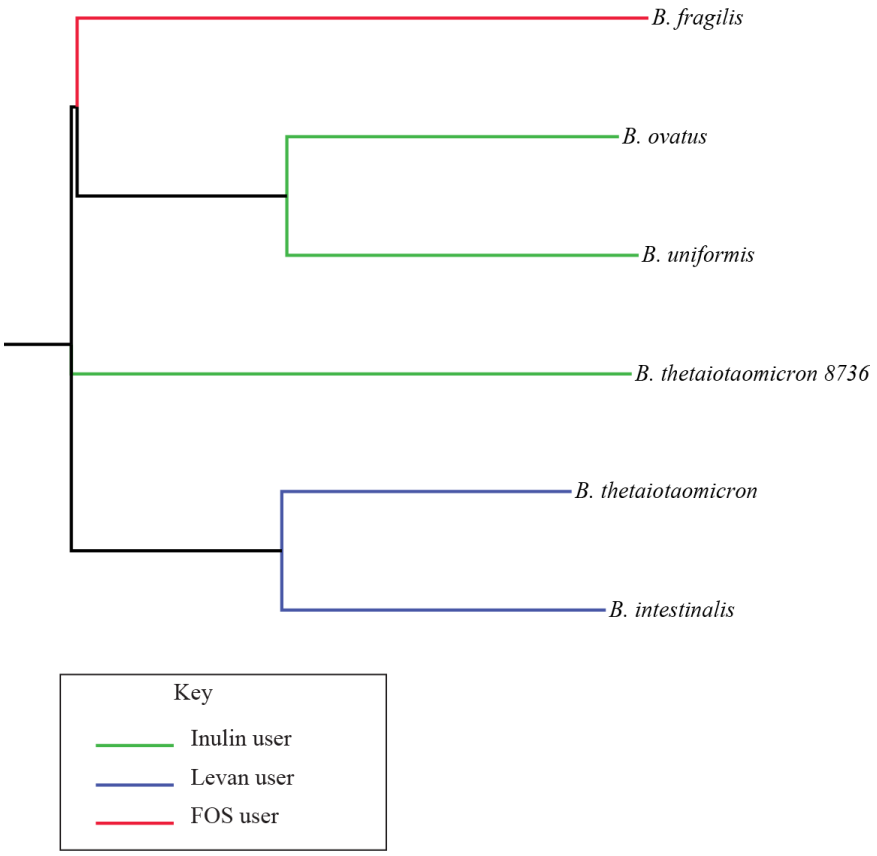


Figure 4.39. Fructan PUL encoded SusD-homologues. SusD-homologues from fructan PUL were aligned. Each are labelled according to the originating species and are colour coded to demonstrate fructan preference. SusD-homologues cluster by fructan specificity. The inulin utilising strain of *B. thetaiotaomicron 8736* has a SusD-homologue closest to *B. thetaiotaomicron*, but this protein is quite divergent as expected as the strain is able to utilise inulin.

4.12.2. Understanding the Role of the GH91 Enzyme

4.12.2.1. Re-evaluating the Role of the BACOVA_04502 CBM

We show that the CBM appended to the GH91 enzyme does contribute slightly to enzyme activity rate on chicory inulin, but that it is not crucial for enzyme activity (Chapter 4.6.4). As discussed in the introduction (Chapter 1.7.1) *Bacteroides* SUS-like systems, including the canonical SUS, often display polysaccharide binding proteins on the cell surface. Many PULs contain a SusE-like binding protein, thought to help the bacterial cell latch onto insoluble substrate, or to tether large polysaccharides to the cell surface prior to, or to assist with, processing and import (Martens *et al.*, 2009; Sonnenburg *et al.*, 2010).

During initial BACOVA_04502 characterisation during a Masters project (Shapiro, MRes thesis, 2012), it was proposed that as this enzyme was inactive and it was functioning solely as a SusE-like binding protein; with the GH91 domain acting as a spacer domain. We now know that this is not the case, as the GH91 domain is indeed active, therefore the role of the CBM may be more similar to other CBM, such as to target insoluble material (Boraston *et al.*, 2004) or the binding role may fulfil the role of a SusE-like gene, and mediate binding of polysaccharide to the cell surface. It is unclear whether the slight retardation of enzyme activity rate (**Figure 4.11**) is due to the CBM binding role, or simply that the lack of this region disrupts the BACOVA_04502/3 protein-protein interaction. Further experiments should be done to characterise this binding domain as if it is a CBM operating through a proximity effect it is the founding member of a novel CAZy CBM family.

Surface polysaccharide binding is important in other Bacteroides SUS-like systems. Cameron *et al.*, 2014, characterised the eight starch binding subsites found across the surface apparatus of the canonical SUS.

It was shown that each binding site fulfilled similar roles during *in vitro* binding assays, however mutant strains lacking individual binding sites demonstrated that several sites had more complex functions during *in vivo* growth assays, either through regulation of the system or compensating for capsule expression. In light of this information, it would be pertinent to further explore fructan binding at the cell surface, particularly in the GH91 (BACOVA_04502/3) CBM.

As the genomic knock-out strain, *Δbacova_04502*, did not display any notable defects compared with wild-type, no conclusions could be drawn regarding the role of this binding domain and indeed the full length enzyme within the inulin utilisation system. The lack of a phenotype suggests that either the apparatus is fine-tuned to out-compete other members of the microbiota such that defects within monocultures cannot be detected, or that the true target for this enzyme has not yet been explored.

Using tagged mutant strains, it would be possible to design a competition assay with a binding mutant of *bacova_04502* CBM against WT, this may reveal whether the binding role confers a competitive advantage.

4.12.2.2. A Putative Model for the GH91, BACOVA_04502/3, Catalytic Mechanism

The GH91 enzyme (BACOVA_04502/3) is the first example of an endo-inulinase in the GH91 family, furthermore it is the example of a hetero-oligomeric enzyme in CAZy (Chapter 4.6). The enzyme is active on linear inulin with a DP > ~6 and creates a range of oligosaccharide products which contain a DFA moiety at what would otherwise be the non-reducing end.

During GH91 characterisation, two catalytic residues were determined through targeted mutagenesis, BACOVA_04502 D185 and BACOVA_04503 E196. Both residues are conserved within the GH91 family.

As BsIFTase shares conserved catalytic apparatus and produced an anhydrous product, it seems likely that a similar mechanism for catalysis will be conserved between the two homologues. Using these data, and the mechanism elucidated by Jung *et al.,* a mechanism for the action of the GH91 enzyme was predicted (**Figure 4.40**).

D185, contributed by BACOVA_04502 co-ordinates the fructose at the +1 subsite. E196, contributed by BACOVA_04503 activates nucleophilic attack from the C3 carbon of F2 on the C1 carbon of F1, forming a new glycoside bond and breaking the bond which was present between the C1 carbon of F1 and the C2 carbon of the fructose at the -1 subsite. As this is an inverting reaction, the conformation of the new glycosidic bond, which was originally β- becomes α-. We predict, based upon BsIFTase data, that DFA III will be produced however the linkage of the secondary glycosidic bond has not been determined. Whilst this mechanism has not been formally elucidated, this model provides a good fit for the observations made both with BACOVA_04502/3 and other members of the GH91 family.

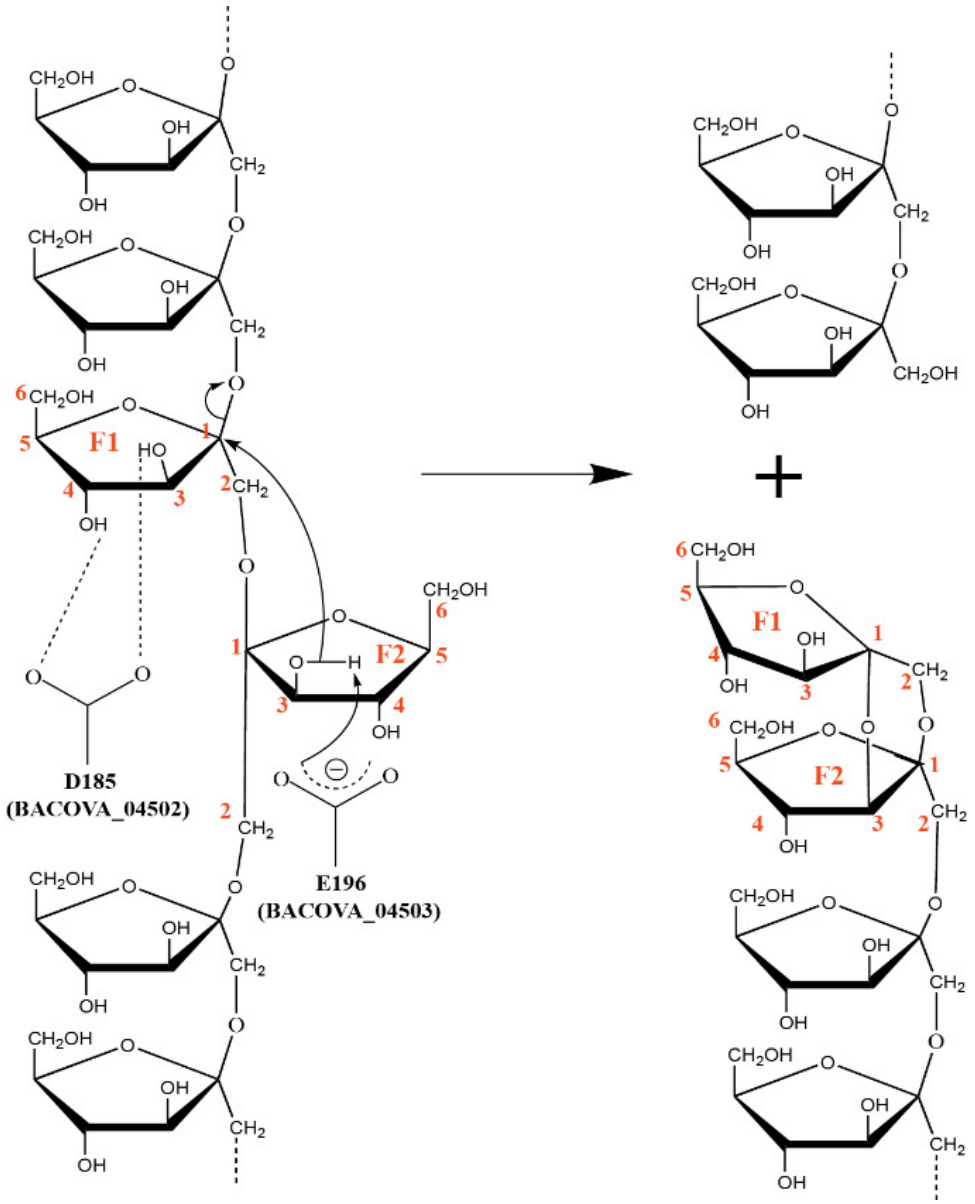


Figure 4.40. Proposed catalytic mechanism for GH91, BACOVA_04502/3. This model is based on the model proposed by Jung *et al.* for BsIFTase as these enzymes belong to the same CAZy family and are likely to have the same mechanism. BACOVA_04502/3 cleaves inulin internally to produce DFA-FOS whereas BsIFTase is exo-acting and produces FOS. Catalysis occurs through the inverting mechanism, with a single nucleophilic attack event. D185, contributed by BACOVA_04502 co-ordinates the fructose at the +1 position (F2). E196, contributed by BACOVA_04503 is the catalytic base, activating the nucleophilic group at C3. A new glycosidic bond is formed at the C1 carbon of the F1 fructose (F1 α1→3 F2), which breaks the glycosidic bond which was present between this carbon and the second carbon at the fructose at the -1 position.

4.12.2.3. The GH91 Enzyme is not Required for Growth on Inulin

The current paradigm for SUS-like systems is that the surface localised endo-active enzyme is required for growth on the polysaccharide, indeed a mutant *B. thetaiotaomicron* strain lacking the endo-levanase, BT1760, is unable to grow on levan (Martens *et al*., 2009; Sonnenburg *et al.,* 2010).

No detectable growth defect is displayed by the *Δbacova_04502* knockout strain or the *Δbacova_04503 E196Q* catalytic mutant on inulin despite loss of surface endo activity (Chapter 4.9.3), which was unexpected. It is likely that either the inulin substrates used during this project are smaller or otherwise not representative of the inulin targeted by the system *in vivo;* or that this enzyme, including the CBM, provide a significant competitive or symbiotic advantage to *B. ovatus in vivo* which cannot be detected with the techniques used in this study. DFA-FOS is produced by the GH91 enzyme (BACOVA_04502/3) and the DFA moiety is inaccessible to *B. ovatus*; this means that *B. ovatus* is losing large quantities of otherwise accessible carbon and energy through the action of the GH91 enzyme. As this enzyme is not required for inulin utilisation it is highly unlikely that this activity would occur without a selective advantage.

One possibility is that large inulin polymers may be present within the intestinal environment and these may only be utilised by after initial processing by endo-acting enzymes, either to break large substrates into smaller chunks which may be imported or to release soluble material from insoluble bulk. In this instance, the production of the waste product, DFA, maybe compensated by access to an otherwise inaccessible glycan.

Another theory is that the GH91 enzymes confer a selective advantage to *B. ovatus* through allowing rapid processing of inulin, simply by generating oligosaccharides inulin can be harvested more quickly and sequestered away from competitors, however as oligosaccharides are apparent in the media, ready for scavenging by competing microbes, this hypothesis seems unlikely.

Finally, the GH91 may confer a selective advantage through beneficial sharing interactions with other endogenous species. *B. ovatus* releases inulin breakdown products into the intestinal lumen which may be accessed by other microbes (Rakoff-Nahoum *et al.*, 2014). We demonstrated that the GH91 inulin breakdown products were preferential substrates for two *Bifidobacterium* species (Appendix I. **Figure I.7**).

Many efficient FOS utilising species are unable to directly access polymeric inulin (Watson *et al*., 2013), we speculate that by generating substrate for FOS utilising strains in the immediate area, *B. ovatus* may dilute out competitors for the polymer, ensuring that a specific niche for an inulin-sharing microbe such as itself is retained. In this scenario, DFA may directly feed a symbiont or may simply be a justifiable loss of energy during the sharing process, as substrate will be lost to competitors through this mechanism by definition. This mechanism is directly opposed to the recently elucidated mechanism adopted by *Ruminoccous gnavus* which creates a 2,7-anhydro-Neu5Ac instead of sialic acid during cleavage of sialic acid from host glycans; by creating an anhydrous product the organism prevents species other than itself from breaking down the liberated sugar (Tailford *et al.*, 2015).

Further experiments should be undertaken to test these hypotheses. The catalytic mutant strain, *Δbacova_04503 E196A* should be tested with insoluble and recalcitrant material to see if a growth defect is observed. Further experiments to test the role of the GH91 *in vivo* could be conducted by inoculating germ-free mice with both wild type and *Δbacova_04503 E196A* strains of *B. ovatus*. If the GH91 mutant is unable to colonise the host in the presence of the wild-type strain during an inulin rich diet, it can be concluded that the GH91 enzyme is vital for normal inulin usage *in vivo*. Competition assays using humanised mice could be undertaken, to demonstrate whether the GH91 enzyme confers an advantage only in the presence of other members of the microbiota (e.g. due to sharing).

4.12.2.4. The Role of DFA in the Gut

Whatever the benefit of the GH91 enzyme, the down-stream effects of its cleavage product, DFA, remain a mystery. There is no evidence that DFA is utilised by other members of the microbiota as our data shows DFA is not broken down by any of the three faecal cultures (**Figure 4.36**).

There may be an underlying symbiotic interaction between *B. ovatus* and the host; DFA has been shown to enhance iron and calcium absorption through the host epithelium (Hara *et al.*, 2008; Mineo *et al*., 2002; Mineo *et al*., 2003). However it is difficult to speculate on how *B. ovatus* may benefit from such an interaction without further investigation into the role of DFA within the gut.

4.12.3. Sucrose Recognition and the SusD-homologue, BACOVA_04504

BACOVA_04504 binding to the terminal sucrose moiety was unexpected, as sucrose is a minor component of the inulin chain, and will be lost after the polysaccharide is broken down into shorter fructo-oligosaccharides e.g. by acid hydrolysis in the stomach or by bacterial enzymes such as the GH91 (BACOVA_04502/3) presented by *B. ovatus*. Other SusD-homologues have been shown to target the breakdown products created by surface localised endo-acting enzymes, and BACOVA_04504 defies this paradigm.

An explanation for this unusual characteristic may be that BACOVA_04504 is recognising short chain fructo-oligosaccharides for direct import. Short FOS are created by many plants and can be seen in the extracts used in this study (onion, wheat and garlic). Short FOS are used by more members of the microbiota than inulin polysaccharide (Watson *et al*., 2013) by targeting short chain FOS for direct import *B. ovatus* may sequester this desirable carbohydrate away from competitors.

Additionally, we speculate that recognition of a sucrose terminus, may allow the SusD-homologue to import short FOS of either β2-1 or β2-6 linkage. Such substrates would be small enough not to require extracellular processing, and could be processed by the periplasmic GH32, BACOVA_04501.

Future work should be undertaken to examine whether BACOVA_04504 is specific to β2-1 linked FOS or whether it is also able to bind to sucrose terminated β2-6 levan oligosaccharides using ITC analysis.
This experiment could not be conducted, as sucrose terminated levan oligosaccharides could not be successfully created. Co-cultures of *Δbacova_04504* with wild-type on inulin and FOS could be tried to assess the impact of the loss of BACOVA_04504 in a more competitive environment.

4.12.4. Evaluating the Role of the putative extracellular GH32, BACOVA_04507

The GH32 BACOVA_04507 does not have a clear role within the *B. ovatus* inulin utilisation locus. The enzyme appears to be presented on the cell surface (**Figure 4.28**) and cleaves short chain fructo-oligosaccharides and sucrose (**Table 4.2**). No phenotype was observed in the genomic knockout strain, *Δbacova_04507 D265A* (**Figure 4.32.**).

The presence of a homologue of BACOVA_04507, BT1765, within the *B. thetaiotaomicron* levan PUL strongly suggests that the role of this protein is non-linkage specific (**Figure 4.2**, **Figure 4.37**).

Both wild-type *B. thetaiotaomicron* (unpublished data) and the *B. ovatus* SusC-homologue knockout (**Figure 4.33**), *Δbacova_04505* are able to utilise inulin after a significant lag phase (>24 hours).

Despite lacking the protein apparatus to import this glycan; it is possible that in both cases the surface located GH32 homologues, BT1765 and BACOVA_04507, are able to release enough free fructose from inulin to support growth, however this does not explain why growth occurs after a lag rather than gradually.

The role of this enzyme is unclear, and relatively unexplored. Sucrose is likely to be completely removed by host absorption in the early G.I. tract and thus seems to be an unlikely target substrate; future experiments should be undertaken to determine whether this protein is targeting a different small fructan. BACOVA_04507 may remove fructose from raffinose, a trisaccharide comprised of $F_{2\beta}\rightarrow_{1\alpha}G_{\alpha 6}\rightarrow_{1}Gal$.
The SusD-homologue, BACOVA_04504 did not bind to raffinose during ITC analysis, which was expected as the galactose unit likely blocks interaction with the sucrose moiety (data not shown). Whilst *B. ovatus* can grow on raffinose, culture supernatant profiles differed between *B. ovatus, B. caccae* and *B. xylanisolvens*, these species contain highly similar inulin PUL (**Figure 4.38**) and different breakdown patterns suggested that the fructan PUL was not involved in raffinose processing (data not shown) and we did not explore this substrate further

4.12.5. Towards a Model for Inulin Utilisation by Bacteroides ovatus

Our data outlined in this chapter allow us to demonstrate a model for inulin utilisation by *B. ovatus* (**Figure 4.41**).

Polymeric inulin bound at the cell surface by the CBM appended to the heteromeric GH91 family enzyme (Shapiro, MRes thesis, 2012). The GH91 enzyme is comprised of two gene products, BACOVA_04502 and BACOVA_04503 and is an endo-acting inulinase which creates an anhydrous terminus (Chapter 4.6). The SusD-homolog, BACOVA_04504, recognises sucrose terminated FOS and inulin at the cell surface, likely targeting desirable FOS for direct import (Chapter 4.7, Chapter 4.12.3).

GH91 breakdown products, polymeric inulin and FOS are imported through the SusC-homolog, BACOVA_04505 (Chapter 4.9.1). Imported fructan is deconstructed to component monosaccharides by a periplasmic GH32 family enzyme, BACOVA_04501. Periplasmic fructose is bound by the periplasmic binding domain of the HTCS which presumably propagates a conformational change through the proteins transmembrane region and cytoplasmic domains, triggering up-regulation of the locus (Sonnenburg *et al.*, 2010; Bolam & Sonnenburg, 2011; **Figure 1.17**; Chapter 4.5). An inner-membrane fructose permease and cytoplasmic fructose kinase import free fructose into the cytosol and phosphorylate it for downstream metabolic use (Chapter 4.4). This model for inulin utilisation is illustrated (**Figure 4.41**).

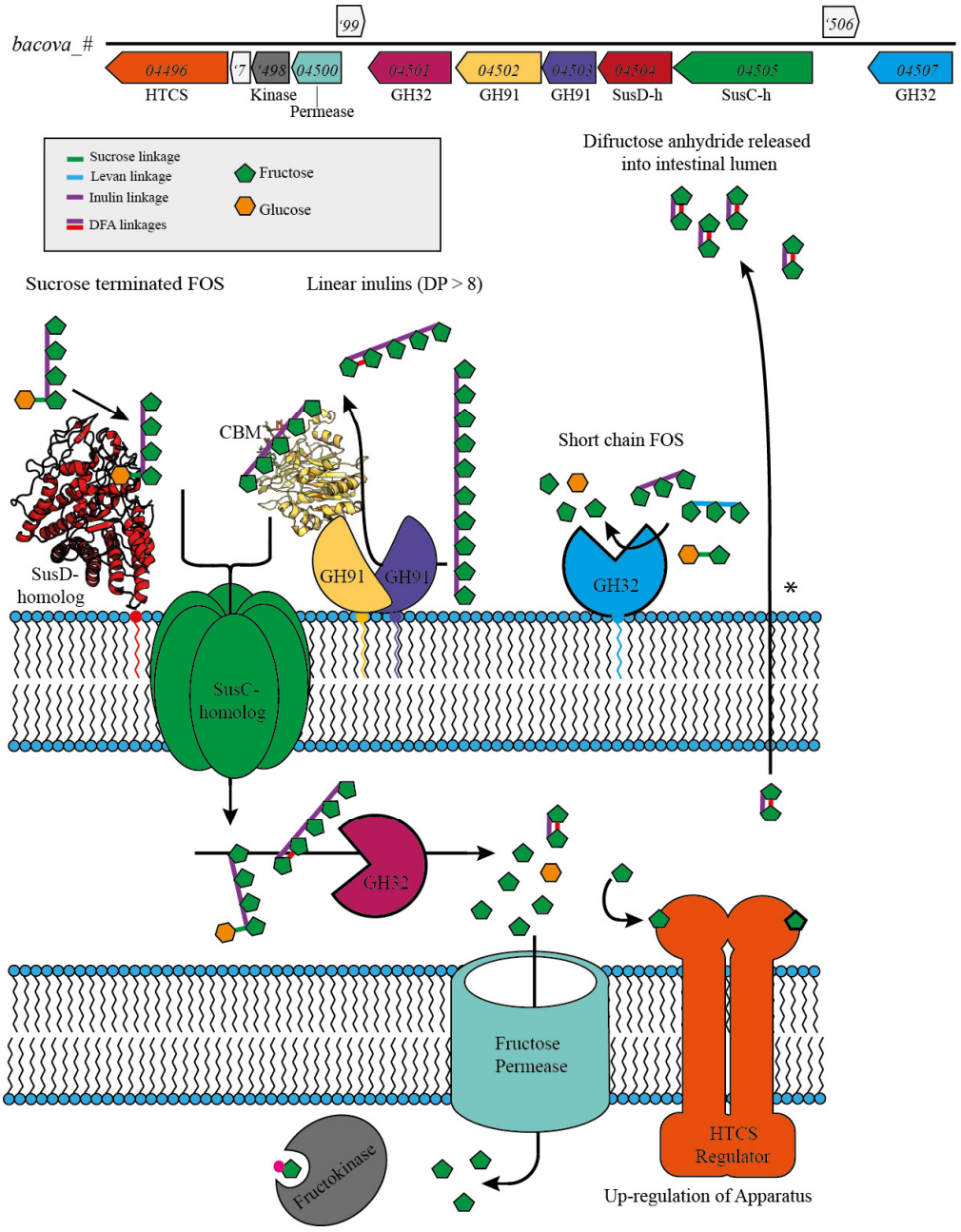


Figure 4.41. Proposed model for inulin utilisation by *Bacteroides ovatus*. Inulin is recognised at the cell surface by an inulin binding domain appended to the GH91 enzyme and is then broken into DFA-FOS by the GH91 enzyme. Sucrose terminated FOS are recognised by the SusD-homologue for direct import. FOS and DFA-FOS is imported through the TonB dependent SusC-homologue into the periplasm, where the periplasmic GH32 depolymerises imported fructans. Fructose is recognised by the periplasmic domain of the HTCS sensor-regulator, and induces PUL expression. Fructose is imported to the cytoplasm through a fructose permease and phosphorylated by the fructokinase for further downstream metabolism by the cell. * DFA is exported across the membrane by an unknown mechanism into the environment.

Chapter 5. Inulin Utilisation by *Bifidobacterium adolescentis*

5.1. Chapter Overview

Bifidobacterium are the most common targets for inulin and FOS prebiotics, which can increase the relative abundance of this genus, despite making up only a small portion of the microbial community (Schaafsma & Slavin, 2015; Eckburg *et al.*, 2005). Members of the *Bifidobacterium* genera are particularly prominent in the healthy infant microbiota and increases in *Bifidobacterium* populations in adults have been associated with therapeutic effects including a lowering of intestinal pH, the production of vitamins and the generation of short chain fatty acids (Gibson & Roberfroid, 1995; Koenig *et al.*, 2011).

Inulin and fructo-oligosaccharides have a bifidogenic effect on the colonic microbiota when consumed, leading to the use of these glycans to modulate the microbiota towards health (Roberfroid *et al,.* 1998; Gibson *et al.*, 2004). Most *Bifidobacterium* species are able to use FOS and inulin is able to stimulate the genus *in vivo*, despite a relatively narrow distribution of *Bifidobacterium* species able to directly utilise the polysaccharide (Schaafsma & Slavin, 2015; Watson *et al.*, 2013). Despite concentrated efforts to stimulate *Bifidobacterium* populations *in vivo* using prebiotics, very little mechanistic insight into inulin utilisation in the genus has been undertaken.

Several glycan utilisation clusters, containing predicted enzymes, substrate binding proteins and ABC permeases, were identified in *Bifidobacterium longum subspecies infantis* (Garrido *et al.*, 2011). Garrido *et al.*, show up regulation of these clusters using qPCR analysis after growth on various glycans, these data demonstrated that these clusters were regulated in response to environmental glycans. It was discovered that this species has a preference for human milk oligosaccharides over plant glycans such as FOS (Garrido *et al*., 2011).

The strategies employed by *Bifidobacterium spp.* to degrade environmental glycans appear varied. Degradation of the HMO lacto-*N*-tetraose in *B. bifidum* requires an extracellular enzyme, lacto-*N*-biosidase, to cleave the tetramer into two disaccharides which are subsequently imported, however *B. longum subspecies infantis* imports the tetramer for internalised degradation (Marcobal & Sonnenburg, 2013; Sela & Millis, 2010; Yoshida *et al.*, 2012). Other species, such as *B. breve* require the action of other members of the intestinal microbiota to degrade HMOs into monosaccharide components before it may import them for energy (Sela & Millis, 2010).

A GH32 enzyme from *Bifidobacterium longum* KN29.1 was identified and characterised, kinetic parameters were elucidated in addition to a crystal structure of the enzyme (Bujacz *et al.*, 2011; Jedrzejczak-Krzepkowska *et al.,* 2011). It was demonstrated that this species could efficiently utilise short chain FOS, a phenotype confirmed by Watson and colleagues in 2013. The GH32 enzyme displayed a β propeller fold typical of the GH32 family, and was able to liberate fructose from non-reducing chain termini through the retaining mechanism (Jedrzejczak-Krzepkowska *et al.,* 2011, Bujacz *et al.*, 2011). It was not clear in these studies if the GH32 was expressed intra- or extracellularly.

Bifidobacterium employ ABC transport systems to import mono- and oligo-saccharides across the membrane. Several ABC transporter associated Extracellular Substrate Binding Proteins (ESBPs) have been identified in *Bifidobacterium* and other Gram positive genera which have been associated with glycan uptake. ESBPs are generally composed of two discrete domains joined by a hinge region, upon ligand binding the ESBP undergoes a significant conformational change as these two domains move relative to one another around the hinge region to form a binding pocket around the ligand (Berntsson *et al.*, 2010). Crystal structures of ESBPs are elucidated in either open form, with no ligand bound, or closed form, generally with a ligand in the active site; both structures of an ESBP must be obtained to observe the conformational changes undergone during ligand binding (Berntsson *et al.*, 2010).

Cockburn *et al.,* demonstrated with the inclusion of a closed-form crystal structure, the role of a starch binding ESBP during starch scavenging by *E. rectale* (Cockburn *et al.*, 2013). The authors demonstrate that initial starch degradation takes place though a peptidoglycan associated enzyme prior to ABC permease mediated oligosaccharide import (Cockburn *et al.*, 2013). Another ESBP structure was elucidated in closed form which is crucial to arabinoxylooligosaccharide uptake by *Bifidobacterium animalis subspecies lactis* for internal degradation (Ejby *et al.*, 2013; Van Den Broek *et al.*, 2005). At present, few ESBPs which bind to oligosaccharides have been characterised and the conformational shift in these proteins has not yet been observed; furthermore no fructan binding ESBP has yet been characterised.

Bifidobacterium adolescentis was identified as able to utilise inulin polysaccharide directly (Watson *et al.*, 2013). This phenotype was of particular interest as direct utilisation of the inulin polysaccharide appears to be relatively uncommon in the genus. Furthermore very little mechanistic insight has been elucidated between the most commonly targeted probiotic genera, *Bifidobacterium,* and the most commonly used prebiotic class, fructans. Understanding how this mechanism works, particularly how inulin is recognised by the cell, may underpin future prebiotic design towards more personalised or more targeted prebiotics.

In this chapter we identify putative fructan associated genes within *B. adolescentis* and show that the expression of several of these genes is induced in the presence of inulin and FOS. Amongst the inulin inducible genes, we identify a GH32 family enzyme (BAD_1325) which is active on inulin, a LacI-homologue (BAD_1326) which bound to fructans during ITC analysis and an ESBP (BAD_1330) which bound to inulin and was successfully crystallised in both open and closed conformation. These data demonstrate the fructan-associated roles within the gene cluster, particularly the BAD_1330 ESBP reveals how long chain inulin is recognised at the cell surface prior to import through the membrane for internalised degradation.

5.2. Objectives

- To identify putative inulin associated genes from *B. adolescentis*
- To characterise the GH32 enzymes deployed during fructan degradation by *B. adolescentis*.
- To further understand how inulin is recognised at the cell surface.
- To determine whether or not inulin is broken up prior to import through the cell membrane.
- To propose a model for inulin capture and utilisation by *B. adolescentis*.

