Understanding nutrient transport across the outer membrane by members of the human gut microbiota

Amy Jane Glenwright

A thesis submitted for the degree of Doctor of Philosophy
September 2017

Institute for Cell and Molecular Biosciences
Faculty of Medical Sciences
Newcastle University

Abstract

The human gut contains a dense group of microbes termed the microbiota, which has been shown to play a major role in health and disease. Despite significant diversity at species level, the microbiota is dominated by only two phyla, the Gram-negative Bacteroidetes and Gram-positive Firmicutes. The Bacteroidetes are able to use a wide range of different complex glycans from both dietary and host sources. Bacteroidetes express groups of co-regulated, cell envelope associated proteins termed polysaccharide utilisation loci (PULs). Each PUL is specific for a different glycan with some species of *Bacteroides,* one of the major genera of the gut, encoding >100 predicted PULs. PULs encodes enzymes, binding proteins, a regulator and a transporter which are localised to the outer-membrane or periplasm for the complete degradation and transport of the target glycan. Following initial, partial degradation of the target polysaccharide at the cell surface, the resulting oligosaccharides are transported into the periplasm. This process involves an outer membrane complex consisting of a substrate binding lipoprotein (SusD-like) and a β-barrel TonB dependent transporter (SusC-like). This SusCD transporter complex is vital to utilisation of glycans by Bacteroidetes as oligosaccharide breakdown to monosaccharides occurs in the periplasm. Despite the importance of this process to microbiota function, the mechanism of SusCD function is unclear. How extracellular substrate binding by the SusD-like protein is coupled to import by SusC is unknown as other classes of TonB dependent transporters do not have a partner lipoprotein.

A SusCD complex with two auxiliary lipoproteins, BT2261-4, was expressed natively and purified directly from *Bacteroides thetaiotaomicron*. The X-ray crystal structure shows the SusC transporters form a homodimer with a SusD-like binding protein capping each barrel like a lid. The structure also shows a linear peptide bound at the interface of the SusC and SusD proteins via interactions with the peptide backbone. Expression levels of the BT2261-4 complex indicated that the proteins were required for growth under nutrient stress conditions which suggests a possible role in peptide scavenging.

A classical glycan targeting SusCD complex, BT1762-3 from the *Bacteroides thetaiotaomicron* levan PUL, was targeted by adding a His-tag to the genomic copy of the SusD-like binding protein BT1762. The His-tag allowed purification of the BT1762-3 complex which led to two further SusCD structures; apo and with a levan oligosaccharide bound. BT1762-3 has the same overall conformation as the peptide importing SusCD suggesting the dimeric SusC transporter and SusD protein 'cap' general structure is conserved across SusCD complexes. MD simulations and electrophysiology experiments allowed us to propose a model for SusCD function where the SusD-like binding protein sits on top of the SusC-like transporter like a lid and is able to open like a pedal bin to allow oligosaccharide binding and uptake.

Acknowledgements

I would first like to thank my supervisors, Bert van den Berg and Dave Bolam for their excellent advice and support throughout the last four years. Bert's expertise and skill as a membrane protein structural biologist has been essential to the success of this project.

I would also like to acknowledge all members of the Bolam, Gilbert and van den Berg labs, past and present, for their helpful ideas and discussions during the course of my PhD. Their advice and support has been crucial therefore I would like to thank, in no particular order, Jonathon Briggs, Ana Luis, Elisabeth Lowe, Fiona Cuskin, Lucy Crouch, Alan Cartmell, Jose Munoz, Justina Briliute, Dave Bulmer, Javier Abellon-Ruiz, Monisha Pathania, Anne Doble, Sarah Shapiro, Max Temple, Aurore Laborel and Immacolata Venditto. I would like to thank our technician, Carl Morland, whose cloning expertise has been essential for obtaining some difficult plasmids used in this project. I would also like to thank Harry Gilbert for interesting discussions and thought-provoking questions regarding my project and research in general.

A large portion of the work presented in Chapter 4 is based on work by Dave Bolam's previous PhD student, therefore I would like to thank Hongjun Zheng for his characterisation of the Levan PUL and for labelling his freezer samples so well.

This project includes several X-ray crystal structures therefore I would like to thank Dr Arnaud Baslé for running the crystallography facility at Newcastle and collecting data at the synchrotron.

I would also like to acknowledge collaborators who contributed to this work. Firstly, Karunakar R. Pothula, of Ulrich Kleinekathöfer's lab (Jacob's University, Bremen), for the molecular dynamics simulations. Secondly, Satya P. Bhamidimarri, of Mathias Winterhalter's lab (Jacob's University, Bremen), for the single channel electrophysiology. Finally, I would also like to acknowledge Dror S. Chorev, of Carol Robinson's lab (Oxford University), for the mass spectrometry.

I would like to sincerely thank the Barbour Foundation for kindly donating money to Newcastle University which funded my PhD. Without the financial support of the Foundation this research would not have been possible.

Finally I would like to thank my friends and family for their love and support over the last four years.

Contents

List of Figures

Chapter 1 – Introduction

Chapter 2 – Materials and Methods

Chapter 3 – Structure of a SusCD complex from *Bacteroides thetaiotaomicron*

Chapter 4 – Acquisition of Levan polysaccharides by *Bacteroides thetaiotaomicron*

Chapter 5 – Final Discussion

List of Tables

Chapter 2 – Materials and Methods Page

Chapter 4 - Acquisition of Levan polysaccharides by *Bacteroides thetaiotaomicron*

Appendix

Abbreviations

B. theta	*Bacteroides thetaiotaomicron*
BHI	Brain Heart Infusion
CFE	Cell Free Extract
DDM	n-Dodecyl β-D-maltoside
DM	n-Decyl-β-D-Maltopyranoside
DUF4480	Domain of Unknown Function – Number 4480
E. coli	*Escherichia coli*
FOS	Fructo-oligosaccharides
GOS	Galacto-oligosaccharides
HGM	Human Gut Microbiota
His_{x6}-tag	Histidine tag of six residues
HMP	Human Microbiome Project
HTCS	Hybrid Two Component System
IEX	Anion-exchange
IM	Inner Membrane
IMAC	Immobilised Metal Affinity Chromatography
ITC	Isothermal Titration Calorimetry
LB	Luria-Bertani Media
LDAO	Lauryldimethylamine-N-oxide
LR	LeMasters-Richards Media
MetaHIT	Metagenomics of the Human Intestinal Tract consortium
MFS	Major Facilitator Superfamily
MM	Minimal Media
MM-Frc	Minimal Media containing 0.5 % (w/v) Fructose
MW	Molecular Weight
OM	Outer Membrane
OMPs	Outer Membrane Proteins
PCR	Polymerase Chain Reaction
PUL	Polysaccharide Utilisation Loci
Sarkosyl	Sodium Lauroyl Sarcosine
SDM	Site Directed Mutagenesis
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEC	Size Exclusion Chromatography
SeMet	Seleno-methionine
SGBP	Surface Glycan Binding Protein
SLIC	Sequence and Ligation Independent Cloning
Sus	Starch Utilisation System
TBDT	TonB Dependent Transporters
TLC	Thin Layer Chromatography
TPRs	Tetratricopeptide Repeats
TYG	Tryptone Yeast Extract Glucose Media
WT	Wild-type

Chapter 1 – Introduction

1.1 Human Gut Microbiota

The human intestine hosts a microbial population consisting of trillions of organisms with a biomass of approximately 1.5 kg (Xu and Gordon, 2003). The vast community of microbes is dominated by bacterial species but also includes eukaryotes and archaea (Eckburg *et al.*, 2003). Some of the micro-organisms are permanent 'residents' of the gut and others are just passing through (Savage, 1977). This microbial population is generally termed the human gut microbiota (HGM) and forms a symbiotic relationship with the host, in which the host provides nutrients and the symbionts positively affect the health of the host. From birth the HGM develops alongside the host, influenced by changes in the human diet from infancy through to adulthood (Ottman *et al.*, 2012).

1.1.1 Composition of the Human Gut Microbiota

Development of high-throughput sequencing techniques has enabled the complex HGM composition to be analysed where previously it was necessary to culture each species for identification (Metzker, 2010). Such advances led to the Human Microbiome Project (HMP) in 2008 with the aim to characterise the human microbiome, including the identification of species within the gut microbiota. The Human Microbiome Project (HMP) and Metagenomics of the Human Intestinal Tract consortium (MetaHIT) studies use stool samples to analyse the HGM composition as the technique is non-invasive unlike alternative methods such as endoscopy. However, stool samples may not give an accurate representation of the overall gut microbiota, containing mostly species from the distal end of the colon. Mice studies using humanised microbiota have shown phylogenetic diversity between the early and late gastro-intestinal (GI) tract compared with mid GI tract where diversity is reduced (Gu *et al.*, 2013). Differences are believed to be due to variations in oxygen along the GI tract, where obligate anaerobes were found in locations more likely to provide anaerobic conditions (Gu *et al.*, 2013).

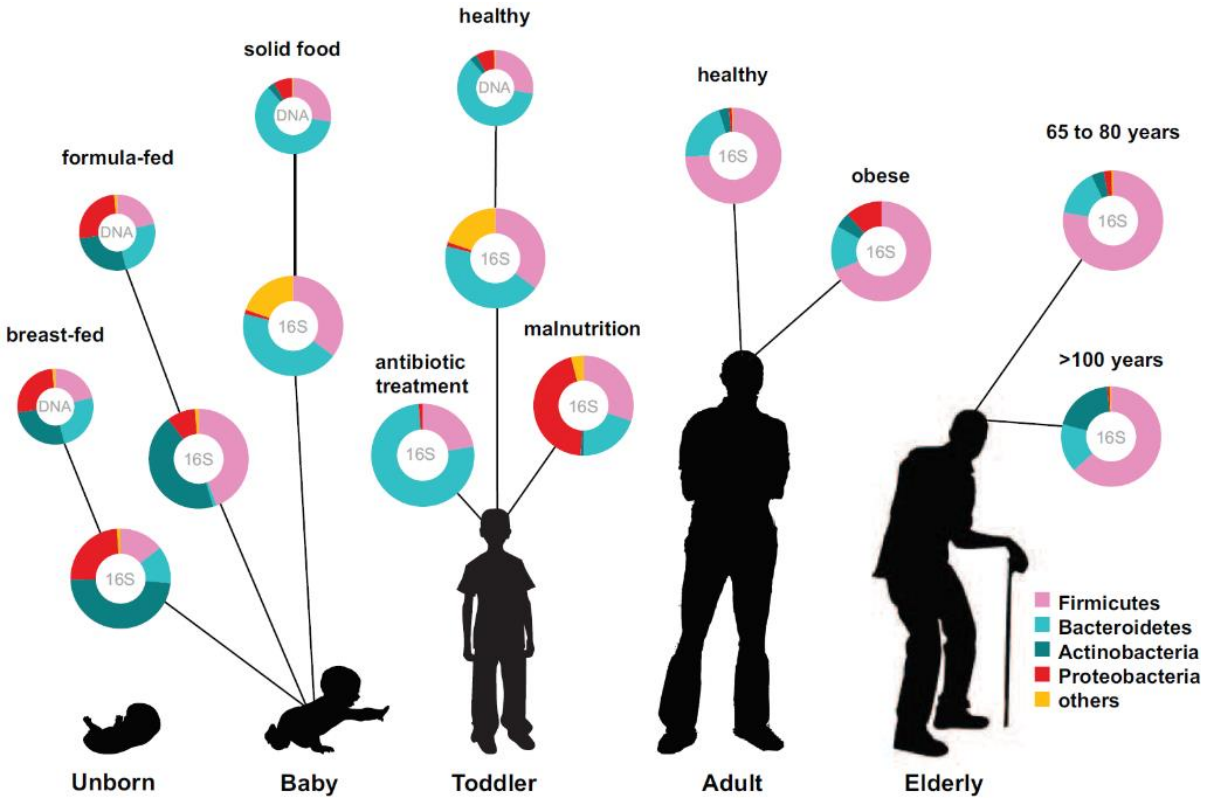


Figure 1.1 Human gut microbiota composition through lifetime
Global overview of the relative abundance of key phyla of the HGM in different stages of life. Measured by 16S RNA sequencing (16S) or metagenomic approaches (DNA). From Ottman *et al.*, 2012.

The HGM is established at birth, starting as a dynamic community dominated by *Bifidobacterium spp*. and differs significantly between breast-fed and formula-fed babies (Figure 1.1) (Schwartz *et al.*, 2012). The HGM community stabilises during the first 2-3 years (Koenig *et al.*, 2011; Scholtens *et al.*, 2012). Diversity of the HGM increases as we mature (Scholtens *et al.*, 2012), reaching peak complexity in adulthood, but dominated by two bacterial phyla Bacteroidetes and Firmicutes (Rajilic-Stojanovic *et al.*, 2009). Each individual's HGM reaches an equilibrium which remains relatively stable during a healthy adult's life (Ottman *et al.*, 2012). The exact species composition is highly variable between individuals, but the overall phylogenetic profile has been categorised during a MetaHIT study into three well-balanced states termed enterotypes which were based on the three dominant genera of the HGM; *Bacteroides, Prevotella* and *Ruminococcus* (Arumugam *et al.*, 2011). However, it has been argued that this is an oversimplification of the highly

complex HGM communities (Knights *et al.*, 2014). During later stages of life the microbiota composition returns to being less diverse and more dynamic, characterised by higher Bacteroidetes to Firmicutes ratio, increased Proteobacteria and decreased Bifidobacterium (Figure 1.1) (Biagi *et al.*, 2010).

Results of HGM identification studies have shown incredible diversity at the species level with thousands of individual species present but remarkably, over 90 % of the HGM are from only two phyla; Bacteroidetes and Firmicutes (Figure 1.2) (Rajilic-Stojanovic *et al.*, 2009; Qin *et al.*, 2010). *Bacteroides* is the most abundant and variable genus (Salyers, 1984; Arumugam *et al.*, 2011).

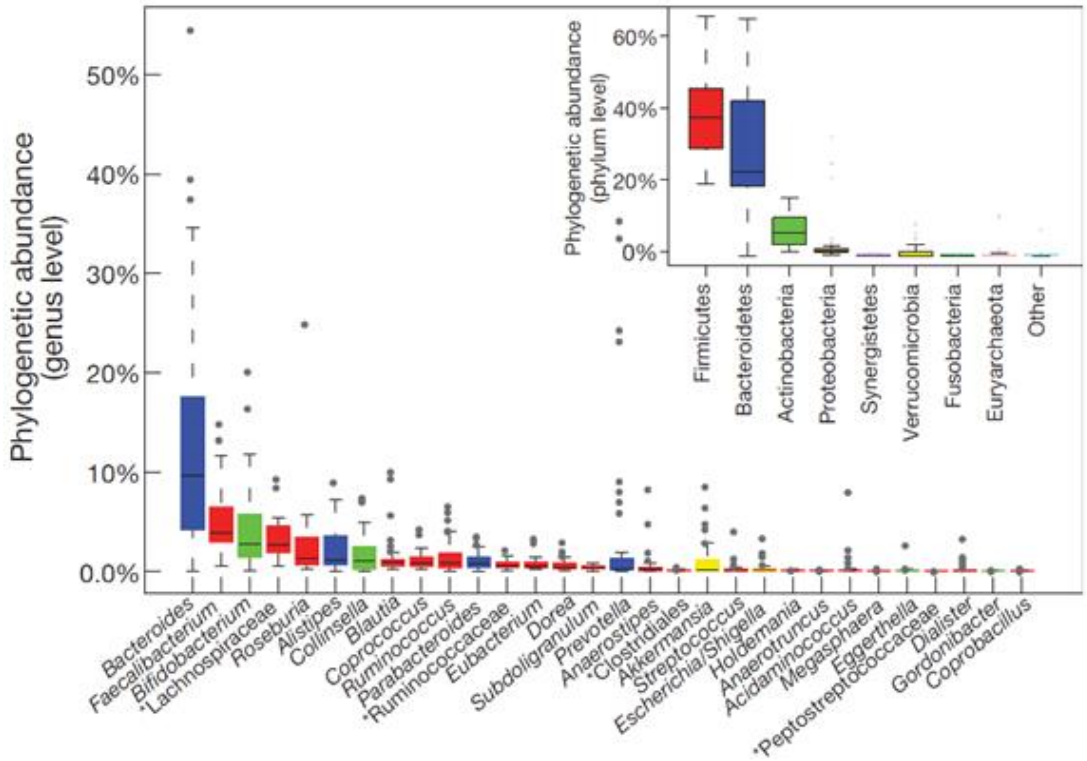


Figure 1.2 Diversity of phyla and genera within the HGM of healthy individuals.
A box plot showing the thirty most abundant genera of the human HGM. Inset shows abundance at the phylum level. Genus and phylum level abundances were calculated using reference genome-based mapping using 85% and 65% cutoff, respectively. Samples were taken from 39 individuals from 6 nationalities. From Arumugan *et al.*, 2011.

The HGM composition is highly sensitive to any change in conditions such as diet or invasion of enteric pathogens (Lupp *et al.*, 2007; David *et al.*, 2014). The resilience of the HGM varies between individuals with some patients returning to the original state and others reaching a new stable state after a disturbance (Lichtman *et al.*, 2016; Sonnenburg *et al.*, 2016). Most data regarding the HGM composition is in context of antibiotic response in humans living in urban areas which have reduced HGM diversity than populations living traditional rural lifestyles (Yatsunenko *et al.*, 2012; Martinez *et al.*, 2015). A recent study of the Hadza hunter-gatherers of Tanzania showed annual reconfiguration of the HGM in which some taxa become completely undetectable but reappear in later seasons (Smits *et al.*, 2017). Interestingly, the most seasonally volatile species of bacteria in the Hadza HGM are missing from industrialised populations, in particular Prevotellaceae sp. of the Bacteroidetes phylum (Smits *et al.*, 2017).

1.1.2 Importance of glycan fermentation by the HGM in health and disease

Humans use fats, proteins, simple sugars and starch which are absorbed in the early GI tract. Polysaccharides and oligosaccharides, commonly referred to as fibre, provide the main source of nutrients for the HGM. Members of the HGM ferment the complex carbohydrates into metabolites which interact with other microbes and the host. Short chain fatty acids (SCFAs) are the main metabolites produced during glycan fermentation and provide many of the microbiota-associated benefits to the host (Figure 1.3). SCFAs provide up to 70 % of the energy required by epthithelial cells of the colon (colonocytes) (Roy *et al.*, 2006). The three most abundant SCFAs are generally found in the human gut at a combined concentration of 50 – 100 mM (Louis *et al.*, 2014). Generally, propionate is produced by Bacteroidetes, while butyrate is generated by Firmicutes such as *Roseburia* species and also by *Bifidobacterium* from the Actinobacteria phylum (Reichardt *et al.*, 2014).

Acetate, the most abundant SCFA, is absorbed in the colon and metabolised by the liver and muscles providing up to 2 kcal/g of energy for the host (Kien, 1996; Topping and Clifton, 2001). Acetate has also been shown to aid prevention of enteropathic infection through stimulation of the intestinal epithelium (Fukuda *et al.*, 2011). Once in circulation acetate can cross the blood-brain barrier where it has been found to reduce appetite via a central homeostatic mechanism (Frost *et al.*, 2014). Acetate has been found to inhibit the action of histone deacetylases, although only in activated T-cells of the human immune system (Park *et al.*, 2015).

Butyrate has been linked to colorectal cancer inhibition by acting as a histone deacetylase inhibitor in cancerous cells, altering expression of genes involved in cell proliferation, differentiation and apoptosis. However, butyrate can be used as an energy source and stimulate growth for non-cancerous cells. This dissonance is referred to as the butyrate paradox (Lupton, 2004). A recent study has shown depletion of butyrate-producing microbes through antibiotic use increased the levels of free oxygen in the lumen of the gut, which allowed harmful bacteria such as *Escherichia coli*

(*E. coli*) and *Salmonella enterica* to multiply, leading to dysbiosis (Byndloss *et al.*, 2017). Butyrate helps restrict growth of such oxygen dependent pathogens by signalling to the cells lining the gut to maximise oxygen consumption which depletes the oxygen levels in the lumen (Byndloss *et al.*, 2017). Propionate delays gastric emptying which promotes satiety in the host and thus is believed to act as an anti-obesity factor (Arora *et al.*, 2011).

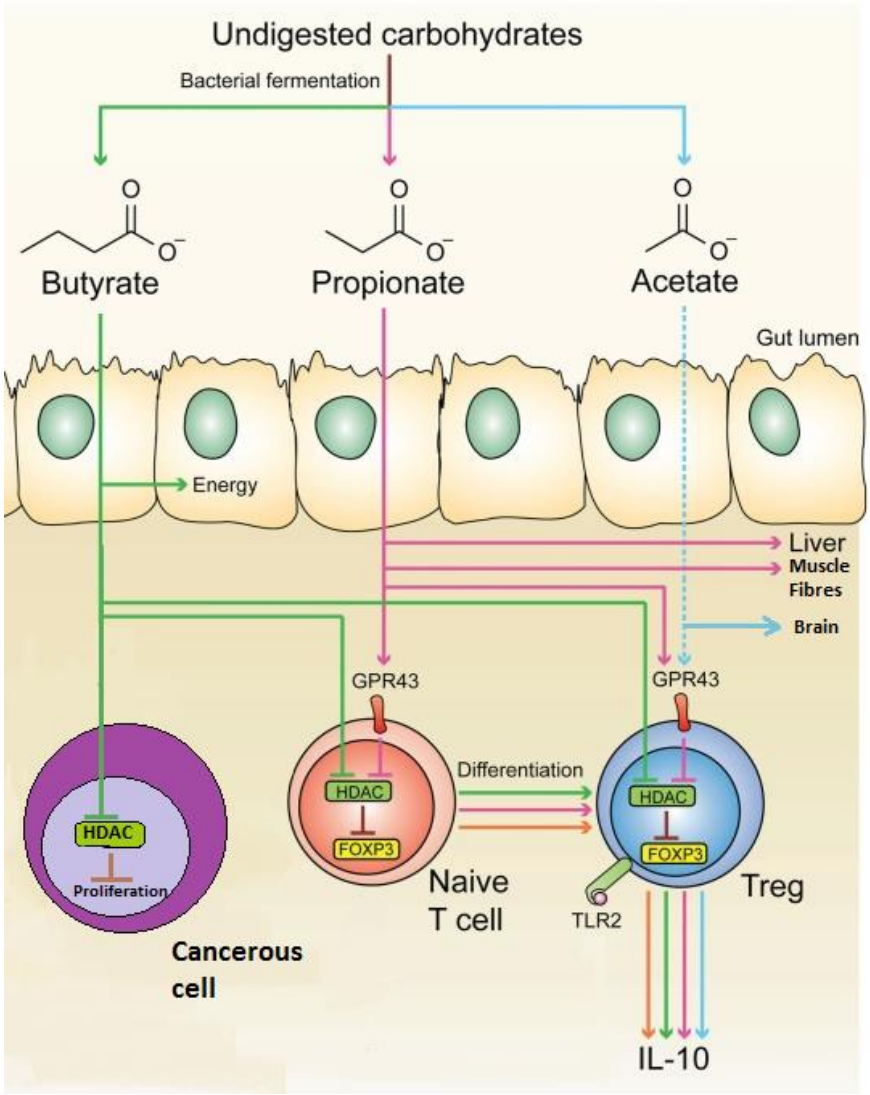


Figure 1.3 Effects of SCFAs produced by the HGM on host cells
Fermentation of carbohydrates in the anaerobic environment of the gut produces SCFAs, the most abundant of which are butyrate, propionate and acetate. All three SCFAs act as histone deacetylase (HDAC) inhibitors in different cell types to produce different downstream effects. Butyrate can be used by colonocytes as an energy source and has been shown to inhibit cancer cell proliferation. Propionate and butyrate can force differentiation of naïve T cells into Treg (regulatory) cells which produce IL-10. Acetate is metabolised by liver and muscle cells but is also able to cross the blood-brain barrier where it induces satiety signals. Adapted from Hoeppli *et al.*, 2015 by Jonathon Briggs (Thesis 2016).

When dietary fibre is limited, microbes can utilise less energetically favourable sources such as amino acids from dietary or endogenous proteins (Cummings and Macfarlane, 1991). These substrates lead to reduced fermentation although SCFAs and some branched-chain fatty acids can be produced (Cummings and Macfarlane, 1991). The resulting metabolites have been linked to insulin resistance in host cells (Newgard *et al.*, 2009).

Reduced dietary fibre has been associated with thinner colonic mucus by several studies (Brownlee *et al.*, 2003; Hedemann *et al.*, 2009; Earle *et al.*, 2015). In the absence of fibre in the diet, at least one common member of the HGM (*Bacteroides thetaiotaomicron*; *B. theta*) can shift from dietary polysaccharides to mucus glycan metabolism (Sonnenburg *et al.*, 2005). A recent study investigated the effect of chronic and intermittent dietary fibre depravation on microbiota physiology and the resulting effects on the mucus barrier (Desai *et al.*, 2016). The experiments involved assembling a synthetic gut microbiota in germ-free mice of 14 commensal members of the HGM. This study showed that a diet deficient of complex plant glycans (fibre) initiated the synthetic gut microbiota to feed on the colonic mucus layer (Figure 1.4). This increased susceptibility of the mice to *Citrobacter rodentium*, a murine pathogen that models human enteric *E. coli* infection, as the mucus layer is a primary barrier against invading pathogens (Desai *et al.*, 2016).

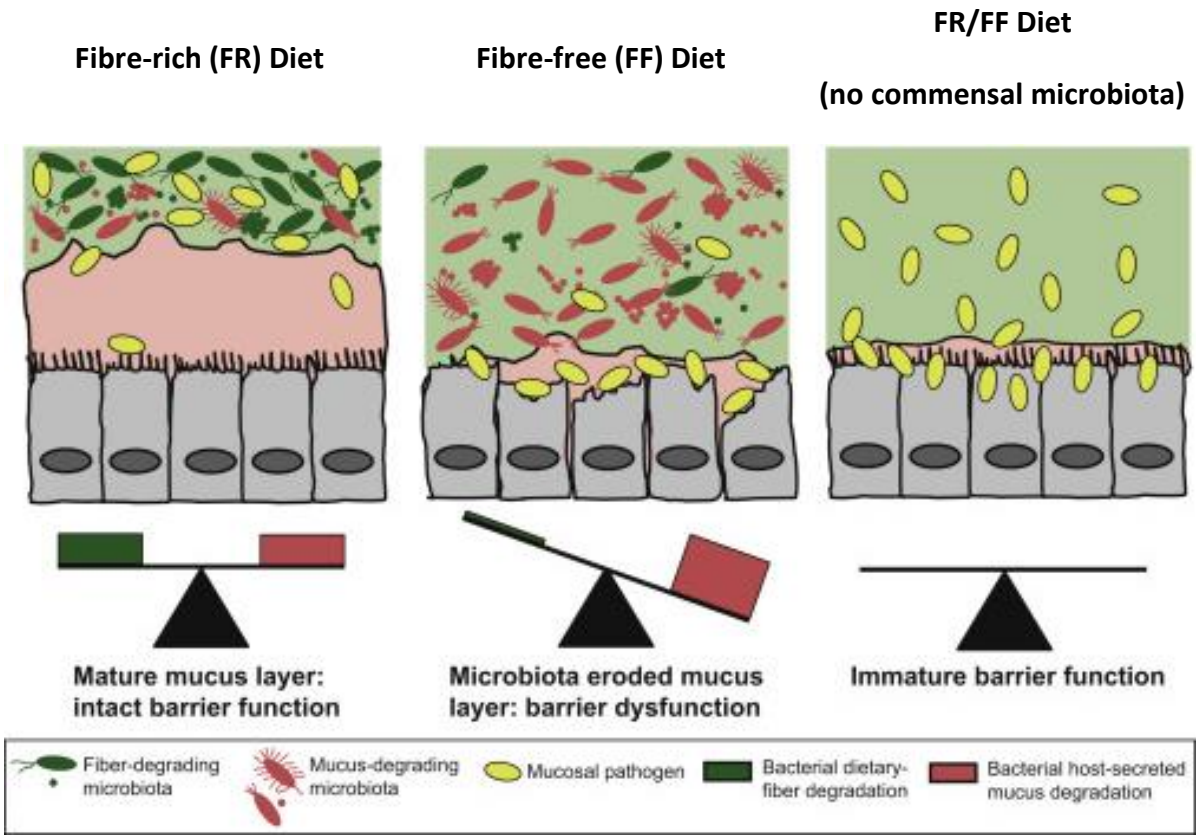


Figure 1.4 Model of how a fibre-deprived gut microbiota mediates degradation of the colonic mucus barrier and heighted pathogen susceptibility

Schematic diagram showing a model of the balance between fibre degradation and mucus degradation in fibre-rich (FR) diet-fed mice. The fibre-free (FF) diet leads to proliferation of mucus-degrading bacteria and degradation of the colonic mucus layer by the microbiota which leads to more severe colitis by *C. rodentium.* Adapted from Desai *et al.* 2016

1.1.3 Manipulation of the HGM

Abnormalities or imbalance in the HGM composition (Sekirov *et al.*, 2010) has been associated with several disease states including; inflammatory bowel disease (IBD) (Ni *et al.*, 2017), obesity (Sonnenburg and Backhed, 2016; Martinez *et al.*, 2017), hypertension (Marques *et al.*, 2017), cardiovascular disease (Lau *et al.*, 2017) and some types of cancer (Chen *et al.*, 2017). The dysbiosis may be caused by increased levels of potentially harmful micro-organisms or an increased flux through harmful metabolic pathways (Koropatkin *et al.*, 2012). Therefore manipulation of the HGM composition can be used to benefit the health of the host.

There are two main treatments for manipulation of the HGM; probiotics and prebiotics. Probiotics are live micro-organisms which can provide a health benefit to the host when administered in adequate amounts (Hill *et al.*, 2014). Prebiotics are complex carbohydrates which enrich a limited number of beneficial bacteria of the HGM (Gibson and Roberfroid, 1995). The definition was later updated to include that the prebiotic must not restrict gastric acidity and must not be absorbed by the host (Gibson *et al.*, 2004). Fructo-oligosaccharides (FOS) and galacto-oligosaccharides (GOS) are used as prebiotics although new research suggests a much wider range of glycans may produce similar positive health benefits (Chapla *et al.*, 2012; Lefranc-Millot *et al.*, 2012). One potential health benefit of prebiotics is related to sepsis. During early infancy, sepsis causes an estimated 1 million deaths worldwide every year and current there is no preventative medicine available (Thaver and Zaidi, 2009). A recent study of 4,556 newborns in rural India showed a reduction in the incidence of sepsis of 40 % when given a prebiotic (FOS) and probiotic (*Lactobacillus plantarum*) combined treatment (Panigrahi *et al.*, 2017).

1.2 Outer Membrane Proteins in Gram-negative Bacteria

Bacterial survival in any niche, such as within the HGM, requires import of nutrients. The additional membrane layer of Gram-negative bacteria, termed the outer membrane (OM), serves as a selective barrier (Figure 1.5) (Raetz and Whitfield, 2002; Nikaido, 2003). Import of required nutrients and solutes across this obstacle is essential for bacterial survival. Movement across the outer membrane is mediated by protein channels which can be divided into three classes; general porins, substrate-specific channels and active transporters (Nikaido, 2003).

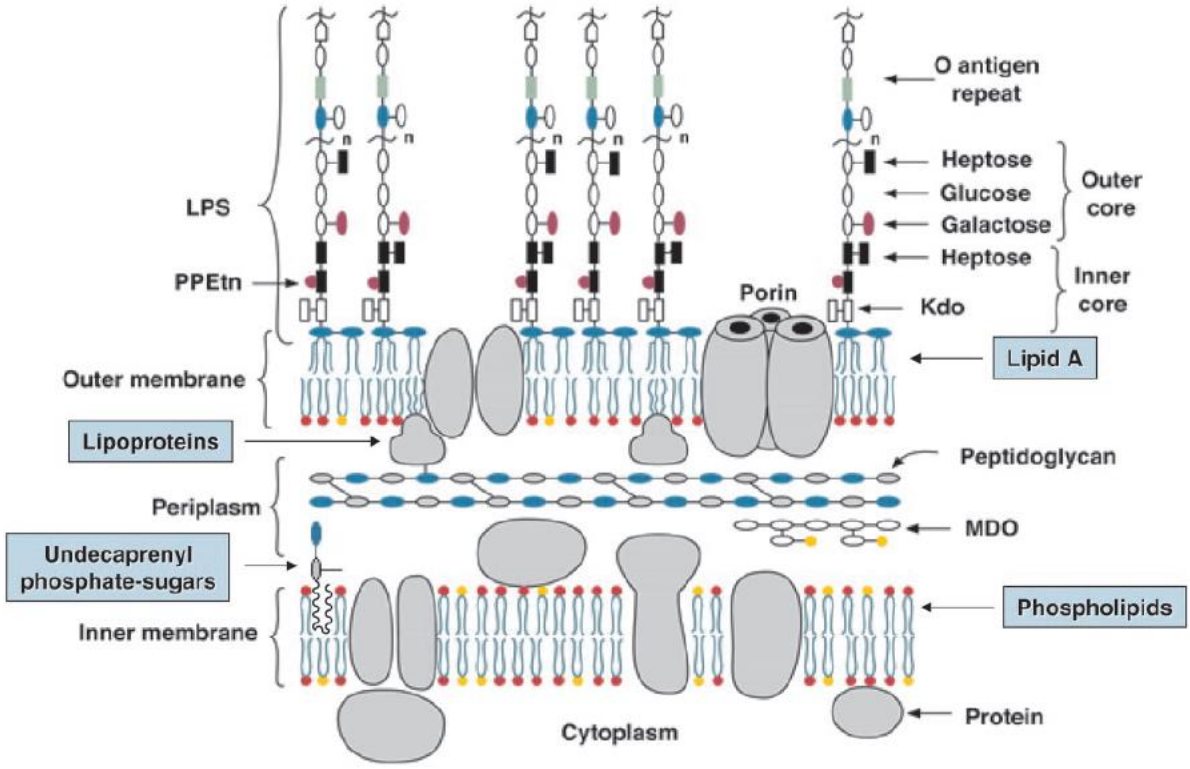


Figure 1.5 Structure of Gram negative bacteria cell walls

The cell wall of Gram negative bacteria consists of an inner and outer membrane with a peptidoglycan layer in periplasm between. The outer membrane is asymmetrical; the inner leaflet consists of phospholipids (similar to the inner membrane) and the outer leaflet is predominantly lipopolysaccharides (LPS). Figure from Raetz and Whitfield, 2002.

1.2.1 General Porins

Passive diffusion across the OM via non-specific porins is utilised by many Gram-negative bacteria. Early characterisation of such proteins from *E. coli* (eg. OmpF and OmpC) formed the basis of our understanding of many porins and so they are often referred to as 'classical porins' (Nikaido, 2003). These form transmembrane channels of anti-parallel β-barrel strands which often self-associate and function as homotrimers (Figure 1.6) (Cowan *et al.*, 1992; Schulz, 2002; Baslé *et al.*, 2006; Yamashita *et al.*, 2008). A common feature of classical porins is the restriction of the channel by loop 3 (L3) which folds back into the barrel (Figure 1.6). The water-filled pores, which are formed by these proteins in the OM, allow substrates to diffuse across the membrane driven by the concentration gradient. Although non-specific, the proteins have a preference for small (<600 Da) hydrophilic substrates with charge (Delcour, 2003). OmpF and OmpC show a preference for cations but some classical porins, such as PhoE, prefer anions. Many antibiotics gain access to cells via non-specific porins and mutations of loop 3 have been associated with antibiotic resistance (Pages *et al.*, 2008; Delcour, 2009).