5.3. Growth of *Bifidobacterium adolescentis* and *Bifodobacterium longum* on Inulin and FOS.

B. adolescentis and *B. longum* were cultured upon minimal media supplemented with glucose, fructose, sucrose, inulin and FOS to confirm the growth phenotypes observed by Watson and colleagues in 2013. Previously, *B. longum* was found to only utilise short chain fructans, whereas *B. adolescentis* was capable of utilising FOS and inulin polysaccharide (Watson *et al*., 2013). Glucose, fructose and sucrose were used as controls to visualise robust growth of each species. We wished to focus on *B. adolescentis* due to the comparatively unusual ability to utilise inulin polysaccharide; however *B. longum* serves as a negative control for this inulin utilisation phenotype and allows for comparison. Our data matched previous findings; *B. adolescentis* was able to utilise inulin, whereas *B. longum* could not. *B. longum* utilised FOS, but did not achieve the same final OD on this substrate as on glucose, suggesting a portion of this FOS could not be utilised (**Figure 5.1 A**). To determine the maximum chain length of FOS which could be accessed by *B. longum* it was grown upon defined oligosaccharides and inulin, our data show that *B. longum* is able to utilise the trisaccharide kestose but not the tetrasaccharide kestotetraose (**Figure 5.1 B**).

Bifidobacterium species proved more difficult than *Bacteroides* to culture within the smaller volumes used during plate reader analysis, and *B. adolescentis* consistently failed to achieve the same final OD as *B. longum* even on the monosaccharide controls, however robust growth comparable to *B. longum* was observed in larger (5 ml) culture volumes.

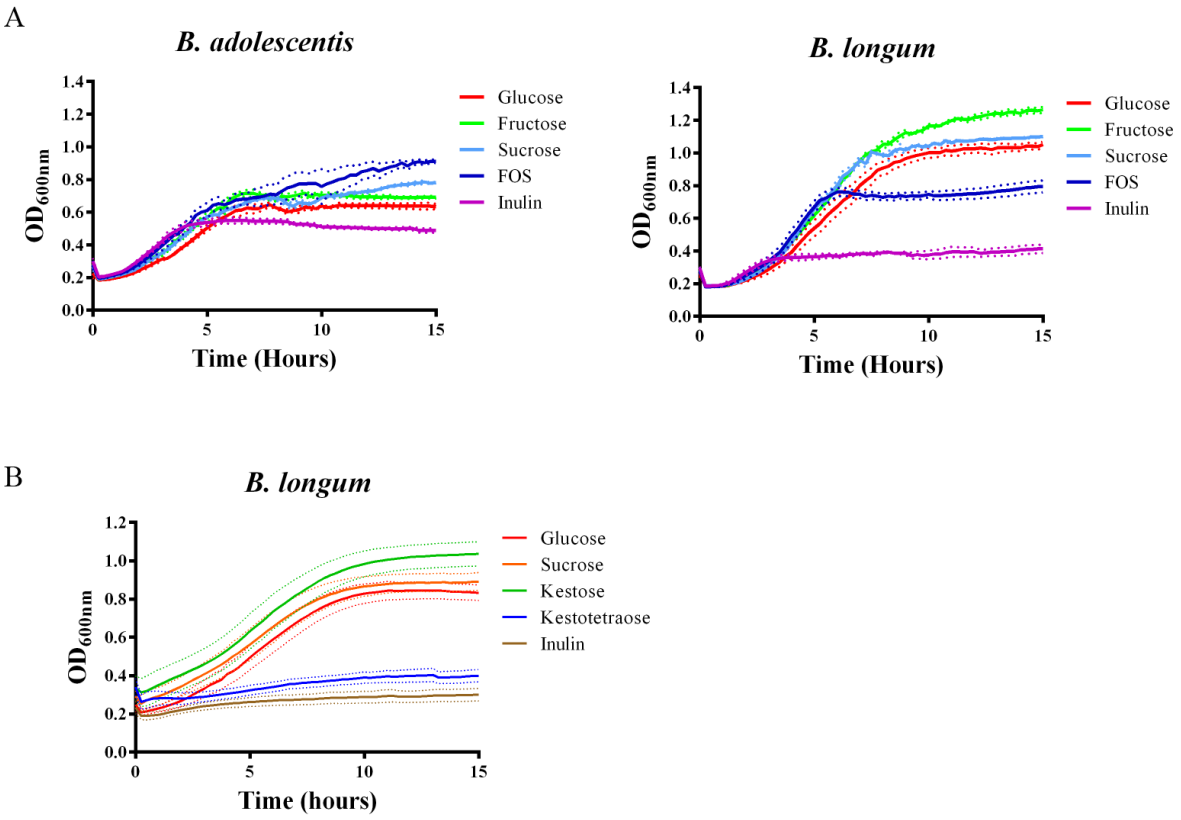


Figure 5.1. The growth of *Bifidobacterium adolescentis* and *Bifidobacterium longum* upon glucose, fructose, sucrose, FOS and inulin. (**A**) *B. adolescentis* and *B. longum* were cultured in minimal media supplemented with 0.5 % glucose, fructose, sucrose, FOS (DP 3-4) or inulin. *B. adolescentis* was able to grow on all substrates in a similar manner, where *B. longum* was unable to utilise inulin, and could not fully utilise FOS. *B. adolescentis* consistently grew to a lower OD than *B. longum* when cultured in the small (250 µl) volumes used in conjunction with the plate reader. (**B**) As *B. longum* grew to a lower final OD on FOS than fructose and glucose we further explored FOS utilisation by growing *B. longum* on FOS of defined length, here we show that *B. longum* can use kestose (DP = 3) but not kestotetraose (DP = 4), explaining the lower final OD observed in panel A. Each experiment was conducted in triplicate and average values plotted. The experiment was repeated with consistent growth phenotypes.

Both species were grown in minimal media supplemented with 0.5 % of either inulin, FOS (DP = 3-4) or fructose (**Figure 5.2**).
After 24 hours the OD_{600nm} of each culture was noted before the spent supernatant was visualised via TLC. Media only controls were set up for comparison and to ensure no contamination occurred during the growth period.

The spent supernatants are in agreement with the growth data and show that *B. longum* is able to utilise FOS but not inulin whilst *B. adolescentis* utilised both. Breakdown products are detected in the spent supernatant, this activity is more pronounced during growth on FOS of both species with comparably little fructose release during inulin processing by *B. adolescentis* (**Figure 5.2**).

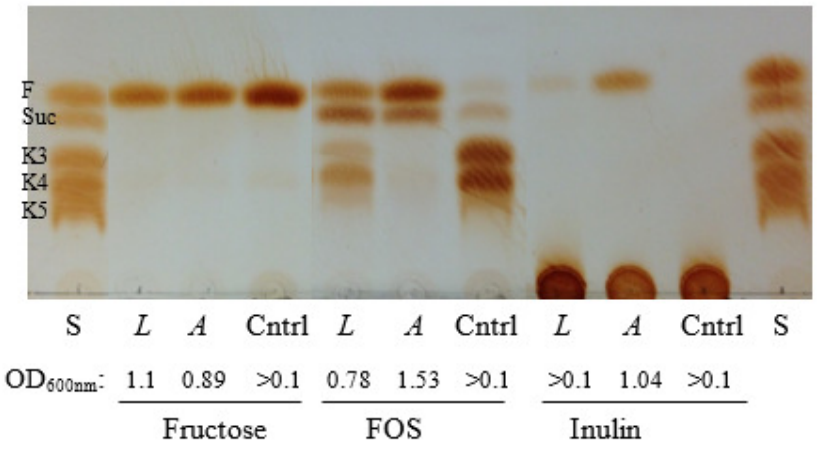


Figure 5.2. Spent culture supernatant fluid from the growth of *B. longum* and *B. adolescentis* on inulin and FOS. *B. longum* (*L*) and *B. adolescentis* (*A*) were cultured upon minimal media supplemented with fructose, FOS (DP = 3-4) or inulin and the spent supernatant fluid visualised using TLC. An increased cell density (OD_{600nm}) correlates with a reduction of glycan remaining in the media as expected. Standards (S) consisting of 2 mg/ml each of fructose (F), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) were visualised for comparison, as were cultures consisting of media and sugar only (Cntrl). *B. adolescentis* is able to utilise both FOS and inulin, whereas *B. longum* is only able to utilise the shorter FOS substrate. Duplicate cultures were used to ensure consistent growth phenotype, only one of these cultures was used for supernatant fluid analysis however no differences were noted between replicate final OD.

5.4. Identification of Putative Inulin Utilisation Genes

5.4.1. Two Putative Gene Clusters May be Involved in Inulin Utilisation

The genomes *Bifidobacterium longum subspecies longum* JCM 1217 and *Bifidobacterium adolescentis* ATCC 15703 were used throughout this chapter, these were avaliable on the Integrated Microbial Genomes database having been sequenced by the HMP Reference Genome Project and Suzuki *et al.* respectively (HMP Consortium, 2010; Suzuki *et al.*, 2006). Whilst only one GH32 was found within *B. longum* (BLLJ_1341); two GH32 enzymes are present within the *B. adolescentis* genome (BAD_1150 and BAD_1325), these enzymes have 28 % identity to each other.
The first GH32, BAD_1150, shared 84 % identity with the only GH32 from *B. longum* (BLLJ_1341); suggesting that these enzymes are likely to perform the same role. The second GH32, BAD_1325, was more divergent from this enzyme, sharing only 29 % identity.

The GH32 enzyme previously characterised by Bujacz and colleagues was from *B. longum subspecies longum KN29.1* (Bujacz *et al.*, 2011). The strain of *B. longum* used in this thesis was *B. longum subspecies longum JCM 1217*, however, the characterised GH32 and the GH32 from retain the strain used in this thesis (BLLJ_1341) had 99.03 % identity with each other and are unlikely to differ significantly.

Both GH32 encoding genes were embedded within a genomic context which had the potential to be geared towards glycan utilisation as genes annotated as ABC permeases and a LacI-homologue were nearby. The first gene cluster (BAD_1148 → BAD_1152) was highly similar (**Figure 5.3**) to the cluster containing the characterised GH32 homologue from *B. longum* (BLLJ_1341). It was therefore hypothesised that this cluster was involved in short chain FOS acquisition and operates in the same manner as the cluster from *B. longum*. The second gene cluster was more divergent from this cluster and was hypothesised to confer the

extra ability to utilise inulin seen in *B. adolescentis* but not *B. longum* (**Figure 5.1, Figure 5.3**).

The first cluster contains 5 genes annotated as a LacI-homologue, a sucrose permease, the GH32 (*bad_1150*), an ABC transporter permease and an extracellular substrate binding protein (ESBP). The second cluster contains six genes; three ABC transporter permeases, an ESBP, a LacI class regulator and the predicted GH32 (*bad_1325*) as illustrated (**Figure 5.3**).

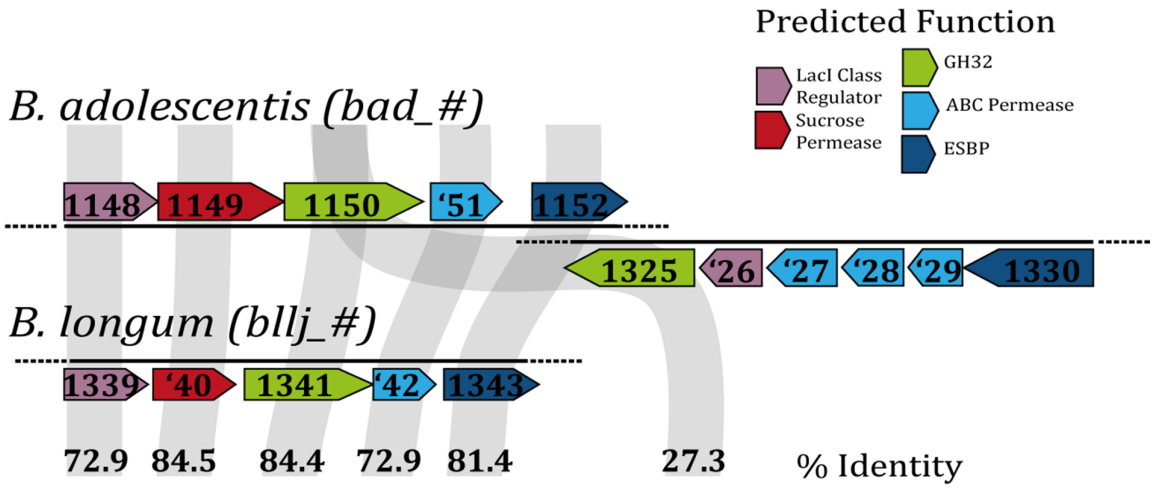


Figure 5.3. Comparison of predicted fructan utilisation loci in *B. adolescentis* and *B. longum*. The two putative gene clusters identified within the *B. adolescentis* genome are displayed (Top panel: *B. adolescentis*, bottom panel: *B. longum*). The first cluster contains 5 genes which retain high similarity to the 5 genes encoded by *B. longum* (ID > 70 %) and were therefore expected to perform the same role. The second cluster contains 6 genes which were divergent with *B. longum* genes. The second, more divergent, gene cluster was hypothesised to be involved in inulin capture and degradation.

5.4.2. qPCR Reveals Substrate Induced Gene Expression

To confirm whether either of these clusters was involved in inulin or FOS acquisition, cultures were grown in the presence of glucose, FOS and inulin before harvest at mid-exponential phase. RNA extracted from these cultures was quantified using qPCR and the fold change between glucose grown cultures and FOS or inulin grown cultures were measured. The data show a clear up-regulation

of the second gene cluster in the presence of inulin and FOS, however neither substrate appears to induce the expression of the first cluster (**Figure 5.4**). Raw data is plotted (Appendix I, **Table I.3**).

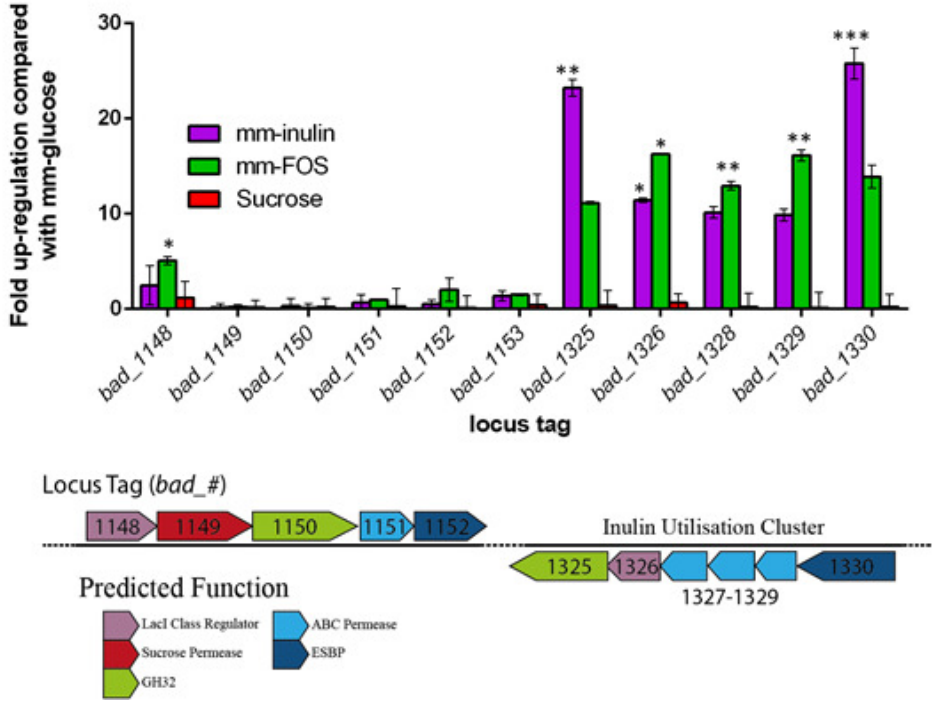


Figure 5.4. qPCR shows the inducible expression of genes from the second gene cluster in the presence of inulin and FOS. *B. adolescentis* was grown on minimal media supplemented with glucose, inulin, FOS or sucrose in triplicates. RNA extracted from these cultures was quantified through qPCR. qPCR was performed in triplicate and the data averaged with standard error. Probe sets for genes within both gene clusters (*bad_1148* → *bad_1153*; *bad_1325* → *bad_1330*) were used, alongside probes targeting the 16S subunit. Cq values were normalised for cell number using the 16S probe set. Each gene under inulin, FOS and sucrose conditions were compared to the same gene under glucose conditions, and the fold change recorded. These data show the second gene cluster (*bad_1325* → *bad_1330*) is up-regulated with all genes undergoing a significant fold change on both FOS and inulin but not sucrose. The first gene cluster however was not upregulated on any substrate tested. Statistical significance was examined using a one-way ANOVA between the glucose mean and the experimental mean for each condition, and is shown (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$)

As only genes from the second gene cluster (*bad_1325* → *bad_1130*) were induced by inulin, this cluster became the focus of the study and is henceforth regarded as the inulin utilisation cluster. The first cluster, (*bad_1148* → *bad_1153*) does not appear to be involved during inulin utilisation in *B. adolescentis* but the GH32 from *B. longum* (BLLJ_1341) is almost certainly

involved in FOS processing as it is the only GH32 in the genome; therefore this cluster is henceforth referred to as the FOS utilisation cluster.

5.5. Characterisation of Two GH32 Enzymes

The two GH32 enzymes, BAD_1150 and BAD_1325, were cloned, expressed and raw kinetic data generated by a summer student, Ms. Harriet Lane, under my supervision. Cloning and expression in recombinant *E. coli* were undertaken as described (Chapters 2.9.7 and 2.11). Proteins were purified using IMAC (Chapter 2.12.1). Further construct details are listed (Appendix I.8). Data generated during this project was analysed by myself.

The FOS-active GH32 (BAD_1150) and the inulin-active GH32 (BAD_1325) were cloned and expressed using pET-28b vector. Enzyme assays were undertaken using fructose detection through a linked enzyme assay kit (**Figure 5.5, Figure 5.6**). Here we show that BAD_1150 is highly active on short chain FOS but has much lower activity on sucrose and inulin, these data are similar to the findings of Bujacz *et al.*, 2011 and confirm the hypothesis that the two highly similar GH32 family enzymes (BLLJ_1341 and BAD_1150) are functionally identical β-fructosidase enzymes which target FOS (**Table 5.1**, **Figure 5.5, Figure 5.6**, Bujacz *et al*., 2011). BAD_1325 has increased activity on FOS and is poor on sucrose. Whilst neither enzyme was inactive on levan, the activity on this substrate was very low in both cases (**Table 5.1**, **Figure 5.5, Figure 5.6**). These data are consistent with the hypothesis that BAD_1325 plays a role in the utilisation of inulin, and BAD_1150 does not.

Table 5.1. Kinetic parameters of the two GH32s (BAD_1150 and BAD_1325)

	BAD_1150			**BAD_1325**		
	K_{cat} (Min-1)	K_m (mM)	K_{cat}/K_m (min^{-1}/mM^{-1})	K_{cat} (Min-1)	K_m (mM)	K_{cat}/K_m (min^{-1}/mM^{-1})
Sucrose	158 ± 9	2.9 ± 0.5	55	325 ± 25	7.5 ± 1.2	44
Kestotetraose	221 ± 6	0.1 ± 0.09	2353	338 ± 17	0.2 ± 0.05	1770
Inulin*	108 ± 5	5.4 ± 0.015	20	335 ± 8.5	1.2 ± 0.12	279
Levan*	3 ± 0.5	0.275 ± 0.05	10.9	N/A	N/A	96

* Concentration of inulin and levan were estimated as described in Chapter 2.13.2.

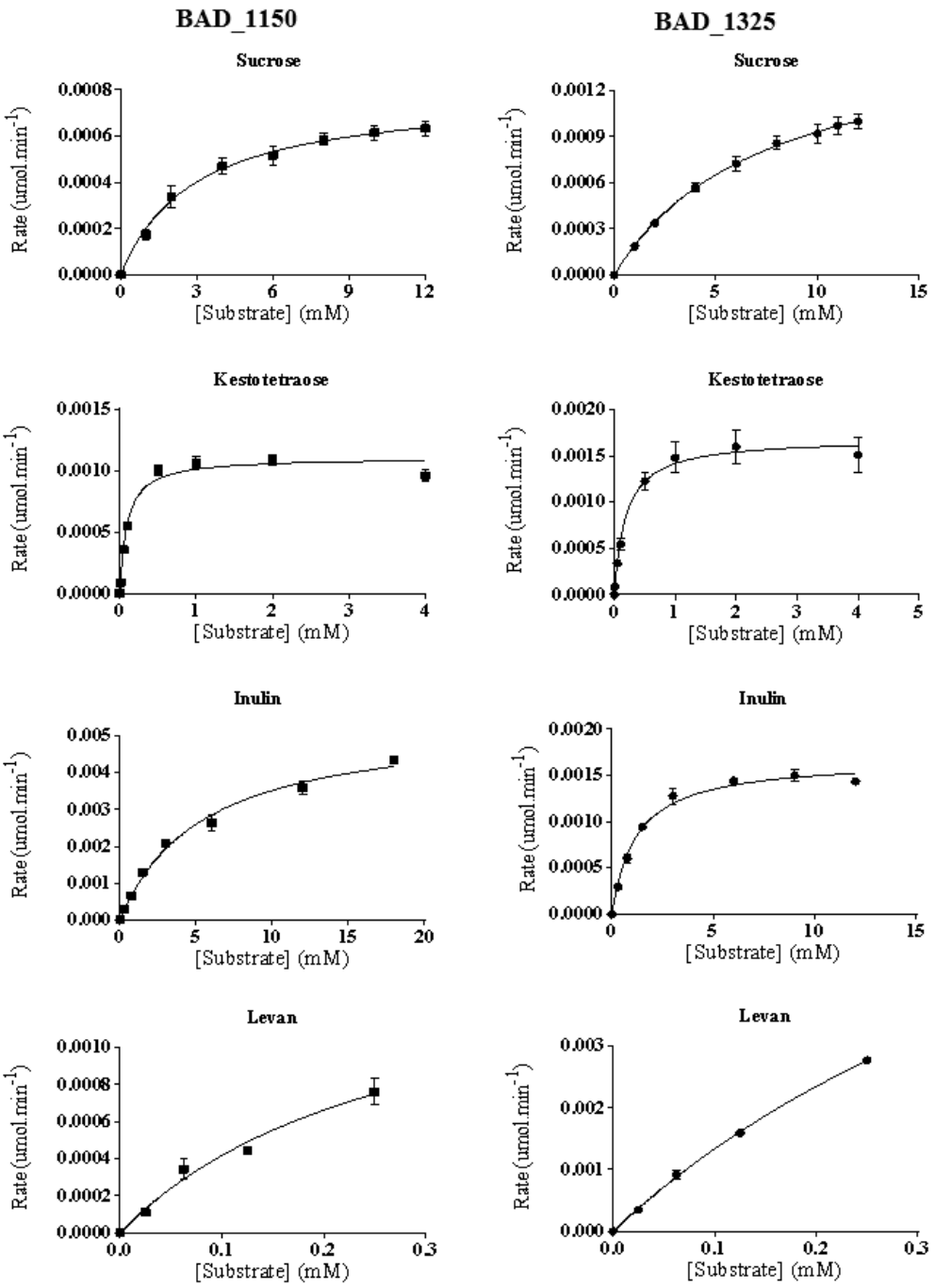


Figure 5.5. Michaelis-Menton plots demonstrate the activity of two GH32s, BAD_1150 and BAD_1325 on fructans. Both BAD_1150 and BAD_1325 were active on sucrose, inulin, FOS and Levan. All enzyme assays were conducted in triplicate. Two separate protein preparations of each enzyme were used to ensure reproducibility. Data were averaged, and standard deviation shown. Concentrations of inulin and levan were calculated as outlined (Chapter 2.13.2).

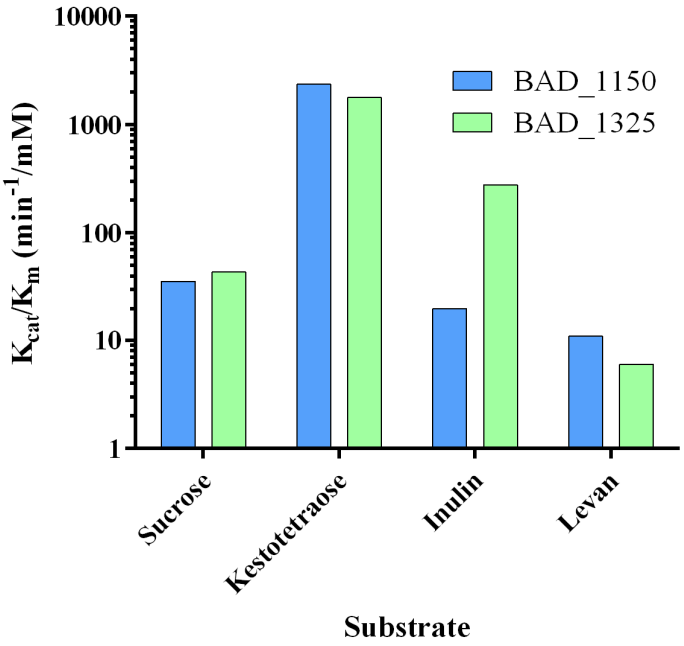


Figure 5.6. Comparison of enzyme rates (K_{cat}/K_m) for GH32 family enzymes. Illustrated are the average specific activities for each GH32 enzyme described in the text (BLLJ_1341, BAD_1150 and BAD_1325) on sucrose, kestotetraose, inulin and levan. BLLJ_1341 and BAD_1150 share 84 % sequence identity and appear to be functionally identical, BAD_1325 has a higher specific activity on longer substrates.

5.6. Specificity of the LacI-homologue, BAD_1326

A final year undergraduate project was undertaken by Mr Alex Bond with the aim of cloning, expressing and generating ITC data for the binding domain of the inulin inducible LacI-homologue regulator, BAD_1326. The project was supervised by Dr David Bolam and myself. Mr. Carl Morland generated replicate ITC data for the purposes of reproducibility. Cloning and expression using recombinant *E. coli* was undertaken as described (Chapters 2.9.7 and 2.11). Proteins were purified using IMAC (Chapter 2.12.1). Further construct details are listed (Appendix I.8).Data analysis was undertaken by myself.

The full length construct of the LacI homologue, BAD_1326 (1 aa – 362 aa), was insoluble during over expression in *E. coli*. A truncated construct, based upon an alignment with LacI was thus generated (73aa – 362aa) in order to remove the putative DNA binding domain and retain only the substrate binding domain (**Figure 5.7**). This construct was expressed successfully and the resultant protein was soluble.

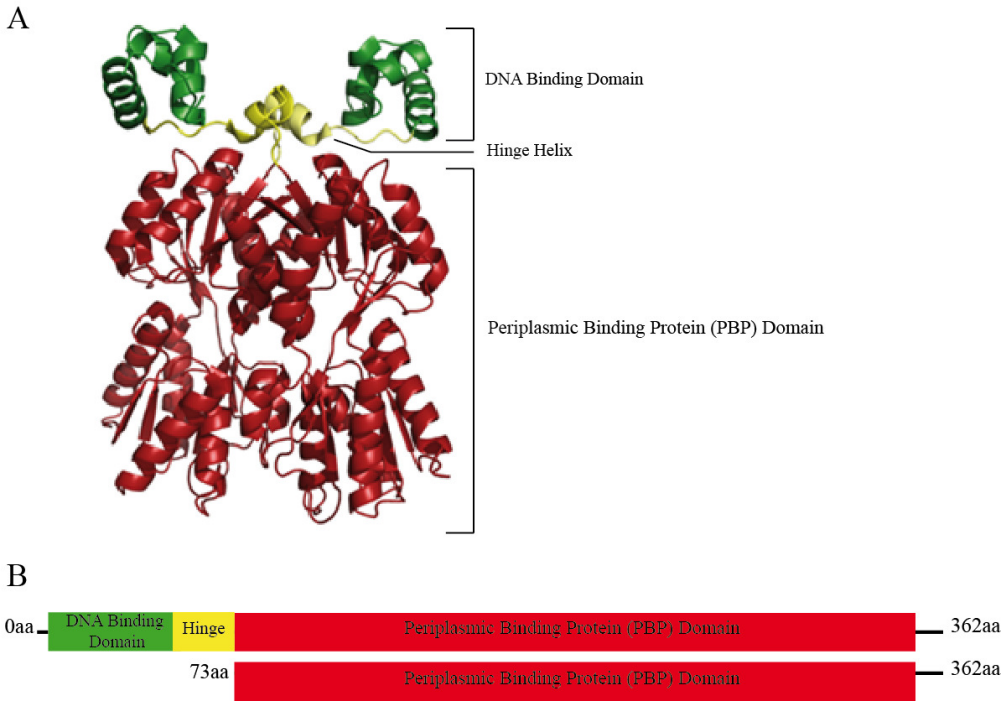


Figure 5.7. Domain structure of the LacI Repressor and LacI homologue, BAD_1326. (**A**) Crystal structure of the LacI Repressor from *E. coli* (PDB ID: 1LBG), showing the DNA binding, hinge and substrate binding domains (Lewis, 2005; Lewis *et al.*, 1996). (**B**) Predicted domain structure of the LacI homologue, BAD_1325 (top) based upon LacI, and the construct used for these experiments (bottom).

The LacI-homologue substrate binding domain bound to FOS (K_d ~ 1 µM), inulin (K_d ~ 7 µM) and levan (K_d ~ 9 µM) but did not bind to fructose or sucrose (**Figure 5.8**). Affinity for oligosaccharides increased as DP increased, suggesting that longer substrates are the preferred binding target (**Table 5.2**).

Table 5.2 Binding Parameters of the LacI homologue substrate binding domain (BAD_1326), from ITC Datasets

Ligand	N	K_a (x10^4 M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	TΔS (kcal mol^{-1})
Fructose			No Binding		
Sucrose			No Binding		
Kestose	1 ±0.1	2.3 ±0.0	-5.9	-4.8 ±0.5	-1.1
Kestotetraose*	0.6 ±0.2	4.7 ±1.5	-6.4	-7.6 ±2.4	-1.2
Kestopentaose	0.8 ±0.4	5.1 ±1.2	-6.4	-12 ±2.9	-5.6
Inulin*+	1 ±0.2	1.5 ±0.2	-5.7	-5 ±1.1	+0.7
Levan*+	0.7 ±0.4	1.1 ±0.3	-5.5	-7.7 ±4.6	-2.2

* In some cases (kestotetraose, inulin, levan) only one ITC trace was generated. In these cases, errors are from the fit to the model. For all other conditions at least 2 datasets were generated and standard deviations were calculated.
+ Concentration of inulin and levan were estimated as described in Chapter 2.13.2.