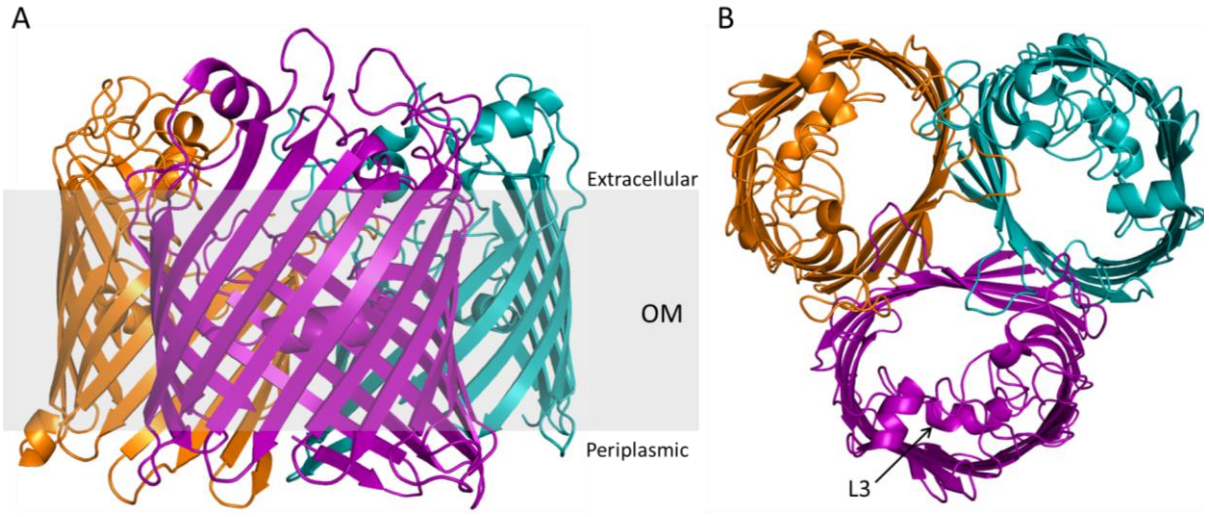


Figure 1.6 Structure of OmpC a non-specific porin from *E. coli*
A: Cartoon representation of the OmpC trimer structure (PDB ID 2J1N) within the outer membrane (OM). The individual monomers are coloured purple, orange and teal. **B:** The OmpC trimer shown from outside the cell, highlighting loop 3 (L3) which restricts the size of the barrel.

1.2.2 Substrate-specific Porins

Diffusion via general non-specific channels is no longer efficient for substrates which are present at low (micromolar) extracellular concentrations. Due to this, substrate-specific porins are especially widespread in bacteria which can survive in nutrient-poor environments such as Pseudomas *spp.* (Eren et al., 2012). Substrate specific channels form very similar structures to their non-specific counterparts, with a β-stranded barrel forming a passage across the membrane (Figure 1.7). However, specific porins such as the maltose channel LamB (Szmelcman and Hofnung, 1975; Schirmer *et al.*, 1995) and nucleoside channel Tsx (Hantke, 1976) from *E. coli* have low-affinity (micromolar to millimolar) binding sites to allow efficient diffusion of specific molecules.

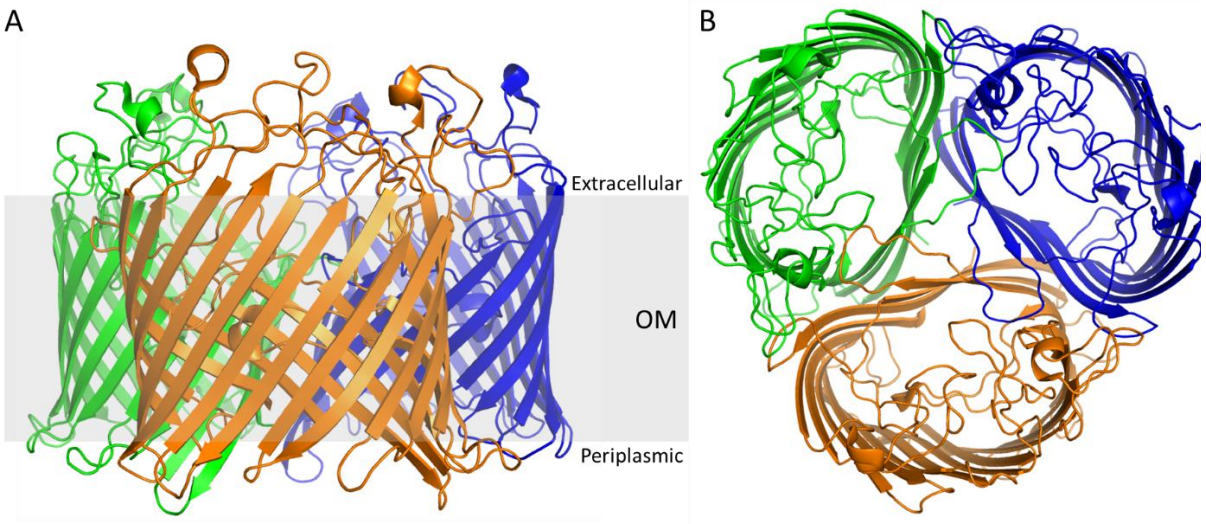


Figure 1.7 Structure of a substrate-specific maltose channel LamB from *E. coli*
A: Cartoon representation of the LamB trimer structure (PDB ID 1MAL) within the outer membrane (OM). The individual monomers are coloured orange, green and blue. **B:** The LamB trimer shown from outside the cell.

1.2.3 TonB Dependent Transporters

Active transport across the OM is complicated by the lack of ATP in the periplasm and there is no membrane potential across the OM. Therefore primary (coupled to ATP hydrolysis) and secondary active transporters commonly found in the inner membrane (IM) or Gram positive bacteria, such as the widely distributed major facilitator superfamily (MFS) transporters, cannot function in the OM (Pao *et al.*, 1998). To circumvent these problems Gram negative bacteria use TonB dependent transporters (TBDTs) which exploit energy from the proton motive force of the IM to drive transport across the outer membrane of Gram-negative bacteria. The TonB-ExbB-ExbD protein complex in the IM is required for energy transduction according to a mechanism that is still not clear (Sean Peacock *et al.*, 2005; Postle, 2007; Celia *et al.*, 2016). Early TBDT structures were of FhuA (Ferguson *et al.*, 1998) and FepA (Buchanan *et al.*, 1999), from *E. coli* and involved in iron-siderophore acquisition (Figure 1.8A and B). These structures highlighted the common features of TBDTs; a 22 stranded β-barrel monomer with an N-terminal plug domain (Noinaj *et al.*, 2010).

Upon ligand interaction with the transporter the TBDT-TonB interaction is mediated by the conserved N-terminal 'TonB box' amino acid sequence of the TBDT and the C-terminal periplasmic domain of TonB (Figure 1.8 C and D). However, the mechanism which allows to TonB to 'remodel' the plug domain of the TBDT is not understood (Hickman *et al.*, 2017). It is widely accepted that the plug domain must undergo a conformational change to allow for import of the substrate. However, the extent of the conformational change, including the possibility of the plug being completely removed from the barrel, is not known (Noinaj *et al.*, 2010).

TBDTs have been identified which transport a wide range of vital nutrients such as maltodextrin (MalA), sucrose (SuxA), metallo-organic compounds including vitamin B12 (BtuB) and haem (HasR and HemR) (Schauer *et al.*, 2008; Noinaj *et al.*, 2010). TBDTs are essential for cell survival, therefore they are considered a virulence factor in pathogenic bacteria (Perkins-Balding *et*

al., 2004; Garcia *et al.*, 2011; Noinaj *et al.*, 2012). Some TBDTs are also exploited for cell entry by colicins and phage (Cascales *et al.*, 2007).

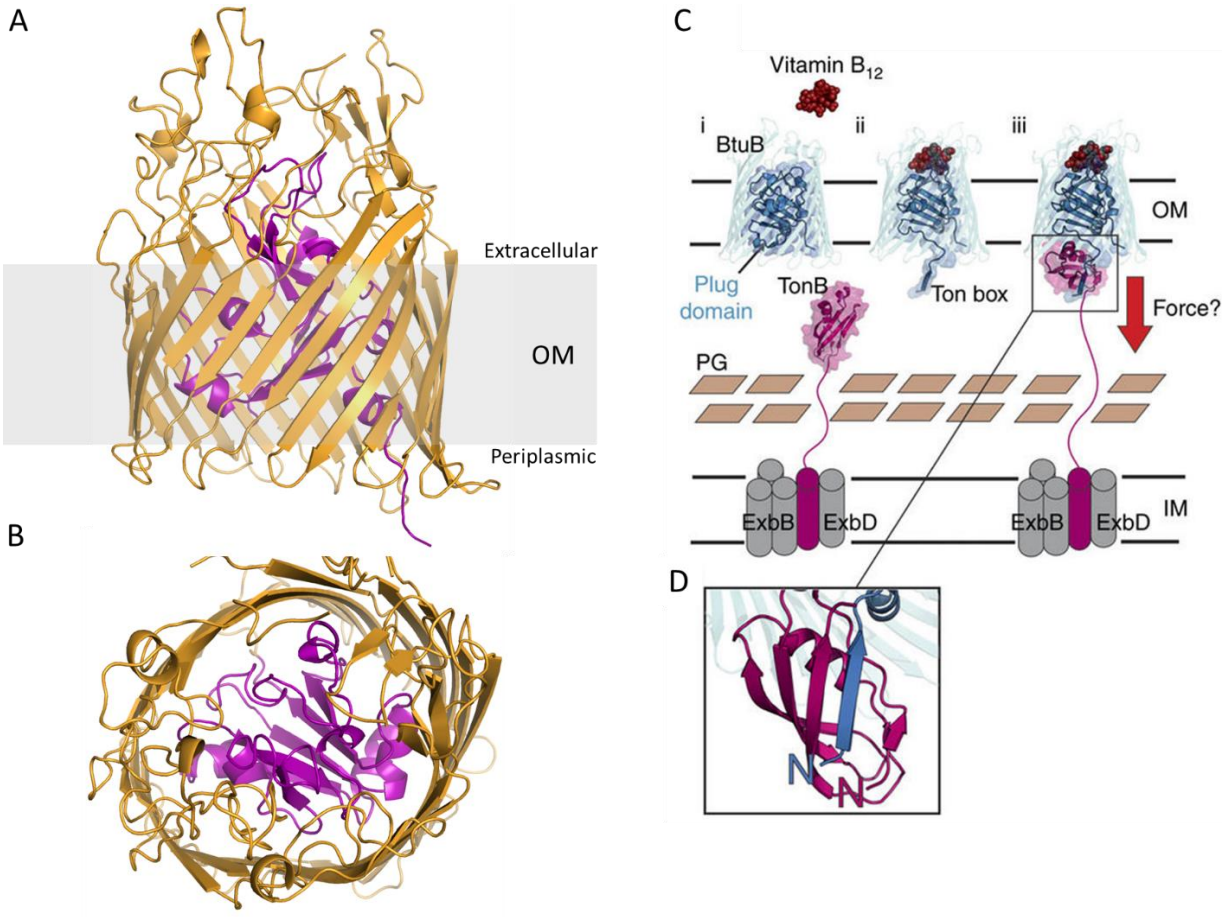


Figure 1.8 Structure and Mechanism of TonB Dependent Transporters
A: Cartoon representation of the *E. coli* TBDT FepA (PDB ID 1FEP) shown from the side within the OM. The N-terminal plug domain is coloured purple. **B:** FepA transporter viewed from the outside of the cell showing occlusion of the channel by the plug domain (purple). **C:** Schematic representation of TonB-dependent vitamin B_{12} transport by BtuB in *E. coli*. Peptidoglycan is labelled PG. (i) The channel of BtuB is occluded by the N-terminal plug domain. (ii) The binding of vitamin B_{12} induces a conformational change of the plug domain, releasing the Ton box into the periplasmic space where is forms a complex with the C-terminal domain of TonB. The N-terminal transmembrane helix of TonB (purple cylinder) forms a complex with ExbB and ExbD (grey cylinders) in the inner membrane (OM). (ii) Linkage of the OM and IM via the TonB – TonB box complex is thought to induce partial or complete unfolding of the plug domain to allow transport. **D:** Close-up of the interactions between TonB (purple) and the TonB box of BtuB (blue). Figure labelled **C** and **D** from Hickman *et al.* 2017.

1.3 Polysaccharide Utilisation by Bacteroidetes

Bacteroidetes are Gram-negative, rod-shaped, anaerobic bacteria and one of the two dominant phyla of the HGM (Rajilic-Stojanovic *et al.*, 2009). Bacteroidetes are of particular interest in regards to human health due to their capacity to use a wide range of glycans as a carbon source. Indeed a recent study showed that only two species of *Bacteroides* could between them use all major dietary and host derived polysaccharides (Martens *et al.*, 2011).

1.3.1 Polysaccharide Utilisation Loci

Bacteroidetes express multi-protein systems to efficiently use many different complex carbohydrates. The genes encoding these proteins are organised in co-regulated Polysaccharide Utilisation Loci (PULs) (Bolam and Koropatkin, 2012). PULs have been identified in every human gut Bacteroidetes sequenced to date as well as many environmental species (Grondin *et al.*, 2017b). PULs have been found targeting every known naturally occurring glycan except cellulose (Sonnenburg *et al.*, 2005; Xu *et al.*, 2007; Martens *et al.*, 2011).

Early characterisation of the first PUL components by the Salyers lab (Salyers *et al.*, 1977; Anderson and Salyers, 1989b) led to the identification of a locus in *Bacteroides thetaiotaomicron* (*B. theta*) targeting starch glycans which was subsequently named the Starch Utilisation System (Sus, Figure 1.9) (Shipman *et al.*, 1999). The PUL encodes eight cell-envelope associated proteins SusRABCDEFG for the binding, degradation and import of starch molecules (Tancula *et al.*, 1992; D'Elia and Salyers, 1996). Target polysaccharides are bound by surface located lipoproteins SusD, SusE and SusF with initial degradation carried out by SusG, a glycoside hydrolase family 13 (GH13) α-amylase lipoprotein (Anderson and Salyers, 1989a; Shipman *et al.*, 1999). The resulting oligosaccharides are then imported to the periplasm via SusC, a TBDT (Reeves *et al.*, 1996). The malto-oligosaccharides are then further digested in the periplasm by a debranching neopullanase GH13 enzyme SusA and a GH97 α-glucosidase enzyme SusB (Anderson and Salyers, 1989a).The periplasmic maltose is sensed by the inner membrane spanning SusR, upregulating the Sus proteins

and the simple sugars are imported into the cytoplasmic space through an undefined transporter (D'Elia and Salyers, 1996). The Sus system is essential for growth on amylose, amylpectin and pullulan (Tancula *et al.*, 1992).

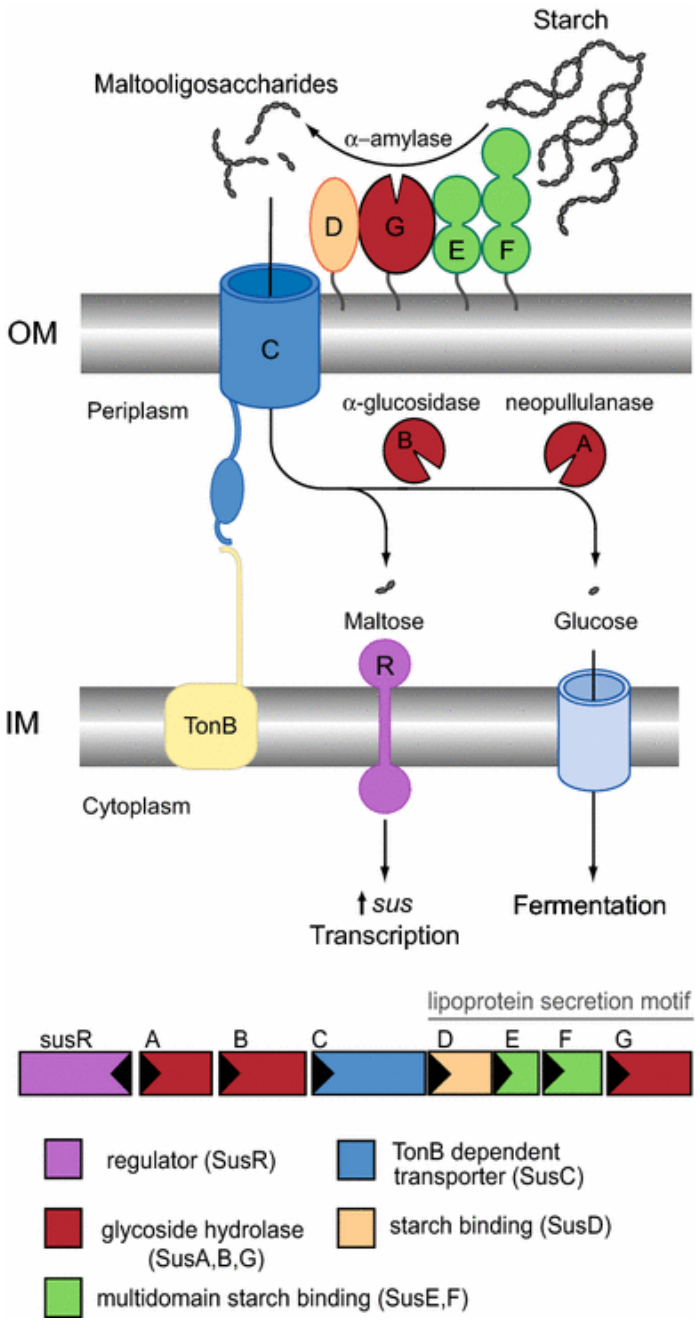


Figure 1.10 Overview of the Starch utilisation system (Sus) from *B. theta*
Starch is bound to the cell surface by outer membrane lipoproteins SusDEF and digested into oligosaccharides by an outer membrane bound glycoside hydrolase (GH) SusG. The malto-oligosaccharides are transported into the periplasm via SusC a TBDT. Two further GHs in the periplasm break down the oligosaccharide into maltose and glucose. The monosaccharide is transported into the cytoplasm via an unknown transporter. Transcription of the SusR inner membrane spanning receptor occurs independently of the rest of the locus. When SusR senses the disaccharide maltose, this induces the PUL and upregulates expression of the other proteins. Figure from Foley *et al*. 2016.

PULs have been characterised for a diverse range of substrates including xyloglucan (Larsbrink *et al.*, 2014), arabinoxylan (Rogowski *et al.*, 2015), α-mannan (Cuskin *et al.*, 2015), inulin (Sonnenburg *et al.*, 2010) and rhamnogalacturonan-II (Ndeh *et al.*, 2017). The number of binding proteins and enzymes involved generally increases with the complexity of the glycan target. Several PULs targeting highly decorated glycans include two SusCD pairs. All PULs encode the following; a SusC-like TBDT alongside a SusD-like binding protein, at least one endo-acting surface-located enzyme and a membrane spanning sensor/regulator (Bolam and Koropatkin, 2012). All identified PULs are expected to function as Sus-like systems (Foley *et al.*, 2016).

1.3.2 Levan Utilisation by *Bacteroides thetaiotaomicron*

The *B. theta* genome includes 88 predicted PULs of which 20 have been characterised. One such well-understood PUL targets levan polysaccharides. Levan is comprised of β,2-6 linked fructose units, often with some β2-1 branches, and is produced extracellularly from sucrose by many bacterial species including *Bacillus* (Kekez et al., 2015)*, Erwinia* (Keith et al., 1991)*, Pseudomonas* (Jathore et al., 2012)*,* and *Zymomonas* (Silbir et al., 2014). Levans are also produced by some plants, including some that are components of the human diet such as wheat and barley (Van den Ende, 2013).

As with the Sus system, the levan PUL of *B. theta* encodes all the proteins required for binding, degrading and importing the target glycan (Figure 1.10). The PUL includes three enzymes, BT1759, BT1760 and BT1765, from glycoside hydrolase family 32 (GH32) which are specific to fructans (Cantarel *et al.*, 2009). The cell-surface exposed BT1760 is essential for growth of *B. theta* on levan polysaccharides (Sonnenburg *et al.*, 2010). A fourth GH32 BT3082, found in isolation in the *B. theta* genome, is also upregulated in response to levan (Sonnenburg *et al.*, 2010).

A schematic model of levan utilisation is shown in Figure 1.10. Levan polysaccharides are recognised by an extracellular surface glycan binding protein (SGBP) BT1761 and degraded by the endo-acting GH32 BT1760 (Sonnenburg *et al.*, 2010; Mardo *et al.*, 2017). The resulting oligosaccharides are bound by the SusD-like BT1762 and imported by the SusC-like BT1763. Further breakdown of the levan oligos occurs in the periplasm mediated by $BT1759^{GH32}$ and $BT3082^{GH32}$. The free monomeric fructose is recognised by the periplasmic domain of the hybrid two component system (HTCS) sensor BT1754 resulting in upregulation of the PUL (Sonnenburg *et al.*, 2010). Fructose is then likely imported by the inner membrane MFS transporter (BT1758) into the cytoplasm and passed along the glycolytic pathway after phosphorylation by the fructokinase BT1757 (Bolam and Sonnenburg, 2011).

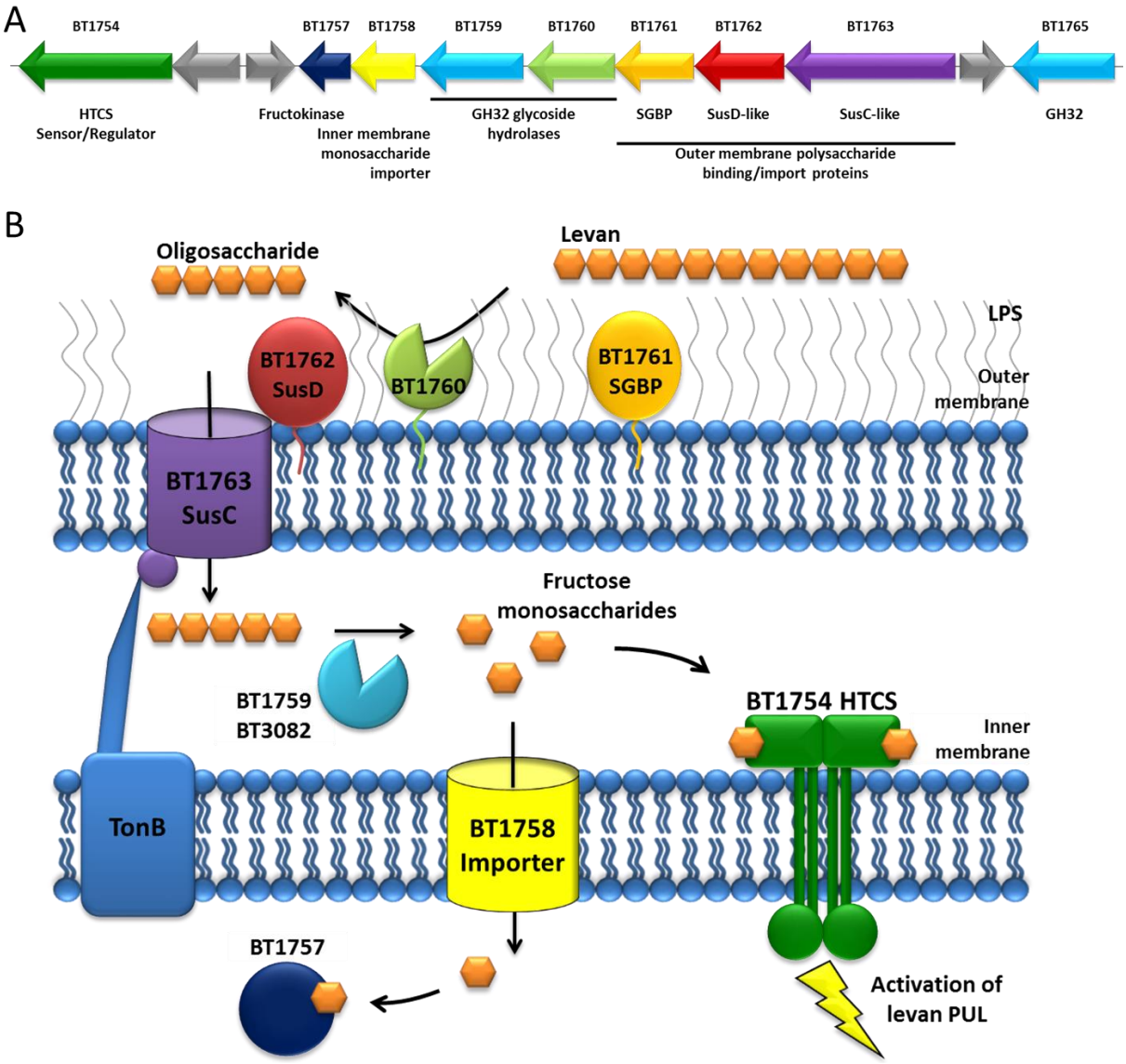


Figure 1.10 Model for levan utilisation by *B. theta* A: Organisation of genes from the levan PUL; arrows shown to scale with the direction of the arrow indicating gene orientation. Proteins of unknown function are shown in grey **B:** Model of levan utilisation. Proteins coloured as with genes in **A**. Levan is bound by the SGBP BT1761 and partially digested by BT1760, an extracellular endo-acting GH32. The oligosaccharides are then bound and imported by the SusCD pair BT1762-3. In the periplasm the oligosaccharides are further degraded by the GH32 exo-acting BT1759 and BT3082. Free fructose binds the periplasmic domain of BT1764 HTCS sensor which activates the protein and results in upregulation of the PUL. Fructose monosaccharides are imported into the peripasm via the BT1758 importer and BT1757 fructokinase phosphorylates the sugar sending it along the glycolytic pathway.

1.3.3 SusC Homologues

Since the majority of enzymes encoded by PULs are expressed in the periplasm, the SusC-like transporters have adapted to import a very wide range of glycans with varying sizes, sugars, linkages, charges and side chains (Porter and Martens, 2017). The SusC homologues are predicted to be 22-stranded β-barrel TonB dependent transporters (TBDTs) (Shipman *et al.*, 2000). SusC homologue domain predictions from Pfam (Finn *et al.*, 2014), SMART (Letunic *et al.*, 2012) and InterPro (Mulder and Apweiler, 2008) databases show the conserved N-terminal TonB box and plug domain with the remaining part of the polypeptide forming the barrel of the transporter (Figure 1.12). Fully characterised TBDTs generally transport small conserved substrates such as iron-siderophores (Section 1.2.4), therefore the mechanisms which allow the SusC-like proteins to function are completely unknown. Many SusC homologues also show a domain of unknown function (DUF4480) at the N-terminus before the TonB box, which is not found in any other TBDTs (Figure 1.11).

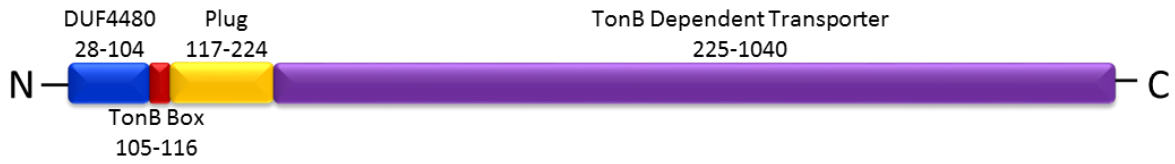


Figure 1.11 Protein domain predictions for BT1763 SusC homologue
BT1763^{SusC} domains predicted by Pfam, SMART and InterPro databases. Numbers represent the residues of the full length protein sequence.

1.3.4 SusD Homologues

The prototypical Sus system requires SusD to bind starch to the cell surface which was shown using ^{14}C-starch binding assays with whole bacteria under aerobic conditions which prevent transport of the substrate into the cells (Shipman *et al.*, 2000). More recent studies have used isothermal titration calorimetry (ITC) to measure binding of specific substrates to SusD homologues (Koropatkin *et al.*, 2008; Sonnenburg *et al.*, 2010).

The structure of SusD was solved both with and without substrate in 2008 (Koropatkin *et al.*, 2008). The protein was expressed in *E. coli* without the lipid anchor and the structure does not include the first 16 residues which likely form a flexible link between the globular protein and the lipid anchor. The structure contains 22 α-helices, 8 of which form 4 helix-turn-helix tetratricopeptide repeats (TPRs) packing together to produce a right-handed superhelix (Allan and Ratajczak, 2011). This motif is structurally conserved between all SusD homologue structures in the PDB and may act as a scaffold for the overall protein structure which is more variable (Bolam and Koropatkin, 2012). TPRs have also been implicated in facilitating protein-protein interactions (Allan and Ratajczak, 2011). There is evidence that the SusC and SusD form a complex in the outer membrane, based on formaldehyde crosslinking experiments and proteolytic tests showing the proteins are more sensitive to degradation in strains lacking the partner SusC or SusD (Shipman *et al.*, 2000; Cho and Salyers, 2001). The interaction between SusC and SusD has been predicted to be via the region surrounding TPR1 (Bakolitsa *et al.*, 2010). It has also been suggested that SusD could be required to mediate interactions between the cell surface Sus proteins in addition to binding large starch molecules (Cameron *et al.*, 2014). Dynamic interactions could be facilitated by TPR 1/2 of SusD which could position SusD for oligosaccharide binding from other Sus lipoproteins and allow transfer of the substrates to the SusC (Figure 1.12) (Bolam and Koropatkin, 2012). Recent live-cell super-resolution imaging has shown the polymeric starch substrate initiates the co-localisation of the OM Sus proteins (Karunatilaka *et al.*, 2014).

Starch recognition by the canonical SusD is shape-dependent with preference for the curved α 1,4-linked glucose over its linear counterpart (Koropatkin *et al.*, 2008). Comparison of the apo- and ligand-bound structures of SusD show a conformational change is required to produce an arched hydrophobic surface of tryptophans and tyrosines complementary to the helical oligosaccharides. This was consistent for maltotriose, maltoheptose and β-cyclodextran which were all shown to be bound by SusD (Koropatkin *et al.*, 2008). However the only other SusD homologue structure solved with and without ligand, BT1043 which targets host mucin O-glycans, shows the binding site in the same location but does not have any conformational change upon substrate binding (Koropatkin *et al.*, 2009).

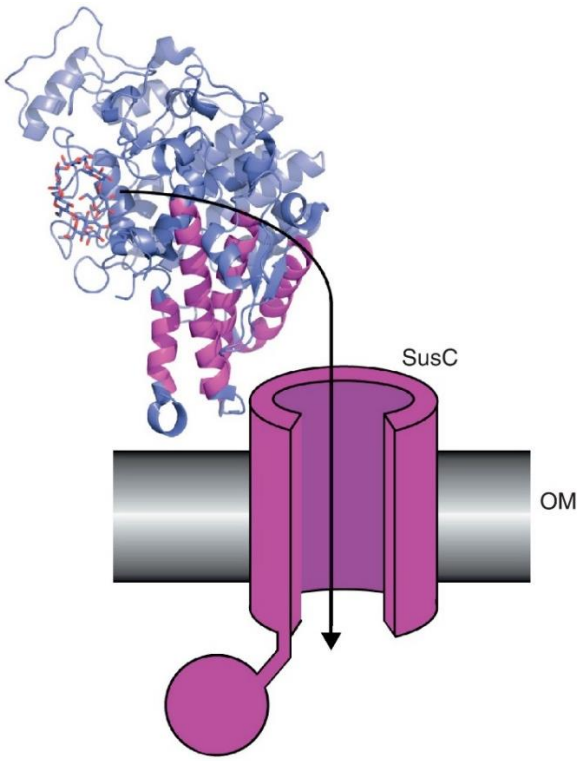


Figure 1.12 Model for SusCD interactions in the outer membrane
SusD is coloured blue with 4 α-helices which form TPR1 and TPR2 coloured purple. Maltoheptaose bound to SusD is shown as sticks. The interaction between SusC and SusD is predicted to occur in the area including TPR1 and TPR2. SusD is thought to 'deliver' oligosaccharides to the SusC for transport across the membrane. Figure from Bolam and Koropatkin, 2012.

SusC homologues and often the SusD-like partner protein are essential for nutrient acquisition by Bacteroidetes. However the import system has not been characterised and the coupling of substrate binding to import is not understood.

1.4 Research Objectives

The main aim of this PhD project was to solve the structure of a SusC homologue, ideally in complex with its cognate SusD. Preliminary work using recombinant expression in *E. coli* was unsuccessful therefore a protocol for endogenous expression and purification of *Bacteroides* outer membrane proteins (OMPs) needed to be developed. The primary focus was to identify highly expressed OMPs and purify proteins of interest in sufficient quantities, directly from *B. theta,* for structural studies (Chapter 3). A further objective was to purify and obtain a structure for a specific SusC homologue, BT1763 from the levan PUL, to provide insight into glycan acquisition by *B. theta* (Chapter 4).

Chapter 2 - Materials and Methods

2.1. Bacterial Strains and Mutants

All bacterial strains are listed in Table 2.1 and mutants produced during this study are in bold.

Name	Genotype Features	Description
BL21 (DE3)	F⁻ *dcm ompT hsdS*(r_B- m_B-) *gal* (DE3)	*E. coli* strain optimised for protein expression using a T7 promoter. Routinely used to over-express recombinant proteins (Studier & Moffat, 1986)
TUNER (DE3)	F⁻ *ompT hsdSB* (r_B^- m_B^-) *gal dcm lacY1* (DE3)	As above for BL21 with added lac permease mutation to allow uniform permeation of IPTG across cells. Linear correlation between IPTG concentration and protein expression level (Novagen)
Top10	F⁻ *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K *rps*L (STR^R) *end*A1 *nup*G λ-	*E. coli* routinely used for cloning and plasmid propagation (Invitrogen)
CC118 λ-pir	Δ(*ara-leu*) *ara*D Δ*lac*X74 *gal*E *gal*K *pho*A20 *thi*-1 *rps*E *rpo*B *arg*E (Am) *rec*A1 λ*pir*	*E. coli* routinely used for cloning and plasmid propagation of pExchange-*tdk (*Herrero *et al.* 1990)
S17.1 λ-pir	*rec*A *pro hsd*R RP4-2 (Tc::Mu;Km::Tn7) (λ pir)	*E. coli* strain used for conjugation of pExchange-*tdk* into *B. theta* (Skorupski and Taylor 1996)
Bacteroides thetaiotaomicron (*B. theta*)	Wild-type	Provided by Martens Lab, University of Michigan
Bacteroides thetaiotaomicron Δtdk	*Δtdk*	Used to produce genome mutations or deletions in *B. theta* through FUdR selection. Provided by Martens Lab, University of Michigan
ΔBT2264	*ΔBT2264*^{SusC}	*B. theta* with knockout of BT2264 SusC homologue
FLAG-BT2263	FLAG-tagged *BT2263*^{SusD}	*B. theta* with FLAG-tag added to the CT of BT2263 SusD homologue
CT-HisBT1762	His-tagged *BT1762*^{SusD}	*B. theta* with His_{x6}-tag added to the CT of BT1762 SusD homologue
ΔBT1762^{SusD}	*ΔBT1762*^{SusD}	*B. theta* with knockout of BT1762 SusD homologue Provided by Justin Sonnenburg, Stanford University (Cell 2010)
W103A-BT1762	W103A of *BT1762*^{SusD}	*B. theta* with single base mutation of $BT1762^{SusD}$ tryptophan 103 to alanine

ΔHinge-BT1763	ΔHinge of *BT1763*^{SusC}	*B. theta* deletion of hinge loop L7 of BT1763^{SusC} replaced with one glycine residue
ΔDUF-BT1763	ΔDUF4480 of *BT1763*^{SusC}	*B. theta* with deletion of DUF4480 of BT1763^{SusC}
CT-HisBT1762/ ΔDUF-BT1763	His-tagged *BT1762*^{SusD} ΔDUF4480 of *BT1763*^{SusC}	*B. theta* with His_{x6}-tagged CT of BT1762^{SusD} and deletion of DUF4480 of BT1763^{SusC}

Table 2.1 Bacterial strains and mutants
Mutants in bold were produced during this study. C-terminal (CT) and N-terminal (NT) have been abbreviated.

2.2. Vectors

Vectors used in this study are listed in Table 2.2.

Plasmid	Size (kbp)	Supplier	Features
pET28a	5.4	Novagen	Kan^r, T7 promoter, *lac* operator, *laciq*, multiple cloning site with integrated NT and CT His_{x6}-tag
pRSETA	2.9	Invitrogen	Amp^r, NT His_{x6}-tag
pET32b	5.9	Novagen	Amp^r, NT His-tag, Trx-tag, rEK cleavage site
pET9	4.3	Susan Buchanan	Kan^r, PelB signal sequence, TEV protease-cleavable His_{x10}-tag
pB22	6.0	Van den Berg Lab plasmid (Guzman *et al.*, 1995)	Amp^r, YtfM signal sequence, TEV protease-cleavable His_{x7}-tag
pExchange-*tdk*	4.2	Nicole Koropatkin (Koropatkin *et al.* 2008)	Amp^r

Table 2.2 Vectors

2.3. Bacterial Growth Conditions

2.3.1. Growth Media Composition

All media compositions are outlined below (Table 2.3). Rich media were used when high cell density was required such as for recombinant protein expression. Minimal media for *B. theta* was used with 0.5 % (w/v) carbon source required to upregulate the PUL or protein of interest.

Medium	Components (per litre)	Description/Method
Luria-Bertani (LB)	25 g LB granules as supplied (Sigma-Aldrich)	Rich media for *E. coli* Dissolved in H_2O and pH adjusted to 7.2 with NaOH. Autoclaved before use.
LeMasters-Richards (LR)	24 g KH2PO4 1 g NaOH 4 g glucose 10 ml LR salts (1.67.5 g/l $(NH_4)_2SO_4$, 30 g/l $MgSO_4$.H2O, 300 mg/l $FeSO_4.7H_2O$, 1 ml/l conc H_2SO_4)	Minimal media for producing Seleno-methionine substituted protein by inhibition of the methionine biosynthesis pathway in *E.coli* Salts and media autoclaved separately. Glucose and salts added immediately before inoculation.
Tryptone Yeast Extract Glucose (TYG)	10 g Tryptone 5 g Yeast Extract 2 g Glucose 0.5 g Cysteine, free base 100 ml 1 M KPO4 pH 7.2 40 ml TYG Salt Solution (MgSO4 0.5 g/l, NaHCO3 10 g/l, NaCl 2 g/l) 4 ml 0.25 mg/ml Resazurin 1 ml 0.4 mg/ml FeSO4 1 ml 1 mg/ml Vitamin K 1 ml 0.8 % CaCl2	Rich media for *B. theta* Components dissolved in H_2O before autoclaving. His-heme* was added immediately before inoculation (1 µl per 1 ml of media)

Minimal Media (MM)	1 g NH_4SO_4 1 g Na_2CO_3 0.5 g Cysteine, free base 100 ml 1 M KPO4 pH 7.2 50 ml MM Salt Solution (NaCl 18 g/l, CaCl2 0.53 g/l, MgCl2 0.4 g/l, MnCl2 0.2 g/l, CoCl2 0.2 g/l) 10 ml 0.4 mg/ml FeSO4 4 ml 0.25 mg/ml Resazurin 1 ml 1 mg/ml Vitamin K 0.5 ml 0.01 mg/ml Vitamin B12	Minimal media for *B. theta* Prepared as with TYG. His-heme* was added immediately before inoculation (1 µl per 1 ml of media) Required carbon source added immediately before inoculation to final concentration of 0.5 % (w/v)
Brain Heart Infusion (BHI)	37.5 g BHI powder, as supplied (Sigma-Aldrich)	Rich media for *B. theta* His-heme* was added immediately before inoculation (1 µl per 1 ml of media)
***His-heme**	1.2 g Hematin 1 l. 0.42 g/l Histidine HCL pH 8	Added to all *B. theta* media (1 µl per 1 ml) after sterilisation Hematin dissolved in Histidine HCL solution by shaking at 37 °C for several hours

Table 2.3 Composition of bacterial growth media

2.3.2. *E. coli* growth Conditions

E. coli strains (Table 2.1) were grown at 37 °C in the required sterile media (Table 2.3) with the appropriate antibiotic. A rotary shaker at 150-180 rpm was used to aerate the cells during the growth. For growth on a solid medium plates were produced by adding 2 % (w/v) Bacteriological agar (No. 1) (Oxoid) to the required media, autoclaving and allowing to cool. The required antibiotic was added immediately before pouring approximately 20 ml per plate into plastic Petri-dishes (Thermo Fisher Scientific). Plates were incubated with lid side down.

2.3.3. *Bacteroides* Growth Conditions

B. theta was grown in either Tryptone Yeast Extract Glucose (TYG) rich media or Minimal media (MM) at 37 °C under anaerobic conditions in a cabinet (Don Whitely Scientific). Brain heart infusion (BHI) with His-heme and 2 % (w/v) agar was used to produce plates for *B. theta* growth. Plates were incubated with lid side down in the anaerobic cabinet.

2.3.4. Selective Media

Antibiotics were used within growth media to select for bacteria containing the desired plasmid. Stock solutions were produced and diluted 1:1000 into media (Table 2.4).

Antibiotic	Stock Concentration (mg/ml)	Working Concentration (µg/ml)	Storage
Kanamycin	100	100	-20°C for 1 month
Ampicillin	50	50	-20°C for 1 month
Gentamycin	0.2	0.2	Freshly prepared
Erythromycin	0.025	0.025	Freshly prepared

Table 2.4 Antibiotic stocks

2.4. Basic Lab Methods

2.4.1. Storage of DNA and Bacteria

All DNA was stored in elution buffer (EB – 10 mM Tris-HCl buffer pH8.5) at -20 °C. Bacterial colonies (*E. coli* or *B. theta*) on agar plates were stored at 4 °C for up to two weeks. For long term storage, bacteria in liquid media was mixed 1:1 with 50 % glycerol and stored at -80 °C.

2.4.2. Sterilisation

Solutions, media and glassware were sterilised by autoclaving at 121 °C under 32 lb/inch^2 of pressure for 20 minutes using either an Astell Hearson 2000 Series Autoclave or a Prestige® Medical Series 2100 Clinical Autoclave. Small volumes of buffer or substrate were filter sterilised through 0.22 µm sterile filter discs (Stupor Acrodisc 3.2 Gelman Sciences) using a suitable sterile syringe.

2.4.3. Centrifugation

Large-scale cultures of bacteria were harvested by centrifugation in 500 ml Nalgene bottles using a Beckman J2-21 centrifuge with JA-10 rotor. The maximum speed used was 8000 rpm (11305 g). Small-scale cultures (1-10ml) were centrifuged at 5000 rpm (2516 x g) at 4 °C in a Hettich Mikro 220R Refrigerated bench top centrifuge with fixed angle rotor. Eppendorf tubes (1.5 ml or 2 ml volume) were centrifuged using a Heraeus Pico 21 benchtop microcentrifuge at 14,000 rpm (21,100 x g).

Ultracentrifugation was carried out at up to 42,000 rpm (204526 x g) using a Beckman L8 – 80 Ultracentrifuge and a 45 Ti rotor with 70 ml polycarbonate bottles (Beckman Coulter).

2.4.4. Plating Bacteria

Approximately 100 µl of bacterial suspension was transferred to an agar plate near a Bunsen burner flame. A metal spreader was immersed in 100 % ethanol and passed through the hottest (blue) part of the Bunsen flame. The spreader was allowed to cool for 20-30 seconds before spreading the bacterial suspension across the surface of the plate.

2.4.5. Transformation of Chemically Competent *E. coli*

Chemically competent *E. coli* were produced with calcium chloride by Mr Carl Morland and stored at -80 °C (Cohen *et al.*, 1972). A 100 µl aliquot of *E. coli* was thawed on ice before 1-5 µl of plasmid or ligation reaction was added. The mixture was incubated on ice for 30 minutes after which the cells were heat shocked at 42 °C for 1 minute before cooling on ice for a further 2 minutes. Cells were plated on LB agar containing an appropriate antibiotic to select for successful transformants. The plates were incubated lid side down at 37 °C for 16 hours.

Ligated plasmids and products of site-directed mutagenesis require a recovery incubation after heat shocking. 500 µl of sterile LB was added and the cells were incubated at 37°C shaking for 1 hour. The *E. coli* were harvested by centrifugation, 500 µl of media removed and the cells resuspended in the remaining media before plating.

2.5. DNA Methods

2.5.1. Rapid, small-scale purification of plasmid DNA from *E. coli*

The required plasmid was used to transform Top10 competent *E. coli* cells as outlined above. A single colony was picked to inoculate 5 ml LB with an appropriate antibiotic and incubated at 37 °C for 16 hours. The cultures were harvested by centrifugation at 2516 x g for 10 minutes. Plasmid purification was then carried out using a QIAspin Prep kit (QIAGEN) following the manufacturer's protocol.

2.5.2. *B. theta* genomic DNA extraction

DNA was extracted from 5 ml cultures using GenElute Bacterial Genomic DNA kit (Sigma-Aldrich) as per the manufacturer's instructions.

2.5.3. Quantification of DNA

DNA concentration was determined using the NanoDrop 2000 benchtop spectrophotometer (Thermo Fisher Scientific). The machine was first blanked using 2 µl EB buffer or Sigma-Aldrich ultrapure water depending on the DNA solution. Absorbance for 2 µl of the DNA sample was measured at 260 nm.

2.5.4. Polymerase Chain Reaction

The polymerase chain reaction (PCR) was used to amplify target DNA (Mullis *et al.*, 1986). A thermostable DNA polymerase KOD (Novagen) was used to catalyse the synthesis of new complementary DNA strands based on a template (plasmid, genomic DNA or cell culture).

Specific regions of DNA were targeted using primers 15-30 bases in length with a melting temperature (T_m) of >45°C (calculated using OligoCalc Table 2.16). Primers were designed to end in G or C bases, when possible, to aid annealing and thus increase efficiency of the reaction. If digestion of the DNA was required the specific target sequence for the endonuclease was included at the 5'-end of the primer. For sequence and ligation independent cloning (SLIC) each primer was extended

by 15-18 bp homologous to the ends produced by digestion of the vector to allow for recombination. Primers were synthesised by MWG-Eurofins and received lyophilised. Primers were resuspended in PCR grade water to a working concentration of 100 pmol/µl.

2.5.4.1. Standard PCR

PCR reaction mixtures were prepared on ice as outlined in Table 2.5 using the KOD DNA polymerase kit (Novagen).

Reagent	Volume per reaction	Amount per reaction
KOD Buffer (10 x)	5 µl	1 x
dNTPs (2 mM each)	5 µl	0.2 mM each
$MgSO_4$ (25 mM)	5 µl	2.5 mM
Forward Primer (100 µM)	5 µl	10 µM
Reverse Primer (100 µM)	5 µl	10 µM
Template DNA	1 µl	50-200 ng
KOD DNA polymerase (1 U/µl)	0.5 µl	0.5 U
PCR grade H2O	Up to total volume 50 µl	-

Table 2.5 PCR reaction mixture

The standard thermocycler program for routine PCR is explained below in Table 2.6.

Temperature (°C)		Time	Function
	95	1 min	Initial denaturing
35 cycles	95	30 s	Denaturing
	55	30 s	Annealing
	68	1 min per kbp	Extension
	68	10 mins	Final extension

Table 2.6 Standard PCR program

After each PCR a sample was analysed for product by gel electrophoresis. Low yield or no PCR product of the correct size was improved by altering the parameters of the reaction such as annealing temperature, extension time or $MgSO_4$ concentration (1-5 µM). Dimethyl sulfoxide (DMSO) was also added to a final concentration of 5 to 10 % for some reactions.

2.5.4.2. Site-Directed Mutagenesis

Single amino acid mutants were produced using site-directed mutagenesis (SDM). The double-stranded recombinant plasmid encoding the wild-type (WT) protein was mutated using two synthetic primers including the desired mutation and 12 bp in each direction complementary to the DNA template. The PCR reaction was prepared as above for routine PCR (Table 2.5) and the thermocycler program in Table 2.7 was used.

Temperature (°C)		Time	Function
	95	1 min	Initial denaturing
20 cycles	95	30 s	Denaturing
	55	30 s	Annealing
	68	1 min per kbp of plasmid	Extension
	68	10 mins	Final extension

Table 2.7 Standard PCR program for SDM

The PCR reaction was purified as above and eluted in 30 µl EB buffer before 1 µl of *Dpn*I (Thermo Fisher Scientific 10 U/µl) and 3 µl Tango Buffer were added. DpnI digests the methylated template DNA, leaving the unmethylated PCR product intact. The digestion reaction was incubated at 37 °C for 2 hours before 5 µl was used to directly transform *E. coli* cells.

2.5.4.3. Sewing PCR

PCR overlap extensions or 'sewing' PCR was used to produce plasmids for homologous recombination to mutate the genome of *B. theta*. This method was used to remove segments of DNA (knockouts), mutate certain residues or to add extra amino acids such as a His-tag. A pExchange-*tdk* plasmid was required encoding the required mutation and a region of DNA homologous to the *B. theta* genome at least 1000 bp upstream and downstream of the mutation site (Figure 2.1A). These adjoining regions are referred to as flank 1 (upstream) and flank 2 (downstream).

Four primers were required to construct the mutated DNA fragment, two outside amplification primers and two internal mutation 'sewing' primers (Figure 2.1A). The outside amplification primers (1 and 4) were designed as for standard PCR with the required restriction sites. The internal primers (2 and 3) were homologous to the genome and the required mutation. For producing knockouts, the mutation extension of the primer was complementary to 20 bp of the opposite flank.

The first step was creation of the two flanks with the mutation extension (Figure 2.1A). Two separate reactions were set up with primers 1 and 2 (Flank 1) alongside primer 3 and 4 (Flank 2) using routine PCR conditions. The second step links the flanks together at the complementary mutation site (Figure 2.1B). To obtain the 2000 bp final product the flanks were used in equal concentrations to avoid preferential amplification of one flank above the other. The first 10 PCR cycles were without primers to allow the flanks to act as template and primer to each other although this was not always required. The outside primers (1 and 4) were then added to amplify the sewn DNA fragment (Figure 2.1C). The final hybrid PCR product was purified and inserted into pExchange-*tdk* vector using a ligation reaction (2.59) or Sequence and Ligation Independent Cloning (2.5.10).

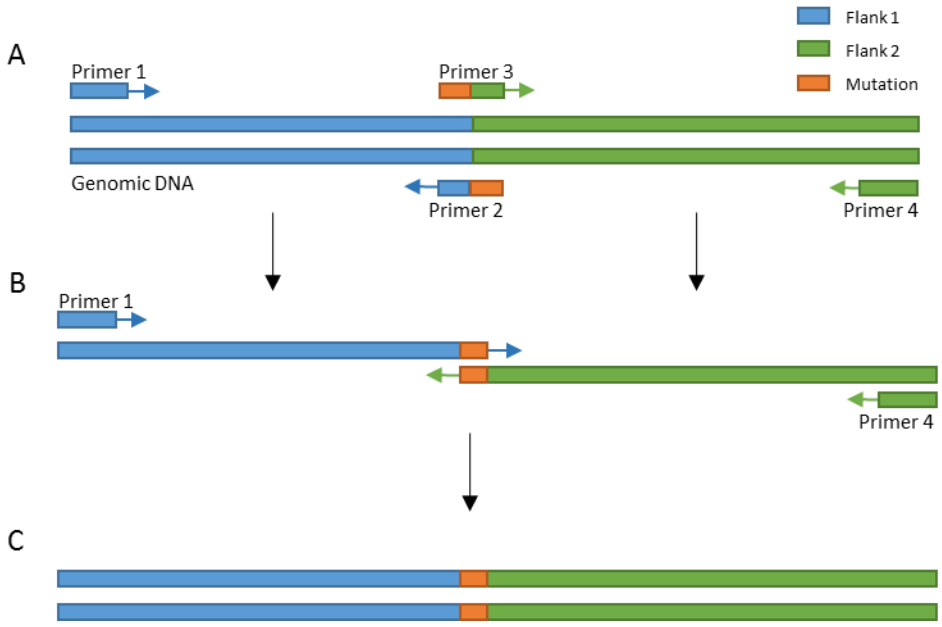


Figure 2.1 Sewing PCR used to produce mutated gene fragments
A: Two 1000 bp flanks upstream and downstream of the mutation site were amplified during individual PCR reactions. The outside half of primer 2 and 3 were homologous to the desired mutation. **B:** The resulting flanks become 'sewn' together by the complementary mutation sequence and are amplified by the outside primers (1 and 4) during a further PCR reaction. **C:** The final product consists of the central mutation and 1000 bp flanking regions.

2.5.4.4. Agarose Gel Electrophoresis

Size of linear DNA samples was determined using gel electrophoresis. A 1 % (w/v) agarose solution was prepared by mixing 500 mg of agarose (Sigma-Aldrich) in 50 ml 1 x TBE buffer (89 mM Tris Base, 89 mM Boric Acid, 2 mM EDTA). The agarose was dissolved completely by boiling, then allowed to cool (<50 °C) before 0.5 µg/ml ethidium bromide was added while the solution was still molten. The gel was poured into a mini gel system mould (Applied Biosystems) with a comb to produce wells for DNA loading. The gel was allowed to set at room temperature for a minimum of 30 minutes before being submerged in 1 x TBE buffer.

DNA samples were mixed with the appropriate volume of 10 x loading dye (0.25 % Bromophenol blue, 50 % glycerol, 10 x TBE) and loaded into the wells of the gel. DNA standards (Hyperladder I, Bioline) were run with the samples to allow estimation of DNA size by comparison.

Migration rate is inversely proportional to the log10 of the size of the fragment. The gels were run at 70 V for 45 – 60 minutes (LKB Bromma 2197 Power Supply).

2.5.5. Visualisation and Photography of Agarose Gels

Following electrophoresis the gels were visualised using a Bio-Rad Gel Doc EZ Imager (Image Lab/ PC Windows software). Photographs were printed using a linked Mitubishi Video Copy Processor (Model P68B) with Mitsubishi thermal paper.

2.5.6. Restriction Digest of DNA

When available, 1000 ng of DNA was used per digest and the reaction carried out in a sterile 1.5 ml Eppendorf. The required DNA (in EB buffer) was mixed with the appropriate volume of 10 x concentrated reaction buffer. The buffer was selected to optimise conditions for the specific endonuclease enzymes used in the reaction (Thermo Fisher Scientific). 0.5 – 1 µl of enzyme was added and the final volume was made up to 40 µl with Ultra-pure water (Sigma-Aldrich). The reaction mixture was incubated at 37 °C in a water bath for 1 hour then the DNA was purified (outlined below) to remove the enzymes and buffer.

2.5.7. Purification of Vector DNA (Gel extraction)

Following linearization of vector DNA using endonucleases, the sample was run on a 1 % (w/v) high purity Seachem Gold Agarose gel. Electrophoresis was carried out as described above. The band required was excised from the gel with a scalpel over a UV transilluminator. DNA was then extracted from the gel slice using a QIAquick Gel Extraction kit (QIAGEN) as per the maufacturer's protocol.

2.5.8. Purification of inserts and PCR products

Digested DNA inserts for ligation and all PCR products were purified using the QIAquick PCR purification kit (QIAGEN) by following the manufacturer's instructions.

2.5.9. Ligation of Insert and Vector DNA

Vector and insert DNA digested with compatible endonucleases were mixed with an insert to vector molar ratio of 3:1. In each reaction 20 ng of vector DNA was used and the required amount of insert DNA was calculated using Equation 1.

$$3 \times \left[\frac{\text{Size of insert (bp)}}{\text{Size of vector (bp)}} \right] \times \text{amount of vector (ng)}$$

Equation 1 - Calculating molar ratio of insert to vector DNA

Ligation reactions using reagents from the Rapid DNA Ligation kit (Thermo Fisher Scientific) were mixed as outlined in Table 2.8 in 200 µl PCR tubes.

Component	Amount per reaction
Vector DNA (digested)	20 ng total DNA
Insert DNA (digested)	Calculated using Equation 1
5 x Ligation Buffer	4 µl
T4 DNA Ligase	1 µl
Sterile H_2O	Up to total 20 µl

Table 2.8 Ligation reaction mixture

The ligation reaction was incubated at room temperature for 30 minutes before 5 µl was used to directly transform chemically competent *E. coli* cells.

2.5.10. Sequence and Ligation Independent Cloning

An alternative method to using ligase is Sequence and Ligation Independent Cloning (SLIC) which uses T4 DNA polymerase (Jeong *et al.*, 2012). This protocol was adapted from the method outlined in Jeong *et al.* 2012. Primers for amplification of the required gene were designed with 15 bp extensions homologous to the ends produced by digestion of the vector. The insert with the 15 bp extensions was generated via standard PCR. Only the vector was digested by endonuclease enzymes and purified as described above. The insert and vector were mixed together in 1:1, 2:1 or 3:1 ratios with DNA amounts calculated using Equation 1. The SLIC reaction was prepared on ice as shown in Table 2.9.

Component	Amount per reaction
Vector DNA (digested)	100 ng
Insert DNA	Calculated using Equation 1
BSA	1 µl
NEB Buffer 2	1 µl
T4 DNA polymerase	0.2 µl
Sterile H_2O	Up to total 10 µl

Table 2.9 Ligation Independent Cloning reaction mixture

The SLIC mixture was incubated at room temperature for 2 minutes 30 seconds before transferring to ice for 10 minutes. 5 µl of the reaction was used to transform chemically competent *E. coli* by heat shock.

2.5.11. Automated DNA sequencing

DNA sequencing was conducted using the Value Read service provided by MWG-Eurofins. Pre-ordered labels provided by the company were applied to 1.5 ml Eppendorf tubes. For each sample, two tubes containing 50-100 ng of DNA (plasmid or PCR product) were sent for sequencing. Each clone or DNA fragment was sequenced in both the forward and reverse direction.

2.6. Microbiology

2.6.1. Mutagenesis of the *Bacteroides thetaiotaomicron* genome

A modified pExchange-*tdk* plasmid was produced containing the mutation and two flanking 1000 bp regions (2.5.4.3.). The plasmid was transformed into S17 λ *pir E. coli* cells referred to as the 'donor' strain, with *B. theta Δtdk* the 'recipient' strain (Koropatkin *et al.*, 2008). The donor and recipient strains were cultured overnight to similar optical densities in 5 ml LB and TYG media respectively (Figure 2.2A). Cells were harvested by centrifugation and equal sized pellets were resuspended in 1 ml TYG (Figure 2.2B). The cells were plated evenly on BHI-heme agar without antibiotic and incubated aerobically for 24 hours. Due to the large volume of liquid the plates were incubated lid side up. A lawn of *E. coli* grew first creating an anaerobic environment underneath which allowed *B. theta* to grow, providing the necessary conditions for conjugation between donor and recipient (Figure 2.2B).

The conjugation plates were scraped and the cells resuspended in 5 ml TYG (Figure 2.2C). 100 µl of this solution and two 10-fold serial dilutions (1:10 and 1:100) were plated onto BHI-heme agar plates containing gentamycin (200 µg/ml) and erythromycin (25 µg/ml). The antibiotics select for the recipient strain and pExchange-*tdk* therefore colonies represent cells which have undergone single recombination of the plasmid with the genomic DNA via one of the flanks (Figure 2.2G). The plates were incubated anaerobically for up to 3 days, then 10 colonies were restreaked onto fresh BHI-heme plates with gentamycin and erythromycin to minimise contamination by WT *B. theta* (Figure 2.2C). 10 colonies were picked and cultured overnight in TYG (Figure 2.2D).

The cells were pooled and 100 µl plated alongside two 10-fold serial dilutions (1:10 and 1:100) onto BHI-heme agar containing FUdR (200 µg/ml) and incubated for up to 3 days (Figure 2.2E). FUdR is toxic to strains which synthesise thymidine; the recipient *B. theta* Δ*tdk* cannot produce thymidine but the knockout gene is complemented by the pExchange-*tdk* plasmid. Therefore FUdR selects for the second recombination event involving the second flank which

eliminates the pExchange *tdk* sequence from the genome (Figure 2.2G). The FUdR resistant colonies were restreaked onto fresh BHI-heme agar containing FUdR to reduce contamination with WT *B. theta* cells (Figure 2.2E).

10 colonies from the second FUdR plate were cultured overnight in 5 ml TYG to allow extraction of genomic DNA (gDNA) and creation of glycerol stocks (Figure 2.2F). The isolated gDNA was screened for successful mutations using diagnostic PCR with the outside primers used to create the plasmid insert (primer 1 and 4 Figure 2.1). Deletions can be detected using agarose gel electrophoresis to compare the PCR product size produced from WT gDNA with the possible mutants. Genomic insertions such as the addition of tag sequences were detected using a specific primer homologous to the mutation therefore PCR will produce no product from WT gDNA. Strains which appeared to be correct after screening were sequenced to check the correct mutation had taken place and was in frame.

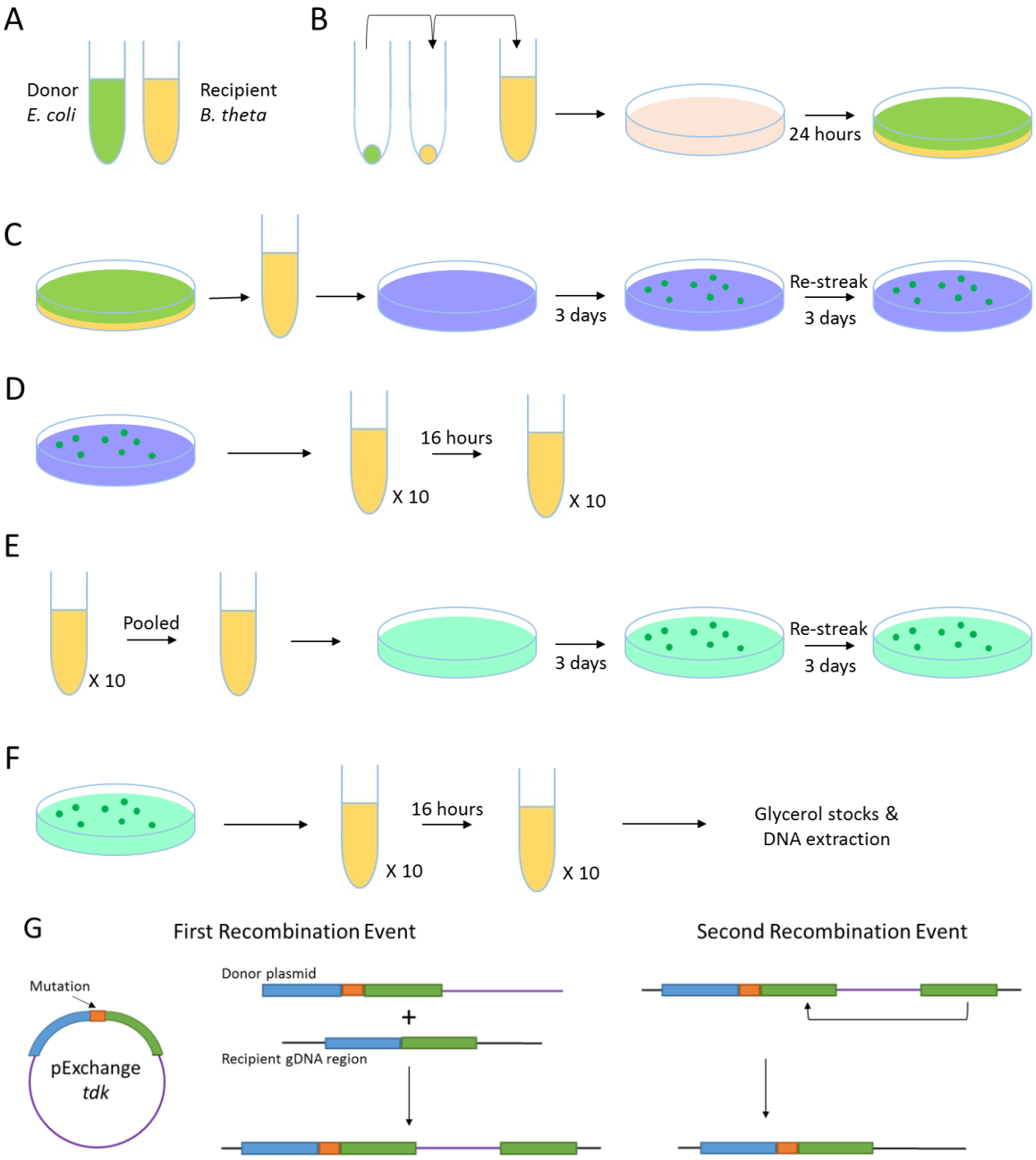


Figure 2.2 Generating mutant strains of *Bacteroides thetaiotaomicron*
A: The donor and recipient strains were cultured overnight in 5 ml LB and TYG media respectively. **B:** Equal size cell pellets were harvested by centrifugation, resuspended and combined in 5 ml TYG. The cells were plated onto BHI-heme agar plates containing no antibiotic (peach) and incubated for 24 hours. **C:** The conjugation plates were scraped and the cells resuspended in 5 ml TYG. This was plated on BHI-heme plates containing gentamycin (200 µg/ml) and erythromycin (25 µg/ml) (lilac). Resistant colonies were re-streaked on fresh plates to minimise WT contamination. **D:** 10 colonies were picked and grown overnight in TYG. These strains have undergone the first recombination event (**G**) **E:** The cultures were pooled and plated onto BHI containing FUdR (200 µg/ml) to select for the second recombination event (cyan). As before, colonies were re-streaked. **F:** 10 colonies were picked and cultured overnight in TYG. Glycerol stocks were prepared and DNA was extracted for analysis. **G:** A schematic diagram of the recombination. pExchange-*tdk* enters the genome by recombination with one flanking region (blue) in the first recombination event. The second recombination event involved the second flank (green) eliminating pExchange-*tdk* from the genome.

2.6.2. CFU growth curves of *B. theta*

As outlined above (2.6.1.) pExchange-*tdk* was used to add a 4xAla linker and FLAG-tag (DYDDDDK) to the C-terminus of $BT2263^{SusD}$ in WT *B. theta*. The resulting mutant strain was grown in TYG rich media and minimal media containing 0.5% fructose anaerobically at 37 °C. 500 µl samples were collected over 24 hours and serial dilutions were plated in duplicate on BHI-heme agar. The dilution plates were incubated at 37 °C under anaerobic conditions for 3 days then the colonies were counted to calculate colony forming units (CFU) for each sample. The 500 µl samples of cells were harvested by centrifugation, resuspended in 100 µl of sterile 1 x PBS (Oxoid) and stored at -20 °C until required for Western blotting (2.1.2.1). The whole cells were diluted to relative CFU using Bugbuster detergent (100 µl per 2 x 10^9 cells). The samples were incubated at room temperature for 20 minutes, centrifuged and the supernatant collected (stored at 4 °C).

2.6.3. Automated monitoring of *B. theta* growth curves

An Epoch microplate spectrometer (Biotek Instruments Ltd.) was used in conjunction with a 96-well Costar culture plate (Sigma-Aldrich) inside an anaerobic chamber at 37 °C (Don Whitely Scientific). 200 µl cultures were used and each condition was prepared in triplicate to be averaged. Media without bacterial inoculum was always used as a control. The plate reader measured and recorded the optical density (at 600 nm) of each well at 15 minute intervals. Data was manipulated in Gen5 2.05 software and later plotted using Prism 6.0 (GraphPad).

2.7. Recombinant Protein Expression and Purification

Soluble proteins were expressed recombinantly in *E. coli* via expression plasmids (Table 2.2).

2.7.1. Protein expression in *E. coli*

Plasmids encoding the required protein were transformed (2.4.5) into the appropriate *E. coli* strain. A single colony was used to inoculate 5 ml LB broth, with the suitable antibiotic, and incubated with shaking at 37 °C for 16 hours. 2 L baffled flasks containing 1 L sterile LB and the required antibiotic were inoculated with 1 ml of the overnight culture. The flasks were incubated shaking at 37 °C until the culture reached an OD_{600nm} of 0.6 – 0.8. The cells were induced using the required reagent for the plasmid and incubated at 16 or 18°C overnight. The specific expression conditions for the proteins used in this study are listed in Table 2.10. Cells were harvested by centrifugation for 10 minutes at 4412 x g using a Beckman J2-21 centrifuge with JA-10 rotor.

Protein	Plasmid	Antibiotic	*E. coli*	Induction	Temp. (°C)
BT2262^{LP}	pET9	Kan	BL21	0.2 mM IPTG	18
BT2263^{SusD}	pET9	Kan	BL21	0.2 mM IPTG	18
BT2263^{SusD}	pET28a	Kan	BL21	1 mM IPTG	16
BT1760^{GH32}	pET32b	Amp	BL21	1 mM IPTG	16
BT1760^{GH32}	pET28a	Kan	BL21	1 mM IPTG	16
BT1761^{SGBP}	pET28a	Kan	BL21	1 mM IPTG	16
BT1761^{SGBP}	pB22	Amp	TUNER	100 mg/l arabinose	16
BT1762^{SusD}	pRSETA	Amp	BL21	1 mM IPTG	16

Table 2.10 Induction and growth conditions for recombinant protein expression
LP is a lipoprotein, SusD is a homologue of SusD, GH32 is a glycoside hydrolase enzyme and SGBP is a surface glycan binding protein

2.7.2. Seleno-methionine incorporation

This method for expressing seleno-methionine (SeMet) substituted protein is based on the inhibition of the methionine biosynthesis pathway protocol for *E. coli* (Van Duyne *et al.*, 1993). Expression plasmids encoding the desired protein were transformed into an appropriate *E. coli* strain. The appropriate antibiotic was added to all growth media to retain the plasmid. The transformed cells were used to inoculate 5 ml LB and the culture was incubated shaking at 37 °C for 6 – 8 hours. 1 ml of the initial culture was used to inoculate 100 ml of LeMasters-Richards (LR) media (Table 2.3) and grown overnight at 37 °C. 1 litre of LR media (in 2 litre baffled flasks) was inoculated with 10 ml of the overnight culture and incubated at 37 °C for approximately 8 hours to OD_{600nm} of 0.6 – 1.0. Amino acids (Lys/Phe/Thr 100 mg/L each and Leu/Ile/Val 50 mg/L each) and SeMet (60 mg/L) were added to the culture and incubated for a further 30 minutes. The cells were then induced and incubated overnight as above (Table 2.10).

2.7.3. Immobilised Metal Affinity Chromatography (IMAC)

All soluble proteins used in this study were expressed with a His-tag comprising of at least six histidine residues. The proteins were purified by immobilised metal affinity chromatography (IMAC) whereby the histidine residues interact with an electropositive transition metal (e.g. nickel or cobalt) immobilised in a column. Imidazole disrupts the interaction by competitively binding the metal ions which forces the His-tagged protein to be eluted from the column.

Harvested cells were resuspended in TALON buffer (20 mM Tris, 300 mM NaCl, pH 8) which could be stored at -20 °C for up to 3 months if required. The cell suspension was transferred to a 50 ml centrifuge tube (Nalgene) and sonicated on ice for 2 minutes using a B. Braun Labsonic U sonicator set at low intensity (~42 watts and 0.5 second cycling). The sample was centrifuged for 30 minutes at 27143 x g and the supernatant, refered to as the cell free extract (CFE) was collected. The CFE was passed through a 0.45 µM syringe filter before loading onto the column.

Gravity-flow columns were prepared with 4 ml TALON resin (Clontech Laboratories Inc.) containing cobalt ions and equilibrated in 40 ml TALON buffer. The CFE was loaded onto the column and the flow through liquid was collected. The column was washed with 2 x 20 ml TALON buffer. The protein was eluted in two fractions; 10 ml TALON buffer with 5 mM imidazole followed by 10 ml TALON buffer with 100 mM Imidazole. SDS-PAGE was used to analyse each stage of the protein purification.