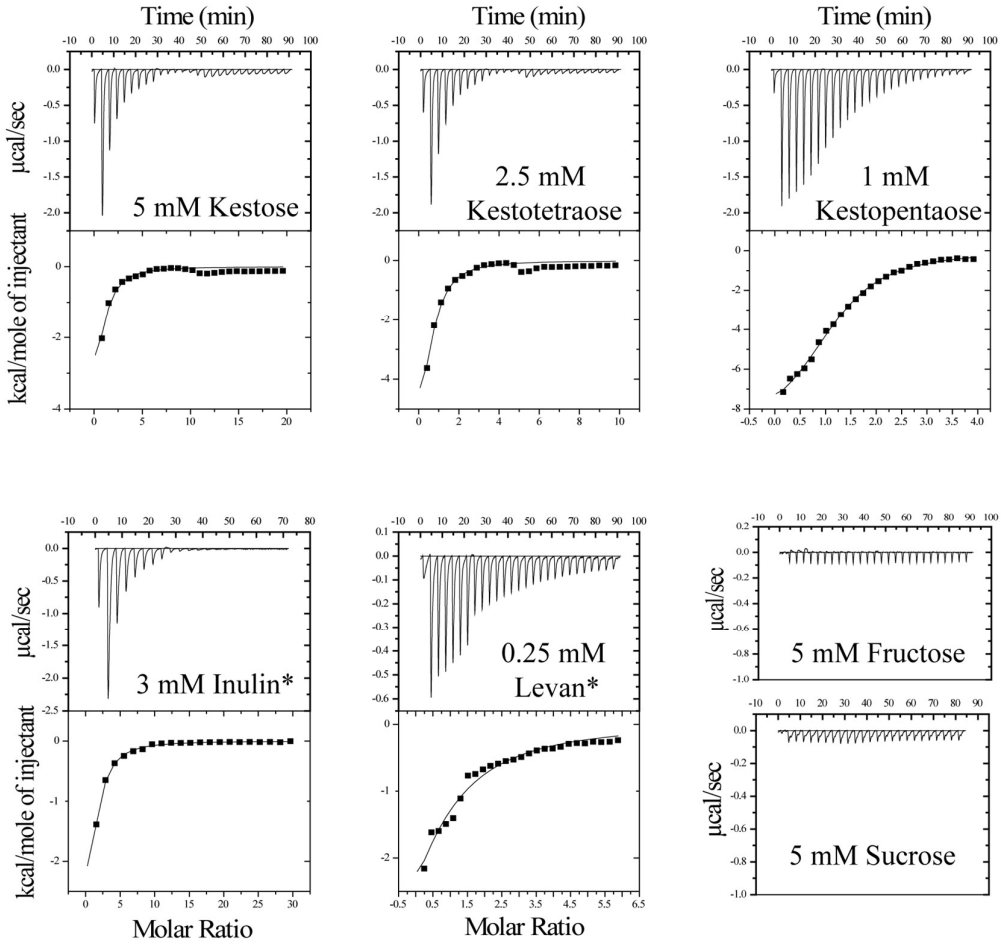


Figure 5.8. ITC analysis of the lacI-homologue substrate binding domain, BAD_1326. ITC analysis of the LacI homologue reveals high affinity binding to FOS and inulin, with weak binding to levan and no binding to sucrose or fructose. ITC traces for each experiment are shown, the top frame shows the raw heats and the bottom shows the integrated peak areas fitted to a one site model using Microcal origin software, apart from fructose and sucrose which did not bind and only the ITC traces are shown. Full binding parameters are shown separately (**Table 5.2**). * Concentrations calculated as shown in Chapter 2.13.2.

5.7. Structural and Functional Insights into the Inulin Binding ESBP, BAD_1330

5.7.1. The ESBP, BAD_1330, Recognises Inulin

The ESBP, BAD_1330, contains a type II signal peptide predicted to facilitate localisation at the cell surface (Bagos *et al.*, 2008), this was removed prior to

cloning and expression using recombinant *E. coli* as described (Chapter 2.9.7 and 2.11). Protein was purified using IMAC (Chapter 2.12.1).

Further construct details are listed (Appendix I.8).

ITC binding analysis (**Figure 5.9**) revealed that the ESBP bound to inulin (K_d ~ 0.04 mM) and FOS, but did not bind to sucrose. It bound more weakly to levan (K_d ~ 0.1 mM). The affinity of the ESBP for kestotetraose and kestopentaose were greater than that of kestose suggesting that at least three fructose units are required for optimal binding. Full binding parameters are listed below (**Table 5.3**).

Table 5.3. Binding Parameters of BAD_1330 from ITC Datasets

Ligand	**N**	**K_a (x 10^4 M^{-1})**	**ΔG (kcal mol^{-1})**	**ΔH (kcal mol^{-1})**	**TΔS (kcal mol^{-1})**
Sucrose			No Binding		
Kestose	1.2 ±0.4	0.5 ±0.05	-5.0	-6.3 ±3.6	-1.3
Kestotetraose	0.9 ±0.4	8.1 ±3.1	-6.7	-9.3 ±2.8	-2.6
Kestopentaose	1 ±0.4	5.1 ±4.5	-6.4	-8.2 ±4.7	-1.8
Inulin+	1.1 ±0.0	4.8 ±1.6	-6.4	-11 ±1	-4.6
Levan+			Binding too weak to fitΔ		

Multiple datasets (3-5) were obtained with at least two freshly prepared protein samples to ensure results were reproducible and to calculate standard deviations
+ Concentration of inulin and levan were estimated as described in Chapter 2.13.2.
Δ An interaction was observed with levan, but this was too weak to fit to the model. Refer to main text.

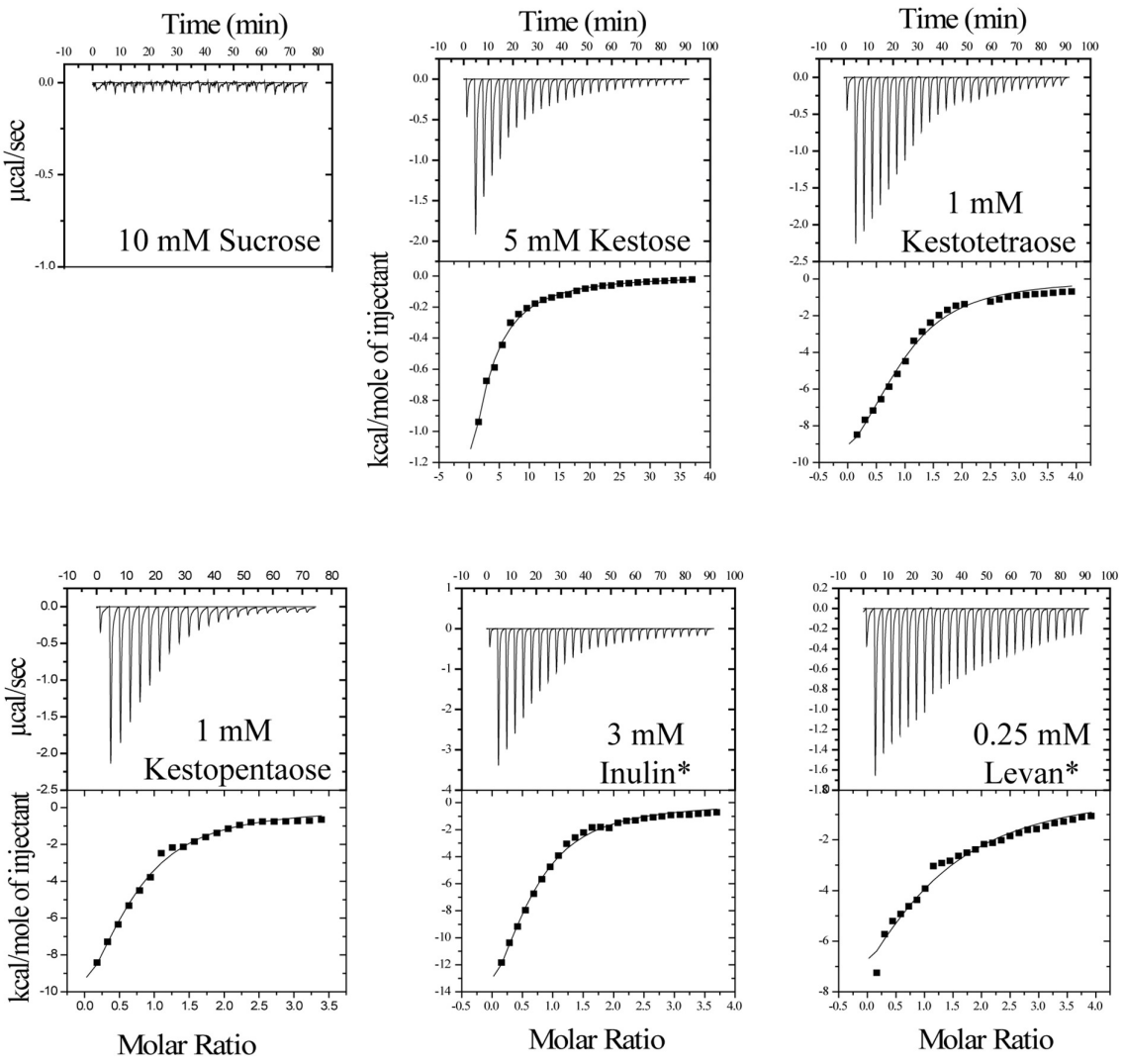


Figure 5.9. ITC analysis of the ESBP, BAD_1330. ITC analysis of the ESBP reveals high affinity binding to FOS and inulin, binding to levan and no binding to sucrose. 2-5 datasets were obtained and representative traces for each experiment are shown. ITC was performed in 20 mM HEPES, pH 7. * Concentrations in mM were calculated as shown in Chapter 2.13.2.

5.7.2. The Crystal Structure of the inulin binding BAD_1330 ESBP

Crystals were harvested, data collected and the structure solved by Dr. Arnaud Baslé as part of a collaboration with the Structural Biology Lab (SBL), Newcastle University. Crystallography data tables generated by Dr. Arnaud Baslé are shown (Appendix I, **Table I.4** and **I.5**)

The ESBP was purified using IMAC followed by Gel Filtration chromatography as outlined (Chapter 2.12.1 and 2.12.3) and crystallised as described (Chapter 2.13.3).

The structure of the ESBP in open form was solved (**Figure 5.10**). The protein consists of two distinct domains (Domain one: 35 – 157, 345-437 aa; Domain two: 158 – 344, 438 – 532 aa) connected by two loops and an alpha helix. Both domains are formed by central β-sheets surrounded by α helices. This structure has a similar overall fold and domain structure with other elucidated class 1 substrate binding proteins (Berntsson *et al.*, 2010).

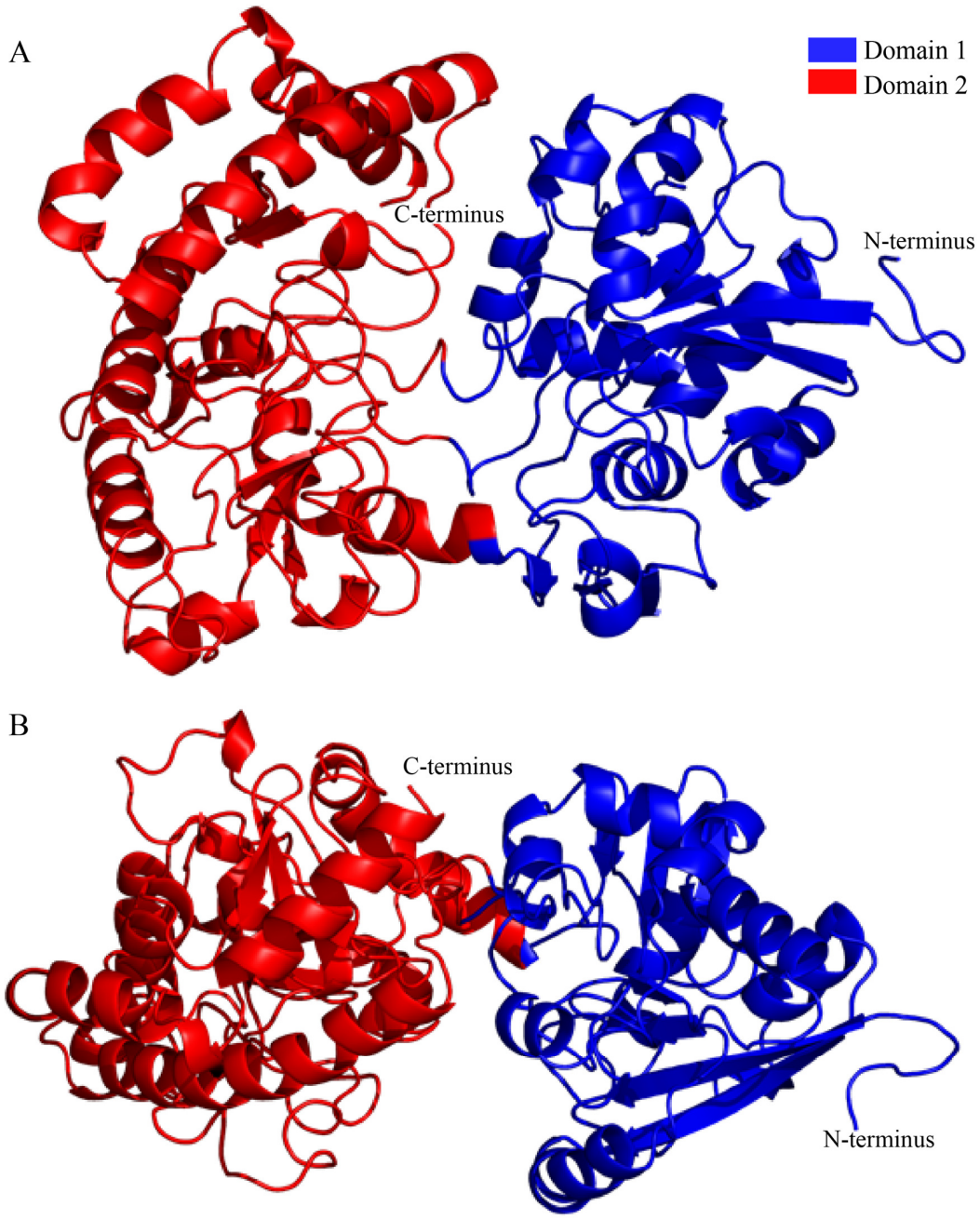


Figure 5.10. The Structure of the ESBP, BAD_1330 in Open Conformation. The crystal structure of the ESBP was solved in open conformation. The protein consists of two distinct domains (domain 1, coloured blue and domain 2, coloured red) joined by two loop regions and an α helix. The protein is displayed from two angles to demonstrate the two domain regions and the three linker regions where the domains meet. Panel B has been rotated 270° along the x axis relative to panel A.

The Dali server was used to identify close structural homologues to the ESBP, BAD_1330 (Holm & Rosenström, 2010). The closest structural homologue to BAD_1330 is BLON_2351 (3OMB), an ESBP from *B. longum subspecies infantis*, which retains 47 % ID, however there is no biochemical characterisation accompanying this structure at present. All other structural homologues have relatively low identity (>25 %).

BAD_1330 retained low identity to the two ESBP structures discussed in the chapter overview, the arabinoxylan oligosaccharide specific ESBP (13 % ID) from *B. animalis* (Ejby *et al.,* 2013) and the starch specific ESBP (14 % ID) from *E. rectale* (Cockburn *et al.,* 2013).

5.7.3. A Significant Conformational Change Occurs Upon Ligand Binding

The ESBP was purified using IMAC followed by Gel Filtration chromatography as outlined (Chapter 2.12.1 and 2.12.3) and crystallised with kestotetraose as described (Chapter 2.13.3). The crystal structure was solved in complex with kestotetraose (**Figure 5.11**)

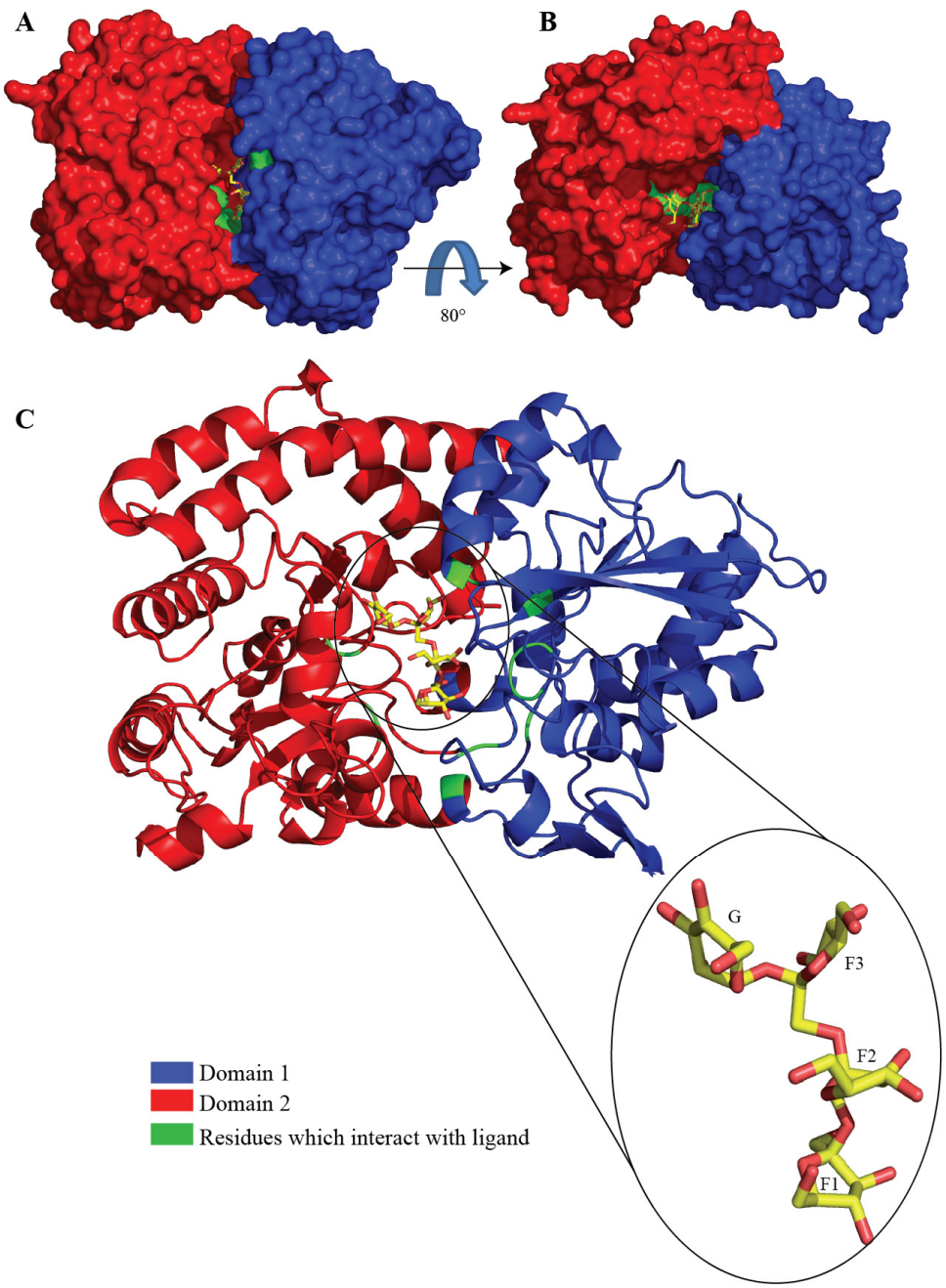


Figure 5.11. The structure of the ESBP, BAD_1330, was solved in complex with kestotetraose. The crystal structure of the ESBP was solved after co-crystallisation with kestotetraose (K4), revealing the protein in a closed conformation with ligand bound. The two domains are coloured (domain 1, blue and domain 2, red). (**A**) BAD_1330 shown in surface representation and K4 (yellow), interacting residues in green. (**B**) As in panel A, with the protein rotated on the x axis by 80° showing a top down view of the binding pocket. (**C**) BAD_1330, with K4 bound, is shown as a cartoon. Inset: The structure of K4 as it appears in the binding pocket, K4 is a tetramer comprised of a glucose (G) as part of a sucrose moiety and two fructose (F) units linked as follows: $G_{\alpha 1} \rightarrow {}_{2\beta}F3_{\beta 2} \rightarrow {}_{1\beta}F2_{\beta 2} \rightarrow {}_{1\beta}F1$.

A significant conformational change upon ligand binding is clearly observed in this structure compared to that of the open form (**Figure 5.12**). DynDom (Poornam *et al.*, 2009) was used to further quantify the change between conformers and to identify hinge regions. The two domains rotated around the hinge region at an angle of 26.5°. Three hinge regions were identified (156-158 aa, 343-349 aa and 437-438 aa) which correspond to the three regions joining the two domains (**Figure 5.12**). The movement can also be quantified by measuring the distance between residues in the closed and open conformers; Q81 moves the greatest distance (8.9 Å) as the two domains are drawn closer together.

The binding pocket is formed upon ligand binding as the two domains move together around the ligand. Most of the interaction takes place at the non-reducing fructose terminus, where the protein forms extensive interactions with the terminal fructose. The pocket opens up into a wide channel in which the glucose end of the ligand is poorly co-ordinated, this widening of the channel presumably allows the ESBP to bind to longer oligosaccharides and polymeric inulin, which can protrude through the channel into the extracellular space.

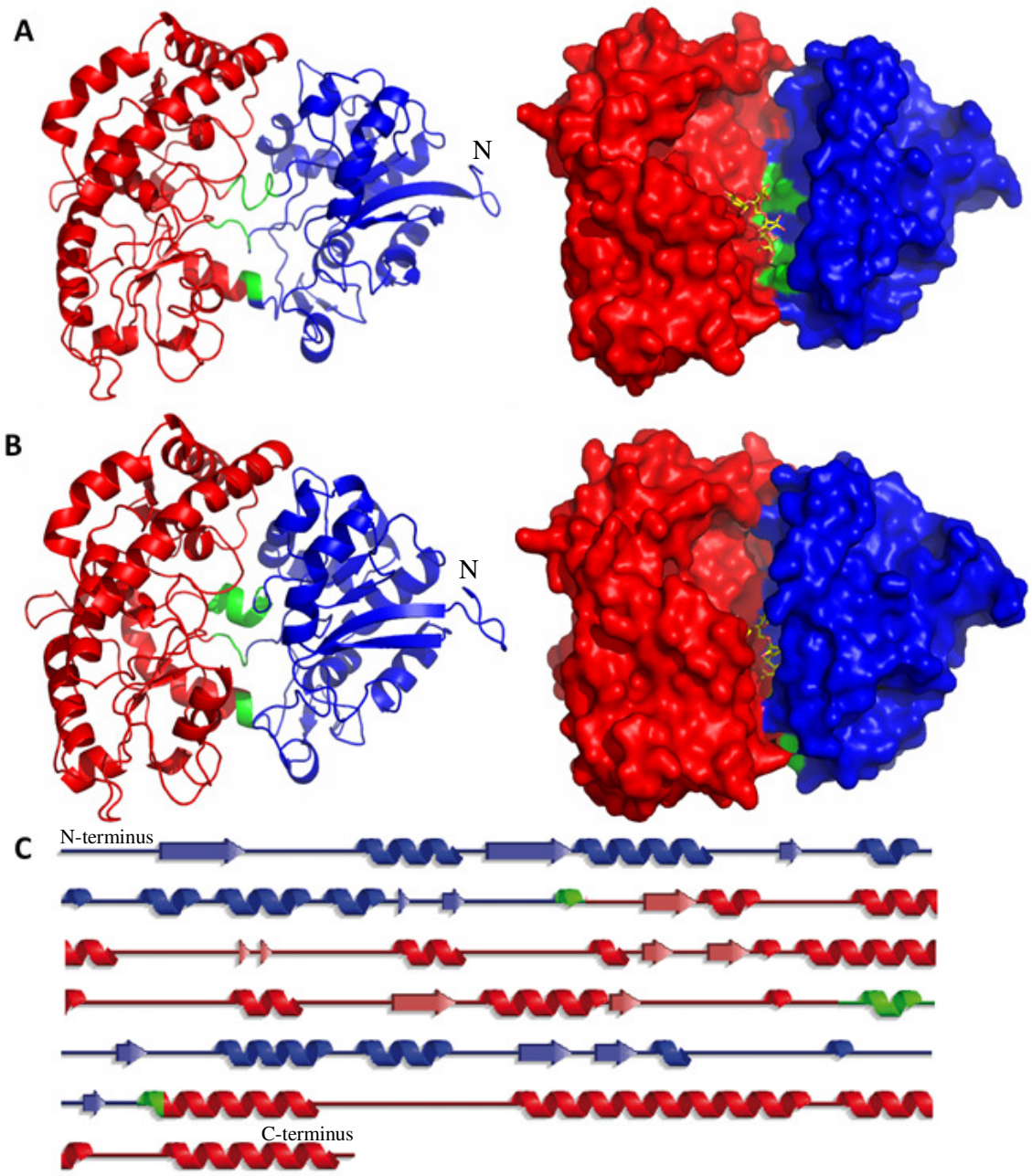


Figure 5.12. The binding site is formed by a change in the relative positions of the two domains. The BAD_1330 binding site is formed upon ligand binding as the two domains move closer together. Domain one (blue) and domain two (red) move around a hinge region (green). N denotes the amine or "N" -terminus of the protein. (**A**) BAD_1330 in open form, displayed as a cartoon (right), and as a surface with kestotetraose (yellow) superimposed (left). (**B**) BAD_1330 in closed form displayed as a cartoon (right) and as a surface in complex with kestotetraose (left). (**C**) Protein secondary structure diagram generated with PDBsum Generate (De Beer, 2014) and coloured to represent the domains and hinge regions.

5.7.4. Insight into Ligand Binding by the Inulin ESBP, BAD_1330

The ESBP structure in complex with kestotetraose allows considerable insight into the binding mechanism of this protein. From the structure, several residues form hydrogen bonds with the kestotetraose ligand (**Figure 5.13**). Most of these form bonds with the non-reducing fructose terminus, F1 (R153, D413, Q420, D421, N438). The subsequent fructose units, F2 and F3 are bound by two residues each (D421 and R422; R45 and D77 respectively). W306 stacks against the F1 ring and W218 protrudes into the space created by the curvature of the ligand, creating a hydrophobic environment energetically favourable with the ligand in this position.

A pocket is formed in which the kestotetraose non-reducing fructose end is co-ordinated. Most of the recognition seems to be driven by the terminal sugar, however the subsequent two fructose units also form interactions with the binding pocket (**Figure 5.13**). It is difficult to analyse binding to the fourth sugar, as this is glucose within the complex but binding to inulin and FOS with a DP > 4 will present a fructose at this position. However the ITC data (**Table 5.3**) shows that maximal affinity may be achieved with kestotetraose, which suggests few further interactions take place; in contrast, there is a marked 10 fold decrease in affinity for kestose (**Table 5.3**).

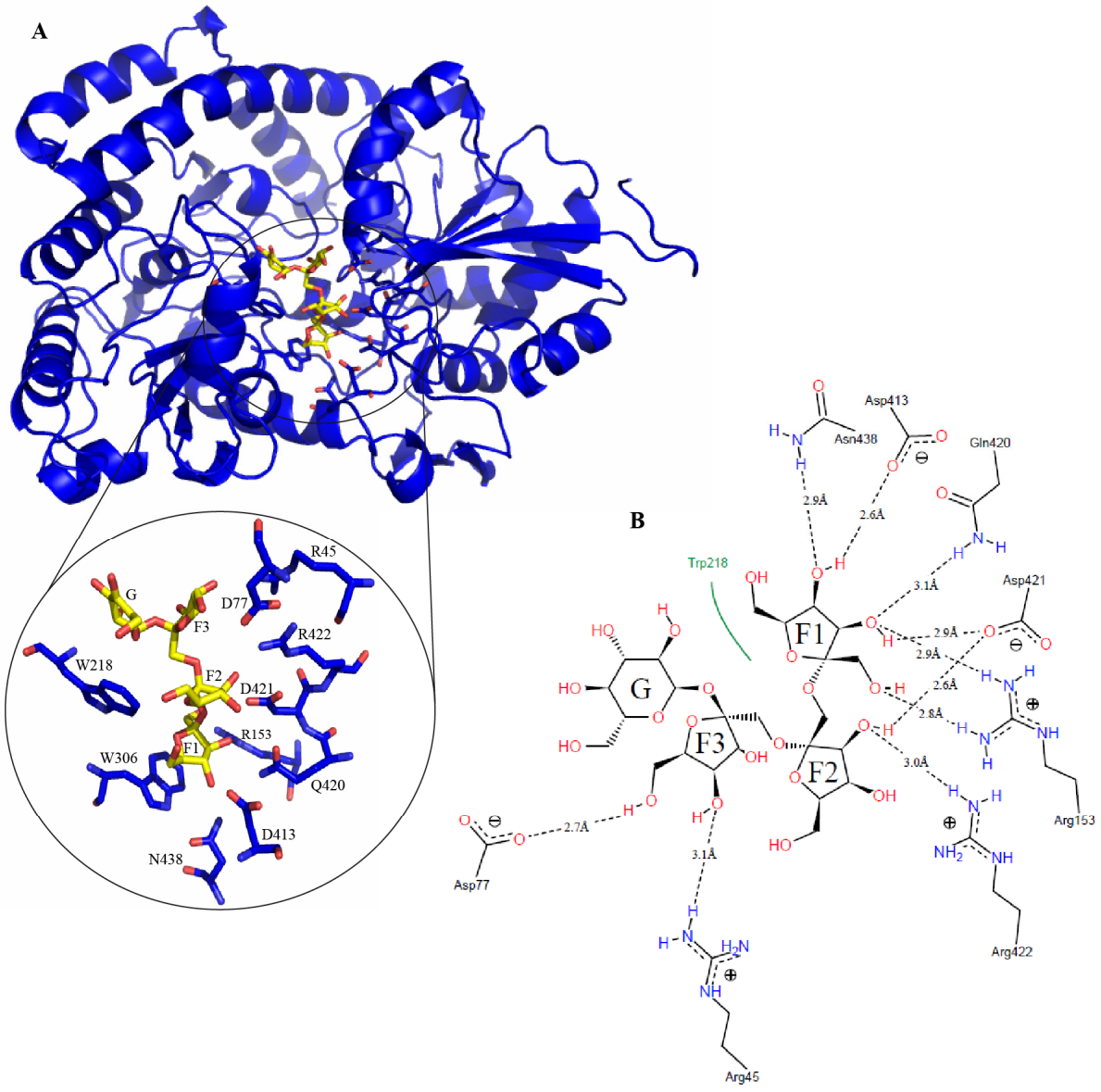


Figure 5.13. The ESBP crystal structure reveals interactions with kestotetraose. (**A**) The BAD_1330 structure showing the location of the binding site. Inset: A 3D representation of the binding pocket. Hydrogen bonding to the non-reducing fructose terminus (F1) is mediated by N438, D413, Q420 and N421. Less hydrogen bonding to subsequent fructose units is observed with N421, R153 and R422 bonding F2, and R45 and D77 bonding to F3. Glucose (G) does not form any interactions. W218 creates an energetically favourable hydrophobic region against the curvature of the ligand. W306 (not shown in panel B) stacks in plane against the F1 ring, creating another hydrophobic interaction. (**B**) PoseView (Stierand & Rarey, 2010) was used to create a 2D diagram of the interactions between key residues and the kestotetraose in the crystal structure. This diagram was then edited for clarity and to show putative bond length measured in PyMOL.

5.8. Discussion

5.8.1. A Model for Inulin Utilisation in *B. adolescentis*

One objective of this research was to identify whether inulin was processed outside of the cell, or whether it was first imported for internal processing. Initially as neither of the GH32 encoding genes (*bad_1326* or *bad_1150*) contained a predicted signal peptide and were suspected to localise to the cytoplasm, we hypothesised that processing occurred internally.

This hypothesis is strengthened by the observation that the LacI-homologue construct, BAD_1326PBP, recognises FOS and inulin, and not fructose monosaccharide. The LacI homologue, BAD_1326, contains a DNA binding domain and a periplasmic binding protein (PBP) domain and is suspected to perform a regulatory function within the cytoplasm. As oligo- and polysaccharides are recognised, it appears clear that these substrates must be imported whole prior to regulation.

In addition, the presence of the inulin binding ESBP BAD_1330 also strongly indicated that inulin is pulled into the cell prior to degradation. ESBP have been shown to mediate the import of carbohydrates into the cell cytoplasm (Ejby *et al*., 2013; Cockburn *et al.*, 2015).

B. adolescentis cells grown on inulin produced small amounts of fructose over the course of growth (**Figure 5.2**) which is consistent with internal inulin degredation. The utilisation of small FOS is less clear as during growth on FOS, fructose and small FOS breakdown products were produced within the culture supernatant suggesting the presence of an extracellular enzyme (**Figure 5.2**). This extra-cellular activity may be the result of cell lysis leading to the release of cytoplasmic GH32 enzyme rather than the co-ordinated secretion of one or both of the GH32 enzymes. We suspect that this extra-cellular activity is artefactual as neither GH32 enzymes contained a signal sequence.