2.7.4. SDS-PAGE

Protein was visualised using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to determine approximate protein size, quantity and purity (Laemmli, 1970). 12.5 % polyacrylamide gels (Acrylogel 3; BDH Electran) were routinely used with AE-6450 apparatus from ATTO Corporation (Genetic Research Instruments) consisting of two 12 cm x 10 cm glass plates sealed with a rubber gasket.

All buffers required for preparation and running of SDS-PAGE gels are outlined in Table 2.11. The resolving gel was pipetted between the plates leaving 2 cm from the top which was covered in water. The gel was allowed to polymerise before the water was removed. The liquid stacking solution was poured on top of the resolving gel and a comb was inserted to form loading wells. Following polymerisation of the second gel layer, the comb and gasket were removed and the gel affixed in the gel tank with running buffer. Loading dye was added to protein samples at a ratio of 1:2 and samples were boiled for 5 minutes when required. Molecular weight (MW) protein standards were loaded into a well alongside the protein samples and a current of 35 A (per gel) was applied. Gels were run for 30 – 45 minutes or until the dye front (from the loading buffer) had reached the end of the gel.

After electrophoresis, the gel was stained in InstantBlue stain (Expedeon) for 15 – 30 minutes to reveal protein bands. The gel was destained overnight using distilled water and an orbital

shaker. Gels were then photographed using Bio-Rad Gel Doc EZ Imager (Image Lab/ PC Windows software).

Component	Reagent	Concentration or Volume
Resolving Gel (per gel)	0.75 M Tris/HCl, pH 8.8 with 0.2 % SDS 40 % Acrylamide (BDH Electran acrylamide, 3 % (w/v) bisacrylamide) d.d. H2O 10 % (w/v) Ammonium persulphate TEMED	2.35 ml 1.45 ml 0.875 ml 22.5 µl 7.5 µl
Stacking Gel (per gel)	0.25 M Tris/HCl, pH 6.8 with 0.2 % SDS 40 % Acrylamide (BDH Electran acrylamide, 3 % (w/v) bisacrylamide) d.d. H2O 10 % (w/v) Ammonium persulphate TEMED	0.938 ml 0.188 ml 0.75ml 15 µl 5 µl
Running Buffer (1 litre)	32 mM Tris / 190 mM glycine, pH 8.3 SDS	350 ml 0.1 % (w/v)
Loading Buffer (10 ml)	0.25M Tris/HCl, pH 8.8 Glycerol SDS β-mercaptoethanol Bromophenol blue dye	5 ml 25 % (w/v) 10 % (w/v) 2.5 ml 0.1 % (v/v)

Table 2.11 Gel and buffer recipes for SDS-PAGE

2.7.5. Removal of His-tag by TEV digest

Recombinant GFP-TEV protease (pET28a) was expressed in BL21 *E. coli* cells. The soluble protein was purified by IMAC and size exclusion chromatography (SEC) before adding 50 % (w/v) glycerol, flash freezing aliquots in liquid nitrogen and storing at -80 °C.

Following initial purification by IMAC, the N-terminal His-tag was removed from protein expressed via pET9 or pB22 vectors (Table 2.2). The protein sample was buffer exchanged (2.7.9) into TEV cleavage buffer (50 mM NaCl, 50 mM Tris pH 8, 0.5 mM EDTA, 0.5 mM TCEP). The total protein amount was calculated using A_{280nm} (2.7.7) and TEV protease was added to a weight ratio of

1:10 (TEV:Protein). The reaction volume was adjusted to 10 ml with TEV cleavage buffer and the sample was incubated, rotating, at 20 °C for 16 – 20 hours.

The sample was buffer exchanged into TALON buffer and a second IMAC purification (2.7.3.) was carried out. The purified protein, lacking a His-tag, was collected in the flow through and wash fractions of the purification. The TEV protease and cleaved His-tag remain bound to the column.

2.7.6. Size exclusion chromatography (SEC)

Protein samples were concentrated to <5 ml for further purification by size exclusion chromatography (SEC). A HiLoad 16/60 Superdex 200pg 120 ml column (GE Healthcare) gel filtration column was used with the ÄKTA pure fast protein liquid chromatography system (GE Healthcare). The column was pre-equilibrated in the required buffer (typically 10 mM HEPES pH 7.5, 100 mM NaCl) before loading the concentrated sample. The proteins were separated on the column at a flow rate of 1.2 ml/min and 2 ml fractions were collected. Fractions were analysed using SDS-PAGE.

2.7.7. Determination of Protein Concentration

Protein concentration was determined using NanoDrop 2000 benchtop spectrophotometer (Thermo Fisher Scientific). 2 µl of purified protein sample loaded onto the detector and $A_{280\text{-}320nm}$ was analysed. The protein concentration was calculated using Equation 2.

$$A = \varepsilon C l D$$

Equation 2 – Calculating molar concentration of a protein solution
Where A = absorbance at 280 nm – absorbance at 320 nm, ε = molar extinction coefficient, C = molar concentration of the sample, l = length of light path (cm), D = dilution factor

The extinction coefficient for each protein was calculated by entering the specific amino acid sequence of the expressed protein into the ProtParam tool (Table 2.16). Alternatively, the

absorbance of a 0.1 % (1 g/l) solution of the specific protein ($\varepsilon^{0.1\%}$) was predicted using ProtParam and used to calculate protein concentration in mg/ml using A_{280nm} and Equation 3.

$$\text{Protein concentration (mg/ml)} = \frac{A}{\varepsilon^{0.1\%}}$$

Equation 3 – Calculating concentration of a protein solution in mg/ml
Where A = absorbance at 280 nm, $\varepsilon^{0.1\%}$ = 0.1% extinction coefficient.

2.7.8. Centrifugal concentration of Proteins

Protein solutions were concentrated as required using 20 ml Vivaspin (Sartorius) or 15 ml Amicon Ultra-15 (Merck) centrifugal concentrators with 30, 50 or 100 kDa molecular weight cut off filters as appropriate. Centrifugation was carried out using a Heraeus Megafuge 16R (Thermo Fisher Scientific) or an Eppendorf 5804R bench-top centrifuge at 4200 rpm (3215 x g) with swing out rotor at 10 °C.

2.7.9. Buffer Exchange

Soluble proteins in small volumes were buffer-exchanged using centrifugal concentrators as above. The sample was concentrated to <1 ml then diluted with the desired buffer up to 20 ml. The concentration and dilution was repeated to ensure complete buffer exchange.

Larger volumes of buffer were exchanged using dialysis. Protein samples were contained within a section of dialysis tubing with a MW cut-off of 12 kDa and both ends sealed with clips. The tubing was submerged in 4 l of the appropriate buffer and stirred at 4°C overnight to allow buffer exchange.

2.8. Native Protein Expression and Purification from *B. theta*

2.8.1. Protein expression in *B. theta*

The required strain of *B. theta* from -80 °C stocks was used to directly inoculate 5 ml TYG rich media with added His-heme (Table 2.3). The culture was incubated at 37 °C overnight under anaerobic conditions (A35 Workstation anaerobic cabinet, Don Whitely Scientific). The desired media (TYG or MM, Table 2.3) for large scale growths was made and decanted into 500 ml DURAN glass bottles. The media was autoclaved and allowed to fully cool before use. 500 µl of His-heme was added to each bottle and 5 ml 50 % fructose if required (MM only). The media was then inoculated with 500 µl overnight culture and incubated at 37 °C under anaerobic conditions for 18 – 20 hours to OD_{600nm} 1.5 – 2.0. Cells were harvested by centrifugation at 11305 x g for 30 minutes. Cell pellets were resuspended in TALON buffer and stored at -20 °C until processed for protein purification.

2.8.2. Selenomethionine incorporation

10 litres of wild type *B. theta* were grown in TYG medium as described above then harvested by centrifuging at 11305 g for 30 minutes. The cells were resuspended and used to inoculate 10 litres of MM containing 0.5 % (w/v) fructose. The cells were incubated at 37°C anaerobically for 30 minutes before the addition of amino acids (Lys, Phe, Thr, Leu, Ile, Val) and selenomethionine, according to a protocol developed for SeMet incorporation in *E. coli* by inhibition of methionine biosynthesis (Van Duyne *et al.*, 1993). The cells were incubated for a further 17 hours to a final OD_{600nm} of 1.7 and harvested as described.

2.8.3. Outer membrane purification

Cells were thawed completely and lysed at a pressure of 23 kilopounds per square inch (KPSI) using a cell disrupter (Constant Systems). The cells were centrifuged a 42,000 rpm (204526 x g) for 1 hour to collect total membranes and the supernatant (CFE) was discarded. The pellets were resuspended using a glass homogoniser in 0.5 % (w/v) Sodium Lauroyl Sarcosine (sarkosyl), Table 2.12. Sarkosyl selectively solubilises inner membrane (IM) proteins of gram negative bacteria (Filip *et*

al., 1973). The sample was incubated stirring at room temperature for 30 minutes and centrifuged at 204526 x g for 30 minutes. The supernatant was discarded and the sarkosyl wash was repeated. The pellets were resuspended and homogenised in a suitable buffer with Lauryldimethylamine-N-oxide (LDAO) (Table 2.12) and incubated stirring at room temperature for 1 hour, or at 4 °C overnight. The sample was centrifuged at 204526 x g for 30 minutes and the supernatant containing solubulised outer membrane proteins (OMPs) was collected.

Buffer	Function	Recipe	pH	Detergent (w/v)
Sarkosyl Buffer	Solubilising inner membranes	20 mM HEPES	7.5	0.5 % Sarkosyl
LDAO Buffer	Solubilising OMPs for anion exchange	10 mM HEPES/ 50 mM NaCl	7.5	1 % LDAO
TALON-LDAO Buffer	Solubilising OMPs for IMAC	20 mM Tris/ 300 mM NaCl	8	1.5 % LDAO

Table 2.12 Buffers for solubilising OMPs from *B. theta*

2.8.4. Determination of total protein concentration

Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) was used to determine the total solubilised protein. 10 µl of sample was added to 500 µl reagent A and 10 µl reagent B. Protein standards were set up containing 0 µg, 2 µg, 5 µg and 10 µg of BSA with the equivalent reagents. All samples were mixed well by vortexing and incubated at 45 °C for 1 hour. The A_{562nm} was taken for each sample using a BioSpectrometer (eppendorf). The BSA standards were used to produce a standard curve of absorbance at 562 nm against total protein (µg). The total protein in the membrane sample was then calculated using the equation for the standard curve.

2.8.5. Anion-exchange chromatography

Anion-exchange (IEX) was used to separate *B. theta* OMPs for identification and was ultimately used for the purification of the BT2261-4 complex. To avoid overloading the column the cells were processed in batches of 2-3 litres. The solubilised total membrane sample was passed through a 0.45 µm filter and loaded onto a Resource Q 6 ml anion exchange column using the ÄKTA pure system (GE healthcare). The proteins were separated with a salt gradient from 50 mM to 1 M in 20 column volumes (Table 2.13). Peak fractions were analysed using SDS-PAGE

Buffer	Function	Recipe	pH	Detergent (w/v)
IEX Buffer A	Equilibration	10 mM HEPES/ 50 mM NaCl	8	0.06 % LDAO
IEX Buffer B	Elution	10 mM HEPES/ 1 M NaCl	8	0.06 % LDAO

Table 2.13 Buffers used for anion-exchange chromatography of *B. theta* OMPs

2.8.6. Immobilised Metal Affinity Chromatography (IMAC)

BT1762-3 was purified using a C-terminal His_{x6}-tag added to the genomic copy of $BT1762^{SusD}$ which allowed separation by IMAC. Gravity-flow columns were prepared with 10 ml TALON resin (Clontech Laboratories Inc.) containing cobalt ions and equilibrated in 100 ml TALON equilibration buffer (Table 2.14). The solubilised membrane sample in TALON-LDAO buffer was loaded slowly onto the column washed with 150 ml TALON wash buffer. The protein was eluted in 20 ml TALON buffer with 100 mM Imidazole. SDS-PAGE was used to analyse the IMAC samples.

Buffer	Function	Recipe	pH	Detergent (w/v)
TALON equilibration buffer	Equilibration	20 mM Tris, 300 mM NaCl	8	0.2 % LDAO
TALON wash buffer	Washing the column	20 mM Tris, 300 mM NaCl 5 mM Imidazole	8	0.2 % LDAO
TALON elution buffer	Elution of bound OMPS	20 mM Tris, 300 mM NaCl 100 mM Imidazole	8	0.2 % LDAO

Table 2.14 Buffers for solubilising OMPs from *B. theta*

2.8.7. Size exclusion chromatography (SEC)

Membrane proteins from anion exchange or IMAC were further purified by size exclusion chromatography (SEC). A HiLoad 16/60 Superdex 200pg 120 ml gel filtration column (GE Healthcare) with the ÄKTA pure system (GE healthcare) was used. The column was first equilibrated in the required buffer with LDAO or n-Decyl-β-D-Maltopyranoside (DM) detergent (Table 2.15). The sample was concentrated to <5 ml and loaded onto the column. The proteins were eluted at 1.2 ml/min and 3 ml fractions were collected. Peak samples of interest were analysed using SDS-PAGE. A second SEC run was used to buffer exchange the protein into a final buffer with 0.4 % tetraethylene glycol monooctyl ether detergent (C_8E_4, Table 2.15) for crystallisation.

Buffer	Function	Recipe	pH	Detergent (w/v)
DM SEC Buffer	Size exclusion column	10 mM HEPES/ 100 mM NaCl	7.5	0.12 % DM
LDAO SEC Buffer	Size exclusion column	10 mM HEPES/ 100 mM NaCl	7.5	0.05 % LDAO
C_8E_4 Buffer	Final protein buffer	10 mM HEPES/ 100 mM NaCl	7.5	0.4 % C_8E_4

Table 2.15 Size exclusion chromatography buffers for OMPs from *B. theta*

2.8.8. SDS-PAGE of *B. theta* OMPs

Samples were prepared with loading buffer (NuPAGE LDS sample buffer (4 x), Thermo Fisher Scientific) and boiled for 5 minutes, if required. Membrane protein samples were generally analysed boiled and non-boiled as many OMPs are heat modifiable (Beher *et al.*, 1980). A NuPAGE 4-12 % Bis-Tris protein gel was inserted into the XCell SureLock Mini-Cell tank which was filled with 1 x NuPAGE MOPS SDS Running Buffer (Gel, tank and buffer - Thermo Fisher Scientific). Up to 15 µl of each sample was loaded onto the gel alongside 5 µl Novex Pre-stained Protein Standard (Thermo Fisher Scientific). Gels were run at 180 V for 75 – 90 minutes or until the dye front (from the loading buffer) had reached the end of the gel.

The gel was removed from the tank, the plates separated and the gel transferred to a staining tray. The gel was covered in water, microwaved for 1 minute, and incubated for 5 – 10 minutes on an orbital shaker. The washing steps were repeated to ensure removal of SDS from the gel. Coomassie Blue stain (0.01 % Coomassie, 10% ethanol, 30mM HCl) was used to cover the gel and incubated for 1 hour. The stain was removed and the gel rinsed in water. The gel was then destained by incubating overnight in water. Gels were photographed using Bio-Rad Gel Doc EZ Imager (Image Lab/ PC Windows software).

2.8.9. Native PAGE

Samples were prepared with 0.5 % n-Dodecyl β-D-maltoside (DDM), 1 x NativePAGE sample buffer and NativePAGE G-250 Additive as per the manufacturer's instructions (Thermo Fisher Scientific). NativePAGE 4-16% Bis-Tris Protein Gels were used with the NativePAGE Anode Buffer and NativePAGE Dark Blue Cathode Buffer which were prepared following the manufacturer's protocol (Thermo Fisher Scientific). The samples were run at 180 V for 90 – 120 minutes using the XCell SureLock Mini-Cell tank (Thermo Fisher Scientific).

The gel was transferred to a staining tray, covered with 100 ml Fix Solution (40 % Methanol, 10 % Acetic Acid) and microwaved for 45 seconds. The gel was incubated on an orbital shaker for 15 minutes before the Fix Solution was removed. 100 ml Destain Solution (8 % Acetic acid) was added to the tray and the gel was incubated overnight to remove the excess stain. Native gels were photographed using Bio-Rad Gel Doc EZ Imager (Image Lab/ PC Windows software).

2.9. Crystallography

2.9.1. Protein crystallisation trials

Membrane and soluble proteins for crystallisation were purified by IMAC or ion exchange chromatography followed by SEC (Sections 2.7 and 2.8). Proteins were concentrated (Section 2.7.8) to 10 – 20 mg/ml in buffer used for SEC. Protein crystallisation was performed using the sitting-drop vapour-diffusion method. Initial screens were set up using an automated Mosquito Crystal nanolitre dispensing robot (TTP Labtech) onto MRC 96-well crystallisation plates (Molecular Dimensions). Commercial crystallisation screens from Molecular Dimensions MemGold 1 and MemGold 2 were used for membrane proteins and complexes. For soluble proteins appropriate screens were selected from a wide range; Structure PACT, Morpheus and JCSG Plus (Molecular Dimensions), Index (Hampton Research), JCSG core (I to IV) Suite and $AmSO_4$ Suite (QIAGEN). Solutions were dispensed 200 nl + 200 nl and 200 nl + 150 nl of protein and crystallisation solution, respectively. The ratio of protein to crystallisation solution was varied when required. Crystallisation plates were incubated at 20 °C.

2.9.2. Optimising protein crystals

Optimisation of initial crystal hits was generally carried out by hand using the hanging-drop vapour diffusion method with XRL or VDX 24-well hanging drop plates (Molecular Dimensions). Original screen conditions were replicated with varied precipitant concentrations. If this was unsuccessful, further optimisations were carried out modifying the concentration of the salt or pH of the buffer. Routinely 0.8 µl + 0.8 µl and 1.2 µl + 0.8 µl of proteins and crystallisation solutions, respectively, were dispensed.

The exception to this method was the BT1760-3 crystals which could not be replicated by manual sitting or hanging drops. The Mosquito was used to dispense 400 nl + 400 nl and 400 nl + 300 nl of protein and crystallisation solution, respectively, onto Intelli 48-3 sitting drop plates.

2.10. Structural Biology

2.10.1. In-house screening of crystals for diffraction

Crystals were harvested in suitably sized loops (Hampton) by Bert van den Berg, cryo-protected in an appropriate buffer and flash frozen immediately in liquid nitrogen. Crystals were mounted on a copper anode Rigaku MSC microfocus generator and data collected at a 1.54 Å on a Raxis IV++ image plate detector. The diffraction images were indexed in MOSFLM. Crystals were ranked based on resolution and diffraction pattern with the most suitable crystals sent for data collection.

2.10.2. Data Collection

Datasets were collected on the most suitable beamline at Diamond Light Source (DLS, Didcot, Oxfordshire, UK) by Dr. Arnaud Baslé.

2.10.3. Solving BT2263 SusD homologue by Molecular Replacement

X-ray Detector Software (XDS) was used to determine the space group, index and scale BT2263 diffraction data (Kabsch, 2010). The data was used to a resolution of 1.9 Å. BT2259 SusD-like was used as the MR model (PDB ID: 4Q69 26 % identity to $BT2263^{SusD}$). MR of the scaled data with the model was carried out in Phaser-MR within Phenix (Adams *et al.*, 2010). One round of refinement was used, followed by Autobuild (Resolve and Buccaneer). The resulting model had R_{work} 0.243 and R_{free} 0.270. Several rounds of building and editing the structure by hand in Coot followed by refinement in Phenix were used to complete the model (Adams *et al.*, 2010). The final R_{free} for the $BT2263^{SusD}$ structure was 0.198.

2.10.4. Structures solved by Bert van den Berg

The remaining structures presented in this study were solved by Bert van den Berg; Soluble BT2262 lipoprotein, BT2261-4 complex, BT1762-3 apo and BT1762-3 with ligand.

2.11. Bioinformatics

Bioinformatics tools used in this study are listed in Table 2.16 below.

Bioinformatic Tool		Website	Description
Genetic and Proteomic tools	BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi Hosted by the National Centre for Biotechnology Information (NCBI)	Finds proteins with conserved regions (homologues). The query sequence is compared to a database of non-redundant protein sequences.
	InterPro	https://www.ebi.ac.uk/interpro/	Prediction of protein domains based on the amino acid sequence
	Pfam 31.0	http://pfam.xfam.org/	Database of protein domains
	KEGG	http://www.genome.jp/kegg/	Kyoto Encyclopedia of Genes and Genomes. Used to search for DNA/protein sequences and organisation of proteins within the genome.
Prediction of signal peptides	Signal P 4.1	http://www.cbs.dtu.dk/services/SignalP/ Hosted by the Center for Biological Sequence Analysis	Prediction of signal peptides in Gram-negative or Gram-positive bacteria
Protein Parameters	ProtParam	http://web.expasy.org/protparam/ From ExPASy Bioinformatics Resource Portal	Calculates protein parameters (molecular weight, isoelectric point and extinction coefficient) based on protein sequence
Primer Design	OligoCalc	http://biotools.nubic.northwestern.edu/OligoCalc.html	Calculates oligonucleotide parameters (melting temp, G/C content etc.) based on DNA sequence.
	WEBcutter 2.0	http://rna.lundberg.gu.se/cutter2/	Finds restriction enzyme sites within a DNA sequence
Alignment tools	Clustal Omega	http://www.ebi.ac.uk/Tools/msa/clustalo/ Hosted at the European Bioinformatics Institute (EBI) website	Multiple alignment of amino acid sequences.
	Multalin	http://multalin.toulouse.inra.fr/multalin/	Multiple alignment of amino acid or DNA sequences.
Structures	PDB	https://www.rcsb.org/pdb/	Protein Data Bank. Database of protein structures

Table 2.16 Bioinformatics tools

2.12. Biochemistry

2.12.1. Western Blotting

Samples were analysed alongside MagicMark XP Western Protein Standard (Thermo Fisher Scientific) using SDS-PAGE as described above (2.7.4.). Protein was then transferred to an Amersham Protran 0.45 Nitrocellulose membrane (GE Healthcare) using transfer buffer pre-cooled to 4 °C (Table 2.16) and the Mini Trans-Blot system (Bio-Rad). The transfer was carried out at 80 V for 90 minutes. The buffer was kept cold during the transfer using a freezer block and a magnetic stirrer.

The membrane was blocked for 1 hour at room temperature or overnight at 4 °C in blocking buffer (Table 2.17). The blot was then incubated with the required primary antibody in 10 ml antibody buffer for 1 hour. Anti-Flag antibody (F7425, Sigma-Aldrich) was used 1:2000 and the Anti-His_{x6}-Peroxidase (11965085001 *ROCHE*, Sigma-Aldrich) antibody was used 1:5000. The primary antibody was removed by washing three times for 5 minutes each with wash buffer (Table 2.17). FLAG western blots required a secondary antibody; Goat Anti-rabbit-HRP antibody (sc-2004, Santa cruz) 1:5000. The secondary antibody was diluted in the antibody buffer, the membrane was incubated for 1 hour and washed as before.

The blot was developed using 500 µl each of the Clarity Western ECL reagents (Bio-rad). An image of the membrane was produced using ChemiDoc XRS+ System (Bio-Rad).

Buffer	Component	Amount	Concentration
Transfer Buffer (1 litre)	Tris - HCl Glycine SDS Methanol Water	3 g 14.3 g 0.5 g 300 ml 200 ml	25 mM 192 mM 1.7 mM 30 % -
Blocking Buffer (10 ml)	1 x PBS (Oxoid, 1 tablet per 100 ml) Milk powder Tween 20	10 ml 0.5 g 50 µl	- 5 % 0.5 %
Antibody Buffer (10 ml)	1 x PBS (Oxoid, 1 tablet per 100 ml) Milk powder	10 ml 0.5 g	- 5 %
Wash Buffer (50 ml)	1 x PBS (Oxoid, 1 tablet per 100 ml) Tween 20	50 ml 250 µl	- 0.5 %

Table 2.17 Buffers required for Western Blotting

2.12.2. Isothermal Titration Calorimetry (ITC)

Isothermal Titration Calorimetry (ITC) can determine the thermodynamic parameters driving macromolecular interactions by titrating ligand into protein. The experiment was carried out using a MicroCal VP-Isothermal Titration Calorimeter at 25 °C. The heat consumed or released in the reaction cell was measured in comparison to a reference cell at stable temperature. The cell and syringe were equilibrated by dialysis (Section 2.7.9) in 20 mM HEPES pH7.5 with the protein and ligand used in the identical buffer. The protein in the cell (40 – 60 µM) was equilibrated to 25 °C. The ligand (5 – 10 mM or 5 – 10 mg/ml) was titrated into the cell, from the syringe, in 28 aliquots of 10 µl. Integrated heat effects were analysed by non-linear regression using a single site-binding model (MicroCal Origin v7) giving the association constant (K_a) and enthalpy of binding (ΔH).

2.12.3. Thin Layer Chromatography (TLC)

Thin Layer Chromatography (TLC) allows a mixture of oligosaccharides to be visualised. TLC plates (Silica gel 60, Sigma-Aldrich) were cut to size and 2 – 6 µl of samples were spotted onto plates 1 cm from the bottom. The spots were dried and the plate placed into a tank containing 1 cm of running buffer (1-butanol, acetic acid and water at 2:1:1). The running buffer migrated to within 1 cm of the top of the plate which was dried and replaced in the tank to migrate again. The plate was dried completely and submerged in developer solution (sulphuric acid, ethanol and water at 3:70:20 with 1% orcinol) for 5 – 10 seconds. The plate was dried and incubated at 80 – 100 °C until sufficiently developed.

2.12.4. Oligosaccharide Production and Purification

Initial small-scale time course digests of levan were carried out and analysed by TLC to find the optimal parameters for a partial digest of the polysaccharide. The reaction was scaled up to produce sufficient oligosaccharides for purification.

500 mg of levan (*Erwinia herbicola*, Sigma-Aldrich) was dissolved in 100 ml of PBS (Oxoid) by warming in a hot waterbath 60 – 80 °C. The solution was allowed to cool before 100 nM BT1760

endo-acting GH32 levanase was added and the reaction was incubated in a 37 °C waterbath for 20 minutes. The reaction was heat inactivated by boiling for 20 min. The sample was incubated at -80 °C for 1 hour and freeze dried using a Christ Alpha 1-2 Freeze Drier at -60 °C. The resulting powder was then resuspended in 5 ml of H_2O for purification.

The partially digested levan products were separated by size exclusion chromatography using P2 Bio-gel (Bio-Rad) matrix packed in 2 Glass Econo-Columns (2.5 cm × 80.0 cm) with a flow adaptor (Bio-Rad). The column was pre-equilibrated in 50 mM Acetic acid and the digest was loaded directly onto the column bed. The oligosaccharides were separated and eluted from the column at 0.25 ml/min in using a peristaltic pump (LKB Bromma 2132 microperpex). The 2 ml fractions were collected continuously for 48 hours using a Bio-Rad model 2110 fraction collector. A 2 µl aliquot of every fifth fraction was analysed by TLC identifying fractions of interest. Freeze drying was used to remove the acetic acid from the fractions.

Chapter 3 – Structure of a SusCD complex from Bacteroides thetaiotaomicron

3.1. Introduction

Proteins in the outer membrane (OM) of Gram-negative bacteria often form highly stable β-barrels which can function as enzymes, signalling proteins or transporters in addition to forming an integral part of the membrane. In contrast to those of other Gram-negative species (e.g. *E. coli*) Bacteroidetes outer membrane proteins (OMPs) have gone largely unstudied.

Bacteroidetes have groups of co-regulated genes organised into numerous polysaccharide utilisation loci (PULs) encoding the specific proteins required for sensing, binding, degrading and importing glycans (Martens *et al.*, 2009). Individual PULs, from Bacteroidetes found in the HGM, have been identified for all known naturally occurring glycans, except cellulose (Martens *et al.*, 2011). This includes a vast range of polysaccharides varying in size, charge and branching.

The general model for polysaccharide utilisation, Figure 3.1, involves a surface glycan binding protein (SGBP) which binds polysaccharides, however, this protein has been shown to be non-essential in many PULs so the exact function is not clear (Tauzin *et al.*, 2016; Cartmell *et al.*, 2017). The complex polysaccharides are digested by an extracellular endo-acting carbohydrate active enzyme, typically a glycoside hydrolase or polysaccharide lyase. The resulting oligosaccharides are bound by a SusD homologue and imported by a SusC homologue, an essential step in the utilisation of complex carbohydrates (Shipman *et al.*, 2000). The imported oligosaccharides are then digested to monosaccharides by exo-acting carbohydrate active enzymes in the periplasm. The monosaccharide products are imported into the cytoplasm where they enter into fermentation pathways yielding carbon and energy for the cell. The import of final products across the inner membrane is not fully understood within the context of PULs. Some PULs encode major facilitator superfamily (MFS) transporters but have yet to be characterised and the size limit for import is unknown (eg. monosaccharides or small oligosaccharides).

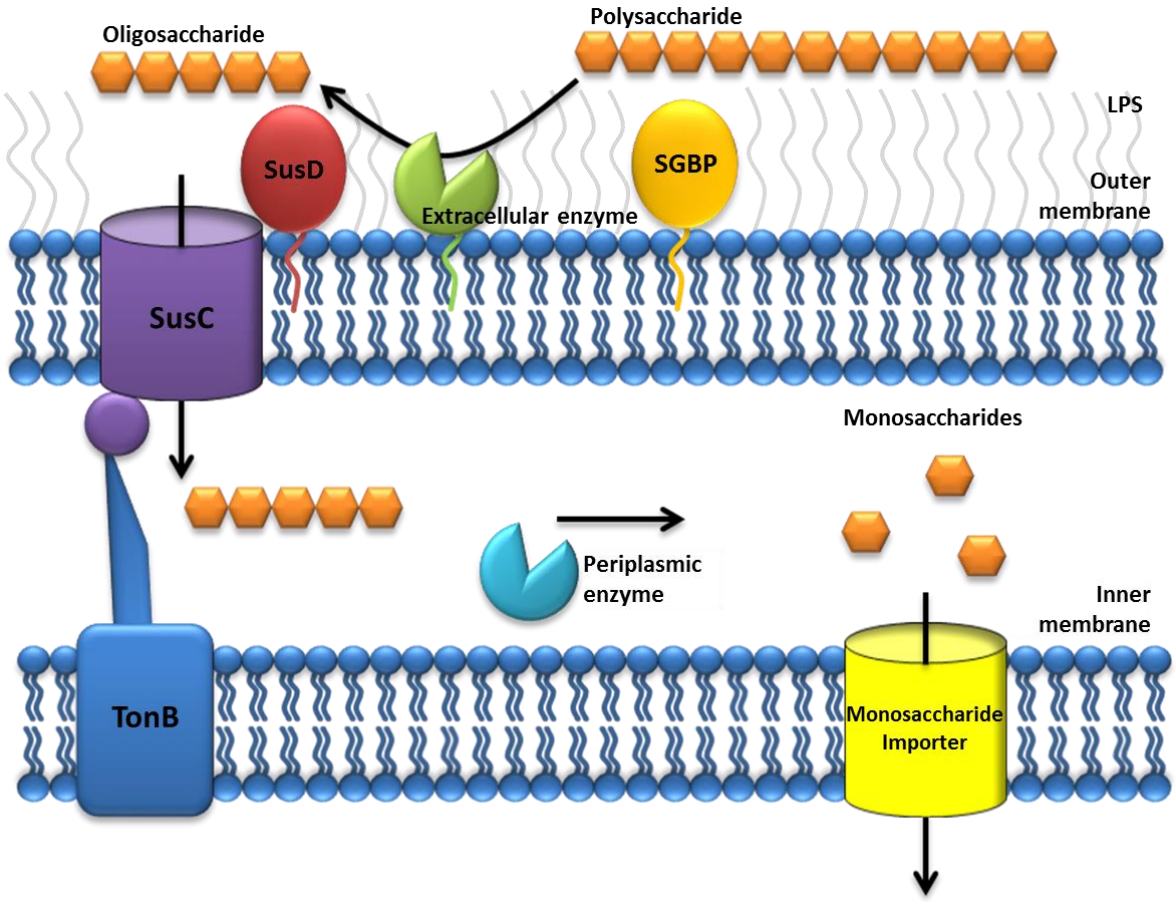


Figure 3.1 General model of polysaccharide utilisation by Bacteroidetes.
The target polysaccharide is bound by a surface glycan binding protein (SGBP) then digested by an extracellular endo-acting carbohydrate active enzyme. The oligosaccharide product is then bound by a SusD homologue and imported by a SusC homologue. The oligosaccharide is digested further by exo-acting carbohydrate active enzymes in the periplasm and monosaccharides are imported into the cytoplasm.

Sequestering oligosaccharides and monosaccharide products to the periplasm is thought to minimise the loss of simple products to other members of the densely populated gut microbiota (Cuskin *et al.*, 2015). However, some Bacteroidetes have been found to 'share' partially digested glycans with other bacteria (Rogowski *et al.*, 2015).

SusC homologues are predicted to be 22 β-stranded TonB dependent transporters (TBDTs) which are thought to interact in the OM with the SusD homologue via the conserved tetratricopeptide repeats (TPRs) (Bolam and Koropatkin, 2012). TBDTs are characterised by an N-terminal globular plug domain which occludes the channel of the 22-stranded β-barrel and a

conserved TonB box sequence of amino acids at the N-terminal which interacts with TonB. The IM protein TonB forms a complex with ExbB and ExbD which altogether provide the energy required for transport by TBDTs (Hickman *et al.*, 2017).

This study focused on *Bacteroides thetaiotaomicron* (*B. theta*), one of many anaerobic, Gram-negative Bacteroidetes which are dominant in the human gut microbiota (HGM). There are 103 SusC homologues encoded by the *B. theta* genome but the structure and exact function of the transporters have not been successfully determined. The findings reported here have implications beyond the HGM as the Bacteroidetes phylum are widely distributed in the environment. This includes sediments, soil, sea water, on the skin, in the mouth and in the gut of many animals (Grondin *et al.*, 2017b). All Bacteroidetes genomes sequenced to date encode SusCD homologues (Grondin *et al.*, 2017a).

3.2. Aims

The aim of this chapter was to identify highly expressed outer membrane proteins from *B. thetaiotaomicron*. The second objective was to purify, crystallise and solve the structure of a SusCD homologue from *B. thetaiotaomicron.*

3.3. Results

3.3.1. *Bacteroides thetaiotaomicron* outer membrane proteins

3.3.1.1. Comparing *E. coli* and *B. theta* outer membrane proteins

Outer membrane proteins from *E. coli* and *B. theta* grown on minimal and rich media were visualised using SDS-PAGE (Figure 3.2). *B. theta* did not express classical trimeric porins at sufficient levels to be visible by SDS-PAGE. Trimeric general porins are abundant in *E. coli* and dissociate to monomers when boiled (OmpF/C highlighted by red boxes). In contrast, *B. theta* expressed many proteins over 100 kDa (Figure 3.2) suggesting a fundamentally different mechanism for nutrient acquisition. This result was expected as the *B. theta* genome does not encode any obvious OMP homologues. There are fewer differences between the two *B. theta* conditions than the *E. coli* which may be because the rich and minimal media used for *B. theta* both contained glucose which likely elicits the same level of catabolite repression.