The data obtained within this chapter allow us to present a model for inulin utilisation within *Bifidobacterium adolescentis* (**Figure 5.14**). Inulin is recognised at the cell surface by the ESBP, BAD_1330, and pulled through the membrane by an associated ABC permease (BAD_1327, BAD_1328 and/or BAD_1329). Once imported to the cytoplasm, the lacI-homologue, BAD_1326, recognises FOS and inulin and propagates the up-regulation of the gene cluster.
Finally, the GH32, BAD_1325, depolymerises imported fructans for use by the cell as energy (**Figure 5.14**).

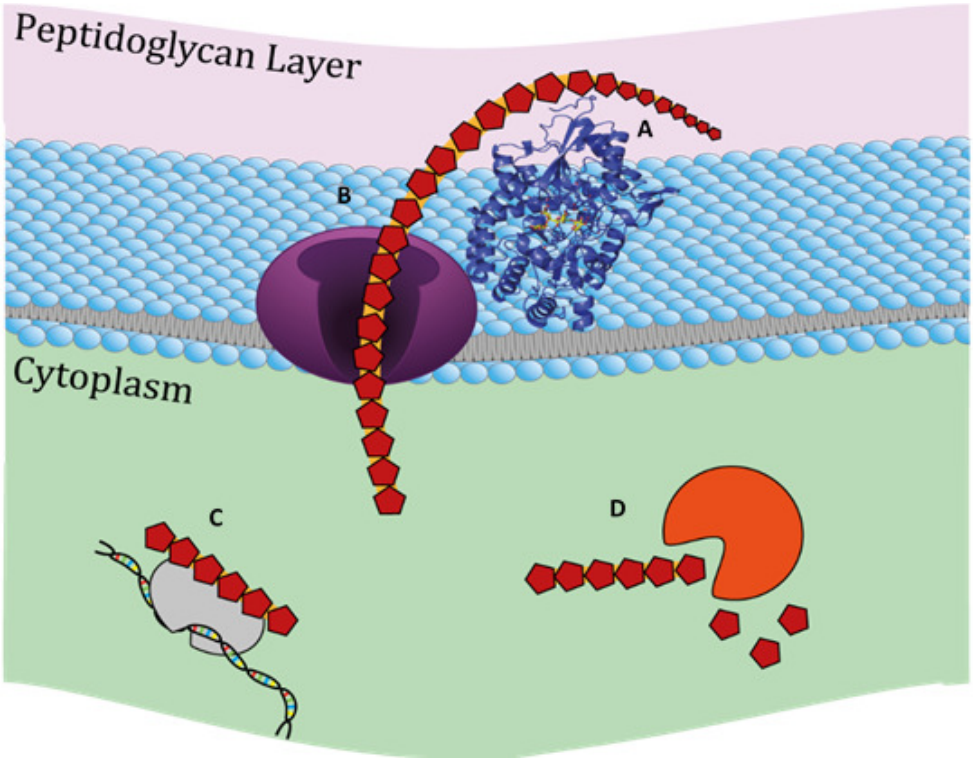


Figure 5.14. A model for inulin utilisation by *Bifidobacterium adolescentis*. The data elucidated from this chapter allow a model for inulin utilisation by *B. adolescentis* to be put forward. (**A**) Inulin and FOS are recognised at the cell surface by the ESBP and imported (**B**) through the associated ABC permeases. (**C**) Cytoplasmic FOS and inulin are bound by the LacI-homologue and the gene cluster is up-regulated. (**D**) The GH32, BAD_1325, mediates the depolymerisation of these substrates for downstream use by the cell.

5.8.2. The FOS-Utilisation Cluster

B. longum contains one GH32 encoding gene (*bllj_1341*) whereas *B. adolescentis* contains two (*bad_1150* and *bad_1325*); *bllj_1341* has high sequence identity to *bad_1150*, and *bad_1325*, from the inulin utilisation cluster, is more divergent (**Figure 5.3**).

As *bllj_1341* is the only fructan active enzyme predicted within the *B. longum* genome, we make the assumption that this enzyme enables growth on FOS (**Figure 5.1**). The genomic context of *bllj_1341* (*bad_1339* → *bad_1343*) is highly similar to the context of bad_1150 (*bad_1148* → *bad_1152*), thus it seems likely that this cluster performs the same role in both species. This cluster was explored with limited success, however, a mechanism of action may be put forward. *B. longum* contains a gene cluster with high sequence identity to the FOS utilisation cluster in *B. adolescentis* (**Figure 5.3**) but does not contain any other GH32 enzymes, or other candidate fructan utilisation systems. We make the assumption, due to this high identity, that these clusters operate in the same manner and this cluster, in *B. longum* is involved in FOS degradation.

B. longum did not reach the maximal cell density on FOS of DP 3-4 as on glucose (**Figure 5.1**) indicating that the full range of oligosaccharides could not be utilised. Growth experiments with this species confirmed that *B. longum* could use kestose but not kestotetraose (**Figure 5.1 B**).

Characterisation of the ESBP from this cluster, BAD_1152, was attempted; however no binding during ITC analysis to fructans was observed (Appendix I, **Figure I.3**). This protein is a different class of ESBP to the characterised ESBP BAD_1330 and is annotated as pfam NMT1, a family associated with thiamine biosynthesis and regulation. The lack of growth of *B. longum* on longer FOS substrates suggests that a bottleneck exists at the import stage.

Furthermore, a permease, *bad_1149*, is encoded within this cluster which may be the sole FOS importer. It seems likely that the permease, *bad_1149*, is responsible for short chain FOS up-take while the ABC permease and associated ESBP are not involved in fructan utilisation at all. It is possible, however, that the expressed ESBP was non-functional during ITC analysis, leading to the negative results obtained.

No upregulation of the FOS cluster was detected on FOS, inulin or sucrose (**Figure 5.4**). We speculate that the relevant genes are either constitutively expressed such that no fold change is observed between glucose and fructan substrates, or that the substrate bound by the lacI class regulator was not tested. The former instance seems unlikely, as lacI class regulators normally propagate an induction of expression, though this protein may not be involved in fructan utilisation or linked to the GH32 enzyme. In the latter instance, it may be that the regulator binds to a specific substrate such as kestose, fructobiose, fructotriose etc. It is unlikely that fructose is recognised as the GH32 enzymes will produce this during inulin and FOS breakdown, and some up-regulation would have likely been observed. Further experiments should be conducted to elucidate which hypothesis (if either) is correct, this could be done through qPCR analysis using the aforementioned substrates, and the lacI class regulator, *bad_1148*, should be characterised.

From these data it is speculated that the ABC transporter and ESBP (BAD_1151 and BAD_1152) are not part of this cluster. As only kestose, sucrose and monosaccharides were utilised well (and there is no recognisable signal peptide), we speculate that the GH32 enzyme is located within the cytoplasm as this enzyme does not demonstrate the tight specificity seen during growth. It is likely that the permease, BAD_1149 facilitates the import of small FOS across the membrane. The LacI class regulator may bind fructose as a proxy for FOS, but is more likely to bind to a di- or tri-saccharide if it is involved within the system.

These speculations are under the assumption that the *B. longum* FOS-utilisation cluster (*bllj_1339* → *bllj_1341*) operates in an identical manner to the homologous cluster from *B. adolescentis* – a dangerous assumption to make, despite high sequence identity of the cluster components (**Figure 5.3**). Indeed, it may be that the acquisition of the inulin utilisation cluster has rendered the FOS utilisation cluster defunct in *B. adolescentis*, and it is no longer required or regulated in response to fructans. The speculated mechanism for FOS utilisation by this cluster are illustrated in the figure below (**Figure 5.15**) however future experiments are required to before conclusions about this system can be drawn.

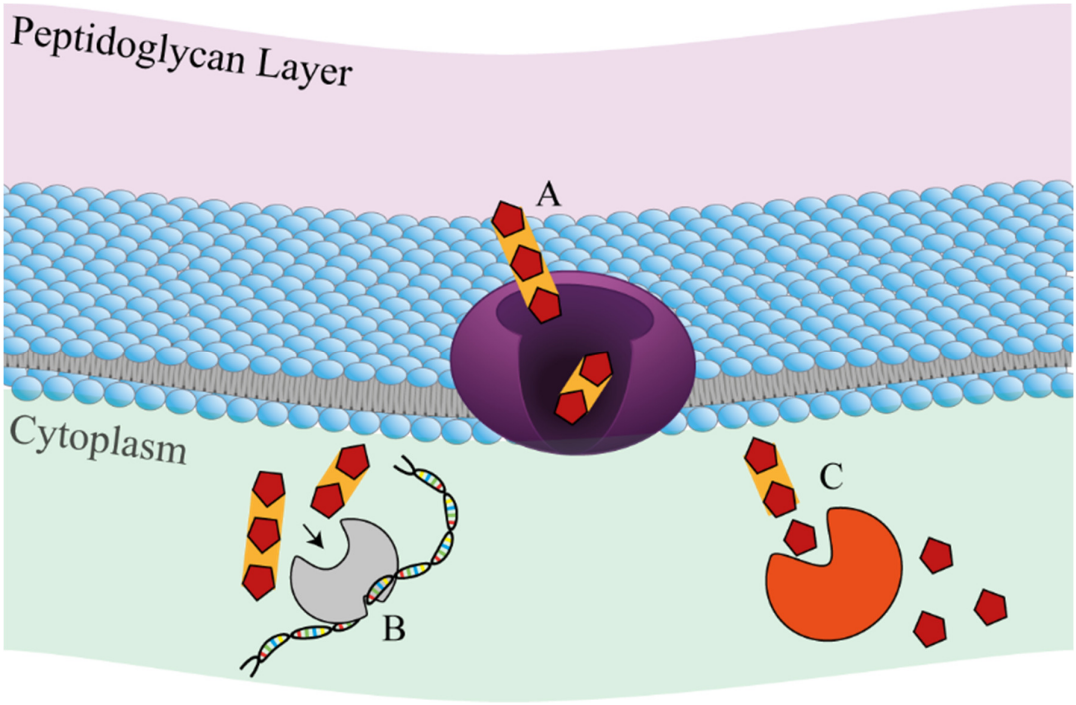


Figure 5.15. A potential mechanism by which the FOS-utilisation cluster may function in *B. adolescentis*. We were unable to satisfactorily characterise the FOS-associated gene cluster (*bad_1148* → *bad_1150*). A potential mechanism by which these genes may operate is shown. (**A**) Short chain FOS are imported through the permease, BAD_1149. (**B**) The LacI class regulator, BAD_1148, recognises imported FOS or a component of this imported FOS to induce the expression of the gene cluster. (**C**) The GH32, BAD_1150, depolymerises this imported FOS.

5.8.3. Insight into Fructan Utilisation by *Bifidobacterium spp*.

There is a paucity of literature regarding the fructan degradative apparatus of *Bifidobacterium species*. It is, therefore, difficult to understand how inulin prebiotics are able to stimulate a bifidogenic affect (Schaafsma & Slavin; 2015, Patel & Goyal, 2012; Dewulf *et al.*, 2013). Here we demonstrate that *B. adolescentis* can utilise inulin directly however most members of the Bifidobacterium genus appear to utilise only short chain FOS (Watson *et al.*, 2013). Other members of the microbiota, such as *B. ovatus*, degrade inulin externally, creating substrate for other microbes (Rakoff-Naouhm *et al*., 2014; Chapter 4.3). We demonstrate that *B. ovatus* breakdown products are the preferred substrate for both *B. longum* and *B. adolescentis* (Appendix I, **Figure I.7**), supporting the hypothesis that glycans may be "shared" or "scavenged" across intestinal microbial communities.

Chapter 6. Final Discussion

6.1. Summary

The aim of this thesis was to explore non-linear fructan and the utilisation of inulin by key members of the microbiota.

In chapter 3 we demonstrate evidence to support the literature showing that fructan from onion, wheat, agave and garlic is non-linear (Chapter 3.0, Chapter 3.3). This includes structures with; a β2-1 backbone with β2-6 branches, a β2-6 backbone with β2-1 branches, bifurcose, and bifurcose-series fructans, though a detailed analysis of linkage types could not be undertaken and suspected linkage types were based upon previous studies (Baumgartner *et al.*, 2000; Arrizon *et al.*, 2010; Velazquez-Martinez *et al.*, 2014; Verspreet *et al.*, 2007; Bancal *et al.*, 1991; Vijn *et al.*, 1997). *Bacteroides ovatus* and *Bacteroides thetaiotaomicron*, have been shown previously to have a mutually exclusive preference for either inulin or levan respectively (Sonnenburg *et al.*, 2010), our data show that non-linear fructans, which we speculate make up a large portion of dietary fructan; can provide carbon and energy to the gut microbiota (Chapter 3.4). The data also demonstrate that *B. ovatus* and *B. thetaiotaomicron* have overlapping niches within the intestine, despite divergent linkage specificity (Sonnenburg *et al.*, 2010, Chapter 3.4).

In chapter 4 we elucidate the mechanism by which *B. ovatus* is able to utilise inulin. *B. ovatus* is a highly prominent Gram negative member of the gut microbiota capable of degrading a wide range of glycans (Martens *et al.*, 2011). We confirmed putative inulin associated genes through qPCR analysis (Sonnenburg *et al.*, 2010; Chapter 4.5). We show that inulin is broken up extracellularly by a heteromeric GH91 enzyme into DFA-FOS (Chapter 4.6).

Sucrose terminated inulin and FOS are targeted for direct import by a SusD-homologue, we show the putative binding site of this SusD-homologue through x-

ray crystallographic studies, targeted alanine mutagenesis and ITC analysis (Chapter 4.7). Extracellular inulin and FOS are imported through a SusC-homologue (Chapter 4.9.1) and degraded in the periplasm by a GH32 (Chapter 4.8). DFA is produced as a waste product during growth on inulin (Chapter 4.6.7), both in *B. ovatus* mono-cultures and in faecal sample batch cultures (Chapter 4.6, Chapter 4.10.1), our data indicate that this DFA is not degraded by other microbiota (Chapter 4.10.2).

In chapter 5 we explore the mechanism by which *Bifidobacterium adolescentis* is able to utilise inulin. *Bifidobacterium* are prominent members of the microbiota and are thought to be health-promoting (Fanning *et al*., 2012a; Fanning *et al*., 2012b; Picard *et al.,* 2005; Fukuda *et al.*, 2011); as such, *Bifidobacterium spp*. are the classic target for prebiotic treatment (Gibson & Roberfroid, 1995). Among *Bifidobacterium spp.*, the ability to utilise inulin polysaccharide is relatively rare despite widespread FOS utilisation in this genus (Watson *et al.*, 2013). We demonstrate that *B. adolescentis* contains a divergent gene cluster which is upregulated in the presence of inulin and FOS (Chapter 5.4.2). Inulin is bound by an ESBP. We characterise through x-ray crystallography the binding site of this ESBP protein, along with the significant conformational change which is undertaken upon ligand binding (Chapter 5.7.3). Inulin is recognised in the cytoplasm by the PBP domain of a LacI-homologue, presumably this protein mediates the up-regulation of inulin associated genes (Chapter 5.6). Inulin degradation takes place internally, through the action of a GH32 (Chapter 5.8.1).

Our data expands our understanding of the types of fructans available to our microbiota and demonstrates detailed analysis of how inulin is captured and utilised by two key members of the microbiota. Our data allows deeper insight into the ecological niches adopted by members of the intestinal microbiota during nutrient acquisition. Understanding how inulin is used at the molecular level may underpin the design and development of novel prebiotic treatments, and enhance our understanding of how microbes are able to respond to prebiotic intervention.

6.2. The Role of Insoluble Inulin and Branched Fructans in the Diet

Ingested MACs move through the digestive tract and by doing so, are exposed to microbial degradation (Flint *et al*., 20012). Theoretically, short, simple and soluble substrates can be more easily accessed whereas insoluble, complex or otherwise recalcitrant structures require more specialised apparatus to degrade and may persist in the lumen longer; therefore a gradient of increasingly recalcitrant carbohydrate structures may exist along the length of the intestinal tract (**Figure 6.1,** Koropatkin *et al.*, 2012). Furthermore, the mucous layer thickens in the late G.I. tract and is thought to provide a better environment for microbes which target host mucins (Koropatkin *et al.*, 2012; Eckburg *et al.*, 2005).

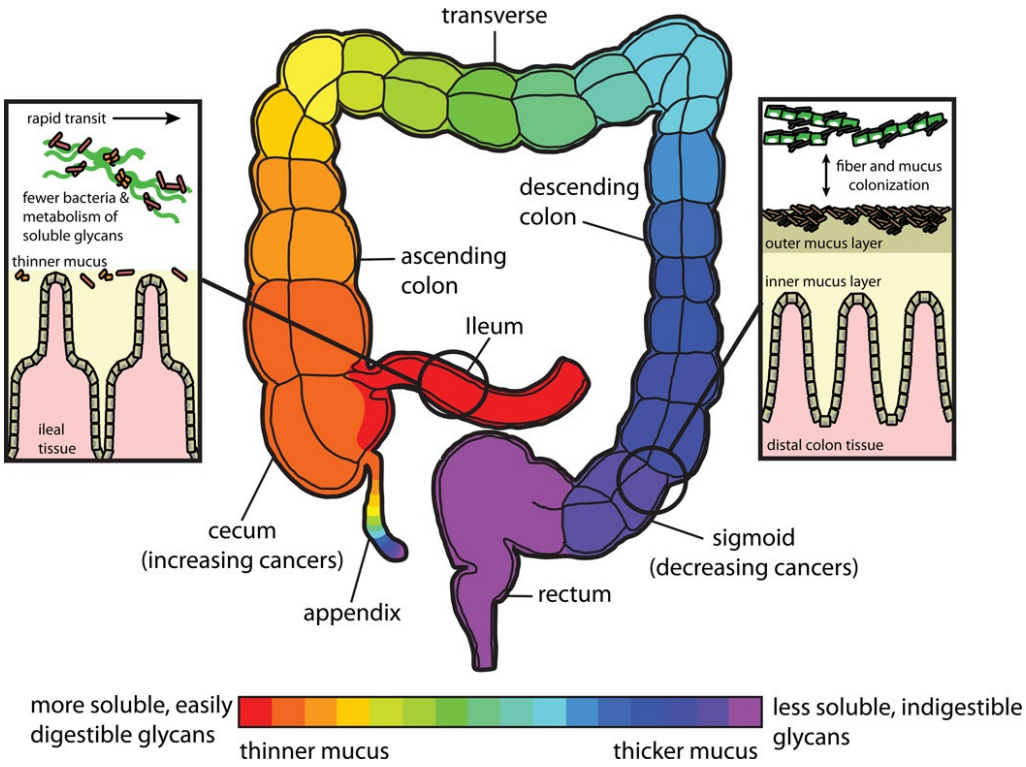


Figure 6.1. Glycan availability changes across the digestive tract (Koropatkin e*t al.*, 2012). A glycan gradient is likely to exist across the host digestive system. Early areas of the G.I. tract move more rapidly, have a thinner mucous layer and contain less recalcitrant glycans than later areas. This potentially creates microenvironments in which distinct microbes reside (Koropatkin e*t al.*, 2012).

We speculate that the non-linear fructans such as those examined in Chapter 3 may be more recalcitrant to microbial degradation and may persist longer in the G.I. tract to form a niche in the mid-late digestive tract. We also speculate that high degree of branching may occlude import apparatus and/or prevent enzymes from cleaving. In *Bacteroides,* PUL systems which tackle branched carbohydrates appear complex often containing enzymes which are able to remove side-chains, or cleave the backbone extracellularly to reduce glycan size prior to import (Cuskin *et al*., 2015; Rogowski *et al.*, 2015). Therefore, to tackle large, highly branched fructans, we would expect to see extracellular processing.
Whilst extracellular DFA-terminated oligosaccharides are produced during growth on inulin by *B. ovatus* through the action of the GH91 enzyme*,* we do not see DFA production on garlic fructan, suggesting that branching occludes the function of the GH91 enzyme (Chapter 3.4; Baumgartner *et al.*, 2000). Experiments have not yet been conducted to show how the comparatively simple import mechanism of *B. adolescentis* copes with non-linear fructan structures, and should be undertaken to further our understanding of the niche non-linear fructan structures provide within the gut.

Insoluble inulin represents another distinct niche; insoluble matter is more recalcitrant to enzymatic degradation and may require co-ordination between binding domains and enzymes (Boraston *et al.*, 2004). Short chain inulin is easily solubilised; however this short chain substrate may not be the full complement of inulin present in the human bowel. Much of the inulin used for commercial and research purposes is obtained by hot water extraction from the plant, commonly chicory root (Gupta *et al*., 2003). Hot water extractions rely on solubilising inulin and removing insoluble matter, therefore by performing these extractions any insoluble inulin is discarded. All substrates used in this study were obtained by hot water extraction and as such any insoluble material present in the original sample was lost, and not tested (**Figure 6.2.**).

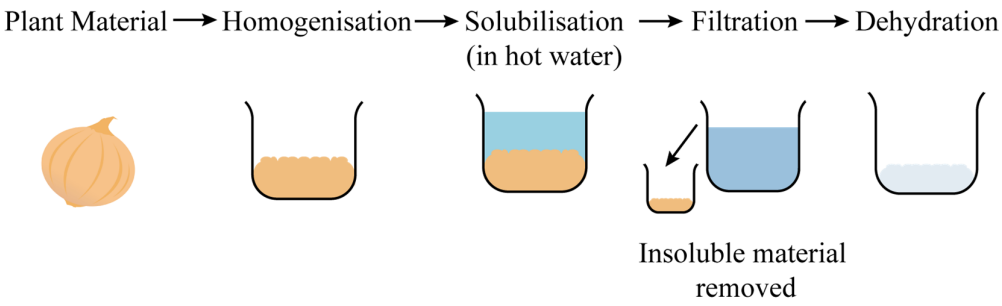


Figure 6.2. Hot water extraction of soluble plant fructans. During this study fructans were acquired from hot water extractions as outlined in Chapter 2.17.1; this technique removes any insoluble matter and thus insoluble fructan was not tested.

Bacteroides ovatus contains apparatus, including the extracellular endo-acting GH91 enzyme, which appears to be redundant and does not contribute much obvious benefit to the degradation of linear, soluble inulin (Chapter 4.9.3). It is possible that *B. ovatus* is able to mediate the breakdown of insoluble inulin through the co-ordinated action of the GH91 enzyme and appended CBM. We propose that by doing this or through the breakdown of non-linear fructans, *B. ovatus* would be able to colonise areas further down the colon, where only the more recalcitrant carbohydrates remain available. It is also speculated that the comparatively simple *B. adolescentis* import apparatus will not be able to handle insoluble or highly branched fructans and will occupy a niche earlier in the digestive tract. In support of this theory, we show that *B. longum* is unable to utilise FOS of DP > 3 (Chapter 5.3) despite using FOS of DP = 3, this suggested that FOS utilisation is constrained at import and is unable to accommodate additions to the glycan chain. To test how well the *B. adolescentis* import mechanisms can handle non-linear fructan, ITC should be conducted with the ESBP against garlic or wheat fructan and growth experiments with these substrates should be conducted.

6.3. Fructan Acquisition Niches in the *Bacteroides* and *Bifidobacterium* genera

Both *B. ovatus* (Chapter 4.3) and *B. thetaiotaomicron* (Sonnenburg *et al.*, 2010) produce extracellular oligosaccharides during growth on inulin and levan respectively. Oligosaccharides released into the intestinal lumen are likely to be scavenged by competing microbiota which have the ability to process short chain FOS, but lack the ability to utilise the polysaccharide (**Figure 6.3**). Rakoff-Naoum and colleagues show this effect well, demonstrating that these oligosaccharides can indeed be utilised by other members of the microbiota (Rakoff-Nahoum *et al.*, 2014). There is a clear advantage to oligosaccharide scavenging; fewer, less specialised enzymes and binding proteins and thus less energy is required to import and degrade simple carbohydrates.

A scavenging system, in the case of fructans, is quite versatile as short chain FOS will exist in the diet in addition to being produced and released by species such as *B. ovatus*; therefore, scavengers may occupy multiple niches along the G.I. tract with minimal expended energy. The advantage to the producer of these oligosaccharides is less clear, as these bacteria lose material to scavengers (**Figure 6.3**). It is possible that by "sharing" resources with scavengers, these bacteria cultivate a niche for themselves occluding the growth of other species able to access the polysaccharide directly which do not release oligosaccharides. It is plausible that other, more complex, relationships exist between producer and scavenger; however these relationships are difficult to quantify and examine.

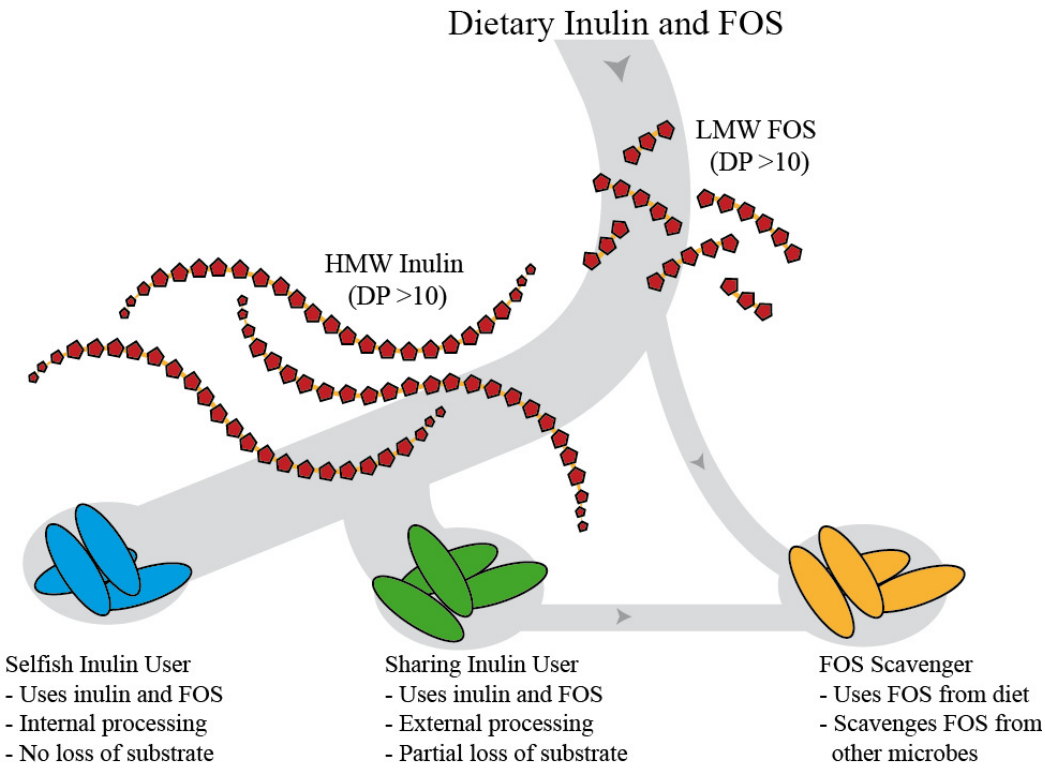


Figure 6.3. Selfish, sharing and scavenging niches. Three mechanisms for utilising inulin are proposed; a selfish mechanism (e.g. *Bif. adolescentis*) which is able to utilise HMW inulin and LMW FOS from the diet without the production of extracellular oligosaccharides and thus does not lose substrate to scavengers, a "sharing" mechanism (e.g. *Bac. ovatus*) which produces extracellular oligosaccharides which may be scavenged by other organisms and, a "scavenging" mechanism (e.g. *Bif. longum*) which is able to use LMW FOS only both from the diet and from sharing microbes.

Not all *Bacteroides* species are able to grow on polymeric inulin or levan; *Bacteroides vulgatus* is able to use FOS only, and whilst Sonnenburg *et al.* demonstrate that *B. fragilis* grew on inulin, we found that it was only able to utilise FOS (Appendix I, **Figure I.6**; Sonnenburg *et al.*, 2010). These FOS users, are potentially able to scavenge oligosaccharides produced by *B. ovatus* and *B. thetaiotaomicron* on polymeric fructan.

Not all inulin using *Bacteroides spp.* produce extracellular oligosaccharides which may be shared; *Bacteroides uniformis* contains a set of four GH32 family enzymes, one of which displays catalytic residues suggestive of an endo-mechanism (Bolam & Sonnenburg, 2011; Lammens *et al.*, 2009; Pouyez *et al*., 2012; Vandamme *et al*., 2013) however when *B. uniformis* whole cells were incubated with inulin, no extracellular oligosaccharides were produced (Appendix I**, Figure I.4.**; Mr. Carl Morland, unpublished data). This suggests that *B. uniformis* might have internalised endo-activity, presumably to prevent loss of glycan to the intestinal environment whilst retaining functionality, such as seen during the growth of *B. thetaiotaomicron* on mannan (Cuskin *et al.*, 2015). This hypothesis is supported by the putative type I (periplasmic) signal peptide present at the start of the predicted endo-inulinase GH32 sequence. It was also shown that the inulin utilising strain *B. thetaiotaomicron 8736* does not produce extracellular oligosaccharides (Appendix I**, Figure I.4.**; Mr. Carl Morland, unpublished data). This data was less surprising as *B. thetaiotaomicron 8736* does not contain a candidate endo-inulinase. Furthermore, the GH91 knockout and mutant strains of *B. ovatus* (*Δbacova_04502, Δbacova_04503 E196Q*) retain WT-like growth on inulin despite loss of endo-activity (Chapter 4.9.3). These data strongly suggest that endo-inulinase activity is not necessary to utilise inulin in *Bacteroides spp.*, or indeed by *Bifidobacterium adolescentis* (Chapter 5). We speculate that the presence of endo-enzymes confers a selective advantage either due to sharing of inulins or by gaining access to insoluble inulin substrate.

The *B. ovatus* SusD homologue, recognises sucrose terminated glycans (Chapter 4.7.3). SusD and other characterised SusD-homologues generally recognise the breakdown products from the endo-acting enzyme, but this is not the case in the *B. ovatus* inulin SusD-homologue. We speculate that the recognition of the sucrose terminus allows the organism to recognise the highly desirable dietary FOS, which contains more sucrose per weight than long chain, branched or insoluble inulin. By targeting this FOS for rapid, direct import, we speculate that *B. ovatus* can compete in niches higher up the digestive tract where these desirable glycans are still available, effectively adopting a “scavenging” niche in these circumstances.

Most *Bifidobacterium* species will utilise short chain FOS but not polymeric fructans (Watson *et al.*, 2013). This is consistent with the concept that *Bifidobacterium* are scavenging species or are able to utilise dietary FOS earlier in the digestive tract. *B. adolescentis* has an expanded ability to utilise inulin due to the ABC transport system with the inclusion of an inulin binding ESBP. The ability to import soluble inulin may confer an advantage to *B. adolescentis* compared to FOS utilising *Bifidobacterium* species as this allows *B. adolescentis* to harvest a wider range of FOS and inulin type fructans. We speculate that members of the *Bifidobacterium* are under strong competition for accessible glycans and likely occupy niches in the early digestive tract where these desirable carbohydrates are likely to be most abundant (Koropatkin *et al.*, 2012).

6.4. A Comparison between *Bacteroides ovatus* and *Bifidobacterium adolescentis* with Regards to Fructan Utilisation

Bifidobacterium adolescentis and *Bacteroides ovatus* display very different glycan utilisation systems. The main contrast between the two organisms is the cell wall; *B. adolescentis* is Gram positive whereas *B. ovatus* is Gram negative. The presence of the additional cell wall is an asset to *B. ovatus* during glycan degradation, as membrane anchored proteins and a periplasmic space lend itself to step-wise degradation and import, PUL encoded systems are incredibly versatile and tightly regulated (Martens *et al*., 2009; Larsbrink *et al*., 2014; Cuskin *et al*., 2015; Rogowski *et al*., 2015). In contrast, *B. longum* does not co-ordinate such extensive apparatus at the cell surface and glycans which cannot be imported directly cannot be accessed. Gram positive species are able to degrade complex glycans; the extracellular cellulosome encoded by species such as *Clostridium thermocellum* is a powerful example of this (Flint *et al*., 2012), however few examples of such complex extracellular glycan degradation structures have been elucidated from human gut microbiota.

Some examples of degradation at the peptidoglycan layer have been elucidated (Cockburn *et al*., 2015) but it is unclear how common and versatile these mechanisms are.

In contrast Gram negative *Bacteroides spp.* are highly versatile and able to degrade a wide variety of glycans (Martens *et al.*, 2011). The findings of both the MetaHIT and HMP metagenomics studies show that within the Gram negative Bacteroidetes phylum only the *Bacteroides* and *Prevotella* genera are highly prevalent; in contrast the Gram positive Firmicutes phylum has a much more varied distribution of genera (Arumugam *et al.*, 2011; Huttenhower *et al.*, 2012). We speculate that Gram negative species are able to target a wide variety of nutrients, whereas Gram positive species are generally specialised to scavenge a smaller subset of glycans.

6.5. Developing Novel Fructan Based Prebiotics

The human microbiota is highly personalised at the species level, this variability likely reflects the variability in diet and lifestyle of individuals (Lozupone *et al.*, 2012). This variability may result in highly divergent and personalised metabolic power as the apparatus encoded by different species of the same genus may be specialised and diverse (Kaoutari *et al.*, 2013). When known pathogenic species are not regarded it becomes clear that genus diversity is lost between healthy individuals and those suffering from digestive diseases (Claesson *et al.*, 2012; Chatelier *et al.*, 2013; Gevers *et al.*, 2014; Qin *et al.*, 2012). More concerning, there appears to be a reduction of diversity in the typical western gut microbiota compared to that of non-western individuals, it has been speculated that this loss of diversity is due to increasing hygiene standards and sanitisation, processing of foods, antibiotic use, C-sections and infant formula (**Figure 6.4.**, Sonnenburg & Sonnenburg, 2014). In light of this, it is pertinent to begin to develop personalised pre- and pro-biotics, to modulate the microbiota toward host health by introduction of relevant microbial metabolic pathways and by increasing the non-pathogenic diversity of the microbiota.

We speculate that currently, probiotic choice is relatively poorly informed as the interactions between beneficial species, the existing microbial community and the host are poorly understood.
More work should be undertaken to identify and utilise species which have the highest impact on host health, through factors such as short chain fatty acid production, effect on local pH and effect on inflammation and the immune system such that more informed choices can be made regarding probiotic administration.

Furthermore, work should be undertaken to examine the relationship between oligosaccharide producers and oligosaccharide scavengers; if the desired probiotic species is supported by other, oligosaccharide producing, species within the gut, then these species in turn may become “probiotic” or targets for novel prebiotics. Indeed we speculate that the absence of the extracellular oligosaccharide producing species may lead to the absence of scavengers, it is possible that reintroduction of this producer species could restore scavenger populations and the probiotic administering of scavengers alone may not be sufficient.

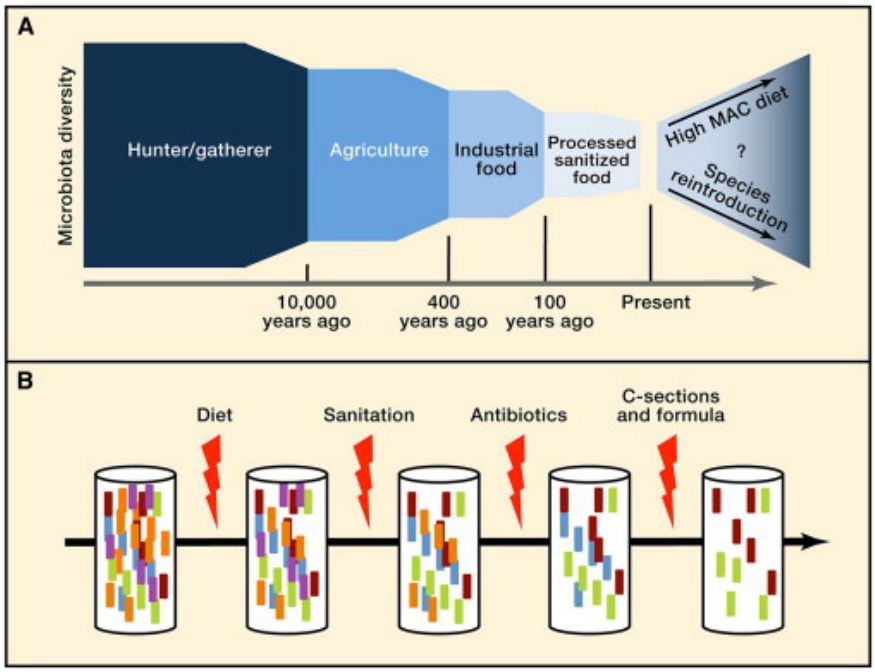


Figure 6.4. Reduction of diversity within the human intestinal microbiota (Sonnenburg & Sonnenburg, 2014). A visual representation of the loss of microbial diversity over the past 10,000 years. (**A**) Proposed timescale and causes for the reduction of microbial diversity. Development of new therapies to reintroduce lost species and recover microbial diversity is needed. (**B**) A visual representation of the potential causes of microbial diversity loss (Sonnenburg & Sonnenburg, 2014).

Probiotics may be vulnerable to acidic conditions in the stomach and viability may decrease during storage and processing (Saunders & Marco, 2010). Prebiotics are, by definition, able to traverse the early G.I. tract intact, and are generally easy to handle and process (Gibson *et al*., 2004). As more information regarding which species target which prebiotics becomes available, it should become easier to rationally manipulate the microbiota using a specific mixture of prebiotic compounds.

It is worth considering the use of longer chain, insoluble or highly branched prebiotics, as these may be more recalcitrant to the microbiota in the early digestive tract and may provide nutrition for species towards the large bowel, potentially shifting the available niches away from host glycan degradation (Koropatkin *et al.*, 2012).

As part of this project, we identified the production of DFA-FOS and DFA during inulin degradation. DFA has been shown to increase absorption of ions at cell junctions in the epithelial layer (Hara *et al.*, 2010; Mineo *et al.*, 2002; Mineo *et al.*, 2004) and may enhance host health. If there is indeed a benefit to the host from DFA production, then DFA-FOS, which is easily produced from the GH91 enzyme, maybe be a good candidate for a novel prebiotic compound; the FOS component will still act as a prebiotic and the released DFA may have a beneficial secondary effect.

6.6. Concluding Remarks

The rapid expansion of our understanding of the human microbiota, both in the intestine and at other body sites, has already yielded many new insights into health and disease. The field, whilst no longer truly in its infancy, is developing at a fast pace and relies on the development and perfection of a suite of novel research tools. For this reason, future work must be conducted with the latest research tools in mind and must be in line with the current generation of literature.

The main goal of most research into the human microbiota is to probe the complex interactions between these organisms and the host in order to understand how these are compromised during disease, with the long term goal of restoring these microbiota-host interactions during the treatment of disease. Pro- and prebiotics have become widely used therapies and prophylactics for the maintenance of gut health, however the choice of pro- or prebiotic is often not well informed and the interactions involved are not well understood. Future work must focus on achieving a deeper understanding of the mechanisms of action of these therapies, and the development of more informed, personalised therapeutic routes for more effective treatment.

References

Aagaard, Kjersti, Jun Ma, Kathleen M. Antony, Radhika Ganu, Joseph Petrosino, and James Versalovic. (2014) "The Placenta Harbors a Unique Microbiome", *Science Translational Medicine*, **6**: 237 37-65.

Akalin, A. S., and D. Erisir. (2008) "Effects of inulin and oligofructose on the rheological characteristics and probiotic culture survival in low-fat probiotic ice cream", *Journal of Food Science*, **73**: M184-M88.

Alberto, F., C. Bignon, G. Sulzenbacher, B. Henrissat, and M. Czjzek. (2004) "The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases", *Journal of Biological Chemistry*, **279**: 18903-10.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. (1990) "Basic Local Alignment Search Tool", *Journal of Molecular Biology*, **215**: 403-10.

Anderson, K. L., and A. A. Salyers. (1989) "Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes", *Journal of Bacteriology*, **171**: 3192-98.

Arora, Tulika, Rajkumar Sharma, and Gary Frost. (2011) "Propionate. Anti-obesity and satiety enhancing factor?", *Appetite*, **56**: 511-15.

Arrizon, J., S. Morel, A. Gschaedler, and P. Monsan. (2010) "Comparison of the water-soluble carbohydrate composition and fructan structures of *Agave tequilana* plants of different ages", *Food Chemistry*, **122**: 123-30.

Arumugam, M., J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J. M. Batto, M. Bertalan, N. Borruel, F. Casellas, L. Fernandez, L. Gautier, T. Hansen, M. Hattori, T. Hayashi, M. Kleerebezem, K. Kurokawa, M. Leclerc, F. Levenez, C. Manichanh, H. B. Nielsen, T. Nielsen, N. Pons, J. Poulain, J. J. Qin, T. Sicheritz-Ponten, S. Tims, D. Torrents, E. Ugarte, E. G. Zoetendal, J. Wang, F. Guarner, O. Pedersen, W. M. de Vos, S. Brunak, J. Dore, J. Weissenbach, S. D. Ehrlich, P. Bork, and H. I. T. Consortium Meta. (2011) "Enterotypes of the human gut microbiome", *Nature*, **473**: 174-80.

Backhed, F., R. E. Ley, J. L. Sonnenburg, D. A. Peterson, and J. I. Gordon. (2005) "Host-bacterial mutualism in the human intestine", *Science*, **307**: 1915-20.

Bäckhed, Fredrik, Josefine Roswall, Yangqing Peng, Qiang Feng, Huijue Jia, Petia Kovatcheva-Datchary, Yin Li, Yan Xia, Hailiang Xie, Huanzi Zhong, Muhammad Tanweer Khan, Jianfeng Zhang, Junhua Li, Liang Xiao, Jumana Al-Aama, Dongya Zhang, Ying Shiuan Lee, Dorota Kotowska, Camilla Colding, Valentina Tremaroli, Ye Yin, Stefan Bergman, Xun Xu, Lise Madsen, Karsten Kristiansen, Jovanna Dahlgren, and Jun Wang. (2015) "Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life", *Cell Host & Microbe*, **17**: 690-703.

Bagos, Pantelis G., Konstantinos D. Tsirigos, Theodore D. Liakopoulos, and Stavros J. Hamodrakas. (2008) "Prediction of Lipoprotein Signal Peptides in Gram-Positive Bacteria with a Hidden Markov Model", *Journal of Proteome Research*, **7**: 5082-93.

Bancal, P., C. A. Henson, J. P. Gaudillere, and N. C. Carpita. (1991) "Fructan Chemical-Structure and Sensitivity to an Exohydrolase", *Carbohydrate Research*, **217**: 137-51.

Baumgartner, Sabine, Thomas G. Dax, Werner Praznik, and Heinz Falk. (2000) "Characterisation of the high-molecular weight fructan isolated from garlic (*Allium sativum L.*)", *Carbohydrate Research*, **328**: 177-83.

Bell, Charles E., and Mitchell Lewis. (2000) "A closer view of the conformation of the Lac repressor bound to operator", *Nat Struct Mol Biol*, **7**: 209-14.

Benigar, Elizabeta, Iztok Dogsa, David Stopar, Andrej Jamnik, Irena Kralj Cigić, and Matija Tomšič. (2014) "Structure and Dynamics of a Polysaccharide Matrix: Aqueous Solutions of Bacterial Levan", *Langmuir*, **30**: 4172-82.

Berntsson, Ronnie P. A., Sander H. J. Smits, Lutz Schmitt, Dirk-Jan Slotboom, and Bert Poolman. (2010) "A structural classification of substrate-binding proteins", *FEBS Letters*, **584**: 2606-17.

Biasucci, G., B. Benenati, L. Morelli, E. Bessi, and G. Boehm. (2008) "Cesarean delivery may affect the early biodiversity of intestinal bacteria", *Journal of Nutrition*, **138**: 1796S-800S.

Bien, Justyna, Vindhya Palagani, and Przemyslaw Bozko. (2013) "The intestinal microbiota dysbiosis and *Clostridium difficile* infection: is there a relationship with inflammatory bowel disease?", *Therapeutic Advances in Gastroenterology*, **6**: 53-68.

Bisgaard, H., N. Li, K. Bonnelykke, B. L. K. Chawes, T. Skov, G. Paludan-Muller, J. Stokholm, B. Smith, and K. A. Krogfelt. (2011) "Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age", *Journal of Allergy and Clinical Immunology*, **128**: 646-U318.

Blake, J. D., M. L. Clarke, P. E. Jansson, and K. E. McNeil. (1982) "Fructan from *Erwinia-herbicola*", *Journal of Bacteriology*, **151**: 1595-97.

Bolam, D. N., A. Ciruela, S. McQueen-Mason, P. Simpson, M. P. Williamson, J. E. Rixon, A. Boraston, G. P. Hazlewood, and H. J. Gilbert. (1998) "*Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity", *Biochemical Journal*, **331**: 775-81.

Bolam, David N., and Justin L. Sonnenburg. (2011) "Mechanistic insight into polysaccharide use within the intestinal microbiota", *Gut Microbes*, **2**: 86-90.

Boraston, Alisdair B, David N Bolam, Harry J Gilbert, and Gideon J Davies. (2004) "Carbohydrate-binding modules: fine-tuning polysaccharide recognition", *Biochemical Journal*, **382**: 769-81.

Brisse, Sylvain, Francine Grimont, and PatrickA D. Grimont. 2006. 'The Genus *Klebsiella.*' in Martin Dworkin, Stanley Falkow, Eugene Rosenberg, Karl-Heinz Schleifer and Erko Stackebrandt (eds.), *The Prokaryotes* (Springer New York).

Brody, Jonathan R., and Scott E. Kern. (2004) "History and principles of conductive media for standard DNA electrophoresis", *Analytical Biochemistry*, **333**: 1-13.

Brück, Wolfram M., Gitte Graverholt, and Glenn R. Gibson. (2002) "Use of batch culture and a two-stage continuous culture system to study the effect of supplemental α-lactalbumin and glycomacropeptide on mixed populations of human gut bacteria", *FEMS Microbiology Ecology*, **41**: 231-37.

Brussow, Harald, and Scott J. Parkinson. (2014) "You are what you eat", *Nat Biotech*, **32**: 243-45.

Buclin, T., A. PechereBertschi, R. Sechaud, L. A. Decosterd, A. Munafo, M. Burnier, and J. Biollaz. (1997) "Sinistrin clearance for determination of glomerular filtration rate: A reappraisal of various approaches using a new analytical method", *Journal of Clinical Pharmacology*, **37**: 679-92.

Bujacz, Anna, Marzena Jedrzejczak-Krzepkowska, Stanislaw Bielecki, Izabela Redzynia, and Grzegorz Bujacz. (2011) "Crystal structures of the apo form of β-fructofuranosidase from *Bifidobacterium longum* and its complex with fructose", *FEBS Journal*, **278**: 1728-44.

Buriti, Flávia C. A., and Susana M. I. Saad. (2013) "Chilled Milk-based Desserts as Emerging Probiotic and Prebiotic Products", *Critical Reviews in Food Science and Nutrition*, **54**: 139-50.

Cameron, E. A., M. A. Maynard, C. J. Smith, T. J. Smith, N. M. Koropatkin, and E. C. Martens. (2012) "Multidomain Carbohydrate-binding Proteins Involved in *Bacteroides thetaiotaomicron* Starch Metabolism", *Journal of Biological Chemistry*, **287**: 34614-25.

Cameron, Elizabeth A., Kurt J. Kwiatkowski, Byung-Hoo Lee, Bruce R. Hamaker, Nicole M. Koropatkin, and Eric C. Martens. (2014) "Multifunctional Nutrient-Binding Proteins Adapt Human Symbiotic Bacteria for Glycan Competition in the Gut by Separately Promoting Enhanced Sensing and Catalysis", *mBio*, **5**: e01441-14.

Campbell, J. M., L. L. Bauer, G. C. Fahey, Ajcl Hogarth, B. W. Wolf, and D. E. Hunter. (1997) "Selected fructooligosaccharide (1-kestose, nystose, and 1(F)-beta-fructofuranosylnystose) composition of foods and feeds", *Journal of Agricultural and Food Chemistry*, **45**: 3076-82.

Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat. (2009) "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics", *Nucleic Acids Research*, **37**: D233-D38.

Cerdeño-Tárraga AM1, Patrick S, Crossman LC, Blakely G, Abratt V, Lennard N, Poxton I, Duerden B, Harris B, Quail MA, Barron A, Clark L, Corton C, Doggett J, Holden MT, Larke N, Line A, Lord A, Norbertczak H, Ormond D, Price C, Rabbinowitsch E, Woodward J, Barrell B, Parkhill J. (2005). "Extensive DNA inversions in the B. fragilis genome control variable gene expression.", *Science*, **307**:1463-1465

Chang, Pamela V., Liming Hao, Stefan Offermanns, and Ruslan Medzhitov. (2014) "The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition", *Proceedings of the National Academy of Sciences*, **111**: 2247-52.

Chassard, Christophe, Eve Delmas, Céline Robert, and Annick Bernalier-Donadille. (2010) "The cellulose-degrading microbial community of the human gut varies according to the presence or absence of methanogens", *FEMS Microbiology Ecology*, **74**: 205-13.

Cho, K. H. & Salyers, A. A. (2001) "Biochemical analysis of interactions between outermembrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*.", *Journal of Bacteriology*, **183**: 7224-7230.

Cimini, Sara, Vittoria Locato, Rudy Vergauwen, Annalisa Paradiso, Cristina Cecchini, Liesbeth Vandenpoel, Joran Verspreet, Christophe M. Courtin, Maria Grazia D'Egidio, Wim Van den Ende, and Laura De Gara. (2015) "Fructan biosynthesis and degradation as part of plant metabolism controlling sugar fluxes during durum wheat kernel maturation", *Frontiers in Plant Science*, **6**: 89.

Claesson, Marcus J., Ian B. Jeffery, Susana Conde, Susan E. Power, Eibhlis M. O'Connor, Siobhan Cusack, Hugh M. B. Harris, Mairead Coakley, Bhuvaneswari Lakshminarayanan, Orla O'Sullivan, Gerald F. Fitzgerald, Jennifer Deane, Michael O'Connor, Norma Harnedy, Kieran O'Connor, Denis O'Mahony, Douwe van Sinderen, Martina Wallace, Lorraine Brennan, Catherine Stanton, Julian R. Marchesi, Anthony P. Fitzgerald, Fergus Shanahan, Colin Hill, R. Paul Ross, and Paul W. O'Toole. (2012) "Gut microbiota composition correlates with diet and health in the elderly", *Nature*, **488**: 178-84.

Clemente, Jose C., Erica C. Pehrsson, Martin J. Blaser, Kuldip Sandhu, Zhan Gao, Bin Wang, Magda Magris, Glida Hidalgo, Monica Contreras, Óscar Noya-Alarcón, Orlana Lander, Jeremy McDonald, Mike Cox, Jens Walter, Phaik Lyn Oh, Jean F. Ruiz, Selena Rodriguez, Nan Shen, Se Jin Song, Jessica Metcalf, Rob Knight, Gautam Dantas, and M. Gloria Dominguez-Bello. (2015) "The microbiome of uncontacted Amerindians", *Science Advances*, **1**.

Cockburn, D. W., N. I. Orlovsky, M. H. Foley, K. J. Kwiatkowski, C. M. Bahr, M. Maynard, B. Demeler, and N. M. Koropatkin. (2015) "Molecular details of a starch utilization pathway in the human gut symbiont *Eubacterium rectale*", *Molecular Microbiology*, **95**: 209-30.

Cordain, L., S. B. Eaton, A. Sebastian, N. Mann, S. Lindeberg, B. A. Watkins, J. H. O'Keefe, and J. Brand-Miller. (2005) "Origins and evolution of the Western diet: health implications for the 21st century", *American Journal of Clinical Nutrition*, **81**: 341-54.

Corpet, F. (1988) "Multiple sequence alignment with hierarchical clustering", *Nucleic Acids Research*, **16**: 10881-90.

Cox, Laura M., Shingo Yamanishi, Jiho Sohn, Alexander V. Alekseyenko, Jacqueline M. Leung, Ilseung Cho, Sungheon Kim, Huilin Li, Zhan Gao, Douglas Mahana, Jorge G. Zárate Rodriguez, Arlin B. Rogers, Nicolas Robine, P'ng Loke, and Martin J. Blaser. (2014) "Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences", *Cell*, **158**: 705-21.

Cuskin, Fiona. (2012) "Mechanisms by which Glycoside Hydrolases Recognize Plant, Bacterial and Yeast Polysaccharides", Newcastle University.

Cuskin, Fiona, James E. Flint, Tracey M. Gloster, Carl Morland, Arnaud Baslé, Bernard Henrissat, Pedro M. Coutinho, Andrea Strazzulli, Alexandra S. Solovyova, Gideon J. Davies, and Harry J. Gilbert. (2012) "How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity", *Proceedings of the National Academy of Sciences*, **109**: 20889-94.

Cuskin, Fiona, Elisabeth C. Lowe, Max J. Temple, Yanping Zhu, Elizabeth A. Cameron, Nicholas A. Pudlo, Nathan T. Porter, Karthik Urs, Andrew J. Thompson, Alan Cartmell, Artur Rogowski, Brian S. Hamilton, Rui Chen, Thomas J. Tolbert, Kathleen Piens, Debby Bracke, Wouter Vervecken, Zalihe Hakki, Gaetano Speciale, Jose L. Munoz-Munoz, Andrew Day, Maria J. Pena, Richard McLean, Michael D. Suits, Alisdair B. Boraston, Todd Atherly, Cherie J. Ziemer, Spencer J. Williams, Gideon J. Davies, D. Wade Abbott, Eric C. Martens, and Harry J. Gilbert. (2015) "Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism", *Nature*, **517**: 165-69.

Davies, Gideon, and Bernard Henrissat. (1995) "Structures and mechanisms of glycosyl hydrolases", *Structure*, **3**: 853-59.

de Beer, Tjaart A. P., Karel Berka, Janet M. Thornton, and Roman A. Laskowski. (2014) "PDBsum additions", *Nucleic Acids Research*, **42**: D292-D96.

De Filippo, Carlotta, Duccio Cavalieri, Monica Di Paola, Matteo Ramazzotti, Jean Baptiste Poullet, Sebastien Massart, Silvia Collini, Giuseppe Pieraccini, and Paolo Lionetti. (2010) "Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa", *Proceedings of the National Academy of Sciences of the United States of America*, **107**: 14691-96.

Dethlefsen, Les, and David A. Relman. (2011) "Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation", *Proceedings of the National Academy of Sciences of the United States of America*, **108**: 4554-61.

Dewulf, Evelyne M., Patrice D. Cani, Sandrine P. Claus, Susana Fuentes, Philippe G. B. Puylaert, Audrey M. Neyrinck, Laure B. Bindels, Willem M. de Vos, Glenn R. Gibson, Jean-Paul Thissen, and Nathalie M. Delzenne. (2013) "Insight into the prebiotic concept: lessons from an exploratory, double blind intervention study with inulin-type fructans in obese women", *Gut*, **62**: 1112-21.

Eckburg, P. B., E. M. Bik, C. N. Bernstein, E. Purdom, L. Dethlefsen, M. Sargent, S. R. Gill, K. E. Nelson, and D. A. Relman. (2005) "Diversity of the human intestinal microbial flora", *Science*, **308**: 1635-38.

Ejby, Morten, Folmer Fredslund, Andreja Vujicic-Zagar, Birte Svensson, Dirk Jan Slotboom, and Maher Abou Hachem. (2013) "Structural basis for arabinoxylo-oligosaccharide capture by the probiotic *Bifidobacterium animalis subsp. lactis Bl-04*", *Molecular Microbiology*, **90**: 1100-12.

Evans, P. (2006) "Scaling and assessment of data quality", *Acta Crystallographica Section D-Biological Crystallography*, **62**: 72-82.

Falony, G., T. Calmeyn, F. Leroy, and L. De Vuyst. (2009) "Coculture Fermentations of *Bifidobacterium Species* and *Bacteroides thetaiotaomicron* Reveal a Mechanistic Insight into the Prebiotic Effect of Inulin-Type Fructans", *Applied and Environmental Microbiology*, **75**: 2312-19.

Fanning, Saranna, Lindsay J. Hall, Michelle Cronin, Aldert Zomer, John MacSharry, David Goulding, Mary O'Connell Motherway, Fergus Shanahan, Kenneth Nally, Gordon Dougan, and Douwe van Sinderen. (2012) "*Bifidobacterial* surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection", *Proceedings of the National Academy of Sciences*, **109**: 2108-13.

Fanning, Saranna, Lindsay J. Hall, and Douwe van Sinderen. (2012) "*Bifidobacterium breve UCC2003* surface exopolysaccharide production is a beneficial trait mediating commensal-host interaction through immune modulation and pathogen protection", *Gut Microbes*, **3**: 420-25.

Flint, Harry J., Edward A. Bayer, Marco T. Rincon, Raphael Lamed, and Bryan A. White. (2008) "Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis", *Nat Rev Micro*, **6**: 121-31.

Flint, Harry J., Karen P. Scott, Sylvia H. Duncan, Petra Louis, and Evelyne Forano. (2012) "Microbial degradation of complex carbohydrates in the gut", *Gut Microbes*, **3**: 289-306.

Franck, A. (2002) "Technological functionality of inulin and oligofructose", *British Journal of Nutrition*, **87**: S287-S91.

Fukuda, Shinji, Hidehiro Toh, Koji Hase, Kenshiro Oshima, Yumiko Nakanishi, Kazutoshi Yoshimura, Toru Tobe, Julie M. Clarke, David L. Topping, Tohru Suzuki, Todd D. Taylor, Kikuji Itoh, Jun Kikuchi, Hidetoshi Morita, Masahira Hattori, and Hiroshi Ohno. (2011) "*Bifidobacteria* can protect from enteropathogenic infection through production of acetate", *Nature*, **469**: 543-47.

Fung, K. Y. C., L. Cosgrove, T. Lockett, R. Head, and D. L. Topping. (2012) "A review of the potential mechanisms for the lowering of colorectal oncogenesis by butyrate", *British Journal of Nutrition*, **108**: 820-31.

Funkhouser, L. J., and S. R. Bordenstein. (2013) "Mom Knows Best: The Universality of Maternal Microbial Transmission", *Plos Biology*, **11**: 9.

Fuse, Haruka, Haruka Fukamachi, Mitsuko Inoue, and Takeshi Igarashi. (2013) "Identification and functional analysis of the gene cluster for fructan utilization in *Prevotella intermedia*", *Gene*, **515**: 291-97.

Garrido, Daniel, Jae Han Kim, J. Bruce German, Helen E. Raybould, and David A. Mills. (2011) "Oligosaccharide Binding Proteins from *Bifidobacterium longum* subsp. *infantis* Reveal a Preference for Host Glycans", *PLoS ONE*, **6**: e17315.

Gasteiger, Elisabeth, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, MarcR Wilkins, RonD Appel, and Amos Bairoch. 2005. 'Protein Identification and Analysis Tools on the ExPASy Server.' in JohnM Walker (ed.), *The Proteomics Protocols Handbook* (Humana Press).

Gevers, Dirk, Subra Kugathasan, Lee A. Denson, Yoshiki Vázquez-Baeza, Will Van Treuren, Boyu Ren, Emma Schwager, Dan Knights, Se Jin Song, Moran Yassour, Xochitl C. Morgan, Aleksandar D. Kostic, Chengwei Luo, Antonio González, Daniel McDonald, Yael Haberman, Thomas Walters, Susan Baker, Joel Rosh, Michael Stephens, Melvin Heyman, James Markowitz, Robert Baldassano, Anne Griffiths, Francisco Sylvester, David Mack, Sandra Kim, Wallace Crandall, Jeffrey Hyams, Curtis Huttenhower, Rob Knight, and Ramnik J. Xavier. (2014) "The treatment-naïve microbiome in new-onset Crohn's disease", *Cell Host & Microbe*, **15**: 382-92.

Gibson, G. R., E. R. Beatty, X. Wang, and J. H. Cummings. (1995) "Selective Stimulation of *Bifidobacteria* in the Human Colon by Oligofructose and Inulin", *Gastroenterology*, **108**: 975-82.

Gibson, G. R., H. M. Probert, J. Van Loo, R. A. Rastall, and M. B. Roberfroid. (2004) "Dietary modulation of the human colonic microbiota: updating the concept of prebiotics", *Nutrition Research Reviews*, **17**: 259-75.

Gibson, G. R., and M. B. Roberfroid. (1995) "Dietary Modulation of the Human Colonic Microbiota - Introducing the Concept of Prebiotics", *Journal of Nutrition*, **125**: 1401-12.

Gilbert, Harry J., J. Paul Knox, and Alisdair B. Boraston. (2013) "Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules", *Current Opinion in Structural Biology*, **23**: 669-77.

Gorrec, Fabrice. (2009) "The MORPHEUS protein crystallization screen", *Journal of Applied Crystallography*, **42**: 1035-42.

Griffin, I. J., P. M. Davila, and S. A. Abrams. (2002) "Non-digestible oligosaccharides and calcium absorption in girls with adequate calcium intakes", *British Journal of Nutrition*, **87**: S187-S91.

Gu, S. H., D. D. Chen, J. N. Zhang, X. M. Lv, K. Wang, L. P. Duan, Y. Nie, and X. L. Wu. (2013) "Bacterial Community Mapping of the Mouse Gastrointestinal Tract", *PLoS ONE*, **8**: 9.

Guaraldi, Federica, and Guglielmo Salvatori. (2012) "Effect of Breast and Formula Feeding on Gut Microbiota Shaping in Newborns", *Frontiers in Cellular and Infection Microbiology*, **2**: 94.

Gunn, Frank J., Christopher G. Tate, and Peter J. F. Henderson. (1994) "Identification of a novel sugar-H+ symport protein, FucP, for transport of L-fucose into *Escherichia coli*", *Molecular Microbiology*, **12**: 799-809.

Gupta, A. K., N. Kaur, and N. Kaur. (2003) "Preparation of inulin from chicory roots", *Journal of Scientific & Industrial Research*, **62**: 916-20.

Hara, Hiroshi, Satoshi Onoshima, and Chie Nakagawa. (2010) "Difructose anhydride III promotes iron absorption in the rat large intestine", *Nutrition*, **26**: 120-27.

Hawrelak, Jason A., and Stephen P. Myers. (2004) "The causes of intestinal dysbiosis: a review", *Alternative medicine review : a journal of clinical therapeutic*, **9**: 180-97.

Hehemann, Jan-Hendrik, Gaelle Correc, Tristan Barbeyron, William Helbert, Mirjam Czjzek, and Gurvan Michel. (2010) "Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota", *Nature*, **464**: 908-12.

Herrero, M., V. de Lorenzo, and K. N. Timmis. (1990) "Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in Gram-negative bacteria", *Journal of Bacteriology*, **172**: 6557-67.

Hervé, Cécile, Artur Rogowski, Anthony W. Blake, Susan E. Marcus, Harry J. Gilbert, and J. Paul Knox. (2010) "Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects", *Proceedings of the National Academy of Sciences*, **107**: 15293-98.

Hoeppli, Romy E., Dan Wu, Laura Cook, and Megan K. Levings. (2015) "The Environment of Regulatory T Cell Biology: Cytokines, Metabolites, and the Microbiome", *Frontiers in Immunology*, **6**: 61.

Holm, Liisa, and Päivi Rosenström. (2010) "Dali server: conservation mapping in 3D", *Nucleic Acids Research*, **38**: W545-W49.

Hosseini, Elham, Charlotte Grootaert, Willy Verstraete, and Tom Van de Wiele. (2011) "Propionate as a health-promoting microbial metabolite in the human gut", *Nutrition Reviews*, **69**:245-58.