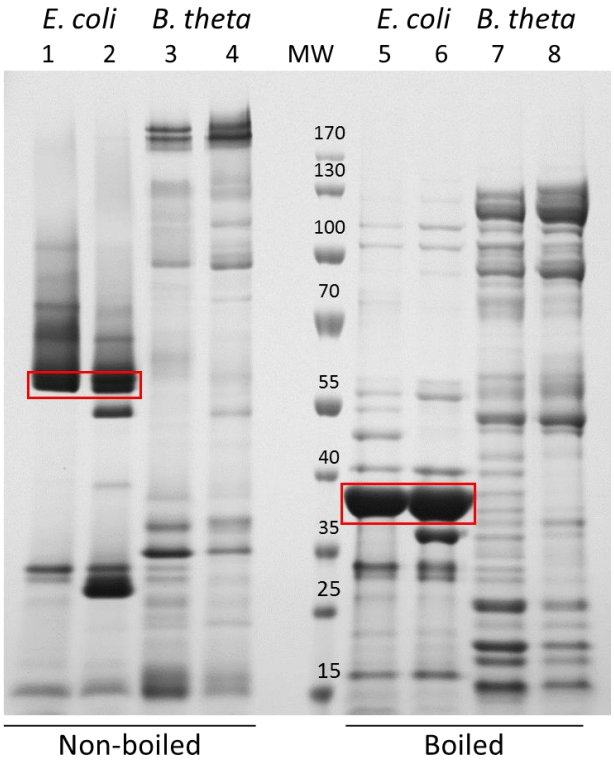


Figure 3.2 SDS-PAGE of total outer membrane proteins from *E. coli* and *B. theta*.
E. coli (lanes 1, 2, 5, 6,) and *B. theta* (lanes 3, 4, 7, 8). Rich medium (even numbered lanes) and minimal medium (odd numbered lanes). Abundant *E. coli* trimeric porins OmpF/C are highlighted by red boxes. Samples 5-8 were boiled before loading and molecular weight (MW) marker indicates approximate protein sizes in kDa. 1 litre of bacteria was grown in each medium. The cells were lysed and processed as outlined in Methods 2.8.3. The samples were then loaded onto an anion exchange Resource Q column, in 0.6 % LDAO, as described in 2.8.5, then eluted in 4 X CV of 1 M NaCl. Approximately 10 ug of the total protein was loaded on the gel.

3.3.1.2. Purification and Identification of *B. theta* OMPs

Wild-type (WT) *B. theta* grown on rich Tryptone Yeast Glucose (TYG) media was processed as described previously (Methods 2.8). A final solubilisation step was carried out using 1 % LDAO detergent before separation by anion exchange using a Resource Q (6 ml) column.

A shallow salt gradient (initially 50 mM to 1 M NaCl over 120 ml) separated the proteins as shown in the chromatogram in Figure 3.3A. The peak samples (labelled A-D) were analysed using SDS-PAGE, boiled (B) and non-boiled (NB, Figure 3.3B). This gel indicated many of the OMPs are heat-modifiable as bands appeared at different MW and extra bands are visible after boiling. The high molecular weight bands, highlighted by the red boxes, were excised and identified by Dr Joe Gray, of Newcastle University, using LCMS-MS.

Bands 1 and 5 contained BT2409 and BT3560, respectively, which were predicted to be TBDTs. Band 3 was identified as BT2260, a SusC homologue from an operon with two predicted enzymes of unknown function and two hypothetical proteins. Bands 2 and 4 were BT1440 and BT2268, respectively. These proteins were both identified as SusC homologues each expressed with a SusD homologue partner. However these SusCD homologues are expressed alone, rather than as part of PULs, and we termed these orphan SusCD homologues.

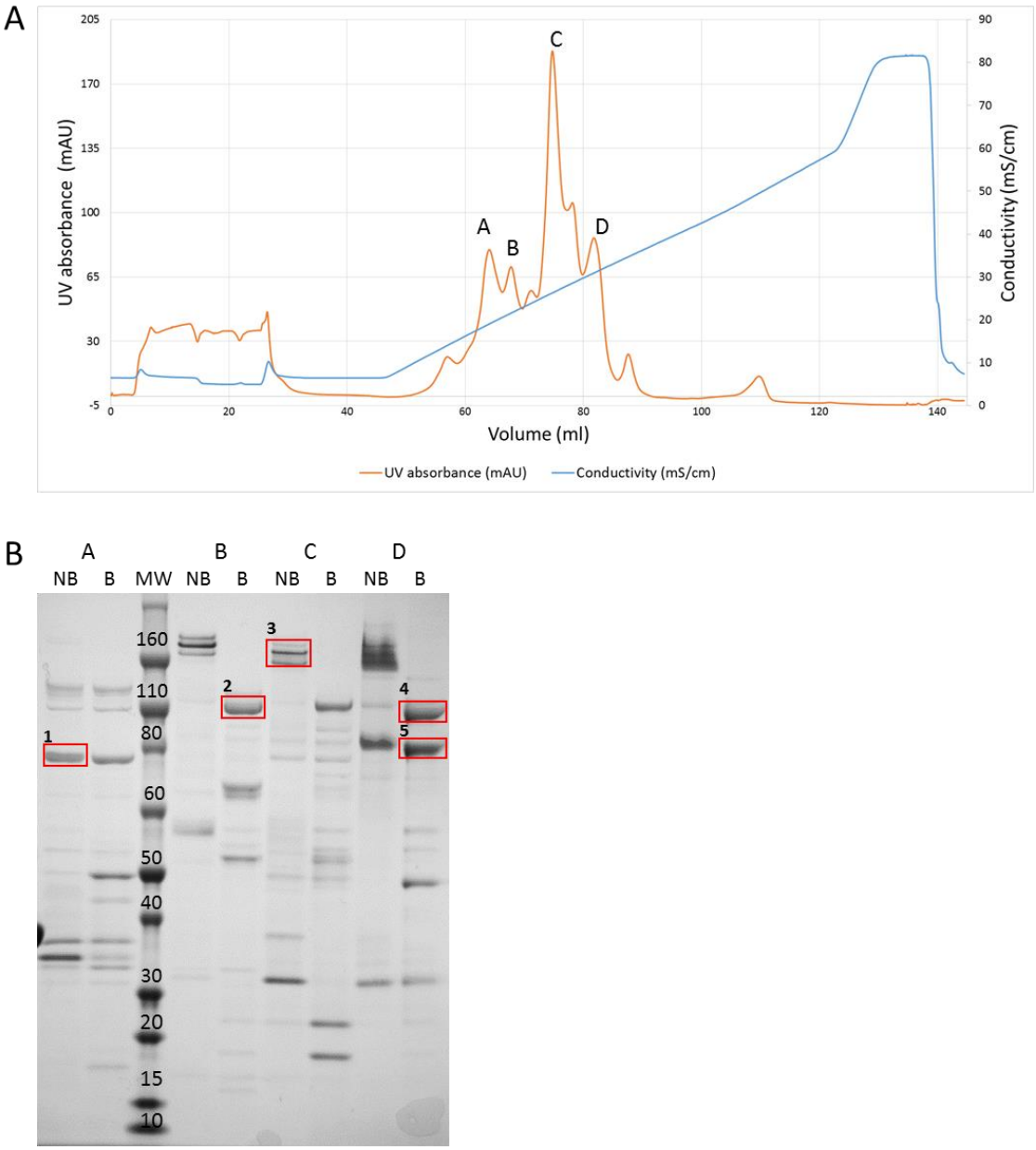


Figure 3.3 Identification of OMPs from *B. theta* grown on TYG rich medium.
A: Anion exchange chromatogram of *B. theta* total outer membrane proteins separated on a ResourceQ column after extraction using LDAO. The main peaks were labelled A-D **B:** SDS-PAGE of peak fractions both boiled (B) and not boiled (NB). Bands highlighted by the red boxes were excised and identified using peptide mass fingerprinting. Band 1: BT2409, Band 2: BT1440, Band 3: BT2260, Band 4: BT2268, Band 5: BT3560. Molecular weight (MW) marker indicates protein sizes in kDa.

A SusC homologue of particular interest was BT1763, from the well characterised *B. theta* PUL which targets the fructo-oligosaccharide levan (Introduction 1.3.2). The PUL is expressed in response to growth on the monosaccharide fructose (Sonnenburg *et al.*, 2010) which is readily available in large quantities making BT1763 a tractable SusC target for large scale native expression. Therefore, the process of purifying and identifying *B. theta* OMPs was repeated using bacteria grown on minimal media containing 0.5 % fructose as the only carbon source (MM-Frc). The anion exchange chromatogram in Figure 3.4A shows proteins eluted as the salt concentration increased. The peaks labelled A and B were further purified using gel filtration. The resulting fractions were analysed boiled (B) and non-boiled (NB) using SDS-PAGE shown in Figure 3.4B. As with the previous samples, the high molecular weight bands highlighted by red boxes were excised and identified by Dr Joe Gray, of Newcastle University, using LCMS-MS.

Band 2 was identified as containing the protein of interest BT1763^{SusC} and another SusC-like BT2268. This was also isolated from the rich media outer membrane purification. Bands 3 and 4 contain the SusD homologues BT1762 and BT2269, respectively. Band 1, which was non-boiled, contained all four of these proteins. This indicated that SusC and SusD homologue pairs interact very strongly such that they co-purify. This interaction is not compromised by strong detergents or SDS and separation of SusCD complex requires boiling. The two SusCD homologue pairs have very similar predicted isoelectric points (5.14 and 5.20), making them very difficult to separate without the introduction of affinity tags fused to the expressed protein.

The peak B sample (Figure 3.4B) showed a single large molecular weight smear around 160 kDa in the non-boiled sample, which became several smaller bands when boiled, indicating a very stable protein complex. The highlighted bands were identified as another SusCD homologue pair BT2264^{SusC} in band 5, and BT2263^{SusD} in band 6. This complex is relatively straightforward to express and purify directly from WT *B. theta* as it binds to the anion exchange column relatively tightly in comparison to the other outer membrane proteins.

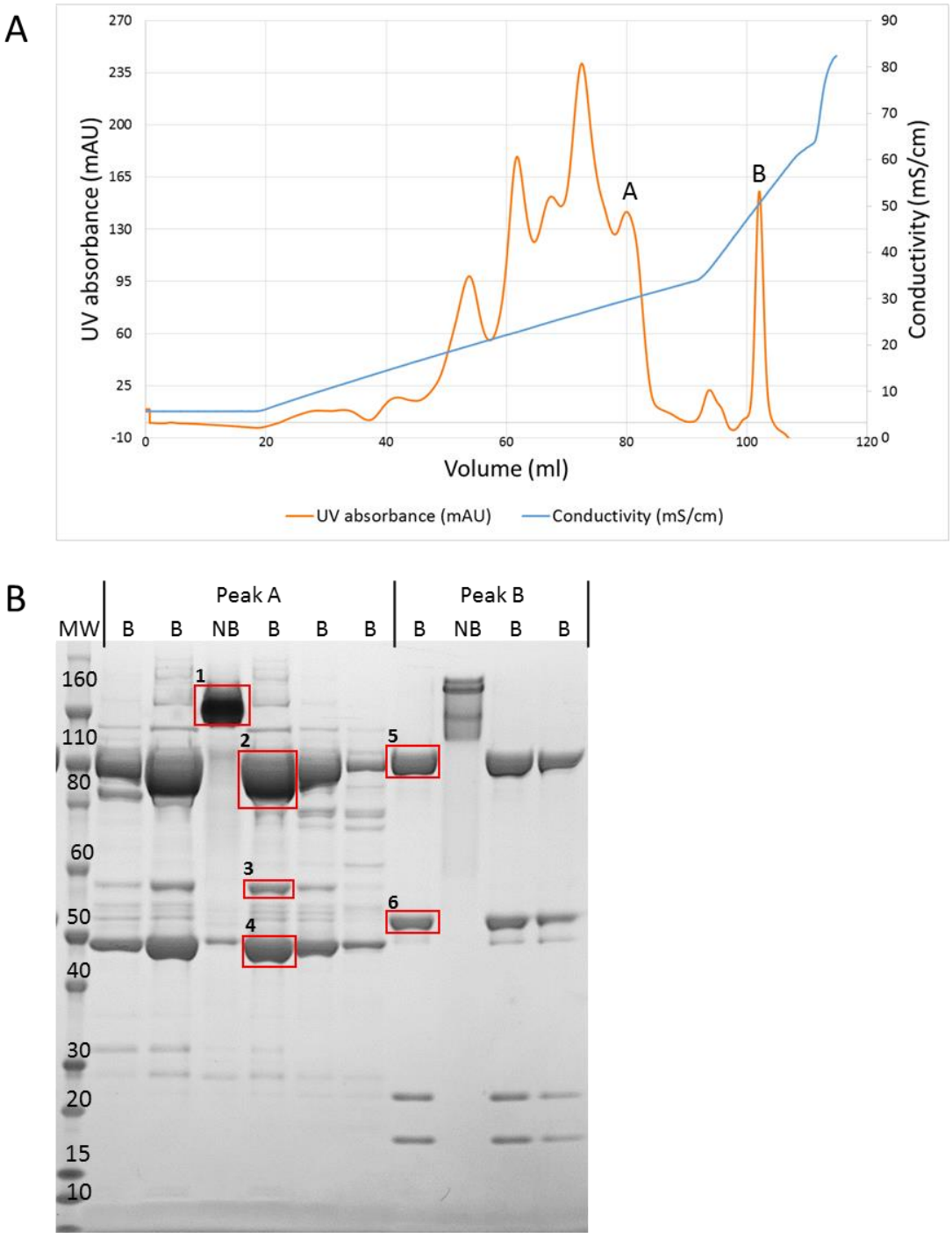


Figure 3.4 Identification of OMPs from WT *B. theta* grown on MM-Frc.
A: Ion exchange chromatogram of WT *B. theta* total OMPs separated on ResourceQ after extraction using LDAO. The peaks labelled A and B were further purified by gel filtration **B:** SDS-PAGE of peaks A and B. Boiled (B) and non-boiled (NB) samples. Bands highlighted by the red boxes were excised and identified using peptide mass fingerprinting. Band 1: BT1762^{SusD}, BT1763^{SusC}, BT2268^{SusC} and BT2269^{SusD}, Band 2: BT1763^{SusC} and BT2268^{SusC}, Band 3: BT1762^{SusD}, Band 4: BT2268^{SusD}, Band 5: BT2264^{SusC}, Band 6: BT2263^{SusD}. Molecular weight (MW) marker indicates protein sizes in kDa.

3.3.2. Obtaining the Structure of the BT2261-4 outer membrane complex

3.3.2.1. Purification and crystallisation of the BT2261-4 complex

Based on the initial purification and identification of outer membrane proteins directly from *B. theta*, BT2264^{SusC} and BT2263^{SusD} were targeted because the proteins were highly expressed and could be purified sufficiently for crystallisation using anion exchange (Peak B, Figure 3.4). WT *B. theta* was grown on MM containing 0.5 % (w/v) fructose (MM-Frc). To avoid overloading the column, two or three litres of culture were used for each purification by anion exchange chromatography (Methods 2.8.3 and 2.8.5). An example chromatogram is shown in Figure 3.5A. Consistent with the purification to identify OMPs from *B. theta* grown on MM-Frc, the trace shows an isolated peak eluting at very high salt concentrations containing BT2263^{SusD} and BT2264^{SusC} (highlighted by red box, Figure 3.5A). The equivalent anion exchange samples from each purification were pooled and further purified by gel filtration. The yield for purification of the complex was up to 1 mg per litre of culture. An example SDS-PAGE, Figure 3.5C, showed that the purified product contained four distinct proteins BT2261-4 from the same locus (Figure 3.5B). The SusCD homologue pair was identified along with two small outer membrane lipoproteins of unknown function BT2261^{LP} and BT2262^{LP} (16 kDa and 24 kDa, respectively) which co-purified with the target proteins BT2263^{SusD} and BT2264^{SusC}. The locus does not contain any further co-regulated proteins, such as enzymes or putative glycan binding proteins, making it impossible to predict the target ligand of the complex.

During one purification of the complex, the protein precipitated following anion exchange, therefore DM detergent was added to the pooled fractions to a final concentration of 1 %. DM was subsequently used instead of LDAO for the first gel filtration purification step of later purifications. The sample was buffer exchanged by a second gel filtration into C8E4 buffer for crystallisation screens.

The anion exchange purification, followed by size exclusion chromatography (SEC) was sufficient to provide enough pure protein for crystal screens. An initial hit in the MemGold2 screen (Molecular dimensions) contained the most promising crystals based on diffraction patterns from Diamond Light Source (DLS). The condition was 19 % PEG3350, 50 mM Tris pH 8 and 0.2 M Magnesium formate dihydrate. Several rounds of optimisation were carried out using the hanging drop vapour-diffusion method (Figure 3.5D). Two data sets were collected at the Diamond synchrotron with two different space groups, P212121 and P1. Data was collected to a resolution of 3.4 Å for a $P2_12_12_1$ space group and 2.8 Å. for a P1 space group.

The PDB already contained a structure of the lipoprotein $BT2261^{LP}$ (PDB ID 3H3I) but $BT2262^{LP}$ had not been solved. Several SusD homologue structures have also been elucidated but not $BT2263^{SusD}$ at this stage. However $BT2261^{LP}$ and $BT2263^{SusD}$,which could be feasibly solved by molecular replacement, are only one third of the overall complex (70 kDa of 202 kDa total). This made solving the overall complex by molecular replacement difficult.

A method was designed to produce BT2261-4 with selenomethionine (SeMet) from WT *B. theta* (Methods 2.8.2) based on a protocol developed for SeMet incorporation by inhibition of methionine biosynthesis (Van Duyne *et al.*, 1993). The protein was purified, screened in various crystal conditions, then optimised as before. The SeMet substituted complex produced diffracting crystals but none of the datasets (SAD and MAD) provided sufficient quality phases to solve the structure. The anomalous signal was also weak suggesting a low level of SeMet incorporation by *B. theta*.

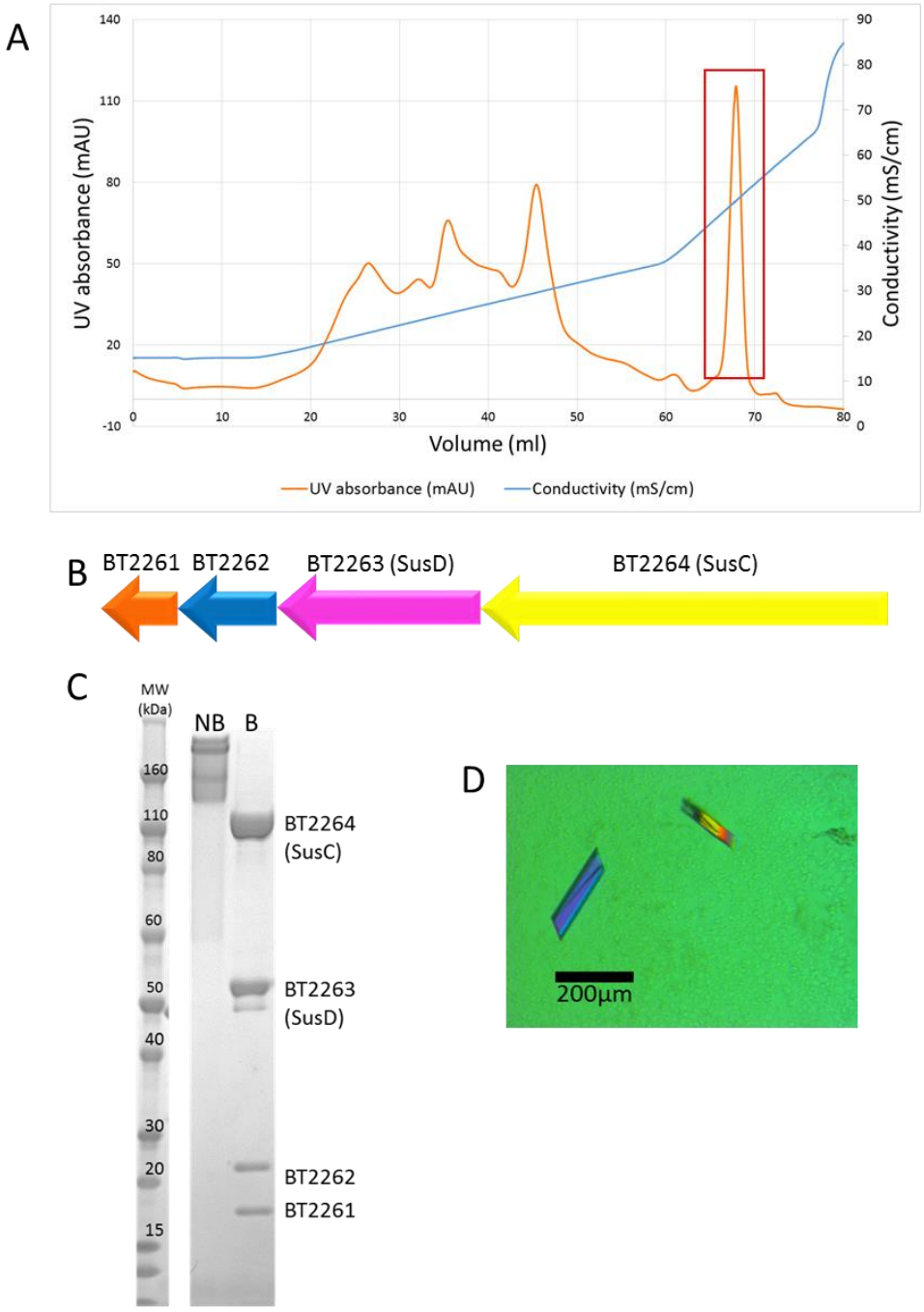


Figure 3.5 Purification of BT2261-4 from WT *B. theta.*
A: Ion exchange chromatogram of *B. theta* total outer membrane proteins separated on ResourceQ after extraction using LDAO. The highlighted peak was further purified by gel filtration **B:** Schematic diagram of BT2261-4 gene locus. Arrows are to scale and the arrow heads indicate gene orientation **C:** SDS-PAGE of the purified protein. Boiled (B) and non-boiled (NB) samples. Molecular weight (MW) marker indicates protein sizes in kDa. **D:** Optimised crystals of BT2261-4 complex.

3.3.2.2. BT2262^{LP} and BT2263^{SusD} structures

To improve the chance of finding a solution for the overall BT2261-4 structure, the individual structures of soluble BT2262^{LP} and BT2263^{SusD} were solved. Each protein was cloned, without the native Type II signal sequence or N-terminal cysteine which is used to append the protein covalently to the outer-membrane, into a pET9 vector for overexpression in the periplasm of *E. coli*. The proteins were purified by immobilised metal affinity chromatography (IMAC) and the His-tag cleaved using TEV protease (Figure 3.6A and B, E1 and +TEV). A second round IMAC was used to remove the tag, and any undigested BT2262^{LP} or BT2263^{SusD} with the cleaved protein collected in the flow through and wash (Figure 3.6A and B, FT+W). A final gel filtration step was used before the proteins were concentrated.

Several crystallisation screens (Molecular Dimensions) were set up using a Mosquito crystallisation robot (TTP Labtech). Initial hits were optimised using vapour diffusion hanging drops and crystals were harvested by Prof. Bert van den Berg. The crystals of BT2262^{LP} were produced in 2 M ammonium sulphate, 0.1 M HEPES (pH 7.5), 2 % (w/v) PEG400 (Figure 3.6C). BT2263^{SusD} crystal conditions were 0.2 M sodium acetate trihydrate, 0.1 M Tris (pH 8.5), 30 % (w/v) PEG4000 (Figure 3.6D). Data sets were collected by Dr Arnaud Baslé at the DLS.

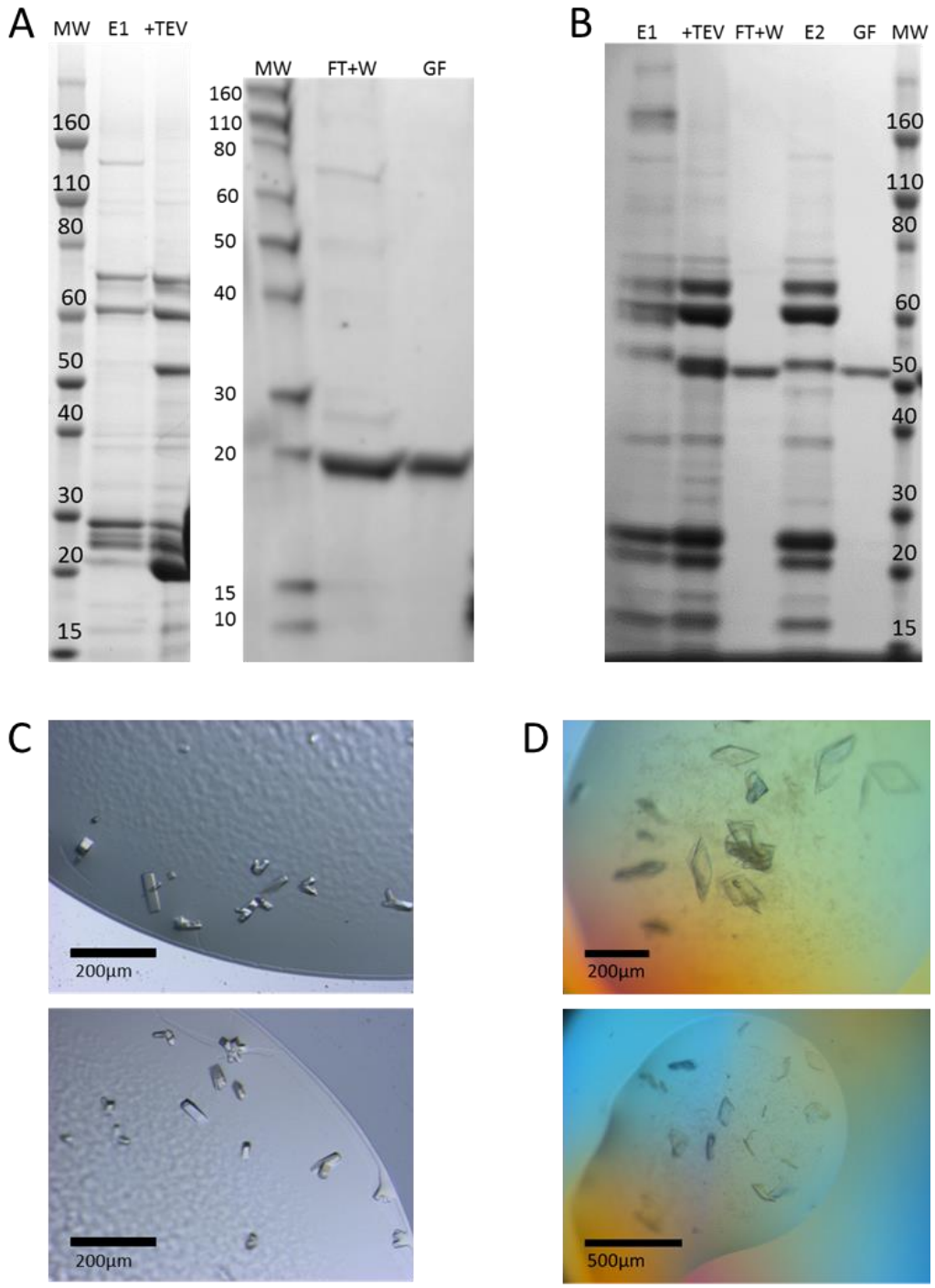


Figure 3.6 Purification and crystallisation of soluble BT2262^{LP} and BT2263^{SusD}
SDS-PAGE analysis of samples from the purification of BT2262 lipoprotein (**A**) and BT2263 SusD homologue (**B**). Initial IMAC elutions (E1) were incubated with TEV protease overnight (+TEV) to cleave the His-tag. Pure protein lacking the His-tag was collected in the flow through and wash (FT+W) of a second IMAC purification. Gel filtration (GF) was used as the final purification step. **C:** Crystals of BT2262 lipoprotein produced in 2 M ammonium sulphate, 0.1 M HEPES (pH 7.5), 2 % (w/v) PEG400 **D:** Crystals of BT2263 SusD homologue produced in 0.2 M sodium acetate trihydrate, 0.1 M Tris (pH 8.5), 30 % (w/v) PEG4000

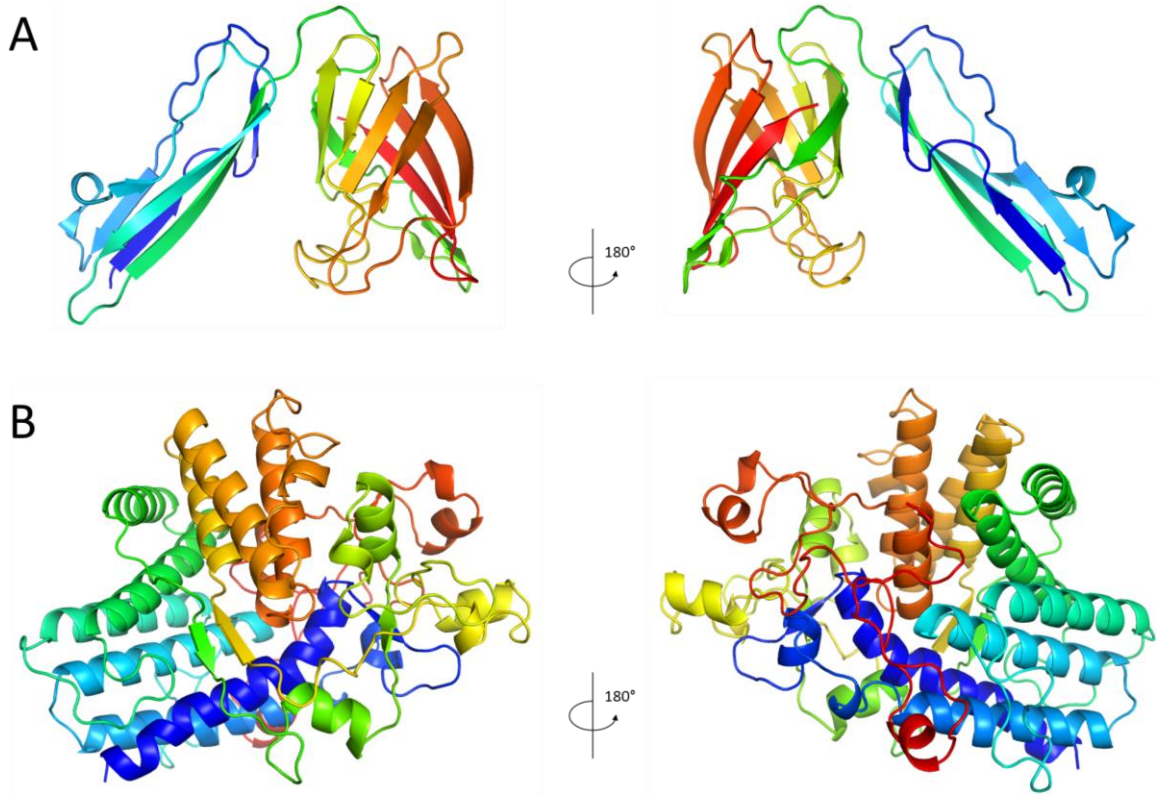


Figure 3.7 Structures of BT2262^{LP} and BT2263^{SusD}
X-ray crystal structures of BT2262^{LP} PDB ID: 5FQ3 (**A**) and BT2263^{SusD} PDB ID: 5FQ4 (**B**) coloured rainbow from blue (N-terminus to red (C-terminal). The structures have been rotated 180° around the vertical axis between the two images.

The NT domain from the BT2261-64 complex structure of the BT2262^{LP} was built manually by Bert van den Berg and consequently used as a MR model to solve the full length recombinant protein structure (Figure 3.7A). BT2262^{LP} is 24 kDa overall and consists of two domains, an N-terminal β-strand domain and a C-terminal eight stranded β-barrel.

BT2263^{SusD} was solved by MR using BT2259 ^{SusD} as a search model (PDB ID: 4Q69 26 % identity to BT2263^{SusD}) to a resolution of 1.9 Å. The structure in Figure 3.7B shows the eight conserved helices forming the four tetratricopeptide repeats (TPRs) typical of SusD homologues. The remaining loops and helices are highly variable as with the other homologue structures.

3.3.2.3. BT2261-4 Complex Structure

Without anomalous data or a suitable molecular replacement model for BT2264^{SusC} solving the overall complex was extremely challenging. The final BT2261-4 structure solved by Bert van den Berg required several rounds of manual building for BT2264^{SusC} and BT2262^{LP}.

MR trials using existing structures within the PDB were unsuccessful; FepA (PDB ID 1FEP 14 % identity to BT2264^{SusC}), BT2259 ^{SusD} (PDB ID 4Q69 26 % identity to BT2263 ^{SusD}) and BT2261^{LP} (PDB ID 3H3I). FepA is a TBDT from *E. coli* which transports iron binding siderophores and the structure was solved in 1999 (Buchanan *et al.*, 1999). Phaser was used to find a definite MR solution for a P1 crystal of the BT2261-4 complex using BT2263 ^{SusD} (Figure 3.7B), and sculptor-modified FepA as search models (Adams *et al.*, 2010). Using the BT2261^{LP} structure, manual building of BT2264 ^{SusC} and the N-terminal of BT2262 ^{LP} along with the 2-fold non-crystallographic symmetry (NCS) the structure was completed with R_{free} ~28 % (Figure 3.8A and B).

The BT2264 ^{SusC} (Figure 3.8C) has a typical, although larger than usual, TBDT structure; a 22-stranded β-barrel and N-terminal plug domain. Two BT2264 ^{SusC} proteins form a dimer and the BT2263 ^{SusD} proteins form an extracellular lid over each barrel to produce the overall complex structure (Figure 3.8A). The dimerization did not appear to be a crystallographic artefact as it was observed in three different crystal forms and was later confirmed by mass spectrometry (carried out by Dror Chorev, Oxford University).

The SusCD pair had a large interface surface area of approximately 3,800 Å^2 which included fifty hydrogen bonds and three salt bridges. BT2263 ^{SusD} did not undergo any conformational changes upon binding to BT2264 ^{SusC} as the recombinant structure of BT2263 ^{SusD} aligned with the complex structure (Figure 3.9A). The interface between the C and D of each pair included the highly variable substrate binding region of SusDs rather than the four conserved tetratricopeptide (TPR) repeats that had been predicted to be the site of interaction (Figure 3.9).

BT2261^{LP} and BT2262^{LP} are small 8-stranded β-barrel lipoproteins located on the peripheral of the complex (Figure 3.8A and B). There is only one poorly ordered copy of BT2262 ^{LP} visible within the structure but there are two BT2261^{LP} copies. BT2261 ^{LP}, BT2262 ^{LP} and BT2263^{SusD} have N-terminal linker segments, leading to their lipid anchors, which remain closely associated with BT2264^{SusC}.

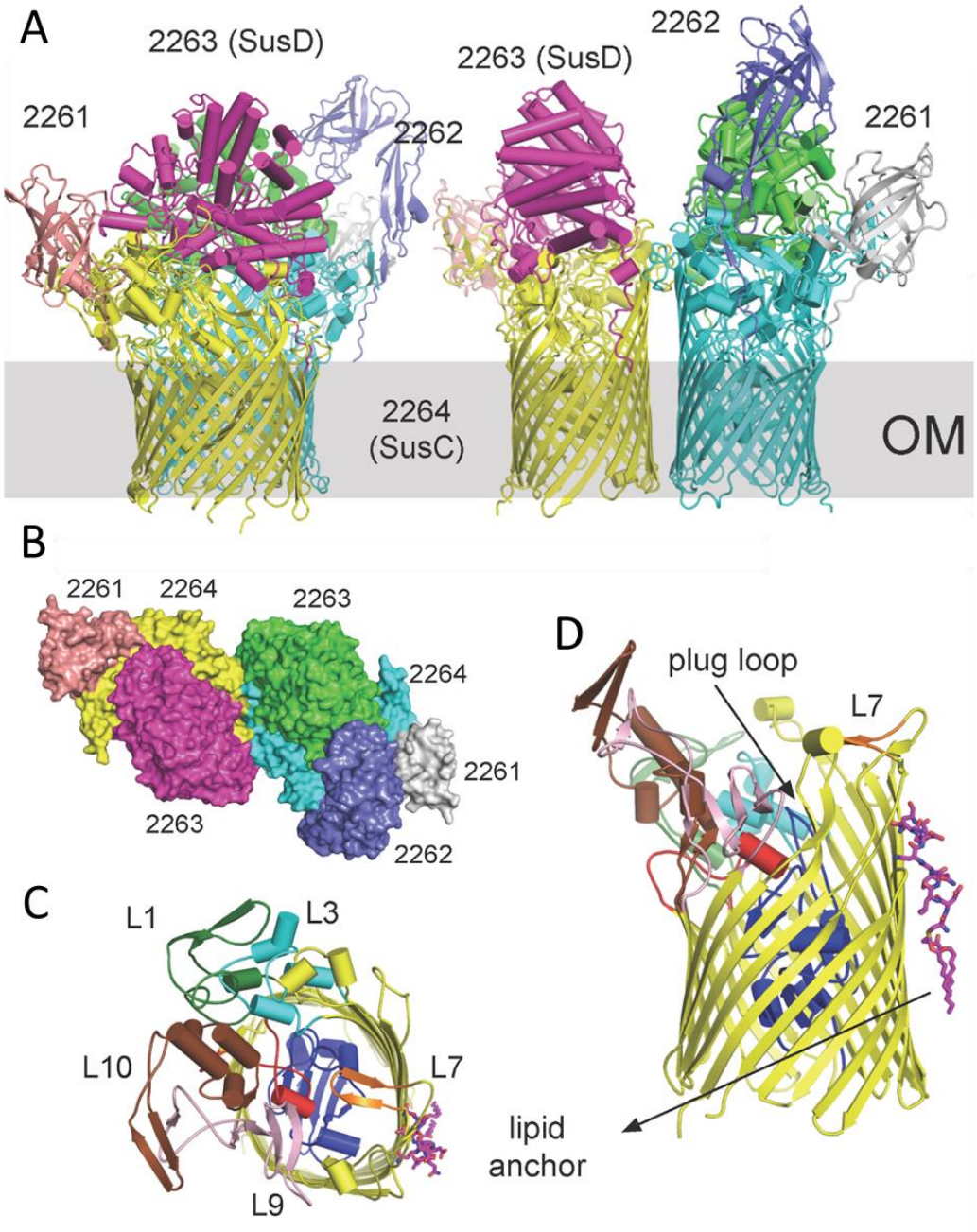


Figure 3.8 Overall architecture of the BT2261-4 complex.
A: Cartoon view of the complex dimer in the outer membrane (OM) PDB ID: 5FQ8 **B:** Surface representation of the complex looking down from outside of the cell **C:** BT2264 ^{SusC} architecture highlighting the plug domain (dark blue) and large extracellular loops; L1 (green), L3 (cyan), L7 (orange), L9 (pale pink), L10 (brown) and L11 (red). The N-terminal 10 residues and lipid anchor of BT2263 ^{SusD} are shown as magenta stick models **D:** The structure of the BT2264 ^{SusC} monomer as in **C** shown from the side view (Modified from Glenwright *et al*. 2017).

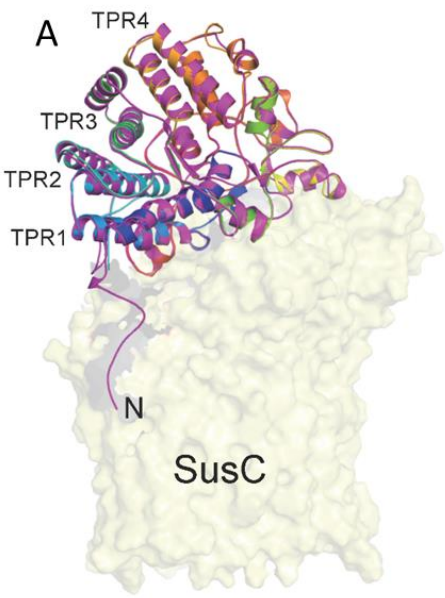
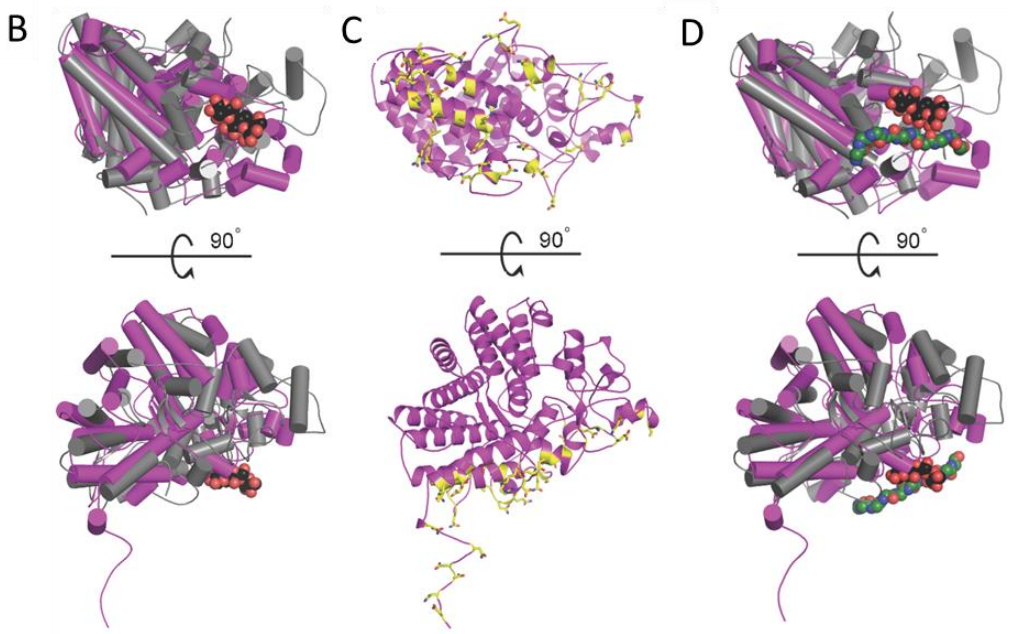


Figure 3.9 The SusC-D interaction involves the ligand-binding face of SusD

A: Cartoon overlay of BT2263 ^{SusD} in isolation coloured rainbow from blue (N-terminal) to red (C-terminal) on BT2263 ^{SusD} within the BT2261-4 complex (magenta). The four TPRs are labelled and BT2264^{SusC} is shown as a translucent surface **B:** Superposition of BT2263 ^{SusD} (magenta) with the archetypal SusD BT3701 (grey, PDB ID: 3CKB) viewed from the bottom (top image) and side (bottom image - as in **A**). Bound maltotriose in BT3701 shown as a space-filling model (carbon black, oxygen red) **C:** BT2263 ^{SusD} with residues forming hydrogen bonds and salt bridges with BT2264 ^{SusC} coloured yellow **D:** Superposition as in **B** with the putative peptide ligand of BT2261-4 included (dark green), showing binding site overlap. Figure from Glenwright *et al.* 2017.



The interface of BT2264^{SusC} and BT2263^{SusD} contained a well-defined stretch of electron density approximately 30 Å in length (Figure 3.10). The density cannot be accounted for by any components of the media or chemicals used for protein purification and crystallisation. The density was modelled well by a ten residue peptide although the sequence is ambiguous at 2.75 Å resolution (Figure 3.10). The weak density for several side chains suggests a heterogeneous group of peptides rather than one specific sequence was bound by the complex. The peptide has been modelled by ten glycine residues in Figure 3.10. The ligand survived several purification steps including the use of strong detergent which indicates the peptide is tightly bound and most likely originated from *B. theta.* The ligand was bound at the BT2264^{SusC} and BT2263^{SusD} interface in a cavity which was not solvent accessible (Figure 3.10A). Nine out of ten residues of the putative peptide ligand backbone interact via hydrogen bonds with either BT2264^{SusC} or BT2263^{SusD} (Figure 3.10B). Interactions with BT2264^{SusC} include the hinge loop (L7 orange) and the plug loop (dark blue).

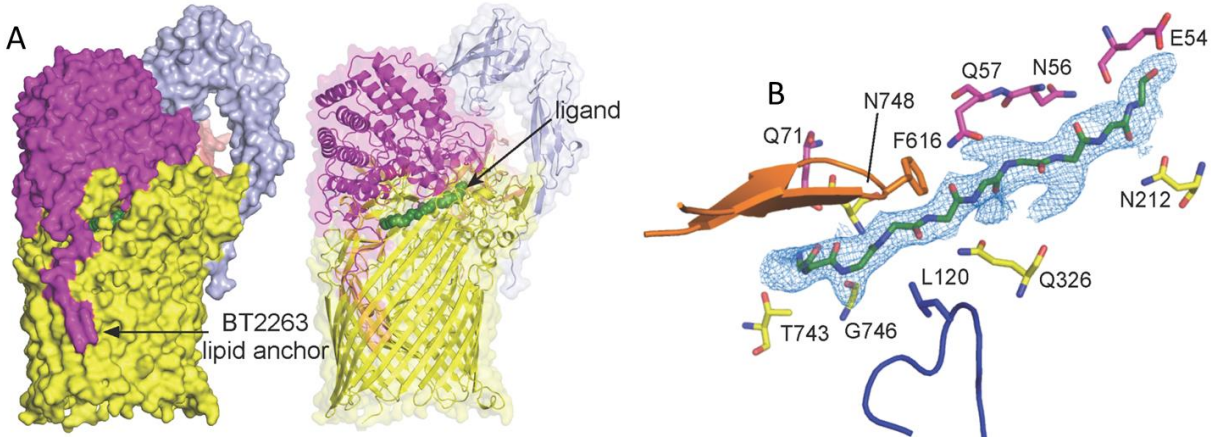


Figure 3.10 Ligand bound to the BT2261-4 complex.
A: BT2261-4 (side view) showing BT2264 ^{SusC} (yellow), BT2263 ^{SusD} (magenta), BT2262^{LP} (lilac) and the putative peptide ligand in green. **B:** Close-up of the ligand binding site showing BT2264 ^{SusC} (yellow) and BT2263 ^{SusD} (magenta) residues forming hydrogen bonds with the peptide backbone. Loop L7 (orange) of BT2264 ^{SusC} and the plug loop (dark blue) also interact. The 2Fo − Fc density for deca-glycine is shown as a blue mesh at 1.5σ (From Glenwright *et al.* 2017).

3.3.3. Investigating BT2261-4 regulation and function

Identification of the ligand was unsuccessful by native mass spectrometry (carried out by Dror Chorev, Oxford University) which supports an ensemble of bound peptides. Several other methods were used to attempt to shed light on the identity of the bound molecule.

3.3.3.1. Attempting to identify the ligand of BT2261-4

The BT2261-64 complex was boiled for 5 minutes and analysed using SDS-PAGE (BOLT Bis-Tris 12 % acrylamide, ThermoFisher) with the aim of separating a low MW band (1-2 kDa) for identification of the peptide by mass spectrometry (Figure 3.11). However there were no obvious protein bands small enough to be the ligand.

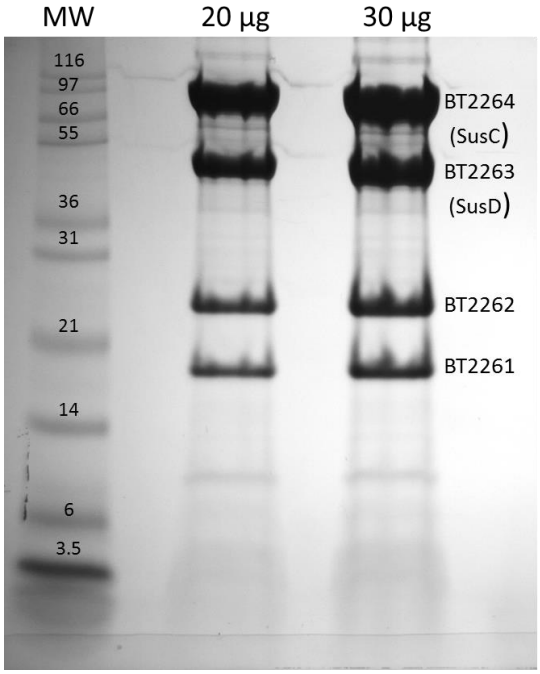
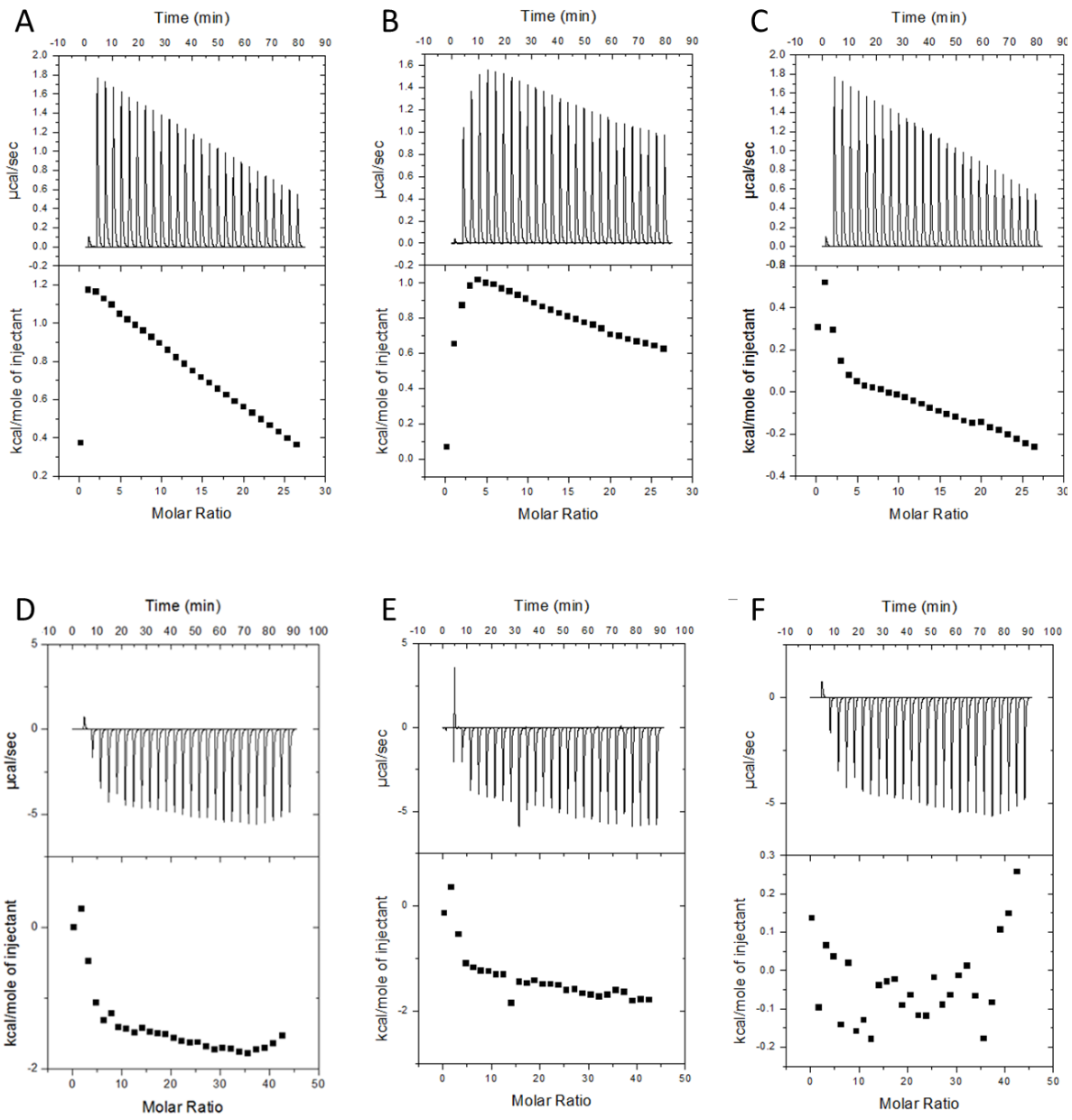


Figure 3.11 SDS-PAGE gel of BT2261-4 complex.
20 µg and 30 µg of the total BT2261-4 complex analysed, after boiling, using a 12 % acrylamide gel (BOLT 12 % Bis-Tris gel with MES Buffer, ThermoFisher).

To further explore the capacity of BT2263^{SusD} to bind peptides isothermal titration calorimetry (ITC) was utilised. This method has been routinely used to measure the binding of proteins, such as SusD homologues, to the target glycans or oligosaccharides. BT2263^{SusD} was cloned without the NT Type II signal sequence into pET28a with an N-terminal His_{x6}-tag and expressed in BL21 *E. coli* cells. Following purification by IMAC, binding of the protein to synthetic peptides was investigated using ITC. The peptides, GSSGGQNEGG and GDSGSKQKKG, were designed according to best fit for the density found in the complex structure (Figure 3.10B). However, BT2263^{SusD}, the respective SusD homologue of the transport complex, showed no affinity for either peptide tested (Figure 3.12).



3.12 BT2263^{SusD} binding affinity to synthetic peptide ligands.
A: 37 µM BT2263^{SusD} against 5mM GSSGGQNEGG **B:** Peptide only control **C:** BT2263^{SusD} against 5mM GSSGGQNEGG minus control data. **D:** 46 µM BT2263^{SusD} against 10mM GDSGSKQKKG **E:** Peptide only control **F:** BT2263^{SusD} against 5mM GDSGSKQKKG minus control data. The top half of each panel shows raw ITC heats of titration with ligand while the bottom half shows integrated peak areas. ITC was carried out in 20 mM HEPES pH 7.5 at 25 °C (Methods 2.12.2).

3.3.3.2. Using BT2264^{SusC} knockout to find protein function

Deleting or disrupting a gene to prevent expression of a protein of interest can provide clues to the function of the protein if there is a growth defect observed. Therefore a knockout of BT2264^{SusC} was produced using pExchange-*tdk*. The gene was removed by 'sewing' together 1000 bp upstream and downstream regions of the genome (Methods 2.5.4.3, Figure 2.1). The resulting hybrid PCR product was then inserted into pExchange-*tdk* which is used to produce *B. theta* mutations by homologous recombination (Methods 2.6.1, Figure 2.2). The *Δbt2264* strain growth on TYG and minimal media with 0.5 % fructose (MM-Frc) was compared with WT *B. theta* (Figure 3.13). There does not appear to be any growth phenotype associated with the deletion of BT2264^{SusC}.

The BT2261-4 complex was highly expressed and subsequently purified from cells grown on MM-Frc however BT2264^{SusC} is not essential under such conditions. This may be due to redundancy in peptide importers as there are 103 SusC homologues expressed by *B. theta* over 80 % of which have not been characterised. Other bacteria have been shown to express multiple peptide transporters, for example, *E. coli* has three distinct peptide permeases with varying specificities; oligopeptide (Opp), tripeptide (Tpp) and dipeptide (Dpp) (Goodell and Higgins, 1987).

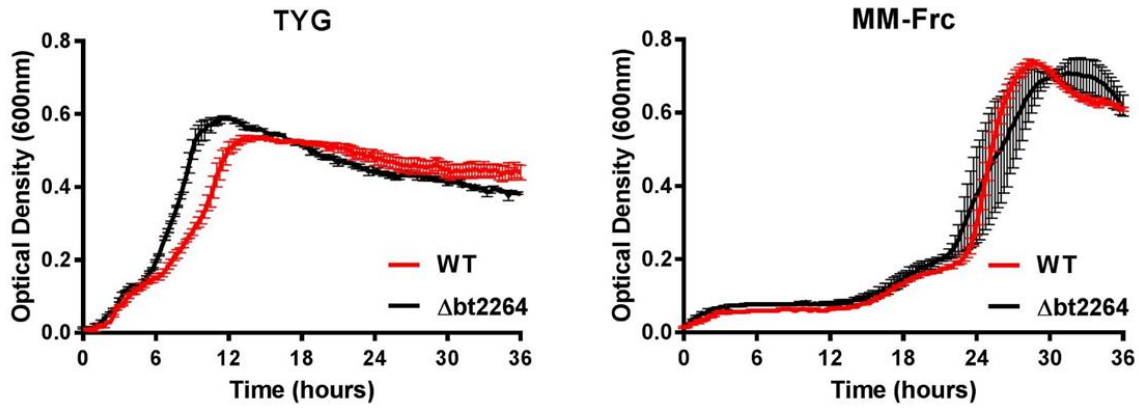
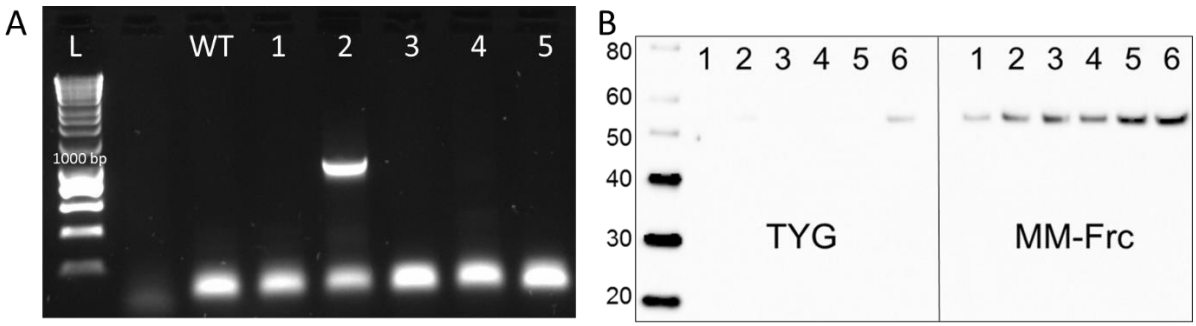


Figure 3.13 Growth curves of WT and Δbt2264 *B. theta*.
Cells grown on TYG rich media and minimal media containing 0.5 % fructose (MM-Frc). Growth was measured by optical density (A_{600nm}) using an automatic plate reader under anaerobic conditions at 37 °C. Triplicate wells have been averaged and the error bars represent the standard deviation from the mean.

3.3.3.3. Investigating upregulation of the BT2261-4 complex

Clarifying the growth stage and conditions when a protein is highly expressed can provide an indication of the function. For this purpose, a FLAG-tag (DYKDDDDK) was added using pExchange-*tdk* to the C-terminus of BT2263^{SusD} (Figure 3.14A) to allow for western blot analysis of the expression of the complex at different stages of growth on TYG and MM-Frc.

BT2261-4 is constitutively expressed during log-phase on MM-Frc and only upregulated on TYG rich media during stationary phase as cell viability falls (Figure 3.14B and C). This expression pattern is in stark contrast to known glycan uptake systems which are greatly upregulated in response to the target glycan (Martens *et al.*, 2011). However the results are supported by previous in vitro *B. theta* data showing upregulation of the BT2261-4 locus under minimal media conditions (with glucose or maltotriose) and downregulation in TYG (Sonnenburg *et al.*, 2005). The expression levels of BT2261-4 suggested the complex is required for growth under nutrient stress conditions.



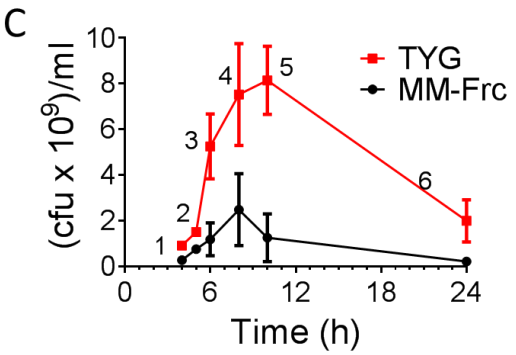


Figure 3.14 Western blot analysis of BT2261-4 complex expression using FLAG-tagged BT2263^{SusD}.
A: Diagnostic PCR to test for successful addition of the FLAG-tag. WT genomic DNA was used as a negative control. L represents the DNA ladder. A product of 1000 bp indicates insertion of the tag as shown in lane 2. **B:** Western blot of samples across the growth curve from cells grown on TYG rich media and minimal media containing 0.5 % fructose (MM-Frc). Lanes 1–6 represent gel samples from the corresponding cfu time points shown in **C**. Corresponding colony-forming unit (cfu) values (n = 3, average ± standard deviation). (Adapted from Glenwright *et al.* 2017).

3.4. Discussion

3.4.1. *Bacteroides thetaiotaomicron* outer membrane proteins

B. theta lack the typical abundant porins and alternatively express very large outer membrane proteins (Figure 3.2) suggesting *Bactroides spp*. use a fundamentally different method for nutrient acquisition than well studied Gram-negative bacteria such as *E. coli*. The genome of *B. theta* encodes 103 predicted SusC homologues of which only ~20 have been characterised and all target specific glycans (Sonnenburg *et al.*, 2010; Martens *et al.*, 2011; Cuskin *et al.*, 2015; Cartmell *et al.*, 2017; Ndeh *et al.*, 2017). Therefore, with so many uncharacterised SusC homologues it is possible *B. theta* could import all of the required nutrients via SusCD pairs. However, this is unlikely for small molecules, such as monosaccharides, which are probably imported passively rather than by TBDTs.

3.4.2. BT2261-4 outer membrane complex

Successful native purification of a highly expressed (~1 mg per 1 l of *B. theta*) outer membrane protein and subsequent structure was incredibly surprising. The high stability of the BT2261-4 complex, even in strong detergent, allowed for many rounds of purification and successful crystallisation. The presence of a ligand almost certainly facilitated the crystallisation of the complex by locking $BT2263^{SusD}$ to $BT2264^{SusC}$ in a closed, less flexible conformation.

The $BT2264^{SusC}$ structure is the first of a SusC homologue and is a large TBDT; 22 β-stranded barrel with an N-terminal plug domain folded into the channel (Figure 3.8). The dimeric structure of $BT2264^{SusC}$ was unexpected and remains unique among TBDTs (Noinaj *et al.*, 2010). The relative positioning of $BT2263^{SusD}$ and contact at the ligand binding site was also surprising as SusCD homologues were predicted to interact via the four conserved TPRs (Bolam and Koropatkin, 2012). TPR domains are commonly associated with protein-protein interactions (Allan and Ratajczak, 2011) therefore may be important for SusD homologues to interact with other proteins such as surface glycan binding proteins (SGBPs) or extracellular enzymes.

3.4.3. BT2261-4 upregulation, ligand and function

BT2261-4 is expected to be a peptide importer based largely upon the electron density of the ligand (Figure 3.10B) and the holo-complex MD simulations using a modelled peptide (Figure 3.15). Binding studies of BT2263^{SusD} against two synthetic ten residue peptides were unsuccessful (Figure 3.12) which may be due to the complex requiring a specific sequence to some extent. Ambiguity in the density of the ligand was most likely a result of different amino acid sequences bound in the crystal. The polymer backbone was clearly well modelled by a peptide, however, which implies polyspecific ligand binding of this transporter complex, a trait that would be advantageous given the assumed function of the protein as a peptide scavenger during nutrient scarcity. The oligopeptide permease (Opp), an ATP-binding cassette transporter from *E. coli*, requires a substrate binding protein OppA to bind and deliver peptides to the transport machinery. OppA binds peptides of any sequence, via interactions with the peptide backbone (Guyer *et al.*, 1986; Tame *et al.*, 1995). Structures of OppA with peptide bound show the ligand is completely enclosed by the protein, as with the binding interface between BT2263^{SusD} and BT2264^{SusC}, accommodating the various peptide side chains in large cavities within the binding protein (Tame *et al.*, 1994).

Another possible reason for lack of binding observed by BT2263^{SusD} to the synthetic peptides is that the protein in complex forms less than half of the binding site. Only 4 residues (E54, Q57, Q56, Q71) from BT2263^{SusD} interact with the peptide backbone and the remainder of the binding pocket consists of 7 residues from BT2264^{SusC} involving the hinge loop (L7) and a loop from the plug domain (Figure 3.10B). Conceivably BT2263^{SusD} in isolation cannot bind the ligand because it does not form enough interactions with the peptide backbone and requires BT2264^{SusC}.

The knockout strain *Δbt2264*^{SusC} shows no growth phenotype on MM-Frc (Figure 3.13), this is most likely due to other transporters targeting peptides. As previously discussed *B. theta* expresses many SusCD homologues which have not been characterised so there is likely to be redundancy between the nutrient acquisition systems. Two SusCD homologues BT3238-9 and BT3240-1 are

found in operons alongside predicted peptidases BT3237 and BT3242, respectively, suggesting the transporters also target peptides. Interestingly, BT3238-9^{SusCD} were isolated from *B. theta* grown on MM-Frc during a purification of BT2261-4 implying this complex could have a similar function. BT3238-9^{SusCD} could not be purified sufficiently for crystallisation.

The BT2261-4 complex is constitutively expressed during log-phase on minimal media and upregulated as cell viability drops (Figure 3.14). This is in contrast to glycan PULs which are selectively upregulated in response to the target polysaccharide (Martens *et al.*, 2011). This lends evidence to BT2261-4 having a role in peptide scavenging. During stationary phase toxic waste products begin to accumulate and nutrient availability decreases, potentially causing cells to lyse, releasing their contents. This includes proteins and peptides which Bacteroidetes can use as a carbon source or an amino acid source (Mysak *et al.*, 2014). Transporter complexes BT2261-4 may provide an advantage to cells to survive these times of stress, the importance of which is underpinned by the apparent redundancy of these systems (Figure 3.13). The scavenging function of BT2261-4 during nutrient stress is supported by whole-genome transcriptional profiling which showed upregulation of the locus in *B. theta* from mice fed a simple sugar diet as opposed to a polysaccharide rich diet (Sonnenburg *et al.*, 2005).

3.4.4. Molecular Dynamics

The ligand-binding interface between BT2264^{SusC} and BT2263^{SusD} was not solvent accessible (Figure 3.10A) therefore BT2263^{SusD} must dissociate from the transporter to capture the substrate. To investigate this hypothesis the structure was used in unbiased molecular dynamics (MD) simulations carried out by Karunakar Pothula at Jacobs University (Bremen, Germany). Simulations were carried out in the presence and absence of the peptide (holo and apo-complexes, respectively) on the central dimer BT2263-4 (SusCD pair) and the BT2261-4 complex (tetramer and (BT2261–64) × 2 octamer). For simplicity the MD simulation snapshots presented here only include the dimer BT2263-4 (Figure 3.15A and B).

During the majority of apo-complex simulations BT2263^{SusD} moved in a hinge-like motion away from BT2264^{SusC} to expose the ligand binding site (Figure 3.15B). The extent of opening varied between individual simulations but in most open states observed BT2263^{SusD} rotates 40-45°. This movement caused all interactions with loops L1, L3-5 and L9-11 of BT2264^{SusC} to be lost. The Cα-Cα distances between BT2264^{SusC} Asn203 and BT2263^{SusD} Thr296 increased from 6 Å in the closed ligand-bound complex to approximately 40 Å in the open ligand-free state (Figure 3.15C).

BT2264^{SusC} loop L7 moves with BT2263^{SusD} in the MD simulations and was required for most of the remaining BT2264^{SusC}-BT2263^{SusD} interactions in the open state (Figure 3.15B). This loop also underwent the biggest conformational change between the apo and holo complex simulations with the hinge point at the base of L7 near the lipid anchor of BT2263^{SusD} (Figure 3.15B), as such the BT2264^{SusC} loop L7 will be referred to as the hinge loop.

All of the holo-complexes were stable for the duration of the simulations although the conformation of the peptide varied substantially between simulations while remaining at the BT2264^{SusC}-BT2263^{SusD} interface. This is unexpected because the electron density of the ligand was well defined in the crystal structure (Figure 3.10B). The most likely explanation for this inconsistency is that the specific peptide sequence used for the simulations is not stable within the binding pocket.

This would imply that the BT2261-4 binding and import system is sequence specific to some extent. MD simulations of the full BT2261–64 octamer show the dimeric structure of BT2264^{SusC} does not obstruct opening of the BT2263^{SusD} 'lid'. Each monomer appears to be able to open and close independently of the other.

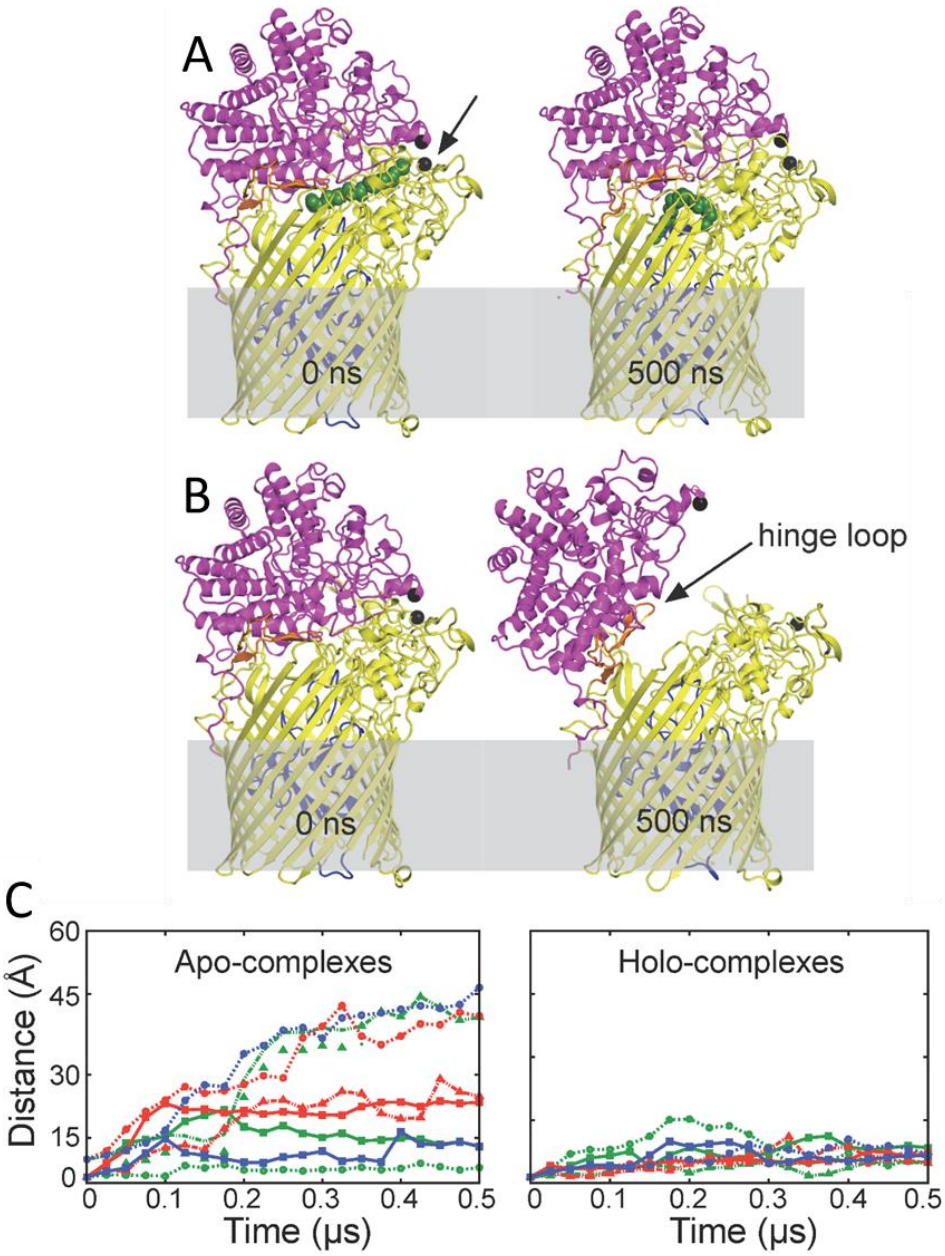


Figure 3.15 Molecular dynamics simulations of BT2263-4. Molecular dynamics simulation snapshots of holo-BT2263-4 (peptide bound) (**A**) and apo-BT2263-4 (**B**) after 0 ns and 500 ns. BT2264^{SusC} is yellow, the plug domain is dark blue, the hinge loop is orange and BT2263^{SusD} is magenta. The peptide ligand is shown as green spheres. BT2264^{SusC} Asn203 and BT2263^{SusD} Thr296 Cα atoms are shown as black spheres. **C:** Average distances versus simulation time between BT2264^{SusC} Asn203 and BT2263^{SusD} Thr296 for dimers (green), BT2261–64 tetramers (red), and the (BT2261–64) × 2 octamer (blue, from Glenwright *et al.* 2017).