The Human Microbiome Jumpstart Reference Strains Consortium. (2010) " A Catalog of Reference Genomes from the Human Microbiome", *Science*, **328**:994-999.

Huttenhower, C., D. Gevers, R. Knight, S. Abubucker, J. H. Badger, A. T. Chinwalla, H. H. Creasy, A. M. Earl, M. G. FitzGerald, R. S. Fulton, M. G. Giglio, K. Hallsworth-Pepin, E. A. Lobos, R. Madupu, V. Magrini, J. C. Martin, M. Mitreva, D. M. Muzny, E. J. Sodergren, J. Versalovic, A. M. Wollam, K. C. Worley, J. R. Wortman, S. K. Young, Q. D. Zeng, K. M. Aagaard, O. O. Abolude, E. Allen-Vercoe, E. J. Alm, L. Alvarado, G. L. Andersen, S. Anderson, E. Appelbaum, H. M. Arachchi, G. Armitage, C. A. Arze, T. Ayvaz, C. C. Baker, L. Begg, T. Belachew, V. Bhonagiri, M. Bihan, M. J. Blaser, T. Bloom, V. Bonazzi, J. P. Brooks, G. A. Buck, C. J. Buhay, D. A. Busam, J. L. Campbell, S. R. Canon, B. L. Cantarel, P. S. G. Chain, I. M. A. Chen, L. Chen, S. Chhibba, K. Chu, D. M. Ciulla, J. C. Clemente, S. W. Clifton, S. Conlan, J. Crabtree, M. A. Cutting, N. J. Davidovics, C. C. Davis, T. Z. DeSantis, C. Deal, K. D. Delehaunty, F. E. Dewhirst, E. Deych, Y. Ding, D. J. Dooling, S. P. Dugan, W. M. Dunne, A. S. Durkin, R. C. Edgar, R. L. Erlich, C. N. Farmer, R. M. Farrell, K. Faust, M. Feldgarden, V. M. Felix, S. Fisher, A. A. Fodor, L. J. Forney, L. Foster, V. Di Francesco, J. Friedman, D. C. Friedrich, C. C. Fronick, L. L. Fulton, H. Y. Gao, N. Garcia, G. Giannoukos, C. Giblin, M. Y. Giovanni, J. M. Goldberg, J. Goll, A. Gonzalez, A. Griggs, S. Gujja, S. K. Haake, B. J. Haas, H. A. Hamilton, E. L. Harris *et al*,. (2012) "Structure, function and diversity of the healthy human microbiome", *Nature*, **486**: 207-14.

Iraporda, Carolina, Agustina Errea, David E. Romanin, Delphine Cayet, Elba Pereyra, Omar Pignataro, Jean Claude Sirard, Graciela L. Garrote, Analia G. Abraham, and Martin Rumbo. (2015) "Lactate and short chain fatty acids produced by microbial fermentation downregulate proinflammatory responses in intestinal epithelial cells and myeloid cells", *Immunobiology*, **220**: 1161-9.

Jacob, François, and Jacques Monod. (1961) "Genetic regulatory mechanisms in the synthesis of proteins", *Journal of Molecular Biology*, **3**: 318-56.

Jedrzejczak-Krzepkowska, M., K. L. Tkaczuk, and S. Bielecki. (2011) "Biosynthesis, purification and characterization of beta-fructofuranosidase from *Bifidobacterium longum KN29.1*", *Process Biochemistry*, **46**: 1963-72.

Jew, S., S. S. AbuMweis, and P. J. H. Jones. (2009) "Evolution of the Human Diet: Linking Our Ancestral Diet to Modern Functional Foods as a Means of Chronic Disease Prevention", *Journal of Medicinal Food*, **12**: 925-34.

Johnson, L. P., G. E. Walton, A. Psichas, G. S. Frost, G. R. Gibson, and T. G. Barraclough. (2015) "Prebiotics Modulate the Effects of Antibiotics on Gut Microbial Diversity and Functioning in Vitro", *Nutrients*, **7**: 4480-97.

Juncker, Agnieszka S., Hanni Willenbrock, Gunnar von Heijne, Søren Brunak, Henrik Nielsen, and Anders Krogh. (2003) "Prediction of lipoprotein signal peptides in Gram-negative bacteria", *Protein Science*, **12**: 1652-62.

Jung, W. S., C. K. Hong, S. Lee, C. S. Kin, S. J. Kim, S. I. Kim, and S. Rhee. (2007) "Structural and functional insights into intramolecular fructosyl transfer by inulin fructotransferase", *Journal of Biological Chemistry*, **282**: 8414-23.

Kabsch, W. (2010) "XDS", *Acta Crystallographica Section D-Biological Crystallography*, **66**: 125-32.

Kaoutari, Abdessamad El, Fabrice Armougom, Jeffrey I. Gordon, Didier Raoult, and Bernard Henrissat. (2013) "The abundance and variety of carbohydrate-active enzymes in the human gut microbiota", *Nat Rev Micro*, **11**: 497-504.

Kawakami, Akira, and Midori Yoshida. (2012) "Graminan breakdown by fructan exohydrolase induced in winter wheat inoculated with snow mold", *Journal of Plant Physiology*, **169**: 294-302.

Kelley, Lawrence A., Stefans Mezulis, Christopher M. Yates, Mark N. Wass, and Michael J. E. Sternberg. (2015) "The Phyre2 web portal for protein modeling, prediction and analysis", *Nat. Protocols*, **10**: 845-58.

Kibbe, Warren A. (2007) "OligoCalc: an online oligonucleotide properties calculator'", *Nucleic Acids Research*, **35**: 3.

Kim, Hyeun Bum, and Richard E. Isaacson. (2015) "The pig gut microbial diversity: Understanding the pig gut microbial ecology through the next generation high throughput sequencing", *Veterinary Microbiology*, **177**: 242-51.

King, D. E., A. G. Mainous, and C. A. Lambourne. (2012) "Trends in Dietary Fiber intake in the United States, 1999-2008", *Journal of the Academy of Nutrition and Dietetics*, **112**: 642-48.

Knights, Dan, Tonya L. Ward, Christopher E. McKinlay, Hannah Miller, Antonio Gonzalez, Daniel McDonald, and Rob Knight. (2014) "Rethinking "Enterotypes"", *Cell Host & Microbe*, **16**: 433-37.

Knol, J., P. Scholtens, C. Kafka, J. Steenbakkers, S. Gross, K. Helm, M. Klarczyk, H. Schopfer, H. M. Bockler, and J. Wells. (2005) "Colon microflora in infants fed formula with galacto- and fructo-oligosaccharides: More like breast-fed infants", *Journal of Pediatric Gastroenterology and Nutrition*, **40**: 36-42.

Koenig, Jeremy E., Aymé Spor, Nicholas Scalfone, Ashwana D. Fricker, Jesse Stombaugh, Rob Knight, Largus T. Angenent, and Ruth E. Ley. (2011) "Succession of microbial consortia in the developing infant gut microbiome", *Proceedings of the National Academy of Sciences of the United States of America*, **108**: 4578-85.

Kolida, S., and G. R. Gibson. (2007) "Prebiotic capacity of inulin-type fructans", *Journal of Nutrition*, **137**: 2503S-06S.

Kornberg, Hans, and Christopher Lourenco. (2006) "A route for fructose utilization by *Escherichia coli* involving the fucose regulon", *Proceedings of the National Academy of Sciences*, **103**: 19496-99.

Koropatkin, N. M., E. A. Cameron, and E. C. Martens. (2012) "How glycan metabolism shapes the human gut microbiota", *Nature Reviews Microbiology*, **10**: 323-35.

Koropatkin, Nicole M., Eric C. Martens, Jeffrey I. Gordon, and Thomas J. Smith. (2008) "Starch Catabolism by a Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices", *Structure*, **16**: 1105-15.

Koropatkin, Nicole M., and Thomas J. Smith. (2010) "SusG: A Unique Cell-Membrane-Associated α-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules", *Structure*, **18**: 200-15.

Koropatkin, Nicole, Eric C. Martens, Jeffrey I. Gordon, and Thomas J. Smith. (2009) "The Structure Of A SusD Homolog, BT1043, Involved In Mucin O-Glycan Utilization In A Prominent Human Gut Symbiont", *Biochemistry*, **48**: 1532-42.

Koshland, D. E., Jr., and S. S. Stein. (1954) "Correlation of bond breaking with enzyme specificity; cleavage point of invertase", *The Journal of Biological Chemistry*, **208**: 139-48.

Kumagai, Erina Yoshida; Haruko Sakurama; Masashi Kiyohara; Masahiro Nakajima; Motomitsu Kitaoka; Hisashi Ashida; Junko Hirose; Takane

Katayama; Kenji Yamamoto; Hidehiko. (2012) "*Bifidobacterium longum subsp. infantis* uses two different β-galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides", *Glycobiology*, **22**: 7.

Laemmli, U. K. (1970) "Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4", *Nature*, **227**: 680-&.

Lammens, W., K. Le Roy, L. Schroeven, A. Van Laere, A. Rabijns, and W. Van den Ende. (2009) "Structural insights into glycoside hydrolase family 32 and 68 enzymes: functional implications", *Journal of Experimental Botany*, **60**: 727-40.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. (2007) "Clustal W and clustal X version 2.0", *Bioinformatics*, **23**: 2947-48.

Larsbrink, Johan, Theresa E. Rogers, Glyn R. Hemsworth, Lauren S. McKee, Alexandra S. Tauzin, Oliver Spadiut, Stefan Klinter, Nicholas A. Pudlo, Karthik Urs, Nicole M. Koropatkin, A. Louise Creagh, Charles A. Haynes, Amelia G. Kelly, Stefan Nilsson Cederholm, Gideon J. Davies, Eric C. Martens, and Harry Brumer. (2014) "A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes", *Nature*, **506**: 498-502.

Lasken, Roger S., and Jeffrey S. McLean. (2014) "Recent advances in genomic DNA sequencing of microbial species from single cells", *Nature reviews. Genetics*, **15**: 577-84.

Le Chatelier, Emmanuelle, Trine Nielsen, Junjie Qin, Edi Prifti, Falk Hildebrand, Gwen Falony, Mathieu Almeida, Manimozhiyan Arumugam, Jean-Michel Batto, Sean Kennedy, Pierre Leonard, Junhua Li, Kristoffer Burgdorf, Niels Grarup, Torben Jorgensen, Ivan Brandslund, Henrik Bjorn Nielsen, Agnieszka S. Juncker, Marcelo Bertalan, Florence Levenez, Nicolas Pons, Simon Rasmussen, Shinichi Sunagawa, Julien Tap, Sebastian Tims, Erwin G. Zoetendal, Soren Brunak, Karine Clement, Joel Dore, Michiel Kleerebezem, Karsten Kristiansen, Pierre Renault, Thomas Sicheritz-Ponten, Willem M. de Vos, Jean-Daniel Zucker, Jeroen Raes, Torben Hansen, H. I. T. consortium Meta, Peer Bork, Jun Wang, S. Dusko Ehrlich, Oluf Pedersen, and H. I. T. consortium additional members Meta. (2013) "Richness of human gut microbiome correlates with metabolic markers", *Nature*, **500**: 541-46.

Leach, J. D., and K. D. Sobolik. (2010) "High dietary intake of prebiotic inulin-type fructans in the prehistoric Chihuahuan Desert", *British Journal of Nutrition*, **103**: 1558-61.

Lewis, M. (2005) "The lac repressor", *Comptes Rendus Biologies*, **328**: 521-48.

Lewis, M., G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Z. Lu. (1996) "Crystal structure of the lactose operon repressor and its complexes with DNA and inducer", *Science*, **271**: 1247-54.

Ley, Ruth E., Daniel A. Peterson, and Jeffrey I. Gordon. (2006) "Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine", *Cell*, **124**: 837-48.

Lombard, Vincent, Hemalatha Golaconda Ramulu, Elodie Drula, Pedro M. Coutinho, and Bernard Henrissat. (2014) "The carbohydrate-active enzymes database (CAZy) in 2013", *Nucleic Acids Research*, **42**: D490-D95.

Lopez, M. G., N. A. Mancilla-Margalli, and G. Mendoza-Diaz. (2003) "Molecular structures of fructans from *Agave tequilana* Weber var. azul", *Journal of Agricultural and Food Chemistry*, **51**: 7835-40.

Louis, Petra, Georgina L. Hold, and Harry J. Flint. (2014) "The gut microbiota, bacterial metabolites and colorectal cancer", *Nat Rev Micro*, **12**: 661-72.

Lowe, Elisabeth C., Arnaud Baslé, Mirjam Czjzek, Susan J. Firbank, and David N. Bolam. (2012) "A scissor blade-like closing mechanism implicated in transmembrane signaling in a *Bacteroides* hybrid two-component system", *Proceedings of the National Academy of Sciences of the United States of America*, **109**: 7298-303.

Lozupone, Catherine A., Jesse I. Stombaugh, Jeffrey I. Gordon, Janet K. Jansson, and Rob Knight. (2012) "Diversity, stability and resilience of the human gut microbiota", *Nature*, **489**: 220-30.

Mancilla-Margalli, N. A., and M. G. Lopez. (2006) "Water-soluble carbohydrates and fructan structure patterns from Agave and Dasylirion species", *Journal of Agricultural and Food Chemistry*, **54**: 7832-39.

Manichanh, C., L. Rigottier-Gois, E. Bonnaud, K. Gloux, E. Pelletier, L. Frangeul, R. Nalin, C. Jarrin, P. Chardon, P. Marteau, J. Roca, and J. Dore. (2006) "Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach", *Gut*, **55**: 205-11.

Marcobal, A., and J. L. Sonnenburg. (2012) "Human milk oligosaccharide consumption by intestinal microbiota", *Clinical Microbiology and Infection*, **18**: 12-15.

Markowitz, Victor M., I. Min A. Chen, Krishna Palaniappan, Ken Chu, Ernest Szeto, Yuri Grechkin, Anna Ratner, Biju Jacob, Jinghua Huang, Peter Williams, Marcel Huntemann, Iain Anderson, Konstantinos Mavromatis, Natalia N. Ivanova, and Nikos C. Kyrpides. (2012) "IMG: the integrated microbial genomes database and comparative analysis system", *Nucleic Acids Research*, **40**: D115-D22.

Martens, E. C., N. M. Koropatkin, T. J. Smith, and J. I. Gordon. (2009) "Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm", *Journal of Biological Chemistry*, **284**: 24673-77.

Martens, Eric C., Elisabeth C. Lowe, Herbert Chiang, Nicholas A. Pudlo, Meng Wu, Nathan P. McNulty, D. Wade Abbott, Bernard Henrissat, Harry J. Gilbert, David N. Bolam, and Jeffrey I. Gordon. (2011) "Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts", *Plos Biology*, **9**: e1001221.

Mazmanian, S. K., J. L. Round, and D. L. Kasper. (2008) "A microbial symbiosis factor prevents intestinal inflammatory disease", *Nature*, **453**: 620-25.

McCarter, John D., and G. Stephen Withers. (1994) "Mechanisms of enzymatic glycoside hydrolysis", *Current Opinion in Structural Biology*, **4**: 885-92.

McFarland, L. V. (1998) "Epidemiology, Risk Factors and Treatments for Antibiotic-Associated Diarrhea", *Digestive Diseases*, **16**: 292-307.

Meng, G. Y., and K. Futterer. (2003) "Structural framework of fructosyl transfer in *Bacillus subtilis* levansucrase", *Nature Structural Biology*, **10**: 935-41.

Metzker, M. L. (2010) "Sequencing technologies - the next generation", *Nature Reviews Genetics*, **11**: 31-46.

Miller, T. L., and M. J. Wolin. (1996) "Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora", *Applied and Environmental Microbiology*, **62**: 1589-92.

Mineo, H., H. Hara, N. Shigematsu, Y. Okuhara, and F. Tomita. (2002) "Melibiose, difructose anhydride III and difructose anhydride IV enhance net calcium absorption in rat small and large intestinal epithelium by increasing the passage of tight junctions in vitro", *Journal of Nutrition*, **132**: 3394-99.

Mineo, Hitoshi, Midori Amano, Hideyuki Chiji, Norihiro Shigematsu, Fusao Tomita, and Hiroshi Hara. (2004) "Indigestible Disaccharides Open Tight Junctions and Enhance Net Calcium, Magnesium, and Zinc Absorption in Isolated Rat Small and Large Intestinal Epithelium", *Digestive Diseases and Sciences*, **49**: 122-32.

Moffatt, Barbara A., and F. William Studier. (1987) "T7 lysozyme inhibits transcription by T7 RNA polymerase", *Cell*, **49**: 221-27.

Moshfegh, A. J., J. E. Friday, J. P. Goldman, and J. K. C. Ahuja. (1999) "Presence of inulin and oligofructose in the diets of Americans", *Journal of Nutrition*, **129**: 1407S-11S.

Muir, J. G., S. J. Shepherd, O. Rosella, R. Rose, J. S. Barrett, and P. R. Gibson. (2007) "Fructan and free fructose content of common Australian vegetables and fruit", *Journal of Agricultural and Food Chemistry*, **55**: 6619-27.

Mullis, Kary B., and Fred A. Faloona. 1987. 'Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction.' in Wu Ray (ed.), *Methods in Enzymology* (Academic Press).

Niness, K. R. (1999) "Inulin and oligofructose: What are they?", *Journal of Nutrition*, **129**: 1402S-06S.

O'Keefe, Stephen J. D., Jia V. Li, Leo Lahti, Junhai Ou, Franck Carbonero, Khaled Mohammed, Joram M. Posma, James Kinross, Elaine Wahl, Elizabeth Ruder, Kishore Vipperla, Vasudevan Naidoo, Lungile Mtshali, Sebastian Tims, Philippe G. B. Puylaert, James DeLany, Alyssa Krasinskas, Ann C. Benefiel, Hatem O. Kaseb, Keith Newton, Jeremy K. Nicholson, Willem M. de Vos, H. Rex Gaskins, and Erwin G. Zoetendal. (2015) "Fat, fibre and cancer risk in African Americans and rural Africans", *Nat Commun*, **6**.

Patel, Seema, and Arun Goyal. (2012) "The current trends and future perspectives of prebiotics research: a review", *3 Biotech*, **2**: 115-25.

Penders, J., C. Thijs, C. Vink, F. F. Stelma, B. Snijders, I. Kummeling, P. A. van den Brandt, and E. E. Stobberingh. (2006) "Factors influencing the composition of the intestinal microbiota in early infancy", *Pediatrics*, **118**: 511-21.

Petersen, Thomas Nordahl, Soren Brunak, Gunnar von Heijne, and Henrik Nielsen. (2011) "SignalP 4.0: discriminating signal peptides from transmembrane regions", *Nat Meth*, **8**: 785-86.

Picard, C., J. Fioramonti, A. Francois, T. Robinson, F. Neant, and C. Matuchansky. (2005) "Review article: *Bifidobacteria* as probiotic agents – physiological effects and clinical benefits", *Alimentary Pharmacology & Therapeutics*, **22**: 495-512.

Poornam, Guru Prasad, Atsushi Matsumoto, Hisashi Ishida, and Steven Hayward. (2009) "A method for the analysis of domain movements in large biomolecular complexes", *Proteins: Structure, Function, and Bioinformatics*, **76**: 201-12.

Pouyez, Jenny, Aurélie Mayard, Anne-Michèle Vandamme, Guillaume Roussel, Eric A. Perpète, Johan Wouters, Isabelle Housen, and Catherine Michaux. (2012) "First crystal structure of an endo-inulinase, INU2, from *Aspergillus ficuum*: Discovery of an extra-pocket in the catalytic domain responsible for its endo-activity", *Biochimie*, **94**: 2423-30.

Praznik, Werner, Renate Löppert, Josè M. Cruz Rubio, Klaus Zangger, and Anton Huber. (2013) "Structure of fructo-oligosaccharides from leaves and stem of *Agave tequilana* Weber, var. azul", *Carbohydrate Research*, **381**: 64-73.

Qin, Junjie, Ruiqiang Li, Jeroen Raes, Manimozhiyan Arumugam, Kristoffer Solvsten Burgdorf, Chaysavanh Manichanh, Trine Nielsen, Nicolas Pons, Florence Levenez, Takuji Yamada, Daniel R. Mende, Junhua Li, Junming Xu, Shaochuan Li, Dongfang Li, Jianjun Cao, Bo Wang, Huiqing Liang, Huisong Zheng, Yinlong Xie, Julien Tap, Patricia Lepage, Marcelo Bertalan, Jean-Michel Batto, Torben Hansen, Denis Le Paslier, Allan Linneberg, H. Bjorn Nielsen, Eric Pelletier, Pierre Renault, Thomas Sicheritz-Ponten, Keith Turner, Hongmei Zhu, Chang Yu, Shengting Li, Min Jian, Yan Zhou, Yingrui Li, Xiuqing Zhang, Songgang Li, Nan Qin, Huanming Yang, Jian Wang, Soren Brunak, Joel Dore, Francisco Guarner, Karsten Kristiansen, Oluf Pedersen, Julian Parkhill, Jean Weissenbach, Peer Bork, S. Dusko Ehrlich, and Jun Wang. (2010) "A human gut microbial gene catalogue established by metagenomic sequencing", *Nature*, **464**: 59-65.

Qin, Junjie, Yingrui Li, Zhiming Cai, Shenghui Li, Jianfeng Zhu, Fan Zhang, Suisha Liang, Wenwei Zhang, Yuanlin Guan, Dongqian Shen, Yangqing Peng, Dongya Zhang, Zhuye Jie, Wenxian Wu, Youwen Qin, Wenbin Xue, Junhua Li, Lingchuan Han, Donghui Lu, Peixian Wu, Yali Dai, Xiaojuan Sun, Zesong Li, Aifa Tang, Shilong Zhong, Xiaoping Li, Weineng Chen, Ran Xu, Mingbang Wang, Qiang Feng, Meihua Gong, Jing Yu, Yanyan Zhang, Ming Zhang, Torben Hansen, Gaston Sanchez, Jeroen Raes, Gwen Falony, Shujiro Okuda, Mathieu Almeida, Emmanuelle LeChatelier, Pierre Renault, Nicolas Pons, Jean-Michel Batto, Zhaoxi Zhang, Hua Chen, Ruifu Yang, Weimou Zheng, Songgang Li, Huanming Yang, Jian Wang, S. Dusko Ehrlich, Rasmus Nielsen, Oluf Pedersen, Karsten Kristiansen, and Jun Wang. (2012) "A metagenome-wide association study of gut microbiota in type 2 diabetes", *Nature*, **490**: 55-60.

Rakoff-Nahoum, Seth, Michael J Coyne, and Laurie E Comstock. (2014) "An Ecological Network of Polysaccharide Utilization among Human Intestinal Symbionts", *Current Biology*, **24**: 40-49.

Ramirez-Farias, C., K. Slezak, Z. Fuller, A. Duncan, G. Holtrop, and P. Louis. (2009) "Effect of inulin on the human gut microbiota: stimulation of *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*", *British Journal of Nutrition*, **101**: 541-50.

Rampelli, Simone, Stephanie L Schnorr, Clarissa Consolandi, Silvia Turroni, Marco Severgnini, Clelia Peano, Patrizia Brigidi, Alyssa N Crittenden, Amanda G Henry, and Marco Candela. (2015) "Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota", *Current Biology*, **25**: 1682-93.

Rastall, R. A., and G. R. Gibson. (2015) "Recent developments in prebiotics to selectively impact beneficial microbes and promote intestinal health", *Current Opinion in Biotechnology*, **32**: 42-46.

Reddy, V. A., and F. Maley. (1990) "Identification of an Active-Site Residue in Yeast Invertase by Affinity Labeling and Site-Directed Mutagenesis", *Journal of Biological Chemistry*, **265**: 10817-20.

Reeves, A. R., J. N. Delia, J. Frias, and A. A. Salyers. (1996) "A *Bacteroides thetaiotaomicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch", *Journal of Bacteriology*, **178**: 823-30.

Reeves, A. R., G. R. Wang, and A. A. Salyers. (1997) "Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*", *Journal of Bacteriology*, **179**: 643-49.

Reichardt, Nicole, Sylvia H. Duncan, Pauline Young, Alvaro Belenguer, Carol McWilliam Leitch, Karen P. Scott, Harry J. Flint, and Petra Louis. (2014) "Phylogenetic distribution of three pathways for propionate production within the human gut microbiota", *ISME J*, **8**: 1323-35.

Rivière, Audrey, Mérilie Gagnon, Stefan Weckx, Denis Roy, and Luc De Vuyst. (2015) "Mutual cross-feeding interactions between *Bifidobacterium longum NCC2705* and *Eubacterium rectale ATCC 33656* explain the bifidogenic and butyrogenic effects of arabinoxylan-oligosaccharides", *Applied and Environmental Microbiology*.

Roberfroid, M. B., J. A. E. Van Loo, and G. R. Gibson. (1998) "The bifidogenic nature of chicory inulin and its hydrolysis products", *Journal of Nutrition*, **128**: 11-19.

Roberfroid, M., G. R. Gibson, L. Hoyles, A. L. McCartney, R. Rastall, I. Rowland, D. Wolvers, B. Watzl, H. Szajewska, B. Stahl, F. Guarner, F. Respondek, K. Whelan, V. Coxam, M. J. Davicco, L. Leotoing, Y. Wittrant, N. M. Delzenne, P. D. Cani, A. M. Neyrinck, and A. Meheust. (2010) "Prebiotic effects: metabolic and health benefits", *British Journal of Nutrition*, **104**: S1-S63.

Roediger, W. E. W. (1980) "Role of Anarobic-Bacteria in the Metabolic Welfare of the Colonic Mucosa in Man", *Gut*, **21**: 793-98.

Rogowski, Artur, Jonathon A. Briggs, Jennifer C. Mortimer, Theodora Tryfona, Nicolas Terrapon, Elisabeth C. Lowe, Arnaud Basle, Carl Morland, Alison M. Day, Hongjun Zheng, Theresa E. Rogers, Paul Thompson, Alastair R. Hawkins, Madhav P. Yadav, Bernard Henrissat, Eric C. Martens, Paul Dupree, Harry J. Gilbert, and David N. Bolam. (2015) "Glycan complexity dictates microbial resource allocation in the large intestine", *Nat Commun*, **6**.

Rosewarne, C. P., P. B. Pope, J. L. Cheung, and M. Morrison. (2014) "Analysis of the bovine rumen microbiome reveals a diversity of Sus-like polysaccharide utilization loci from the bacterial phylum Bacteroidetes", *Journal of Industrial Microbiology & Biotechnology*, **41**: 601-06.

Salazar, Nuria, Evelyne M. Dewulf, Audrey M. Neyrinck, Laure B. Bindels, Patrice D. Cani, Jacques Mahillon, Willem M. de Vos, Jean-Paul Thissen, Miguel Gueimonde, Clara G. de los Reyes-Gavilán, and Nathalie M. Delzenne. "Inulin-type fructans modulate intestinal *Bifidobacterium species* populations and decrease fecal short-chain fatty acids in obese women", *Clinical Nutrition*, **34**: 501-07.

Sanders, Mary Ellen, and Maria L. Marco. (2010) "Food Formats for Effective Delivery of Probiotics", *Annual Review of Food Science and Technology*, **1**: 65-85.

Schaafsma, G., and J. L. Slavin. (2015) "Significance of Inulin Fructans in the Human Diet", *Comprehensive Reviews in Food Science and Food Safety*, **14**: 37-47.

Schell, Mark A., Maria Karmirantzou, Berend Snel, David Vilanova, Bernard Berger, Gabriella Pessi, Marie-Camille Zwahlen, Frank Desiere, Peer Bork, Michele Delley, R. David Pridmore, and Fabrizio Arigoni. (2002) "The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract", *Proceedings of the National Academy of Sciences*, **99**: 14422-27.

Schnorr, Stephanie L., Marco Candela, Simone Rampelli, Manuela Centanni, Clarissa Consolandi, Giulia Basaglia, Silvia Turroni, Elena Biagi, Clelia Peano, Marco Severgnini, Jessica Fiori, Roberto Gotti, Gianluca De Bellis, Donata Luiselli, Patrizia Brigidi, Audax Mabulla, Frank Marlowe, Amanda G. Henry, and Alyssa N. Crittenden. (2014) "Gut microbiome of the Hadza hunter-gatherers", *Nat Commun*, **5**.

Scott, Karen P., Jenny C. Martin, Christophe Chassard, Marlene Clerget, Joanna Potrykus, Gill Campbell, Claus-Dieter Mayer, Pauline Young, Garry Rucklidge, Alan G. Ramsay, and Harry J. Flint. (2011) "Substrate-driven gene expression in *Roseburia inulinivorans*: Importance of inducible enzymes in the utilization of inulin and starch", *Proceedings of the National Academy of Sciences*, **108**: 4672-79.

Segain, J. P., D. Raingeard de la Blétière, A. Bourreille, V. Leray, N. Gervois, C. Rosales, L. Ferrier, C. Bonnet, H. M. Blottière, and J. P. Galmiche. (2000) "Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease", *Gut*, **47**: 397-403.

Segata, Nicola. (2015) "Gut Microbiome: Westernization and the Disappearance of Intestinal Diversity", *Current Biology*, **25**: R611-R13.

Sela, D. A., J. Chapman, A. Adeuya, J. H. Kim, F. Chen, T. R. Whitehead, A. Lapidus, D. S. Rokhsar, C. B. Lebrilla, J. B. German, N. P. Price, P. M. Richardson, and D. A. Mills. (2008) "The genome sequence of *Bifidobacterium longum subsp infantis* reveals adaptations for milk utilization within the infant microbiome", *Proceedings of the National Academy of Sciences of the United States of America*, **105**: 18964-69.

Sela, David A., and David A. Mills. (2010) "Nursing our microbiota: molecular linkages between *Bifidobacteria* and milk oligosaccharides", *Trends in microbiology*, **18**: 298-307.

Seydel, Anke, Pierre Gounon, and Anthony P. Pugsley. (1999) "Testing the '+2 rule' for lipoprotein sorting in the *Escherichia coli* cell envelope with a new genetic selection", *Molecular Microbiology*, **34**: 810-21.

Shapiro, Sarah. (2012) "Complex Carbohydrate Recognition by Our Resident Gut Microbiota", Newcastle University.

Shipman, J. A., J. E. Berleman, and A. A. Salyers. (2000) "Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*", *Journal of Bacteriology*, **182**: 5365-72.

Shipman, Joseph A., Kyu Hong Cho, Hilary A. Siegel, and Abigail A. Salyers. (1999) "Physiological Characterization of SusG, an Outer Membrane Protein Essential for Starch Utilization by *Bacteroides thetaiotaomicron*", *Journal of Bacteriology*, **181**: 7206-11.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins. (2011) "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega", *Molecular systems biology*, **7**: 539.

Singh, Nagendra, Ashish Gurav, Sathish Sivaprakasam, Evan Brady, Ravi Padia, Huidong Shi, Muthusamy Thangaraju, Puttur D Prasad, Santhakumar Manicassamy, David H Munn, Jeffrey R Lee, Stefan Offermanns, and Vadivel Ganapathy. (2014) "Activation of Gpr109a, Receptor for Niacin and the Commensal Metabolite Butyrate, Suppresses Colonic Inflammation and Carcinogenesis", *Immunity*, **40**: 128-39.

Singh, Ramkrishna D., Jhumur Banerjee, and Amit Arora. (2015) "Prebiotic potential of oligosaccharides: A focus on xylan derived oligosaccharides", *Bioactive Carbohydrates and Dietary Fibre*, **5**: 19-30.

Skorupski, Karen, and Ronald K. Taylor. (1996) "Positive selection vectors for allelic exchange", *Gene*, **169**: 47-52.