The MD simulations show a possible mechanism for peptides to become bound by the complex (Figure 3.15). In the absence of ligand BT2263^{SusD} opens like a lid from BT2264^{SusC} exposing the binding site. A loop from the plug domain of BT2264^{SusC} interacts with the peptide via Leu120, potentially this could signal to the TonB complex that the binding site is occupied, which then leads to TonB induced transport of the peptide into the periplasm (Hickman *et al.*, 2017). The system would then require energy to reset back to the original dynamic opening state. Therefore the energy from the ExbBD-TonB system is equivalent to stepping on the pedal of the SusCD bin (Figure 3.16).

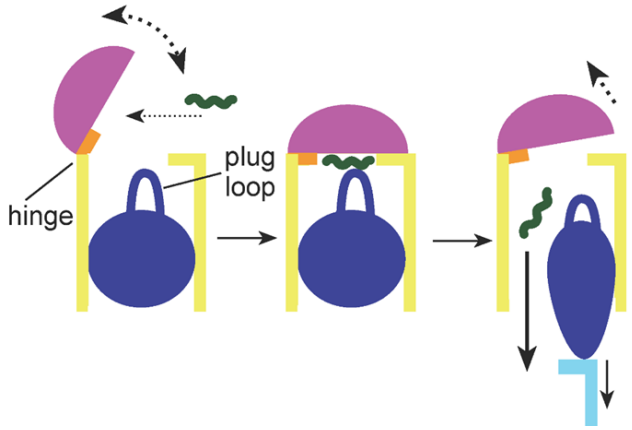


Figure 3.16 'Pedal bin' mechanism for nutrient import by SusCD homologues. SusD is coloured magenta, SusC β-barrel is yellow, hinge loop is orange and plug domain is dark blue. The ligand is shown as a wavy green line and the CT domain of TonB which interacts with the SusC is coloured cyan.

Chapter 4 - Acquisition of Levan polysaccharides by *Bacteroides thetaiotaomicron*

4.1 Introduction

The solved BT2261-4 complex is not a classical glycan importer, therefore it is not clear if the overall SusCD configuration is relevant for other homologues. To explore this, a well characterised easily upregulated PUL was targeted.

Levan is a β 2,6-linked fructan polysaccharide, which is found in plants and the capsule of many bacteria. *B. theta* expresses a PUL (BT1754-65) encoding all of the proteins required to use levan as a carbon source (Bolam and Sonnenburg, 2011). Of these proteins, four are extracellular, producing levan oligosaccharides and transporting these across the outer membrane (Figure 4.1). Levan polysaccharides are bound by the surface glycan binding protein (SGBP) BT1761 and broken into oligosaccharides, by the endo-acting GH32 levanase enzyme BT1760 (Sonnenburg *et al.*, 2010). The oligosaccharide products of the enzyme are then bound and imported into the periplasm by the SusD homologue BT1762 and the SusC homologue BT1763, respectively.

$BT1760^{GH32}$ and $BT1762^{SusD}$ have been previously characterised and both structures solved (Zheng, 2009; Sonnenburg *et al.*, 2010; Mardo *et al.*, 2017). The endo-acting enzyme, $BT1760^{GH32}$, is a β2,6-fructan (levan) specific cell surface lipoprotein and essential for *B. theta* growth on levan (Sonnenburg *et al.*, 2010). $BT1762^{SusD}$ binds levan, and a $\Delta BT1762^{SusD}$ strain lacking the levan oligosaccharide binding SusD homologue has a major growth defect on levan (Sonnenburg *et al.*, 2010). The levan SGBP BT1761 has been shown to bind levan (Sonnenburg *et al.*, 2010), however the affinity of the interaction has not been investigated. Currently there are only three SGBP structures in the PDB; SusE from the starch system, BT4661 which targets heparin and BACOVA_02650 from *B. ovatus* which binds xyloglucan (Cameron *et al.*, 2012; Tauzin *et al.*, 2016; Cartmell *et al.*, 2017). Therefore characterising $BT1761^{SGBP}$ and obtaining the structure could be very informative with regards to glycan acquisition at the cell surface.

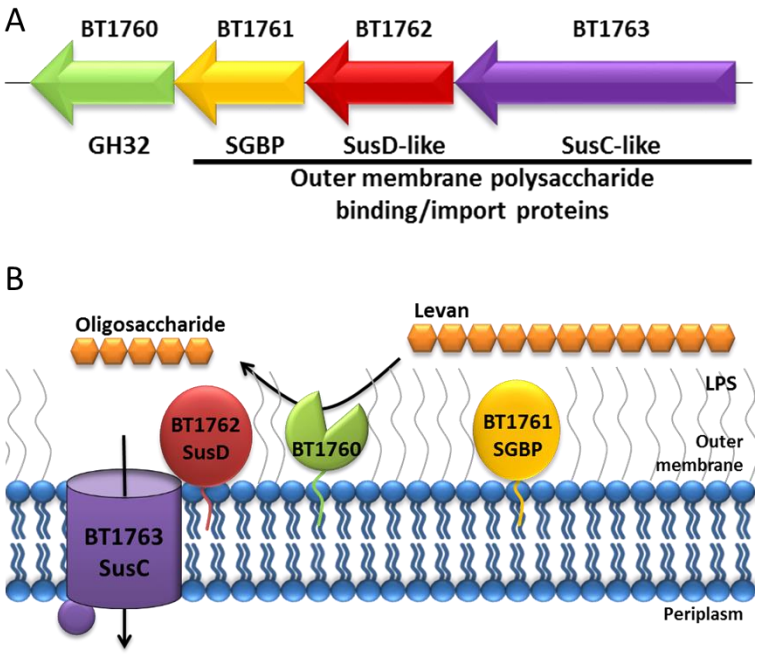


Figure 4.1 Levan acquisition across the outer membrane of *B. theta.*
A: Genetic organisation of outer membrane genes of the levan PUL; arrows shown to scale with the direction of the arrow indicating gene orientation. **B:** Model of levan degradation and transport at the cell surface. Proteins coloured as with genes in **A**. Levan is bound by the SGBP BT1761 and partially digested into oligosaccharides by BT1760, an extracellular GH32 (glycoside hydrolase family 32) lipoprotein. The oligosaccharides are then bound and imported into the periplasm by the SusCD homologue BT1762-3. The levan is further digested into fructose in the periplasm and upregulates expression of the PUL by interacting with BT1754, a hybrid two-component system (HTCS) in the IM.

As with other PULs from the Bacteroidetes phylum, the SusC homologue BT1763 has never been studied due to the relative difficulty of working with integral membrane proteins. How binding of levan oligosaccharide by BT1762^{SusD} is coupled to import via the BT1763^{SusC} is also unknown. BT1762-3 is an ideal target for studying SusCD homologues within the context of glycan utilisation, as the PUL has been characterised and the locus is upregulated by fructose which is cheap and readily available.

4.2 Aims

The main aim of this chapter was to purify, crystallise and solve the structure of the SusCD complex BT1762-3 from *B. thetaiotaomicron.* The second objective was to express, purify, crystallise and solve the structure of the surface glycan binding protein BT1761 and investigate binding to levan. The final aim of this chapter was to solve the structure for the full outer membrane complex involved in levan acquisition BT1760-3.

4.3 Results

4.3.1 BT1762-3 SusCD complex

4.3.1.1 Addition of a His_6-tag to $BT1762^{SusD}$

As described in Chapter 3 (section 3.3.1.2), the levan targeting SusCD homologues BT1762-3 were purified from WT *B. theta* cells grown on MM-Frc. However, the proteins co-purified with another SusCD-like pair and could not be isolated by ion exchange or size exclusion chromatography due to almost identical size and isoelectric points (Figure 3.4). Preliminary work on SusC homologues by Bert van den Berg was unsuccessful in expressing the proteins recombinantly in *E. coli* and currently there is not a suitable expression plasmid for *Bacteroides spp*. Expression of OM proteins within their native bacteria is beneficial because the protein is folded and inserted in the membrane in the natural environment. This was advantageous for the purification of BT2261-4 as it was possible to purify $B2263^{SusD}$ and two lipoproteins alongside $BT2264^{SusC}$ (Chapter 3).

To separate BT1762-3 from the other OMPS a His_{x6}-tag, in frame with the protein of interest, was added to BT1763 in the *B. theta* genome. Initially, a His_{x6}-tag was added to the N-terminus of the genomic copy of $BT1763^{SusC}$ after the signal peptide. However, the cells were unable to grow on full length levan polysaccharides, suggesting the addition of the tag compromised the function of the transporter. Purification of OMPs directly from *B. theta* (Chapter 3) has shown that SusC and SusD pairs interact very strongly and co-purify therefore tagging the SusD protein should also allow extraction of the cognate SusC. Therefore a His_{x6}-tag with a 4xAla linker was added to the C-terminus of the genomic copy of $BT1762^{SusD}$ of WT *B. theta.* The successful addition of a His_{x6}-tag was shown using PCR with a His_{x6} primer (Figure 4.2). The 1000bp PCR products produced from genomes 1, 2, 4, 5 and 6 indicated incorporation of the C-terminal His_{x6}-tag into the *B. theta* genome. Sequencing was used to confirm the addition of the tag was in-frame with the protein. The resulting tagged strain will henceforth be referred to as *CT-His$BT1762^{SusD}$*.

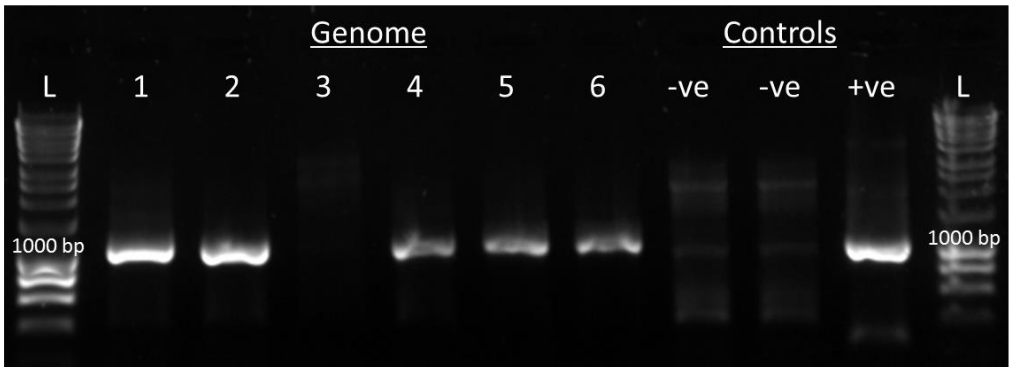


Figure 4.2 PCR analysis to test for successful addition of a His-tag to BT1762^{SusD}. PCR from mutated *B. theta* genomes using a 6xHis reverse primer (5'-GTGGTGGTGGTGGTGG-3') and a primer 1000 bp upstream of the mutation (5'- AAAGAAGATAACATTCGAGTCGACTGCCCTTACCGGAATCAAC-3'). A 1000 bp PCR product indicates successful addition of a His-tag to the C-terminus of BT1762. Negative (-ve) controls were WT *B. theta* genomes and the positive (+ve) control was pExchange containing the His-tagged BT1762 insert. The correct size of 1000 bp has been labelled on the DNA ladder (L).

4.3.1.2 Expression, purification and crystallisation of BT1762-3

The *CT-HisBT1762*^{SusD} strain of *B. theta* was grown on a minimal media containing 0.5 % fructose (MM-Frc). The addition of the His_{x6}-tag did not affect growth on MM-Frc or MM-Levan (Figure 4.14). Following lysis of the cells, the cell free extract was discarded and the membrane proteins were solubulised in LDAO overnight (Methods 2.8.3). The sample containing solubulised membrane proteins was then collected by ultracentrifugation.

The tagged BT1762^{SusD} was co-purified with BT1763^{SusC} using IMAC followed by SEC (Methods 2.8.6 and 2.8.7). As with BT2261-4 the complex was stable in 2 % SDS and separated into the component proteins upon boiling before SDS-PAGE. However, unlike with the BT2261-64 complex, the other OM lipoproteins from the PUL (BT1760 and BT1761) did not co-purify with the SusCD pair (Figure 4.3A). The isolated BT1763^{SusC} (after boiling) appears as a double band indicating there is some degradation of the protein. IMAC of *B. theta* membrane proteins is highly specific as only the tagged complex BT1762-3 is visible by SDS-PAGE. WT *B. theta* does not produce any protein bands on SDS-PAGE following membrane solubilisation and IMAC.

BT1762-3 was concentrated to 10 mg/ml and tested for potential crystallisation conditions using Molecular Dimensions MemGold1 and MemGold2 screens. Positive conditions were further optimised by hand using hanging drop vapour diffusion. Crystals were optimised by varying the precipitant concentration in 12-22 % w/v PEG 3350, 0.1 M magnesium formate dihydrate and 0.1 M MOPS, pH 7 (Figure 4.3B). The crystals were cryo-protected with 20 % glycerol, and a 3.1 Å resolution dataset was collected at Diamond Light Source (DLS). A different crystal form (Figure 4.3C) was produced in 18-20 % PEG 3350, 0.2 M ammonium nitrate and 50 mM HEPES, pH 7.5. A second dataset from this crystal form was collected to a resolution of 3.0 Å.

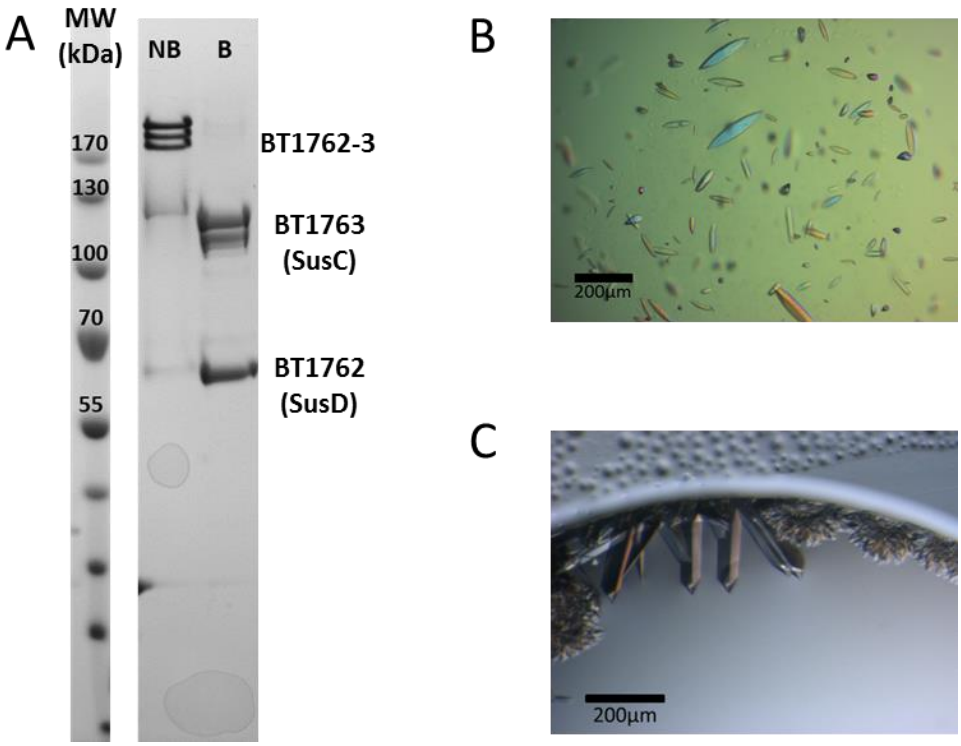


Figure 4.3 Purification and crystallisation of BT1762-3 SusCD complex
A: SDS-PAGE of BT1762-3 purified by IMAC followed by size exclusion chromatography from the CT-HisBT1762^{SusD}strain. The sample is shown boiled (B) and not boiled (NB). Molecular weight (MW) marker sizes are shown in kDa **B, C:** Optimised crystals of BT1762-3 in 0.1 M Magnesium formate dihydrate, 0.1 M MOPS (pH 7), 17% w/v PEG 3350 (**B**) and 50 mM HEPES (pH 7.0), 0.2 M ammonium nitrate, 19% (w/v) PEG 3350 (**C**).

4.3.1.3 Apo structure of BT1762-3

The crystal structure of BT1762-3 was solved by Bert van den Berg using the previously solved soluble BT1762^{SusD} structure (Zheng, 2009) and a sculptor-generated BT2264^{SusC} model (21 % identity to BT1763^{SusC}). The structure showed the same relative orientations of BT2261-4 although, as expected from the SDS-PAGE (Figure 4.3A), there were no accessory lipoproteins (Figure 4.4). BT1763^{SusC} formed a homodimer where BT1762^{SusD} acted as a 'lid' capping each transporter. As with the BT2263-4 SusCD structure, binding to BT1763^{SusC} did not require/induce a conformational change in BT1762^{SusD} and involved the binding interface of the SusD homologue rather than the conserved TPRs (Figure 4.4D). The hinge loop (L7) highlighted in BT2264^{SusC} was conserved in BT1763^{SusC} suggesting it may be essential for successful opening of the SusD 'lid' as shown in the molecular dynamics simulations (Chapter 3, section 3.3.3.5). There was no levan oligosaccharide ligand bound in the BT1762-3 structure which was to be expected due to the lack of substrate (levan) during cell growth. The BT1762^{SusD} cap is closed, like the BT2261-4 ligand bound structure, almost certainly because crystallisation favors compact, stable protein states.

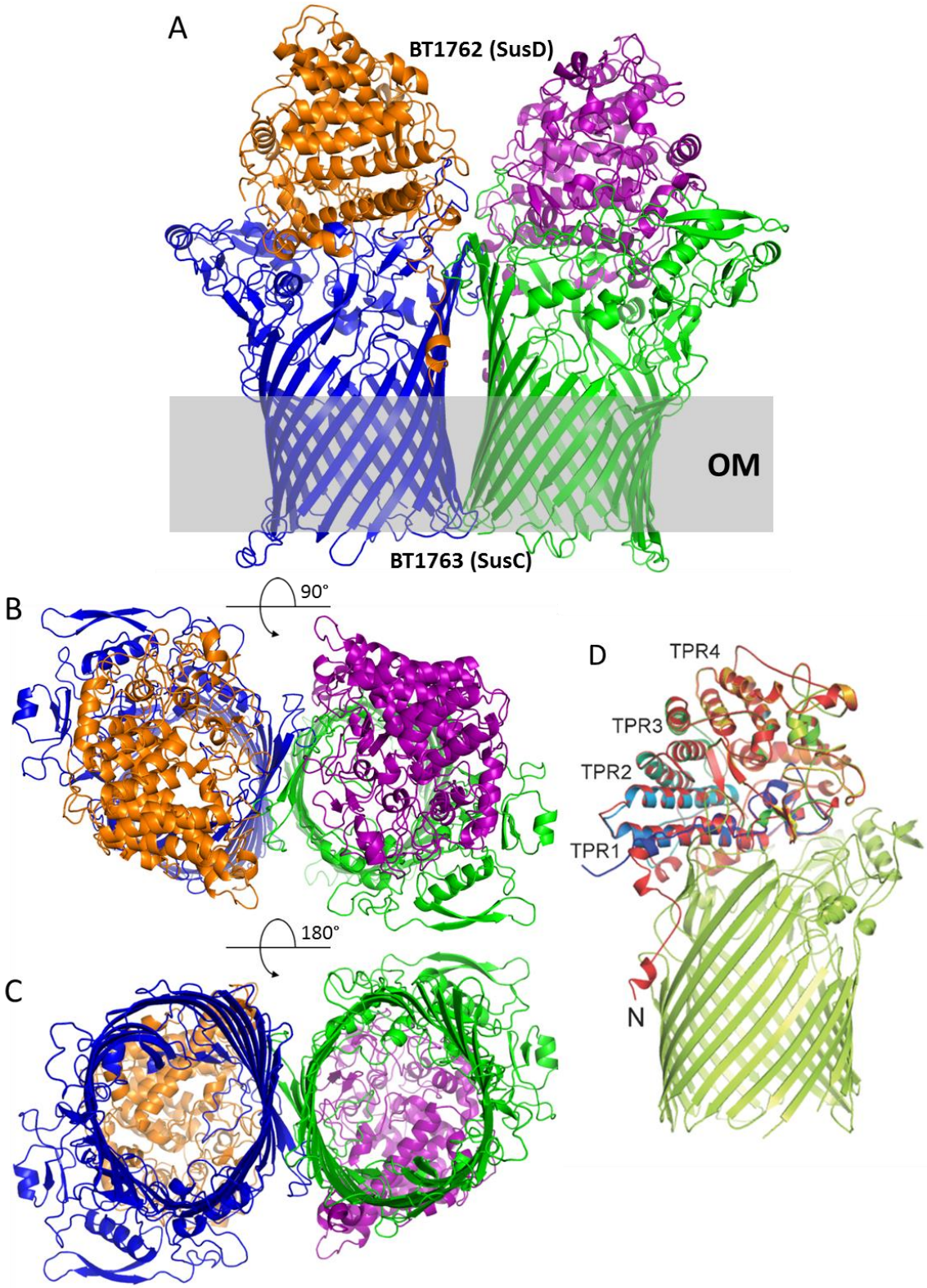


Figure 4.4 Apo structure of BT1762-3 SusCD complex.
A: Cartoon views of the overall BT1762-3 structure from the side within the OM PDB ID: 5T3R **B:** BT1762-3 from the outside of the cell (rotated 90 °C from **A**) **C:** BT1762-3 from inside the cell (rotated 180 °C from **B**) In **A-C** $BT1762^{SusD}$ is shown in orange/purple and $BT1763^{SusC}$ is shown in dark blue/green. **D:** Cartoon overlay of soluble $BT1762^{SusD}$ structure (rainbow) and $BT1762^{SusD}$ within BT1762-3 (red). $BT1763^{SusC}$ is shown in green. TPR motifs in $BT1762^{SusD}$ are labelled. Adapted from Glenwright *et al.* 2017

Surprisingly, there was no electron density for the plug domain of BT1763^{SusC} (Figure 4.4C). SDS-PAGE comparison of three different preparations of BT1762-3 (Figure 4.5A) showed degradation of the BT1763^{SusC} of the sample which produced diffracting crystals (Figure 4.5A, lane 3). SDS-PAGE of crystal-containing drops (Figure 4.5B) showed the protein had a molecular weight approximately 30 kDa lower than full length BT1763^{SusC} (Figure 4.5A). Prolonged incubation of freshly isolated complex at room temperature also led to spontaneous degradation of the plug domain and DUF4480, suggesting the plug of the SusC was cleaved by a co-purified protease contaminant.

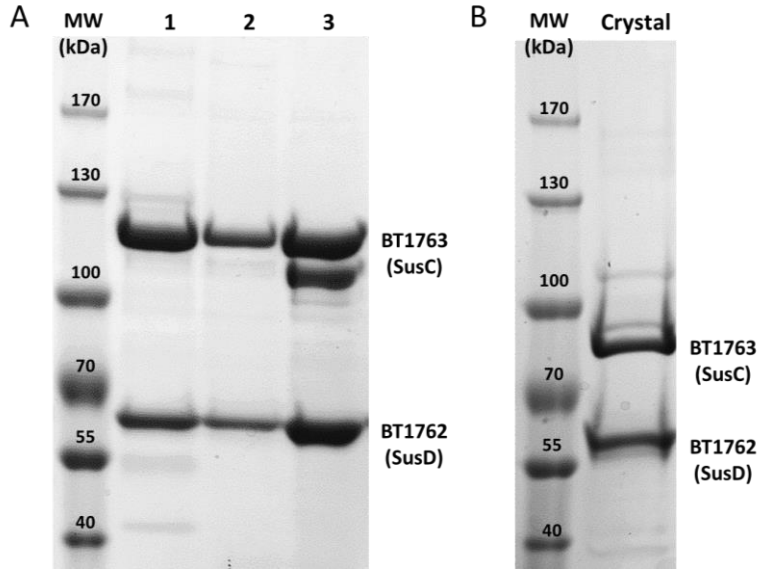


Figure 4.5 SDS–PAGE of BT1762-3 SusCD showing degradation of the BT1763^{SusC}.
All samples were boiled to separate the two proteins of the complex. Molecular weight (MW) marker shows approximate protein sizes in kDa. **A:** SDS-PAGE comparison of three BT1762-3 protein samples from different purifications. All protein was flash frozen in liquid nitrogen and stored at -80 °C. Lane 3 shows BT1763^{SusC} degradation which had not occurred with the other samples. The complex in lane 3 produced the well diffracting crystals (Figure 4.3) **B:** BT1762-3 protein from a drop containing crystals showing severe degradation of BT1763^{SusC}

4.3.2 BT1762-3 SusCD complex with bound levan oligosaccharide.

4.3.2.1 Growth of *CT-HisBT1762^{SusD}* with added Levan

BT2261-4 with ligand bound, and apo BT1762-3 protein were successfully purified directly from *B. theta*. Following this, the aim was to purify BT1762-3 with levan oligosaccharides bound, by a similar method. Initially, *CT-HisBT1762^{SusD}* cells were grown on MM-Frc for 18 – 20 hours as before. Levan (0.05 %) was then added to the cultures, under anaerobic conditions, and the cells were incubated for a further 10 minutes at 37°C. Supernatant from the cultures was boiled and analysed by TLC to ensure levan remained in the growth medium (Figure 4.6). The cultures were harvested and purified using IMAC as previous. This process was repeated several times but failed to return sufficient protein for crystallisation (<10 % of expected yield). The cells were stickier than usual which could suggest that the addition of the levan polysaccharide may have altered the capsule of *B. theta* (Tzianabos *et al.*, 1992) and disrupted the purification. Alternatively, the presence of levan may have recruited other outer membrane proteins from the PUL (BT1760^{GH32} and BT1761^{SGBP}) to produce a complex which could have obstructed the His_{x6}-tag on BT1762^{SusD}.

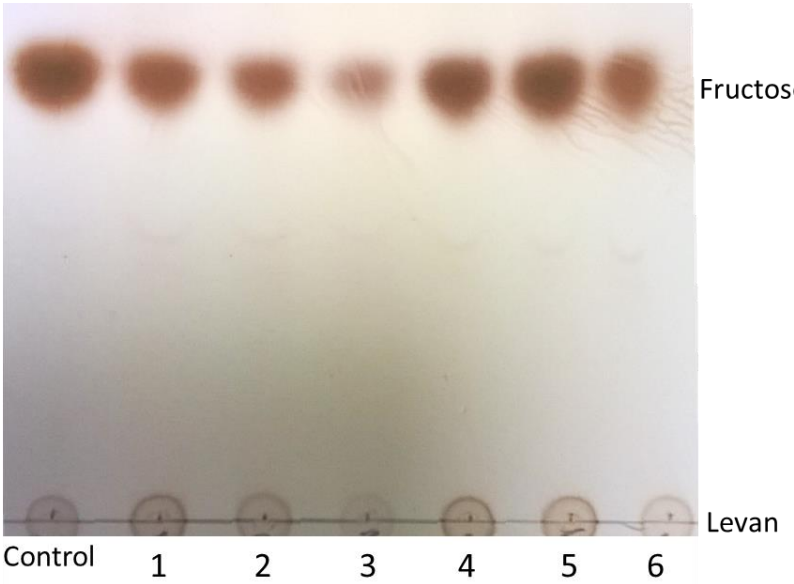


Figure 4.6 TLC showing the supernatant from *B. theta* grown on MM-Frc plus levan

CT-HisBT1762^{SusD} cells were grown on MM-Frc for 18 – 20 hours. 0.05 % levan was added and the cells were incubated for 10 minutes before harvesting. The supernatant was boiled and 2 µl was analysed by TLC. The control sample was taken immediately following addition of levan and numbers 1-6 represent different culture bottles. The remaining fructose and levan in the samples are labelled.

4.3.2.2 Production and purification of levan oligosaccharides

A practical alternative to purifying BT1762-3 with substrate bound was to purify the protein and oligosaccharides separately and co-crystallise. The BT1760^{GH32} endo-acting levanase from the *B. theta* levan PUL was expressed in *E. coli* and purified using IMAC (Figure 4.7A). 100 nM of the enzyme was then used to partially digest levan polysaccharides from *Erwinia herbicola* (Sigma-Aldrich) by incubating at 37 °C for 20 minutes (Figure 4.8B, Methods 2.12.4). BT1760^{GH32} appears to preferentially produce fructotriose (Figure 4.7B) which has been shown previously (Sonnenburg *et al.*, 2010). The resulting oligosaccharides were separated using size exclusion chromatography (P2 Biogel, Biorad) and every fifth fraction was analysed by TLC (Figure 4.7C). Previous ITC data had shown BT1762^{SusD} in isolation requires at least seven fructose molecules for levan binding, although affinity for levanheptaose is very weak (Zheng, 2009). The TLC standards facilitated estimation of oligosaccharide length in each fraction. Samples containing approximately 6-14 fructose units from fractions 90-125 (highlighted by box, Figure 4.7C) were freeze-dried and pooled for co-crystallisation with BT1762-3.

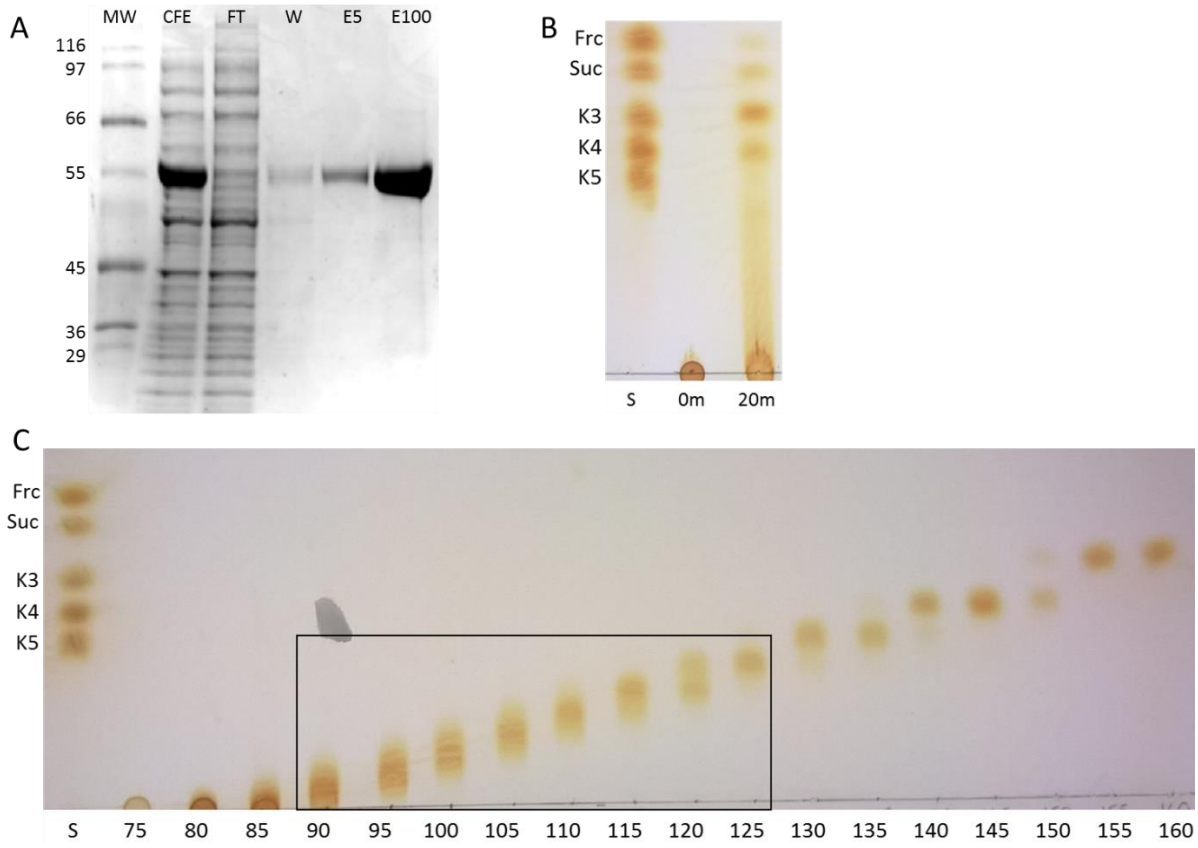


Figure 4.7 Production and purification of levan oligosaccharides.
A: Example SDS-PAGE of BT1760^{GH32} samples following overexpression in *E. coli* BL21 and IMAC purification. Cell free extract (CFE), flow through (FT), wash (W), elution with 5 mM imidazole (E5) and elution with 100 mM imidazole (E100). The major band at ~57kDa in CFE and E100 lanes is the expected MW of BT1760^{GH32}. Molecular weight (MW) marker shows approximate protein sizes in kDa. **B:** 500 mg (5 mg/ml) Levan from *E. herbicola* shown before (0m) and after 20 minute incubation (20m) with 100 nM BT1760^{GH32}. Standards (S) shown are 2 mg/ml (each) fructose (Frc), sucrose (Suc), kestose (K3), kestotetraose (K4) and kestopentaose (K5) used as comparison. K3, K4 and K5 are 2,1-linked fructo-oligos because 2,6-linked levan oligos cannot be purchased. **C:** Partially digested levan from **A** separated by size exclusion chromatography. TLC shows every fifth fraction and standards (S) as in **A**. The fractions highlighted by the box (90-125) were pooled for co-crystallisation.

4.3.2.3 Co-crystallisation of BT1762-3 with levan oligosaccharides

A large-scale MM-Frc growth of *CT-HisBT1762*^{SusD} *B. theta* and purification of BT1762-3 was carried out as before with the addition of protease inhibitors (cOmplete, EDTA-free Protease Inhibitor Cocktail Tablets, ROCHE) to the cells prior to lysis and during membrane solubilisation (Figure 4.8A). The protein was concentrated to 12 mg/ml and incubated with 2.5 mM levan oligosaccharides for 1 hour before crystal screens were set up. Additional detergent (0.2 % LDAO) was added to half of the crystal drops and produced crystal hits (Figure 4.8B), which were then optimised. Crystals grown in 0.15 M sodium formate, 0.1 M HEPES pH 7.2, 18 % PEG 3350, showed diffraction when tested in-house and were sent to DLS for data collection.

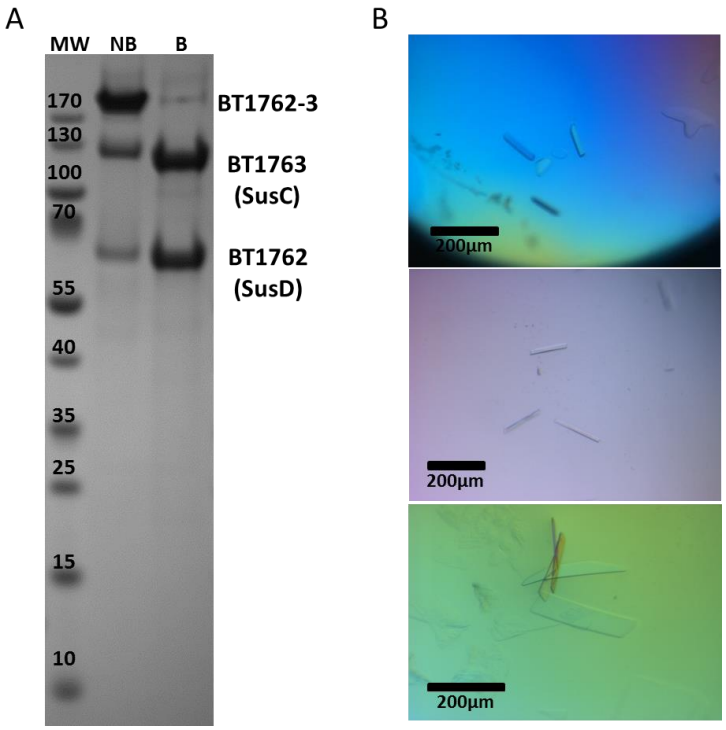


Figure 4.8 Co-crystallisation of BT1762-3 with levan oligosaccharides
A: SDS-PAGE of BT1762-3 purified by IMAC followed by size exclusion chromatography. The sample is shown boiled (B) and not boiled (NB). Molecular weight (MW) marker sizes are shown in kDa **B:** Initial crystal hits of BT1762-3 (12 mg/ml) with 0.2 % LDAO and 2.5 mM levan oligosaccharides. Crystallisation conditions; 0.05 M NaCl, 0.1 M Sodium phosphate pH 6, 16 % PEG 4000 (top panel), 0.15 M sodium formate, 0.1 M HEPES pH 7.2, 18 % PEG 3350 (middle panel), 0.1 M Sodium cacodylate pH 6.5, 24 % (w/v) PEG 1500.