Slavin, Joanne. (2013) "Fiber and Prebiotics: Mechanisms and Health Benefits", *Nutrients*, **5**: 1417-35.

Smith, Donald B., and Kevin S. Johnson. (1988) "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase", *Gene*, **67**: 31-40.

Smith, Patrick M., Michael R. Howitt, Nicolai Panikov, Monia Michaud, Carey Ann Gallini, Mohammad Bohlooly-Y, Jonathan N. Glickman, and Wendy S. Garrett. (2013) "The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis", *Science*, **341**: 569-73.

Sonnenburg, E. D., H. J. Zheng, P. Joglekar, S. K. Higginbottom, S. J. Firbank, D. N. Bolam, and J. L. Sonnenburg. (2010) "Specificity of Polysaccharide Use in Intestinal *Bacteroides Species* Determines Diet-Induced Microbiota Alterations", *Cell*, **141**: 1241-U256.

Sonnenburg, Erica D, and Justin L Sonnenburg. (2014) "Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates", *Cell Metabolism*, **20**: 779-86.

Soyucen, Erdogan, Aynur Gulcan, Ayse Cigdem Aktuglu-Zeybek, Hasan Onal, Ertugrul Kiykim, and Ahmet Aydin. (2014) "Differences in the gut microbiota of healthy children and those with type 1 diabetes", *Pediatrics International*: n/a-n/a.

Spies, Thomas, Werner Praznik, Andreas Hofinger, Friedrich Altmann, Ernst Nitsch, and Reinhold Wutka. (1992) "The structure of the fructan sinistrin from *Urginea maritima*", *Carbohydrate Research*, **235**: 221-30.

Stierand, Katrin, and Matthias Rarey. (2010) "Drawing the PDB: Protein–Ligand Complexes in Two Dimensions", *ACS Medicinal Chemistry Letters*, **1**: 540-45.

Studier, F. William, and Barbara A. Moffatt. (1986) "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes", *Journal of Molecular Biology*, **189**: 113-30.

Subramanian, Sathish, Sayeeda Huq, Tanya Yatsunenko, Rashidul Haque, Mustafa Mahfuz, Mohammed A. Alam, Amber Benezra, Joseph DeStefano, Martin F. Meier, Brian D. Muegge, Michael J. Barratt, Laura G. VanArendonk, Qunyuan Zhang, Michael A. Province, William A. Petri Jr, Tahmeed Ahmed, and Jeffrey I. Gordon. (2014) "Persistent gut microbiota immaturity in malnourished Bangladeshi children", *Nature*, **510**: 417-21.

Suzuki,T., Tsuda,Y., Kanou,N., Inoue,T., Kumazaki,K., Nagano,S., Hirai,S., Tanaka,K. and Watanabe,K.. (2006) "*Bifidobacterium adolescentis* complete genome sequence", Published only in database: NCBI reference sequence NC_008618.1

Szabo, L., S. Jamal, H. F. Xie, S. J. Charnock, D. N. Bolam, H. J. Gilbert, and G. J. Davies. (2001) "Structure of a family 15 carbohydrate-binding module in complex with xylopentaose - Evidence that xylan binds in an approximate 3-fold helical conformation", *Journal of Biological Chemistry*, **276**: 49061-65.

Tailford, Louise E., C. David Owen, John Walshaw, Emmanuelle H. Crost, Jemma Hardy-Goddard, Gwenaelle Le Gall, Willem M. de Vos, Garry L. Taylor, and Nathalie Juge. (2015) "Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation", *Nat Commun*, **6**.

Terrapon, N., V. Lombard, H. J. Gilbert, and B. Henrissat. (2015) "Automatic prediction of polysaccharide utilization loci in Bacteroidetes species", *Bioinformatics*, **31**: 647-55.

Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon. (2007) "The Human Microbiome Project", *Nature*, **449**: 804-10.

Untergasser, Andreas, Harm Nijveen, Xiangyu Rao, Ton Bisseling, René Geurts, and Jack A. M. Leunissen. (2007) "Primer3Plus, an enhanced web interface to Primer3", *Nucleic Acids Research*, **35**: W71-W74.

Vaishampayan, Parag A., Jennifer V. Kuehl, Jeffrey L. Froula, Jenna L. Morgan, Howard Ochman, and M. Pilar Francino. (2010) "Comparative Metagenomics and Population Dynamics of the Gut Microbiota in Mother and Infant", *Genome Biology and Evolution*, **2**: 53-66.

Valluru, R., and W. Van den Ende. (2008) "Plant fructans in stress environments: emerging concepts and future prospects", *Journal of Experimental Botany*, **59**: 2905-16.

Van de Wiele, Tom, Nico Boon, Sam Possemiers, Heidi Jacobs, and Willy Verstraete. (2004) "Prebiotic effects of chicory inulin in the simulator of the human intestinal microbial ecosystem", *FEMS Microbiology Ecology*, **51**: 143-53.

van den Broek, LambertusA M., RuthM Lloyd, Gerrit Beldman, JanC Verdoes, BarryV McCleary, and AlphonsG J. Voragen. (2005) "Cloning and characterization of arabinoxylan arabinofuranohydrolase-D3 (AXHd3) from *Bifidobacterium adolescentis* DSM20083", *Applied Microbiology and Biotechnology*, **67**: 641-47.

Van der Meulen, Roel, Lefteris Makras, Kristof Verbrugghe, Tom Adriany, and Luc De Vuyst. (2006) "In Vitro Kinetic Analysis of Oligofructose Consumption by *Bacteroides* and *Bifidobacterium spp.* Indicates Different Degradation Mechanisms", *Applied and Environmental Microbiology*, **72**: 1006-12.

Vandamme, Anne-Michèle, Catherine Michaux, Aurélie Mayard, and Isabelle Housen. (2013) "Asparagine 42 of the conserved endo-inulinase INU2 motif WMNDPN from *Aspergillus ficuum* plays a role in activity specificity()", *FEBS Open Bio*, **3**: 467-72.

Vanloo, J., P. Coussement, L. Deleenheer, H. Hoebregs, and G. Smits. (1995) "On the Presence of Inulin and Oligofructose as Natural Ingredients in the Western Diet", *Critical Reviews in Food Science and Nutrition*, **35**: 525-52.

Velazquez-Martinez, J. R., R. M. Gonzalez-Cervantes, M. A. Hernandez-Gallegos, R. C. Mendiola, A. R. J. Aparicio, and M. L. A. Ocampo. (2014) "Prebiotic Potential of *Agave angustifolia* Haw Fructans with Different Degrees of Polymerization", *Molecules*, **19**: 12660-75.

Vereyken, Ingrid J., J. Albert van Kuik, Toon H. Evers, Pieter J. Rijken, and Ben de Kruijff. (2003) "Structural Requirements of the Fructan-Lipid Interaction", *Biophysical Journal*, **84**: 3147-54.

Verhoef, S. P. M., D. Meyer, and K. R. Westerterp. (2011) "Effects of oligofructose on appetite profile, glucagon-like peptide 1 and peptide YY3-36 concentrations and energy intake", *British Journal of Nutrition*, **106**: 1757-62.

Verspreet, J., S. Cimini, R. Vergauwen, E. Dornez, V. Locato, K. Le Roy, L. De Gara, W. Van den Ende, J. A. Delcour, and C. M. Courtin. (2013) "Fructan Metabolism in Developing Wheat (Triticum aestivum L.) Kernels", *Plant and Cell Physiology*, **54**: 2047-57.

Vijn, I., and S. Smeekens. (1999) "Fructan: More than a reserve carbohydrate?", *Plant Physiology*, **120**: 351-59.

Vijn, I., A. Van Dijken, N. Sprenger, K. Van Dun, P. Weisbeek, A. Wiemken, and S. Smeekens. (1997) "Fructan of the inulin neoseries is synthesized in transgenic chicory plants (*Cichorium intybus L.*) harbouring onion (*Allium cepa L.*) fructan:fructan 6G-fructosyltransferase", *The Plant Journal*, **11**: 387-98.

Vincze, T., J. Posfai, and R. J. Roberts. (2003) "NEBcutter: a program to cleave DNA with restriction enzymes", *Nucleic Acids Research*, **31**: 3688-91.

Wagner, W., F. Keller, and A. Wiemken. (1983) "Fructan Metabolism in Cereals: Induction in Leaves and Compartmentation in Protoplasts and Vacuoles", *Zeitschrift für Pflanzenphysiologie*, **112**: 359-72.

Walker, Alan W., Jennifer C. Martin, Paul Scott, Julian Parkhill, Harry J. Flint, and Karen P. Scott. (2015) "16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice", *Microbiome*, **3**: 26.

Walter, J., and R. Ley. 2011. 'The Human Gut Microbiome: Ecology and Recent Evolutionary Changes.' in S. Gottesman and C. S. Harwood (eds.), *Annual Review of Microbiology, Vol 65* (Annual Reviews: Palo Alto).

Watson, D., M. O'Connell Motherway, M. H. C. Schoterman, R. J. Joost van Neerven, A. Nauta, and D. van Sinderen. (2013) "Selective carbohydrate utilization by *Lactobacilli* and *Bifidobacteria*", *Journal of Applied Microbiology*, **114**: 1132-46.

Wen, L., R. E. Ley, P. Y. Volchkov, P. B. Stranges, L. Avanesyan, A. C. Stonebraker, C. Y. Hu, F. S. Wong, G. L. Szot, J. A. Bluestone, J. I. Gordon, and A. V. Chervonsky. (2008) "Innate immunity and intestinal microbiota in the development of Type 1 diabetes", *Nature*, **455**: 1109-U10.

Wen Z., Fiocchi C. (2004) "Inflammatory bowel disease: autoimmune or immune-mediated pathogenesis", *Clin. Dev. Immunol.*, **11**: 195-204.

Wikoff, William R., Andrew T. Anfora, Jun Liu, Peter G. Schultz, Scott A. Lesley, Eric C. Peters, and Gary Siuzdak. (2009) "Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites", *Proceedings of the National Academy of Sciences*, **106**: 3698-703.

Yan, Nieng. (2013) "Structural advances for the major facilitator superfamily (MFS) transporters", *Trends in Biochemical Sciences*, **38**: 151-59.

Yatsunenko, Tanya, Federico E. Rey, Mark J. Manary, Indi Trehan, Maria Gloria Dominguez-Bello, Monica Contreras, Magda Magris, Glida Hidalgo, Robert N. Baldassano, Andrey P. Anokhin, Andrew C. Heath, Barbara Warner, Jens Reeder, Justin Kuczynski, J. Gregory Caporaso, Catherine A. Lozupone, Christian Lauber, Jose Carlos Clemente, Dan Knights, Rob Knight, and Jeffrey I. Gordon. (2012) "Human gut microbiome viewed across age and geography", *Nature*, **486**: 222-27.

Xu, J., Bjursell,M.K., Himrod,J., Deng,S., Carmichael,L.K., Chiang,H.C., Hooper,L.V. and Gordon,J.I. (2003) “A genomic view of the human-Bacteroides thetaiotaomicron symbiosis”, *Science*, **299**:2074-2076

Zheng, Hongjun. (2009) "Nutrient Acquisition in a Human Gut Symbiont: Molecular Analysis of the Carbohydrate Utilisation Apparatus of *Bacteroides thetaiotaomicron*", Newcastle University.

Zitta, S., W. Schrabmair, G. Reibnegger, A. Meinitzer, D. Wagner, W. Estelberger, and A. R. Rosenkranz. (2013) "Glomerular Filtration Rate (GFR) determination via individual kinetics of the inulin-like polyfructosan sinistrin versus creatinine-based population-derived regression formulae", *Bmc Nephrology*, **14**.

Appendix I - Supplemental Data

I.1. Appendix Overview

The data contained within this appendix relates to the central results of this thesis (contained within Chapters 3-5). These data were not central to the chapter and were not included, however are displayed here to provide further insight into the results obtained or the interpretations made either as part of the chapter discussion sections or as part of the final discussion. Each figure or table is accompanied by a reference to the main body of text in which it is discussed, in addition to a short description of the data.

I.2. qPCR Raw Data

Tables I.1., I.2 and **I.3** display the raw data obtained during qPCR analysis of genes within putative fructan PULs (Chapters 4.5 and 5.4.2). The tables show the Cq value for each condition, normalised for cell density using the 16S primers for the same condition. The data is analysed as described (Chapter 2.15).

Table I.1. Up-regulation of SusC-homologues on Inulin and Fructose and Xylan Compared to Glucose (Chapter 4.5)

Probe Set	Substrate	Normalised Cq Value	Difference (Glucose – Substrate Cq)	Fold Change (2^Difference)	Standard Error
Xylan associated SusC-homologue *(bacova_04393)*	Glucose	21.29	0	1	1.46
	Fructose	20.12	-0.38	0.86	0.21
	Inulin	20.11	-0.37	1.10	0.46
	Xylan	10.42	9.32	660.56	127.64
Xylan associated SusC-homologue *(bacova_04505)*	Glucose	20.00	0	1	0.22
	Fructose	16.62	3.38	27.33	18.35
	Inulin	12.35	7.66	259.92	95.33
	Xylan	18.83	1.16	2.25	0.12

Table I.2. The putative *B. ovatus* fructan locus is upregulated in the presence of inulin and fructose (Chapter 4.5)

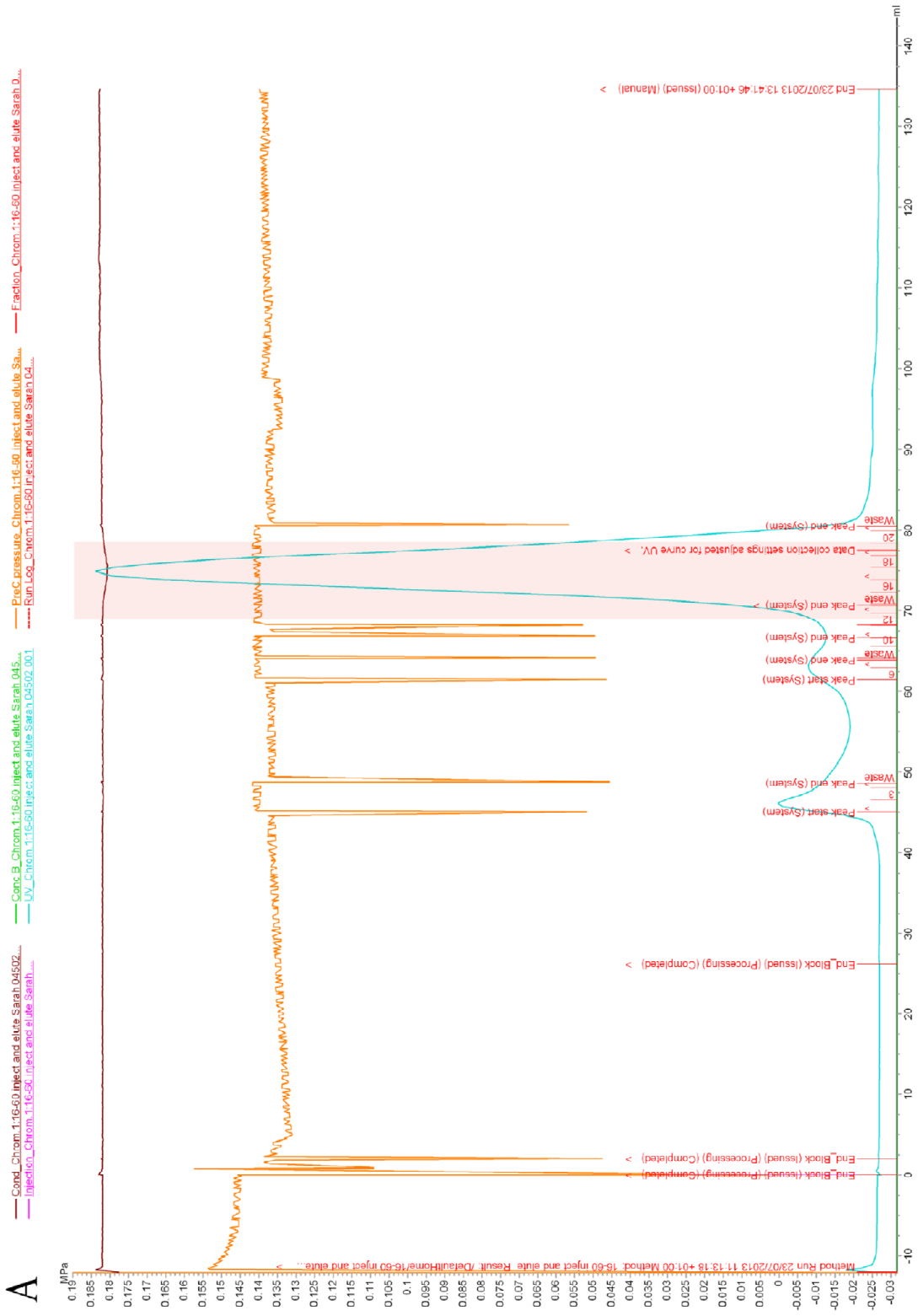
Probe Set (Locus Tag)	Substrate	Normalised Cq Value	Difference (Glucose – Substrate Cq)	Fold Change (2^Difference)	Standard Error
	Glucose	18.27	0	1	1.56
bacova_04496	Fructose	15.79	2.48	5.58	0.49
	Inulin	13.94	4.33	20.12	0.28
	Glucose	22.26	0	1	2.12
bacova_04498	Fructose	17.42	4.84	28.61	0.60
	Inulin	15.61	6.65	100.46	0.14
	Glucose	20.68	0	1	2.00
bacova_04500	Fructose	14.13	6.55	93.76	0.05
	Inulin	16.10	4.57	23.83	2.71
	Glucose	20.22	0	1	2.22
bacova_04501	Fructose	14.26	5.97	62.61	0.04
	Inulin	13.02	7.20	147.50	0.20
	Glucose	20.79	0	1	2.29
bacova_04502	Fructose	15.58	5.21	37.04	0.53
	Inulin	14.32	6.47	88.54	0.24
	Glucose	20.67	0	1	2.03
bacova_04503	Fructose	15.58	5.84	57.41	0.39
	Inulin	14.32	7.67	203.42	0.31
	Glucose	20.67	0	1	2.27
bacova_04504	Fructose	14.83	5.57	47.61	0.47
	Inulin	13.00	7.35	163.33	0.13
	Glucose	21.14	0	1	2.62
bacova_04505	Fructose	15.78	5.36	41.11	0.63
	Inulin	13.57	7.57	189.96	0.40
	Glucose	19.33	0	1	1.61
bacova_04506	Fructose	17.12	2.21	4.64	1.26
	Inulin	16.75	2.58	6.00	1.60
	Glucose	19.07	0	1	1.56
bacova_04507	Fructose	15.02	4.05	16.56	0.04
	Inulin	13.76	5.31	39.73	0.32

Table I.3. The Second Putative Inulin Associated Gene Cluster is Upregulated in the Presence of Inulin and FOS (Chapter 5.4.2)

Probe Set (Locus Tag)	Substrate	Normalised Cq Value	Difference (Glucose – Substrate Cq)	Fold Change (2^Difference)	Standard Error
	Glucose	20.07	0	1	2.43
bad_1148	FOS	17.73	2.34	5.05	0.41
	Inulin	18.77	1.31	2.47	2.04
	Glucose	11.75	0	1	0.21
bad_1149	FOS	14.09	-2.34	0.20	0.21
	Inulin	14.28	-2.53	0.17	0.34
	Glucose	13.02	0	1	0.24
bad_1150	FOS	16.20	-3.17	0.11	0.40
	Inulin	14.75	-1.72	0.30	0.79
	Glucose	16.22	0	1	0.46
bad_1151	FOS	16.29	-0.07	0.95	0.07
	Inulin	16.84	-0.63	0.65	0.82
	Glucose	12.85	0	1	0.51
bad_1152	FOS	11.85	1.00	2.00	1.23
	Inulin	13.79	-0.95	0.52	0.39
	Glucose	15.13	0	1	0.19
bad_1153	FOS	14.55	0.58	1.49	0.10
	Inulin	14.71	0.43	1.34	0.52
	Glucose	16.78	0	1	0.89
bad_1325	FOS	13.31	3.47	11.10	0.15
	Inulin	12.25	4.54	23.19	0.89
	Glucose	16.33	0	1	0.6
bad_1326	FOS	12.31	4.02	16.23	0.07
	Inulin	12.82	3.51	11.39	0.23
	Glucose	15.55	0	1	0.37
bad_1328	FOS	11.85	3.69	12.92	0.45
	Inulin	12.21	3.34	10.12	0.60
	Glucose	17.22	0	1	0.26
bad_1329	FOS	13.21	4.01	16.10	0.57
	Inulin	13.92	3.30	9.85	0.61
	Glucose	17.05	0	1	1.12
bad_1330	FOS	13.25	3.79	13.87	1.19
	Inulin	12.36	4.69	25.77	1.63

I.3. FPLC Chromatographs

FPLC Chromatographs were obtained when attempting to investigate the interaction between the two GH91 monomers (Chapter 4.6.8). The enlarged graphs are shown here for clarity. Both GH91 monomers, BACOVA_04502 and BACOVA_04503 eluted in defined peaks, regardless of whether these proteins were loaded separately (**Figure I.1**) or incubated for 10 minutes and loaded together (**Figure I.2**).


A
Cond_Chrom.1:16-60 inject and elute Sarah 04502...
Conc.B_Chrom.1:16-60 inject and elute Sarah 045...
PreC pressure_Chrom.1:16-60 inject and elute Sa...
Fraction_Chrom.1:16-60 inject and elute Sarah 0...
Injection_Chrom.1:16-60 inject and elute Sarah ...
Uv_Chrom.1:16-60 inject and elute Sarah 045002.001
Run Log_Chrom.1:16-60 inject and elute Sarah 04...
MPa
ml
Method Run 23/07/2013 11:13:18 +01:00 Method: 16-60 inject and elute Result /DefaultHome/16-60 inject and elute...
End_Block (Issued) (Processing) (Completed)
End_Block (Issued) (Processing) (Completed)
End_Block (Issued) (Processing) (Completed)
Peak start (System)
Peak end (System)
Waste
Peak start (System)
Peak end (System)
Waste
Peak end (System)
Waste
Peak end (System)
Data collection settings adjusted for curve UV...
Peak end (System)
Waste
End 23/07/2013 13:41:46 +01:00 (Issued) (Manual)

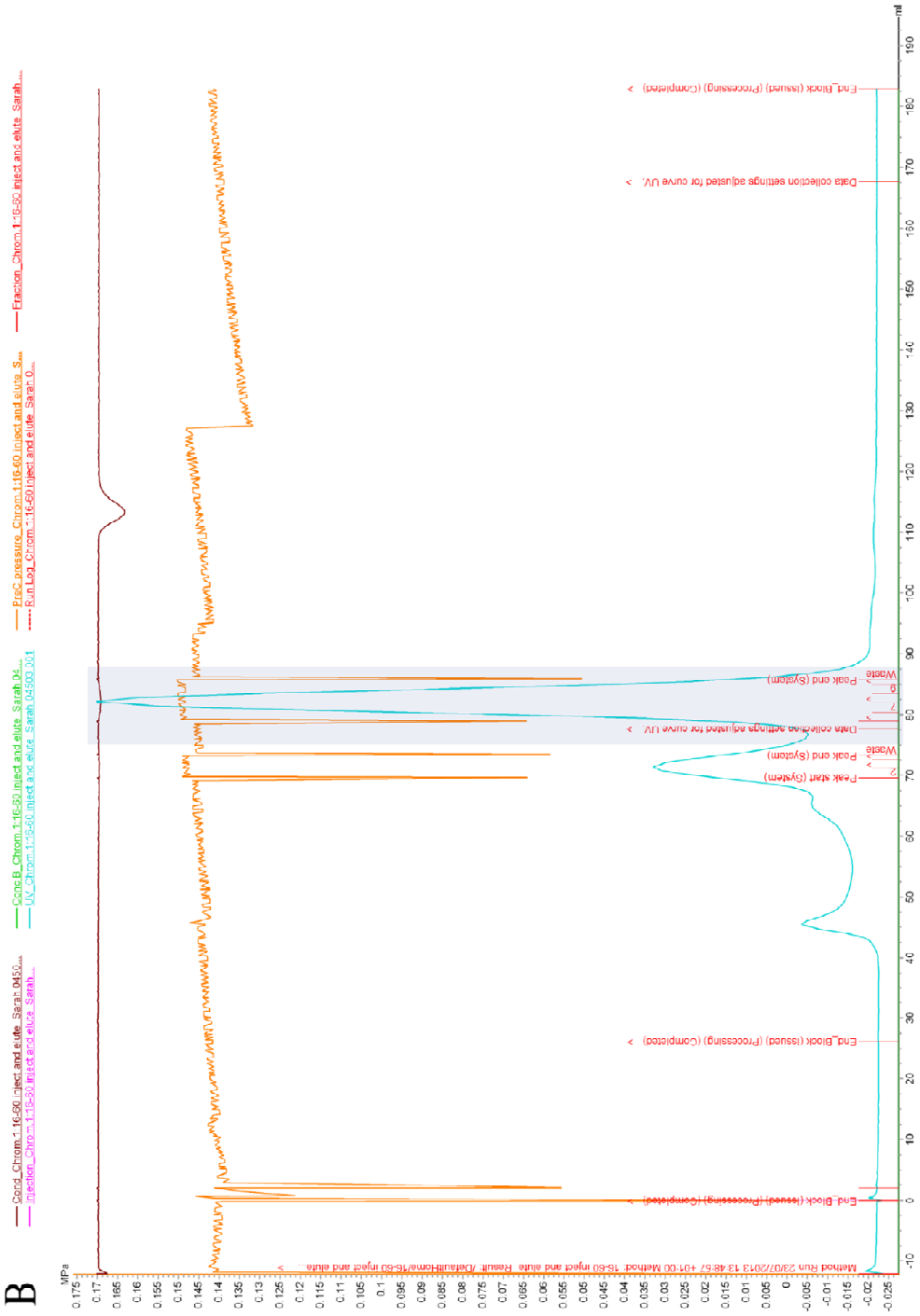


Figure I.1. FPLC Chromatograms of the two GH91 Enzymes. The two GH91 proteins (**A**) BACOVA_04502 and (**B**) BACOVA_04503 elute from the column at 70-80ml, and BACOVA_04503 77.5-86ml respectively when applied to the column individually.

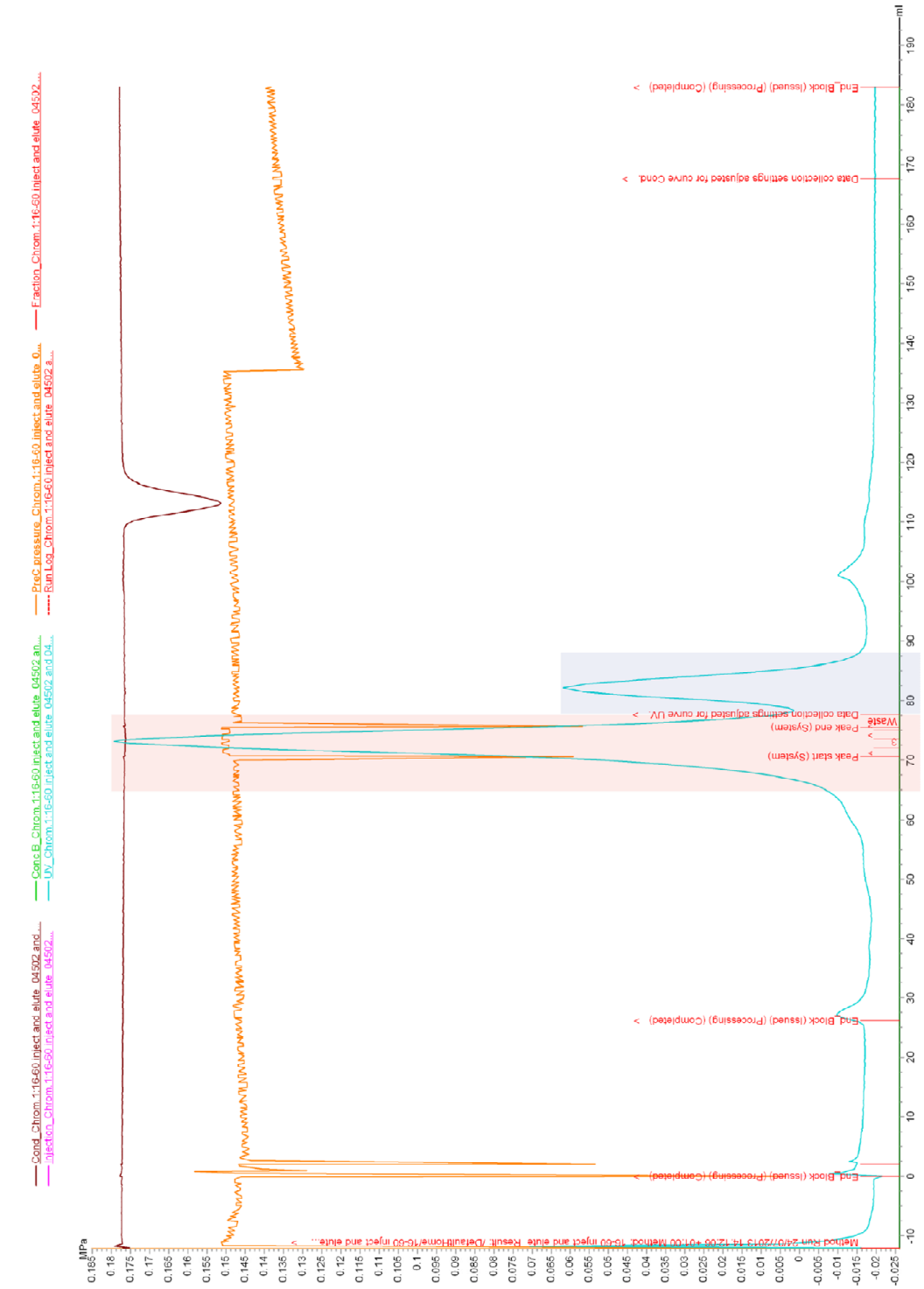


Figure I.2. FPLC Chromatogram showing the elution of the two GH91 enzymes after an incubation period. Each elutes in monomeric state, BACOVA_04502 between 70-80ml, and BACOVA_04503 between 77.5-86ml, despite pre-incubation of both enzymes together prior to application to the column.

I.4. Structural Models

Crystals were harvested, crystallography datasets were collected, the phase problem solved and models generated by Dr. Arnaud Baslé. Dr. Arnaud Baslé was instrumental during crystallography studies, three structures were obtained as part of this thesis, the SusD-homologue, BACOVA_04504 (Chapter 4.7.4) and both the open and closed form of the ESBP, BAD_1330 (Chapter 5.7.2). Here we show the data statistics (**Table I.4.**) and refinement details (**Table I.5**) for each of these structures.

Table I.4. Data statistics

	Data statistics*		
	SusD-homologue (BACOVA_04504)	ESBP – OPEN (BAD_1330o)	ESBP – CLOSED (BAD_1330c)
Beamline	I02	I02	I02
Date	26/01/14	25/01/15	25/01/15
Wavelength (Å)	0.97949	0.9749	0.9749
Resolution (Å)	48.89-1.80 (1.83-1.80)	48.47 – 1.35 (1.37 - 1.35)	48.86 – 1.55 (1.58 - 1.55)
Space group	$P2_1$	$P22_12_1$	$P2_1$
Unit-cell parameters			
a (Å)	97.78	56.39	97.7
b (Å)	65.72	106.24	108.04
c (Å)	97.78	118.51	101.85
$\alpha = \beta = \gamma$ (°)	90 90 90	90 90 90	90 113.4 90
Unit-cell volume ($Å^3$)	628312	709967	986970
Solvent content (%)	51	57	38
No. of measured reflections	369373 (17152)	1121421 (37355)	1052069 (51623)
No. of independent reflections	114809 (5569)	156497 (7600)	280325 (13822)
Completeness (%)	99.7 (99.6)	100.0 (99.2)	95.3 (94.4)
Redundancy	3.2 (3.1)	7.2 (4.9)	3.8 (3.7)
Rmerge (%)	9.3 (53.0)	5.9 (58.1)	7.2 (62.2)
$\langle I \rangle / \langle \sigma(I) \rangle$	6.5 (1.6)	13.9 (1.9)	8.1 (1.6)

*(Values in parenthesis are for the highest resolution shell).

Table I.5. Refinement details.

Refinement statistics*			
	SusD-homologue (BACOVA_04504)	ESBP – OPEN (BAD_1330o)	ESBP – CLOSED (BAD_1330c)
Rwork (%)	16.20	15.03	15.50
Rfree[#] (%)	19.43	17.81	20.27
No. of non-H atoms			
No. of protein, atoms	8587	3949	15791
No. of solvent atoms	986	428	1442
No. of ligand atoms	0	0	180
R.m.s. deviation from ideal values			
Bond angle (°)	1.4	2.53	1.6
Bond length (Å)	0.012	0.03	0.012
Average B factor (Å^2)			
Protein	21.7	17.0	25.3
Solvent	25	25.6	34.7
Ligand	N. A.	N.A.	20.4
Ramachandran plot[+], residues in most favoured regions (%)	100	100	100

*(Values in parenthesis are for the highest resolution shell).
[#]5% of the randomly selected reflections excluded from refinement.
[+]Calculated using MOLPROBITY.

I.5. *B. adolescentis* FOS-Utilisation Cluster

Putative fructan-associated genes close to the candidate FOS utilising GH32 within this cluster in *B. adolescentis* were investigated. The ESBP and associated permease was a candidate for FOS import, however no binding was observed between the annotated ESBP, BAD_1152 and fructose, sucrose or kestose (**Figure I.3**).

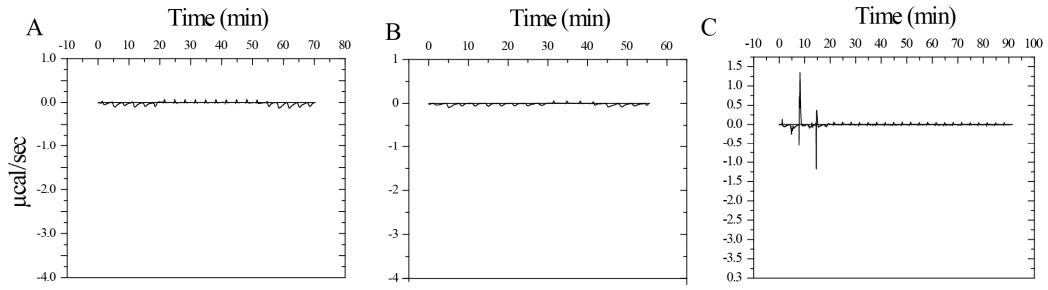


Figure I.3. No Binding to Fructose, Sucrose or Kestose was observed with BAD_1152 ESBP. ITC characterisation was attempted with BAD_1152 ESBP, however no binding of this protein to (A) fructose, (B) sucrose or (C) kestose was observed. This experiment was repeated with two protein batches and was run with and without calcium in case the protein was metal dependent however no binding was detected.

I.6. Investigation of Bacteroides spp.

Several *Bacteroides* species were available for study during the course of the project. As discussed (Chapter 6.3.) these species are likely to adopt various divergent niches in the human gut during inulin degradation. Mr. Carl Morland demonstrated that *B. uniformis* does not produce extracellular oligosaccharides through whole cell assays incubated on inulin with cells prepared after growth on inulin (**Figure I.4 A**), this is in contrast to *B. ovatus* (**Figure I.4 B**) which does produce oligosaccharides as described (Chapter 4.3). As discussed (Chapter 4.12.1, Chapter 6.3.) *B. uniformis* contains a putative endo-active GH32 family enzyme which appears to localise to the cell periplasm, rather than the cell surface as in other fructan SUS-like systems, such as *B. ovatus* inulin PUL and *B. thetaiotaomicron* PUL.

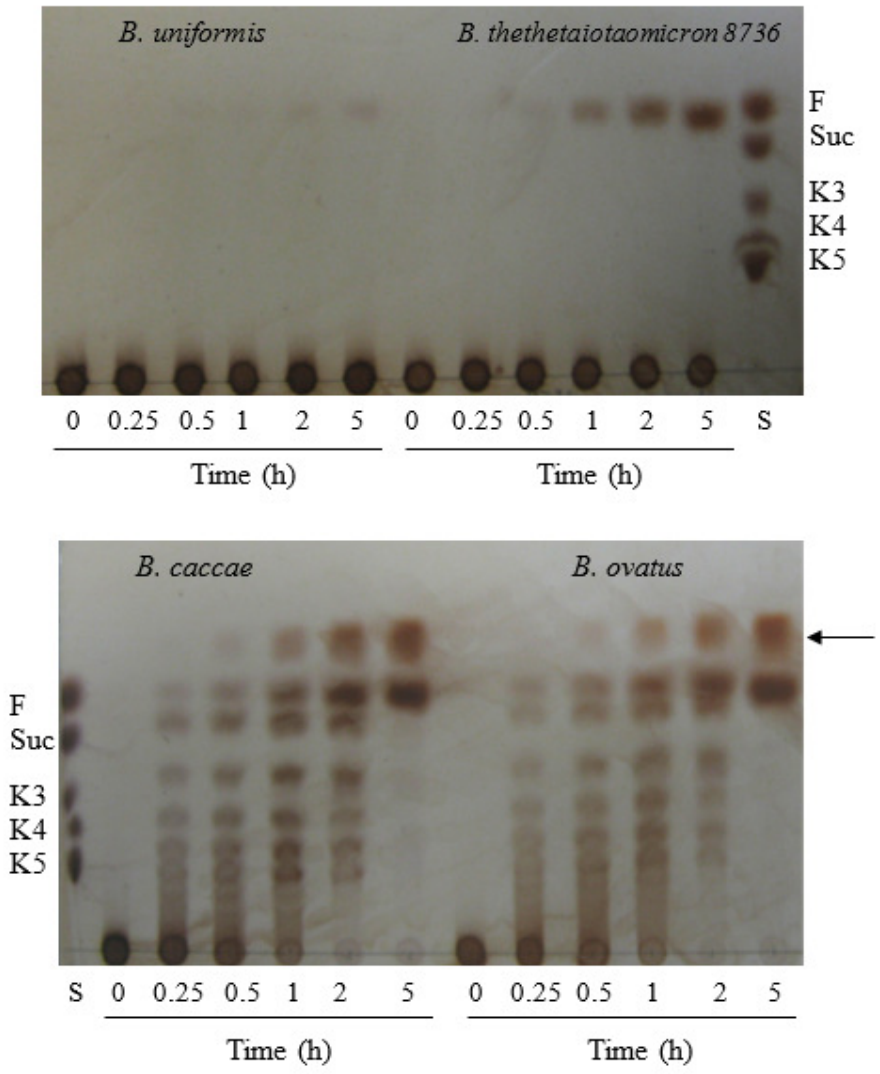


Figure I.4. Production of extracellular oligosaccharides by *Bacteroides species*. This experiment was performed by Mr. Carl Morland. *B uniformis*, *B. thetaiotaomicron 8736*, *B. caccae* and *B. ovatus* were grown on inulin and whole cells prepared. Whole cells were incubated with inulin for 0 – 5 hours to examine cell surface activity. *B. uniformis* and *B. thetaiotaomicron 8736* demonstrate no endo-inulinase activity at the cell surface, however *B. ovatus* and *B. caccae* whole cell assays show endo-inulinase activity characteristic of the heteromeric GH91 (BACOVA_04502/3 in *B. ovatus*). DFA (arrowed) is produced by both *B. ovatus* and *B. caccae*. Standards consisting of fructose (f), sucrose (suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) were run for comparison.

Glucose was used as a control for *Bacteroides spp.* growth experiments, this is because glucose supported the growth of all species tested well and allowed us to ensure that cultures were viable. We demonstrate this robust growth on glucose in *Bacteroides ovatus* and *Bacteriodes thetaiotaomicron* (**Figure I.5**) as discussed (Chapter 3.4)

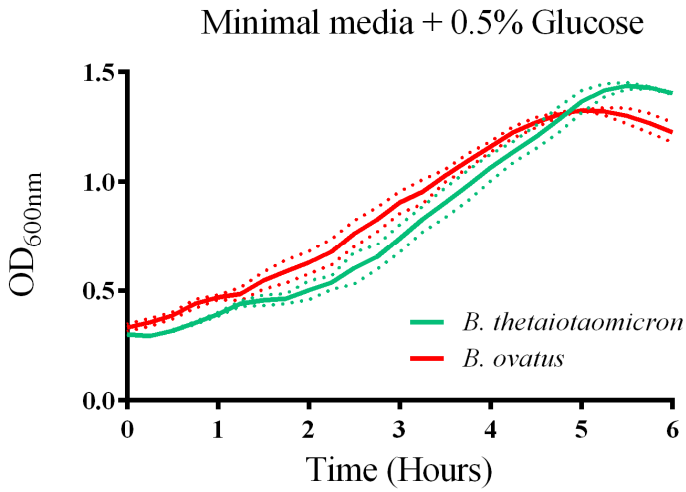


Figure I.5. *Bacteroides thetaiotaomicron* and *Bacteroides ovatus* are both able to utilise glucose. Glucose was used as a control during growth experiments using Bacteroides species. Here we show one of these controls conducted in triplicate, this data is representative.

As discussed (Chapter 4.12.1, Chapter 6.3) growth experiments were conducted with various available *Bacteroides species* on a number of fructan substrates (**Figure I.6**). Glucose controls are displayed within the figure to demonstrate robust growth, and media-only controls are also displayed as some substrates result in an increased baseline optical density at 600nm (**Figure I.6**).

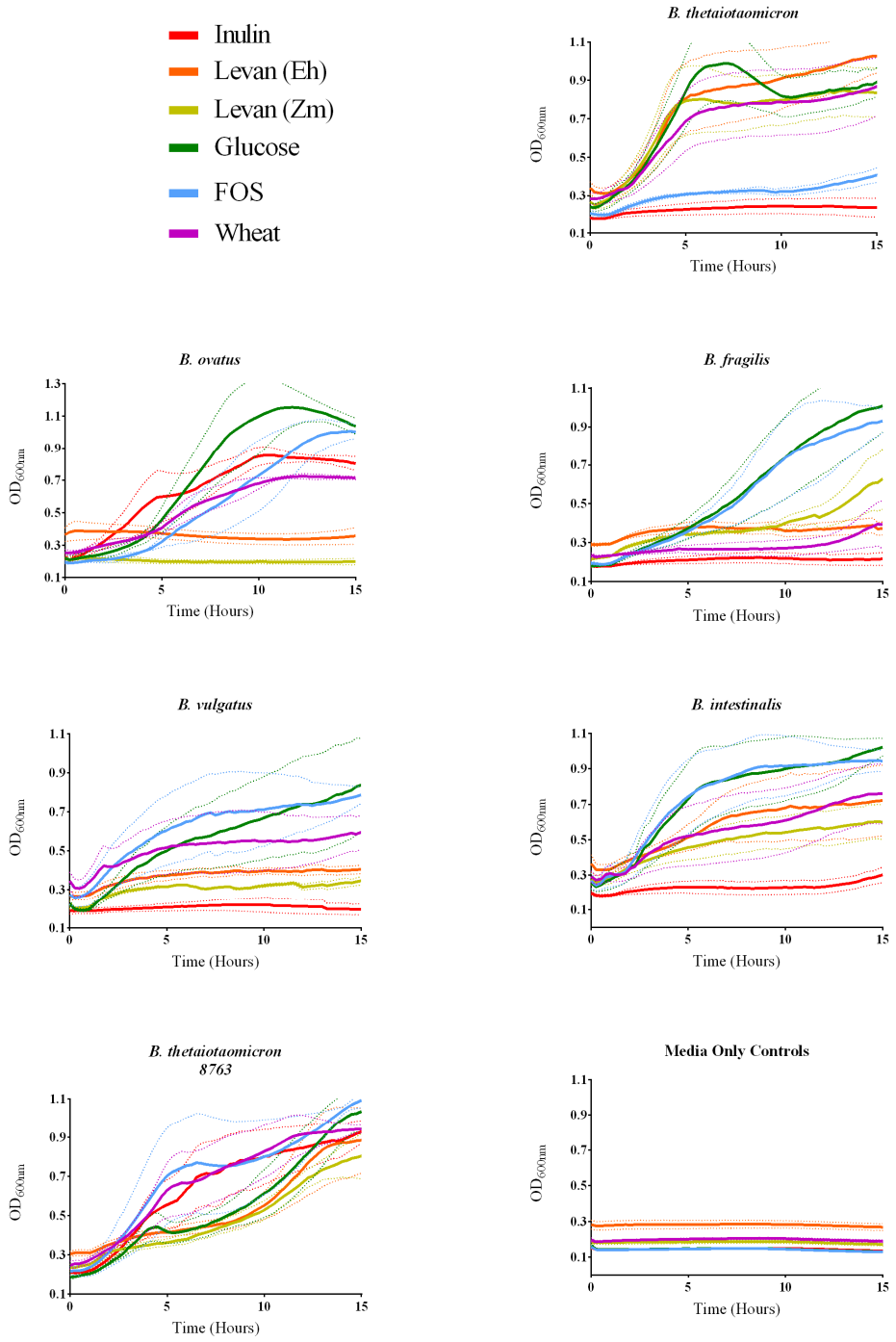


Figure I.6. The growth of various Bacteroides species upon Fructans. Each species was grown in triplicate on 0.25% inulin (chicory), Levan from E. herbicola (Eh), Levan from Z. mobilis (Zm), Glucose, FOS (DP 3-4) or wheat fructan and the data averaged. Three datasets were obtained to ensure consistency, this dataset is representative. Each dataset was conducted in duplicate.

The growth data shows that there was no significant difference between Eh and Zm levan with any of the species tested. As shown previously (Sonnenburg *et al.*, 2010; Chapter 3.4) *B. thetaiotaomicron* is a levan user and does not use inulin, whilst *B. ovatus* is an inulin user and does not use levan. Wheat fructan is able to support the growth of both *B. ovatus* and *B. thetaiotaomicron*. FOS was able to support the growth of all species except *B. thetaiotaomicron*, which was not supported well.

Of the *Bacteroides* species used, most display a preference for linkage type where polymeric fructans are utilised (*B. thetaiotaomicron*, *B. ovatus*, *B. intestinalis*). Only the *B. thetaiotaomicron 8763* strain was able to utilise both levan and inulin, though a large lag period was seen during growth on both levans tested. *B. intestinalis* appears to grow on all substrates apart from inulin, these data suggest a preference for levan, yet with better performance on FOS than *B. thetaiotaomicron*. Sonnenburg and colleagues showed that *B. fragilis* and *B. vulgatus* used fructose, but both contained GH32 family enzymes, suggesting fructans, potentially LMW FOS were degraded (Sonnenburg *et al.*, 2010). Here we demonstrate that these species are able to utilise the shorter FOS substrates. Furthermore, *B. vulgatus* appears to be able to use wheat fructan.

B. uniformis did grow on inulin and did not grow on levan, however, this phenotype was only seen using 5 ml culture volumes and could not be reproduced within the plate reader assays.

We speculate that *Bacteroides* generally fit into a levan utilisation or inulin utilisation niche, potentially due to the sensing of fructose as a proxy for the polymer (Bolam & Sonnenburg, 2011). By sensing fructose, and not a component of the polysaccharide chain, information regarding linkage type is lost such that regulation to either inulin or levan is impossible with this system. It may be metabolically expensive to encode extra proteins, including a SusC/D pair, to deal with both linkage types simultaneously. It is unclear whether *B. thetaiotaomicron* 8763 is able to do this, whilst our data was reproducible this scenario seems unlikely and further experiments should be performed to confirm this phenotype.

I.7. Bifidobacterium spp. Prefer GH91 products to HMW Inulin

To demonstrate whether it was likely that *B. ovatus* "shares" inulin with *Bifidobacterium spp.* GH91 products or inulin were used as the sole carbon source during the growth of *B. longum* and *B. adolescentis* (**Figure 1.7**). These data demonstrate that both *Bifidobacterium species* grow more rapidly and to a higher final density on GH91 products compared with inulin; indeed *B. longum* appears to only utilise the breakdown products and cannot grow on inulin.

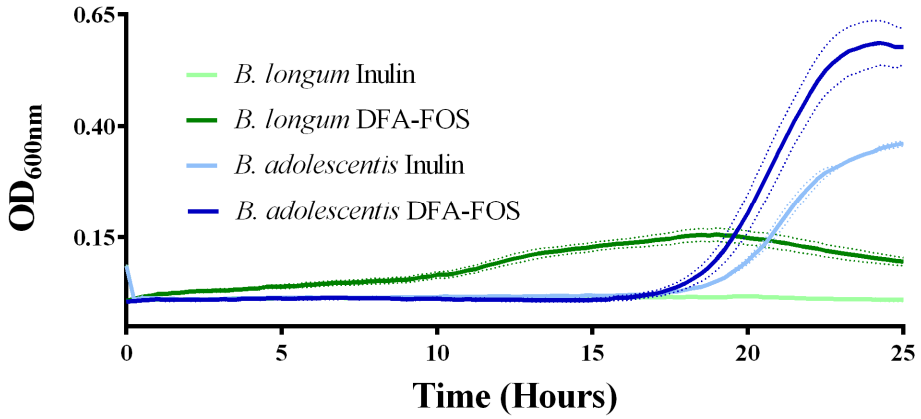


Figure I.7. *B. adolescentis* and *B. longum* prefer GH91 breakdown products to inulin as a source of carbon and energy. Inulin or GH91 products were used as the sole carbon source during growth of *B. adolescentis* or *B. longum.* Here we show that GH91 breakdown products are the preferred energy and carbon source for both species; *B. adolescentis* grows more rapidly and to a higher final OD. *B. longum* cannot grow on inulin, but some growth is supported by the GH91 products.

I.8. Protein Constructs and Primers used in this Study

The following figure shows example gels from the purification of each protein construct described in this thesis (**Figure I.8**). Key protein constructs are described (**Table I.6.**) and key primer sets listed (**Table I.7.**). All proteins were expressed with a his tag for initial studies, and all example gels show his tagged constructs purified using IMAC. For further information on his tag purification using IMAC, and on interpreting the gels, please refer to Chapter 2.12.1.

Table I.6. Protein constructs used throughout this project

Protein (Locus tag)	Vector (Restriction Enzymes)	Features	Primer Set ID	Construct Size (kDa)
BACOVA_04501 (GH32)	pET-28a (NheI, XhoI)	N-Terminal His-tag	04501	66
BACOVA_04502 (GH91)	pET-28a (NheI, XhoI)	N-Terminal His-tag	04502	69
BACOVA_04502 NTD (GH91)	pET-28a (NheI, XhoI)	N-Terminal His-tag	04502_NTD	39
BACOVA_04502 CTD (CBM)	pET-28a (NheI, XhoI)	N-Terminal His-tag	04502_CTD	30
BACOVA_04503 (GH91)	pET-28a (NheI, XhoI)	N-Terminal His-tag	04503	38
BACOVA_04504 T3 pET28a (SusD-h)	pET-28a (Nhe, XhoI)	N-Terminal His-tag	04504_pET28a	65
BACOVA_04504 T3 pGEX6P1	pGEX-6P-1 (BamHI,XhoI)	N-Terminal GST-tag	04504_pGEX6P1	82
BACOVA_04507 (GH32)	pET-28a (NheI, XhoI)	N-terminal His-tag	04507	68
BAD_1325 (GH32)	pET-28a (NheI, XhoI)	N-terminal His-tag	1325	70
BAD_1326_PBP (LacI-h PBP)	pET-28a (NcoI, XhoI)	C-terminal His-tag	1326_PBP	32
BAD_1330 (ESBP)	pET-28b (NcoI, NdeI)	C-terminal His-tag	1330	57

Table I.7. Primers used throughout this project

Primer Set ID	Forward Primer	Reverse Primer
	Cloning Primers (5' – 3')	
	see Table I.6 for construct details	
04501	CTC**GCTAGC**ACTTCTTCCTTATTAATAAAAG	CTC**CTCGAG**TTATGGCTTCAACTTATGAAC
04502	CTC**GCTAGC**CCTTCCGAAGCGCAGGATTTGACA	CTC**CTCGAG**CTATTTCTTAGCGCTTAGATAATG
04502_NTD	CTC**GCTAGC**CCTTCCGAAGCGCAGGATTTGACA	CTC**CTCGAG**TTTCACCTTTATATTTTCTTCCGTC
04502_CTD	CTC**GCTAGC**CCTTCCGAAGCGCAGGATTTGACA	CTC**CTCGAG**CTATTTCTTAGCGCTTAGATAATG
04503	CTC**GCTAGC**GATAGTGATGATAATCTGTTGTGT	CTC**CTCGAG**TTATTCTACTATGCTCACCTTATC
04504_T3_pET28	CTC**GCTAGC**GTTACTTCAGCTTATGCAGG	CTC**CTCGAG**CTAGTTAAACTTACCGTATC
04504_T3_pGEX6P1	CTC**GGATCC**GTTACTTCAGCTTATGCAGG	CTC**CTCGAG**CTAGTTAAACTTACCGTATC
04507	CTC**GCTAGC**CACACCACTAAAAGCAAG	CTC**CTCGAG**TTATAAACCCAACCGATACG
1325	CTC**GCTAGC**ATGTCGAATACCTATTCCATC	CTC**CTCGAG**TTAGCGCCAAATACCGTTCAATG
1326_PBP	CTC**CCATGGCC**AGCAGGTCGTATGTCCTCGG	CTC**CTCGAG**GGAAACGATCGAATCGCGTTC
1330	CTC**CCATGG**CGCAGGCCACCACAGACGAG	CTC**GCTAGC**CTTGGTGTACTTGTCGTACCAC

Table I.7. Primers used throughout this project (continued)

Primer Set ID	Forward Primer(s)	Reverse Primer(s)
	Sewing Primers (5' – 3')	
	used with pExchange (XbaI, BamHI) see Figure 2.3.	
04502_KO	Primer 1: CTC**TCTAGA**AATAGTGCTATTGCTTCCTTG	Primer 2: GATAAAAAAAATGTTACCGATGGTGTTACTATATTAATAATG
	Primer 3: CATTATTAATATAGTAACACCATCGGTAACATTTTTTTTATC	Primer 4: CTC**GGATCC**AATACCATGATCCAGCGTTTG
04503_E196Q_KO	Primer 1: CTC**TCTAGA**CAAGGAAAGGCATATACTC	Primer 2: GGTGTATGTTGAGGAGGATTTATATTTAAGGAATTTATAC
	Primer 3: GTATAAATTCCTTAAATATAAATCCTCCTCAACATACACC	Primer 4: CTC**GGATCC**AGCCACCTGAATGGTGTTCG
04504 _KO	Primer 1: CTC**TCTAGA**TGCTTTCTTTGACAACCGTTTG	Primer 2: TTATATTTAAGGAATTTATAAATGTTCTGTATTTAGGATA
	Primer 3: TATCCTAAATACAGAACATTTATAAATTCCTTAAATATAA	Primer 4: CTC**GGATCC**AGTATAAGCCCATTCACATG
04505 _KO	Primer 1: CTC**TCTAGA**TTGCAGGTGGAGACCAAAGAC	Primer 2: GTTCTGTATTTAGGATAAAAATAGTTTGATGTTATTAAATTAAAAG
	Primer 3: CTTTTAATTTAATAACATCAAACTATTTTTATCCTAAATACAGAAC	Primer 4: CTC**GGATCC**GACTGGGTAAGAAAAATCCG
04507_D265A _KO	Primer 1: CTC**TCTAGA**CCGATTTTCTCCGCTACGCTTC	Primer 2: CTATTGTGTAACTTCTATTTCTAGGCAACAAAGTATTCCTAG
	Primer 3: CTAGGAATACTTTGTTGCCTAGAAATAGAAGTTACACAATAG	Primer 4: CTC**GGATCC**CTCACGGAGCTGAACTCTGC

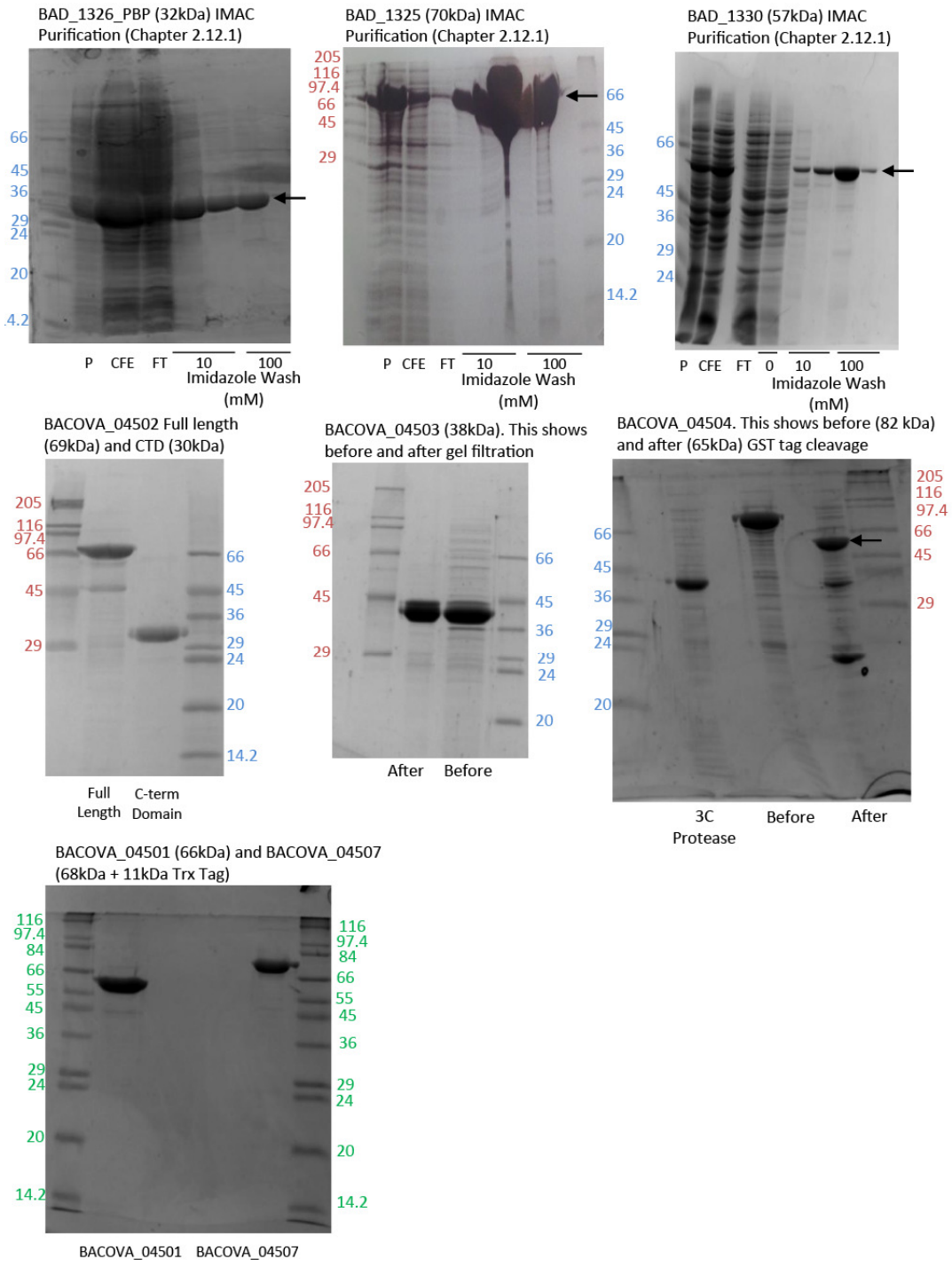


Figure I.8. Visualisation of Protein Constructs used in this Thesis. Included in this figure are example IMAC purification of each protein construct characterised in this thesis as described (Chapter 2.12.1). Purified target protein is arrowed and the size indicated for each example, molecular weight ladders are labelled for reference. For further information on IMAC purification and how to interpret these gels, please refer to Chapter 2.12.1.

Appendix II – Chemicals, Media, Enzymes and Kits

II.1. Chemicals

Amersham-Boehringer Mannheim

- 2'-Deoxyadenosine 5'-triphosphate (dATP)
- 2'-Deoxycytidine 5'-triphosphate (dCTP)
- 2'-Deoxyguanosine 5'-triphosphate (dGTP)
- 2'-Deoxythymidine 5'-triphosphate (dTTP)

BioGene

- Electrophoresis grade agarose

British Drug Houses (BDH)

- Acetic Acid (Glacial)
- Acrylamide solution (40% w/v; Electran)
- Boric acid
- Bromophenol blue
- Citric Acid
- Calcium Chloride
- Chloroform
- Dimethylformamide
- Ethanol (industrial grade)
- Hydrochloric acid
- Isopropanol
- Magnesium Chloride
- Magnesium sulphate
- Methanol

Polyethelene glycol MW 400 (PEG-400)

Polyethelene glycol MW 550 (PEG-5500)

Polyethelene glycol MW 1000 (PEG-1000)

Polyethelene glycol MW 20000 (PEG-20000)

Sodium acetate

Sodium Chloride

Sulphuric acid

Fisions

46/48% w/v NaOH

Sodium acetate trihydrate

James Burrough (F.A.D.) Ltd

Ethanol

Megazyme

Inulin from chicory

Fructan from onion

Fructan from wheat

Fructan from agave

Neosugar (FOS)

Raftilose (FOS)

Kestose

Kestotetraose

Kestopentaose

Fructotriose

Melford Laboratories

Isopropyl-β-D-thiogalactosidase (IPTG)

HEPES

G.E. Healthcare

Agarose (ultrapure)

Sigma-Aldrich

3,5-Dinitrosalasylic acid (DNSA)

Ammonium persulphate

Ampicillin

Bis tris propane

Bovine serum albumin, fraction V (BSA)

Chloramphenicol

Coomassie brilliant blue G

D-Glucose

di-Sodium hydrogen phosphate

Ethelene diamine tetra-acetic acid, disodium salt (EDTA)

Ethidium bromide

Ethylene glycol

Glycerol

Imidazole

Kanamycin

Levan from *E. herbicola*

Levan from *Z. mobilis*

N,N,N',N,-Tetramethylethylene diamine (TEMED)

Nicotinamide adenine dinucleotide-reduced

Phenol

Polyethylene glycol MW 3350 (PEG-3350)

Sodium bicarbonate

Sodium carbonate

Sodium dihydrogen orthophosphate

Sodium dodecyl sulphate (SDS)

Sucrose (nuclease free)

Trizma base (Tris)

B-Mercaptoethanol

II.2. Media

Difco

Bacto®tryptone

Bacto®yeast extract

Oxoid

Bacteriological Agar

Sigma-Aldrich

LB Broth

II.3. Enzymes

MBI Fermentase

DNA restriction endonucleases

Invitrogen

Bacteriophage T4 DNA ligase

Novagen

KOD HotStart DNA polymerase

Stratagene

Dpn1 restriction endonuclease

II.4. Kits

Qiagen

Plasmid Mini Kit

Plasmid Midi Kit

Qiaquik Gel Extraction Kit

Qiaquik PCR Purification Kit

QuantiTect Reverse Transcription Kit

RNA Protect/RNeasy RNA Extraction Kit

Sigma-Aldrich

GeneElute™ Bacterial Genomic DNA Kit

Stratagene

QuikChange™ Site-Directed Mutagenesis Kit

Megazyme

D-Mannose/D-Fructose/D-Glucose assay Kit

Roche

LightCycler® 480 SYBR Green I Master