4.3.2.4 Structure of BT1762-3 with levan oligosaccharide bound

Co-crystallisation of BT1762-3 with the levan oligosaccharides was successful and led to a structure with the ligand bound. The structure was solved by Bert van den Berg using MR with the apo BT1763^{SusC} structure and the plug domain of BT2264^{SusC} used as search models (Figure 4.9). The overall structure, solved using data to 2.9 Å resolution, was almost identical to the apo structure, although the plug domain was intact within the barrel of BT1763^{SusC} (Figure 4.9). An overlay of the two BT1763^{SusC} structures, showed small variations in the positioning of extracellular loops but no significant conformational change upon binding the levan oligosaccharide (Figure 4.10). The root mean square deviation of atomic distances (RMSD) for the two structures is 1.35 Å (all atoms).

The levan binding site contains five β2,6-linked fructose units (Figure 4.11). The binding site is completely enclosed between the two proteins (Figure 4.11A). Interactions with levan include seven residues from BT1762^{SusD} (D59, N61, D85, W103, C316, R386, Y413) and six residues from BT1763^{SusC} (E404, D431, R432, F674, N926, V927).

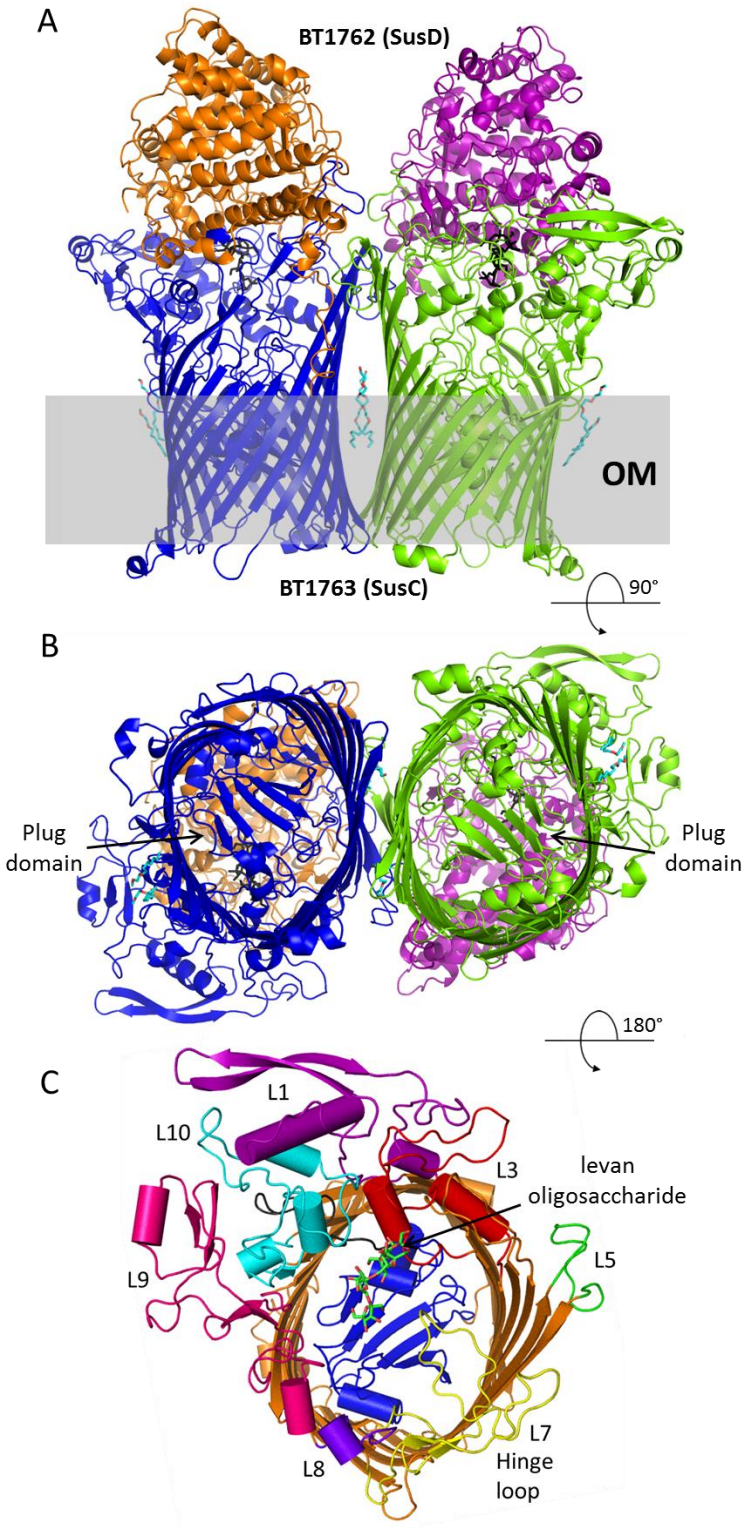


Figure 4.9 Overall BT1762-3 structure with levan oligosaccharide bound.
A: Cartoon view of the overall BT1762-3 structure from the side within the OM. $BT1762^{SusD}$ shown in orange/purple and $BT1763^{SusC}$ shown in dark blue/green. The bound levan oligosaccharide is shown in black. **B:** BT1762-3 from inside of the cell showing the plug domain intact (coloured as in **A**). **C:** $BT1763^{SusC}$ monomer shown from outside the cell (rotated 180 °C from **B**) with large extracellular loops labelled. The barrel of the transporter is shown in orange, the plug domain is dark blue and the hinge loop (L7) is yellow. Helices are shown as cylinders. The bound levan oligosaccharide is shown as green sticks.

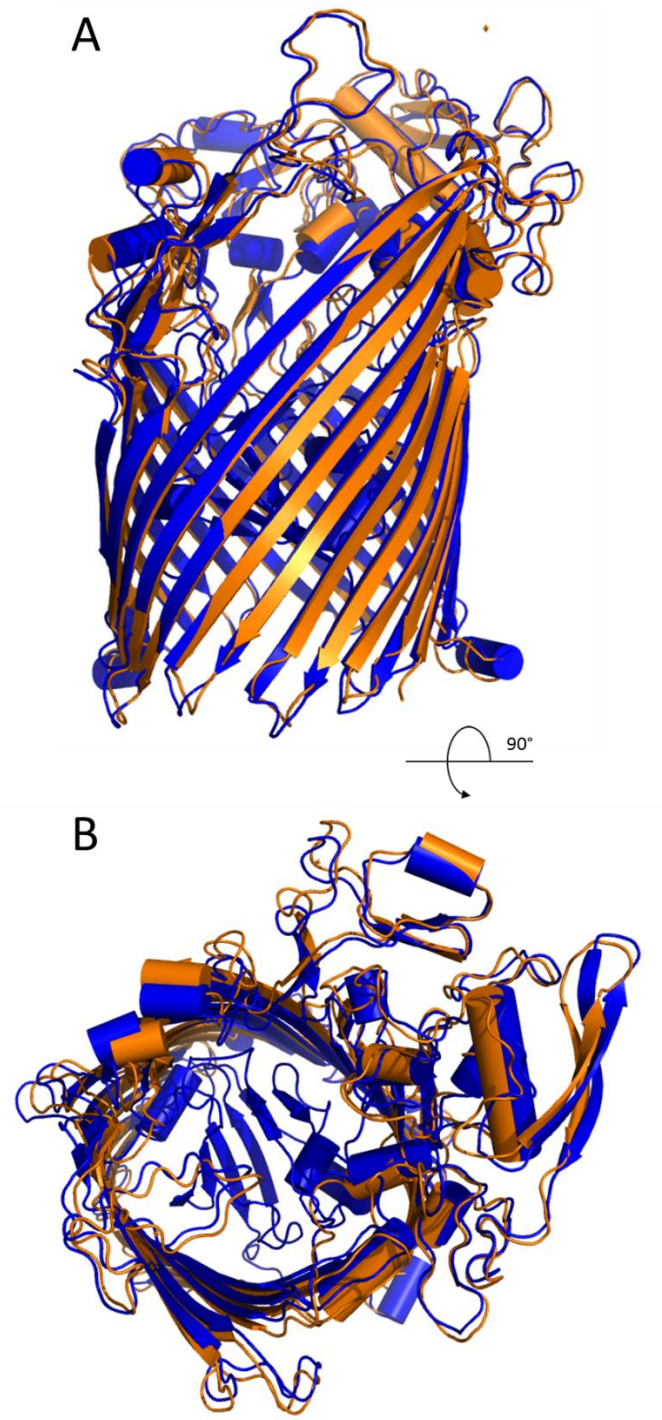


Figure 4.10 Overlay of BT1763^{SusC} apo and ligand bound structures.
A: BT1763^{SusC} truncated apo structure (orange) aligned with the ligand bound BT1763^{SusC} including the intact plug domain (blue) shown from the side. **B:** The overlay shown from the extracellular side top down view, coloured the same as in **A.** Helices are shown as cylinders.

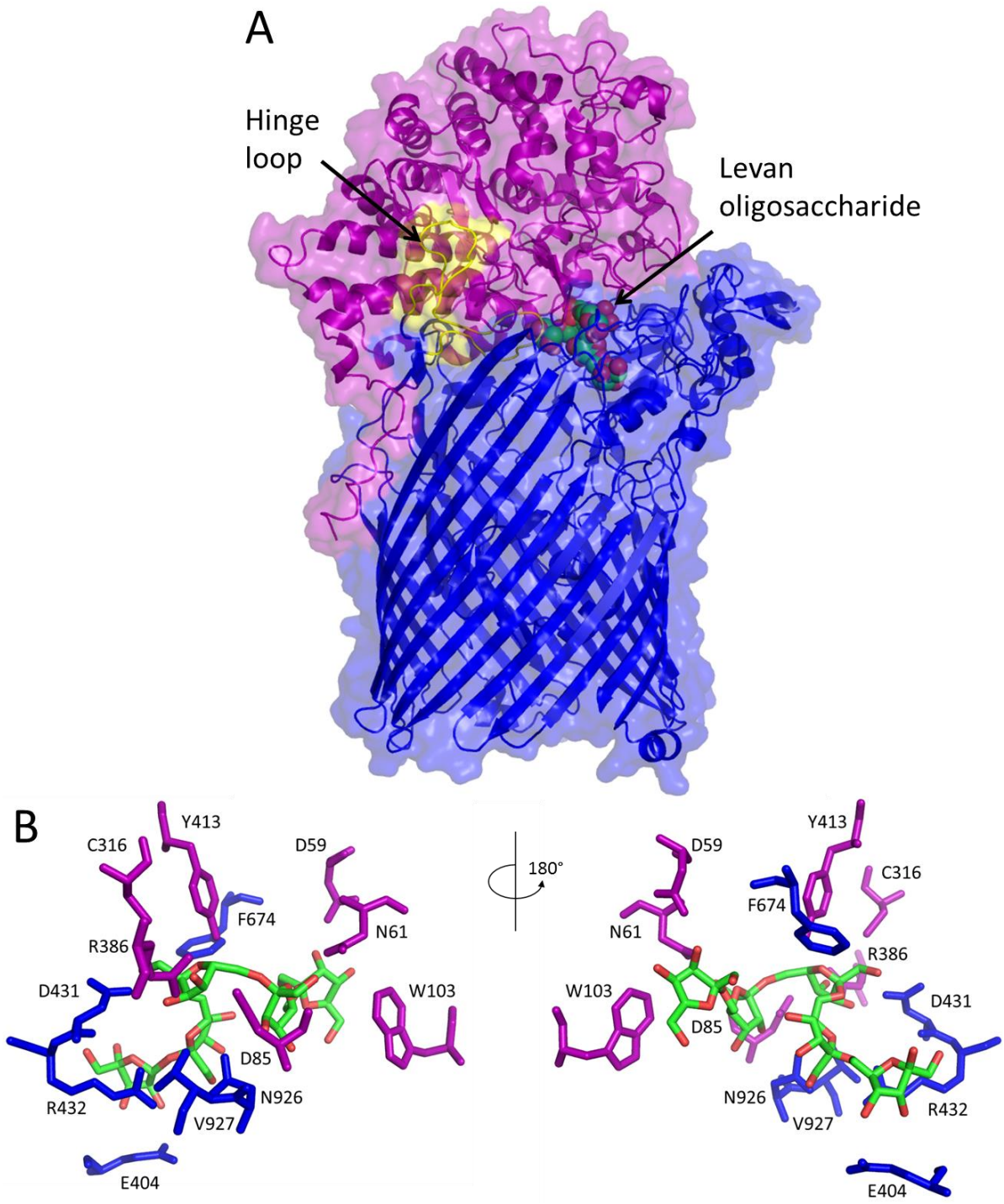


Figure 4.11 Levan binding site within BT1762-3.
A: Levan oligosaccharide shown as spheres within BT1762-3. BT1762^{SusD} is purple and BT1763^{SusC} is blue with the hinge loop shown in yellow. **B:** Close-up of the levan binding site within the SusCD complex. Levan shown in green, binding residues from BT1762^{SusD} are purple and residues from BT1763^{SusC} are dark blue. All residues are numbered as for full length proteins including signal peptides. The binding site has been rotated 180° around the vertical axis between the two images.

4.3.3 Mutating BT1762-3 to investigate function

4.3.3.1 BT1762^{SusD} binding site mutants

Previously, Hongjun Zheng produced mutations of BT1762^{SusD} (W103A, C316A, Y413A, W414A, W497A) based on the putative binding site of the recombinant structure. The mutants were protein was expressed recombinantly and purified using IMAC (Methods 2.7) then tested for binding with levan using ITC (Glenwright *et al.*, 2017). Based on the structure of the SusCD complex BT1762-3 with bound levan (Figure 4.11B), the remaining interacting residues of BT1762^{SusD} were mutated to alanine using site directed mutagenesis (Methods 2.5.4.2) and sequenced to ensure each mutation. These mutants were also expressed in *E. coli*, purified and tested for binding to *Bacillus* levan by ITC (Figure 4.12). Six mutations completely disrupt the binding ability of the protein (D59, N61, D85, W103A, C316A and R386) and Y413A significantly reduces the affinity of BT1762^{SusD} for levan (Table 4.1). Two control mutations of tryptophan residues, which were not expected to interact with levan, were tested; one close to the binding site (W414A) and one on the opposite side of BT1762^{SusD} (W497A). Neither mutation significantly affected the interaction with levan, therefore the lack of binding by the other mutants is due to disruption of the binding site rather than affecting the overall structure of the protein.

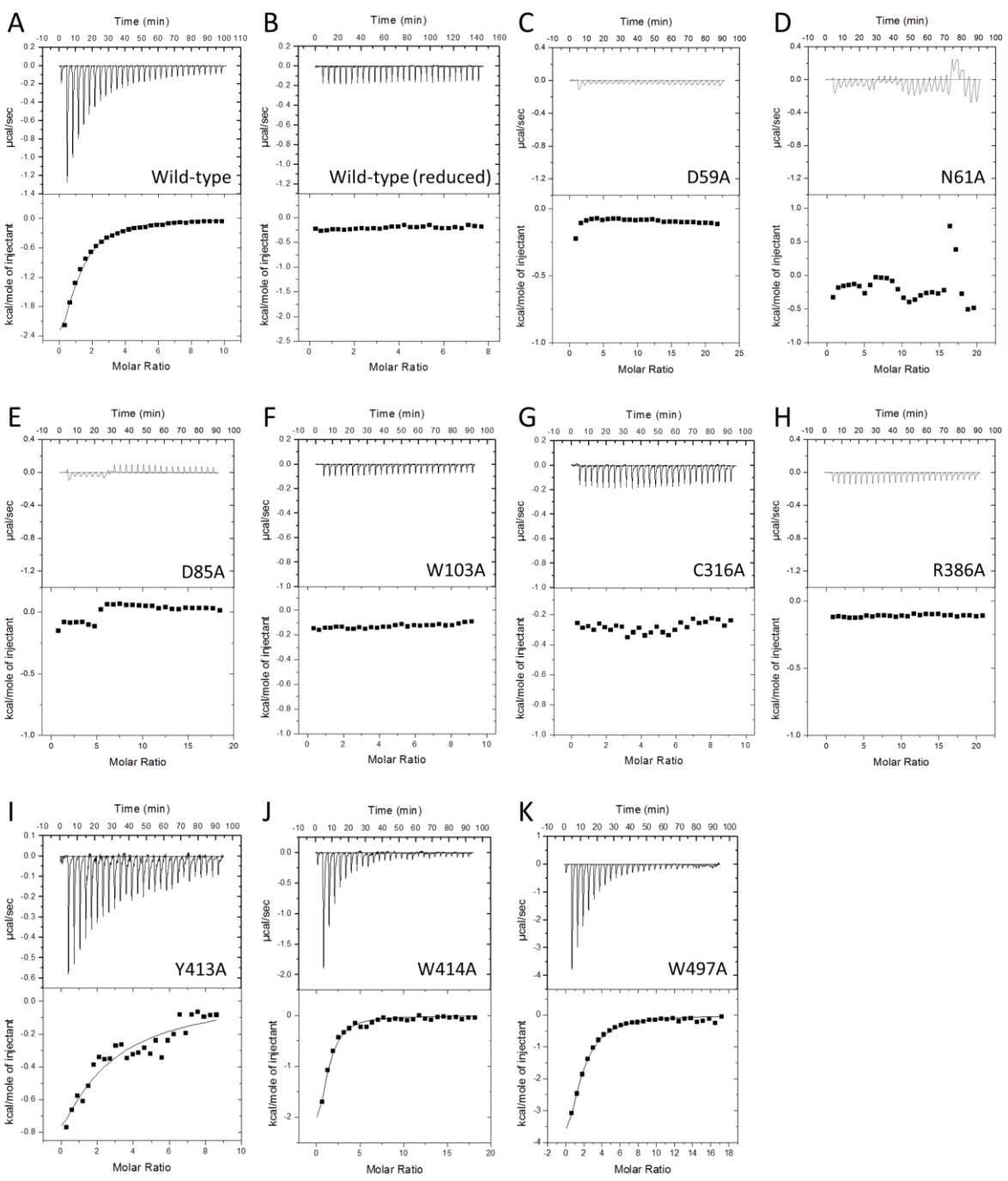


Figure 4.12 BT1762^{SusD} mutants binding to levan measured by ITC.
Bacillus levan (5 mg/ml) was titrated against WT and mutant forms of BT1762 (all ~50µM). All ITC was carried out in 20 mM HEPES pH7.5 at 25 °C. Where possible, data was fit using MicroCal Origin 7 software. The reduced wild-type titration was carried out in the presence of 1 mM Tris 2-carboxyethyl phosphine (TCEP). Figure includes data collected by Hongjun Zheng (WT, W103A, C316A, Y413A, W414A, W497A) from Glenwright *et al.* 2017.

BT1762^{SusD}	Ka x 10^4 (M^{-1})	Kd x 10^{-5} (M^{-1})
Wild-type	1.9 ±0.05	5.26
Wild-type (reduced)	NB	-
D59A	NB	-
N61A	NB	-
D85A	NB	-
W103A	NB	-
C316A	NB	-
R386A	NB	-
Y413A	0.3 ±0.05	33.3
W414A	1.9 ±0.4	5.26
W497A	1.5 ±0.4	6.66

Table 4.1 ITC data for BT1762^{SusD} mutants binding to levan polysaccharides
Data from ITC shown in Figure 4.13. No binding observed is abbreviated to NB.

4.3.3.2 Introducing BT1762-3 mutations into the *B. theta* chromosome

B. theta requires functional BT1762-3 for growth on levan (Sonnenburg *et al.*, 2010), suggesting it may be possible to investigate the importance of components of the complex by introducing mutations into genomic copies of BT1762^{SusD} or BT1763^{SusC} and analysing growth on the fructan.

The hinge loop (L7 Figure 4.9C and 4.11A) is conserved in both SusCD complexes and appears to be important in the molecular dynamic simulations of BT2263-4 (Chapter 3, section 3.3.3.5). To explore the importance of the hinge loop in SusCD function, the region (residues 667-696, Figure 4.13A) was removed in the genomic copy of BT1763^{SusC} and replaced with a single glycine residue to join the β-strands of the barrel. The *B. theta* mutant lacking the hinge loop will be referred to as *ΔHinge-BT1763*.

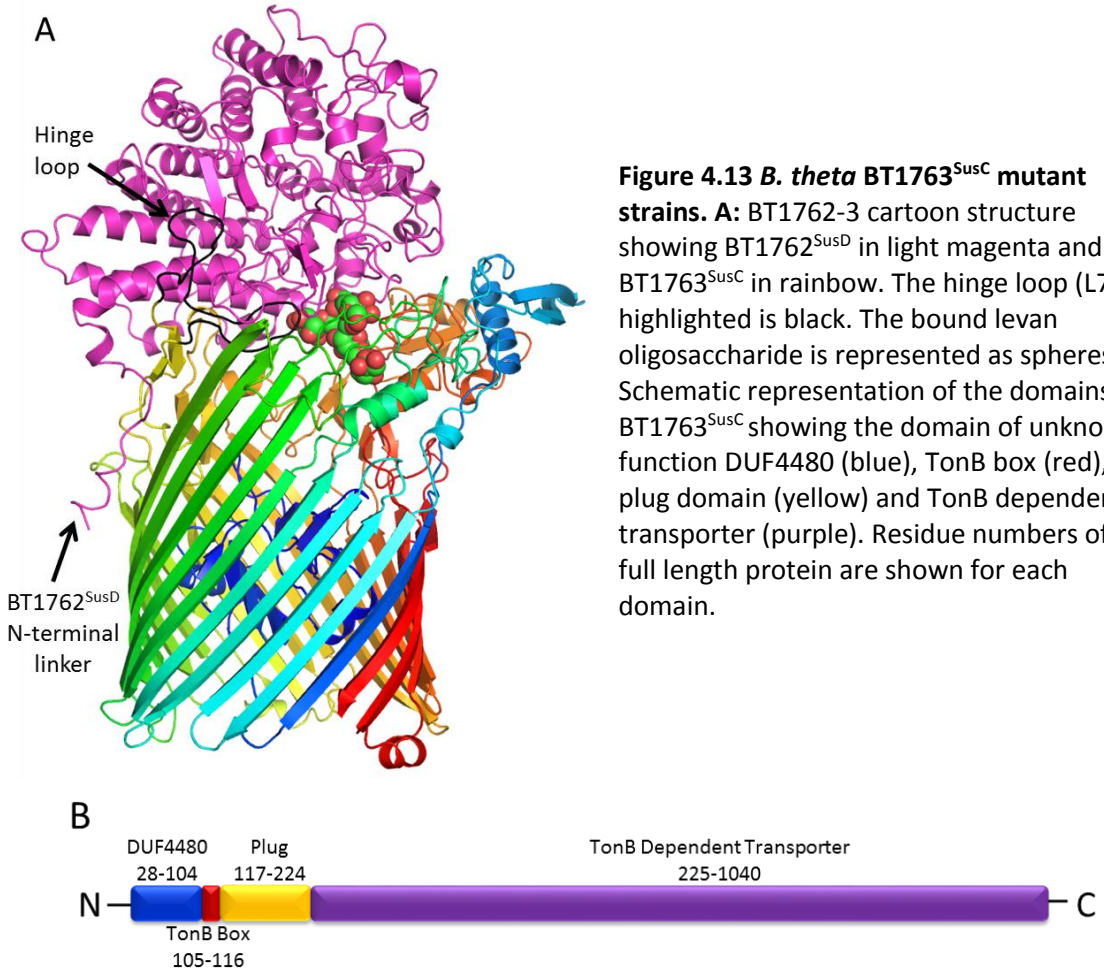


Figure 4.13 *B. theta* BT1763^{SusC} mutant strains. A: BT1762-3 cartoon structure showing BT1762^{SusD} in light magenta and BT1763^{SusC} in rainbow. The hinge loop (L7) highlighted is black. The bound levan oligosaccharide is represented as spheres. **B:** Schematic representation of the domains of BT1763^{SusC} showing the domain of unknown function DUF4480 (blue), TonB box (red), plug domain (yellow) and TonB dependent transporter (purple). Residue numbers of the full length protein are shown for each domain.

Over 80 % of *B. theta* SusC homologues contain a domain of unknown function DUF4480, which is unique among TBDTs. DUF4480 is at the N-terminus, immediately before the TonB box (Figure 4.13B). The domain did not appear in either of the SusCD complex structures suggesting it was disordered and flexible or proteolytically removed. A mutant strain of *B. theta* without residues 30-103 of BT1763^{SusC} was produced, and named *ΔDUF-BT1763*.

Based on ITC data, W103 is vital for binding of isolated BT1762^{SusD} to levan (Figure 4.13F), therefore a strain with the W103A mutation was produced, *W103A-BT1762*. The BT1762 deletion strain (*ΔBT1762^{SusD}*; Sonnenburg *et al*. 2010), was used for comparison in the growth assay (Figure 4.14).

All of the strains tested grew similarly on the control minimal media with 0.5 % fructose and the addition of the His_{x6}-tag did not affect growth of *CT-HisBT1762* on levan (Figure 4.14). Surprisingly, the growth of *W103A-BT1762* on levan was also unaffected by the mutation, suggesting redundancy within the binding site. Unexpectedly, *ΔHinge-BT1763* only showed a slight increase in lag-phase before achieving similar cell density in levan as the WT, suggesting the function of the complex is largely unaffected by the loss of the hinge loop (Figure 4.14). Finally, the *ΔDUF-BT1763^{SusC}* mutant showed the most severe phenotype, where no growth was observed after 48 hours. This was an even more extreme phenotype than the *ΔBT1762^{SusD}* strain which eventually grew after a 36 hour lag (Figure 4.14).

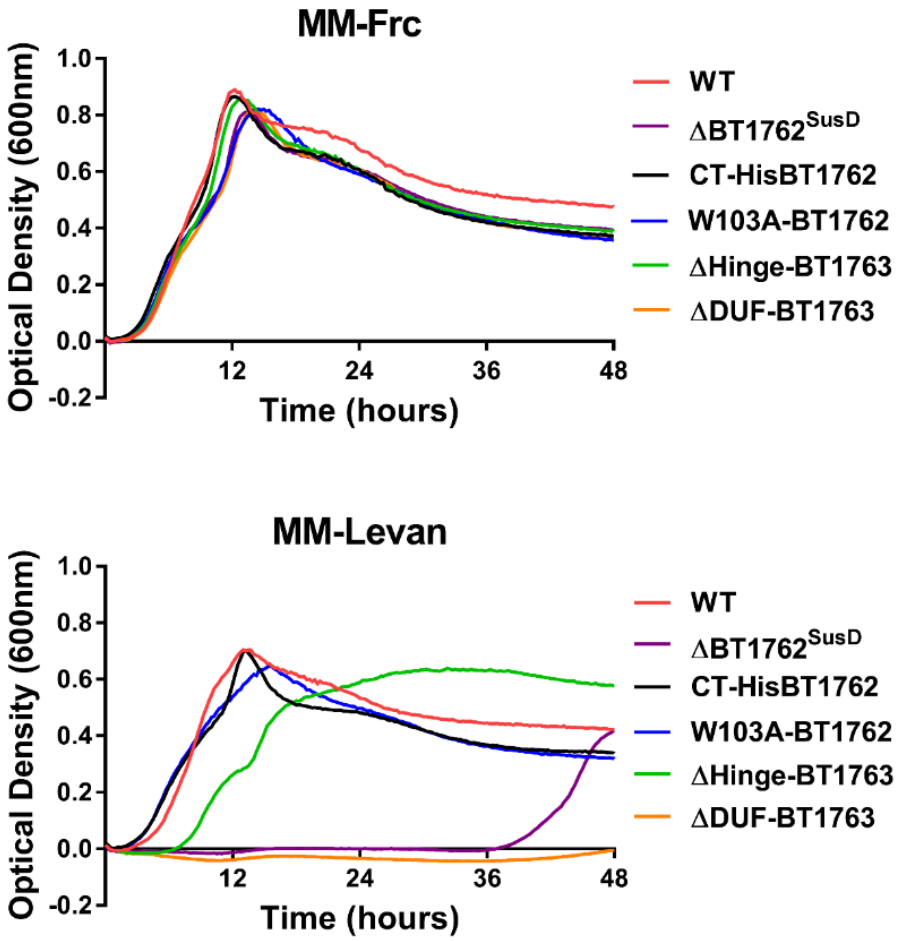


Figure 4.14 Growth of mutant *B. theta* strains on levan.
Top panel shows cells grown on minimal media containing 0.5 % fructose. The bottom panel shows growth curves on minimal media containing 0.5 % levan. The graphs show WT *B. theta* (red), ΔBT1762^{SusD} (purple), CT-HisBT1762 (black), W103A-BT1762 (blue), ΔHinge-BT1763 (green) and ΔDUF-BT1763. Data was collected in triplicate and averaged for each trace. Results are representative of at least four individual repeats, errors not shown for clarity. Growth assays were performed anaerobically at 37 °C in an automatic plate reader.

The level of BT1762-63 SusCD expression was analysed in the mutant strain *ΔDUF-BT1763* due to the severity of the growth phenotype (Figure 4.14). For this purpose a DUF knockout strain was produced which also possessed the C terminal His_6-tag fusion with BT1762^{SusD}. For both expression tests the original *CT-HisBT1762^{SusD}* strain was used as a positive control (Figure 4.15). Western blot analysis using Anti-His_6 antibodies showed expression of BT1762^{SusD} by *ΔDUF-BT1763* at levels comparable with the positive control (Figure 4.15A). A small scale (1 litre) purification of BT1762-3 was carried out, as previously described, using IMAC and the resulting samples were

analysed using SDS-PAGE (Figure 4.15B). The *ΔDUF-BT1763* strain expresses both BT1762^{SusD} and BT1763^{SusC} in the OM. While the mutant appeared to express less BT1762-3 than the control strain, this was most likely due to lower total protein from the purification as expression levels of BT1762^{SusD} were equivalent between the strains by Western Blot (Figure 4.15). These data indicate that the major growth defect observed on MM-Levan by *ΔDUF-BT1763* was due to a non-functional BT1763^{SusC} rather than lack of expression of the complex in the OM.

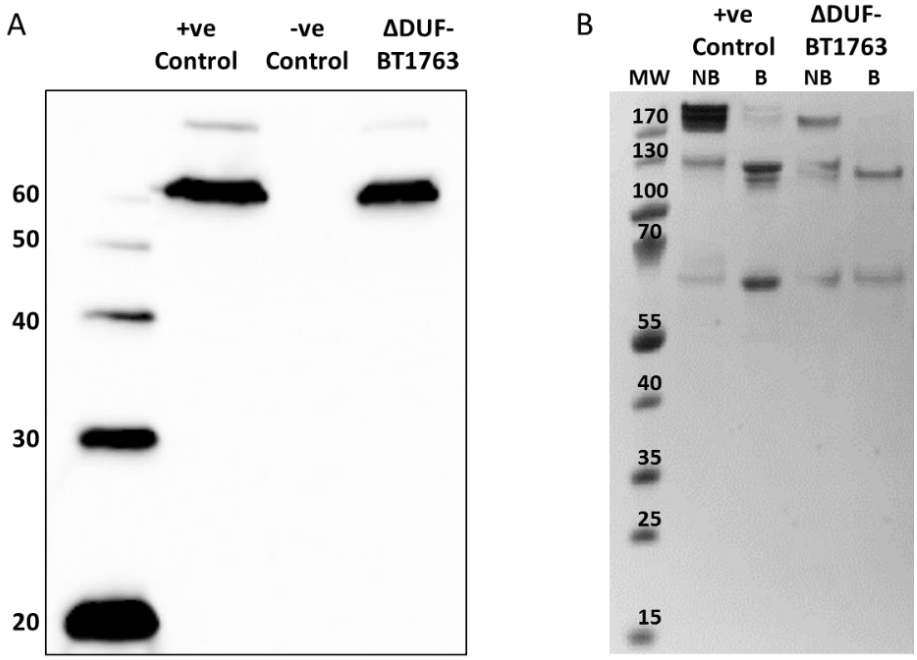


Figure 4.15 Analysing expression of BT1762-3 SusCD by the *ΔDUF-BT1763 B. theta* strain. **A:** Western blot analysis of BT1762^{SusD} expression using Anti-His_6 antibodies. The positive (+ve) control was CT-HisBT1762^{SusD} strain and the negative (-ve) control is WT *B. theta.* **B:** SDS-PAGE analysis of BT1762-3 expression from *ΔDUF-BT1763* strain compared with the positive (+ve) control (CT-HisBT1762^{SusD}). Cells were grown on MM-Frc for 16 hours and protein was purified from the OM fraction by IMAC. Protein standard sizes shown are in kDa on both panels.

4.3.4 BT1761 surface glycan binding protein

4.3.4.1 ITC of BT1761^{SGBP} with levan

While SGBPs share the same function they do not show sequence homology, making them difficult to identify by bioinformatics techniques (Cameron *et al.*, 2012; Cartmell *et al.*, 2017). Genes encoding SGBPs are typically found immediately adjacent to the SusD homologue of the PUL and show no homology to known carbohydrate-active enzymes (Martens *et al.*, 2009). BT1761 was predicted to be a surface glycan binding protein (SGBP) based on the position within the levan PUL and interactions with levan were observed previously by affinity gel (Sonnenburg *et al.*, 2010). An N-terminal tagged pET28a construct, purified by IMAC (Figure 4.16), was used for ITC binding analysis against microbial levan. BT1761^{SGBP} bound *Bacillus* levan polysaccharides and partially digested *Erwinia* levan (Figure 4.17). This confirmed that BT1761 is a levan binding protein. BT1761^{SGBP} bound *Bacillus* levan with an 8-fold higher affinity than BT1762^{SusD} (Figure 4.17). There also appeared to be some interaction between BT1761^{SGBP} and *Erwinia* polysaccharide levan although the ITC trace does not show definitive binding (Figure 4.17). By comparison, BT1762^{SusD} absolutely does not interact with full length *Erwinia* levan in a manner which can be measured by ITC. BT1762^{SusD} shows some possible binding to partially digested *Erwinia* levan, but the affinity was too low to fit (Figure 4.17).

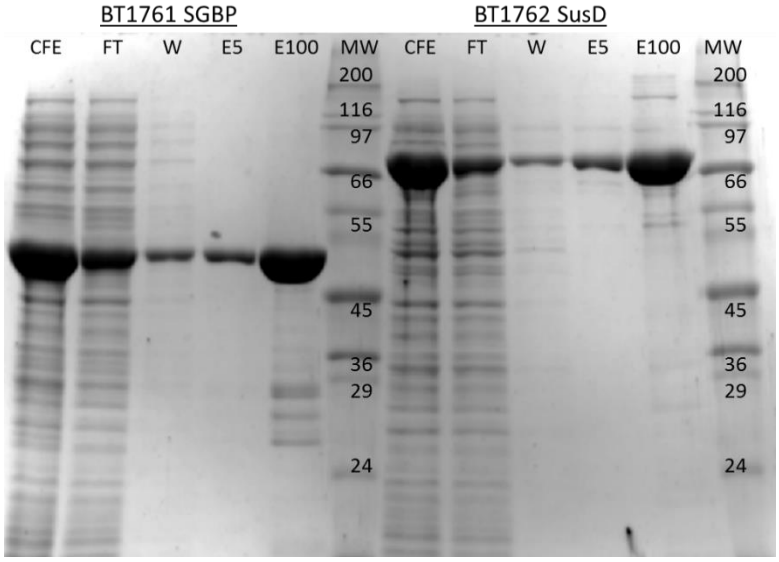
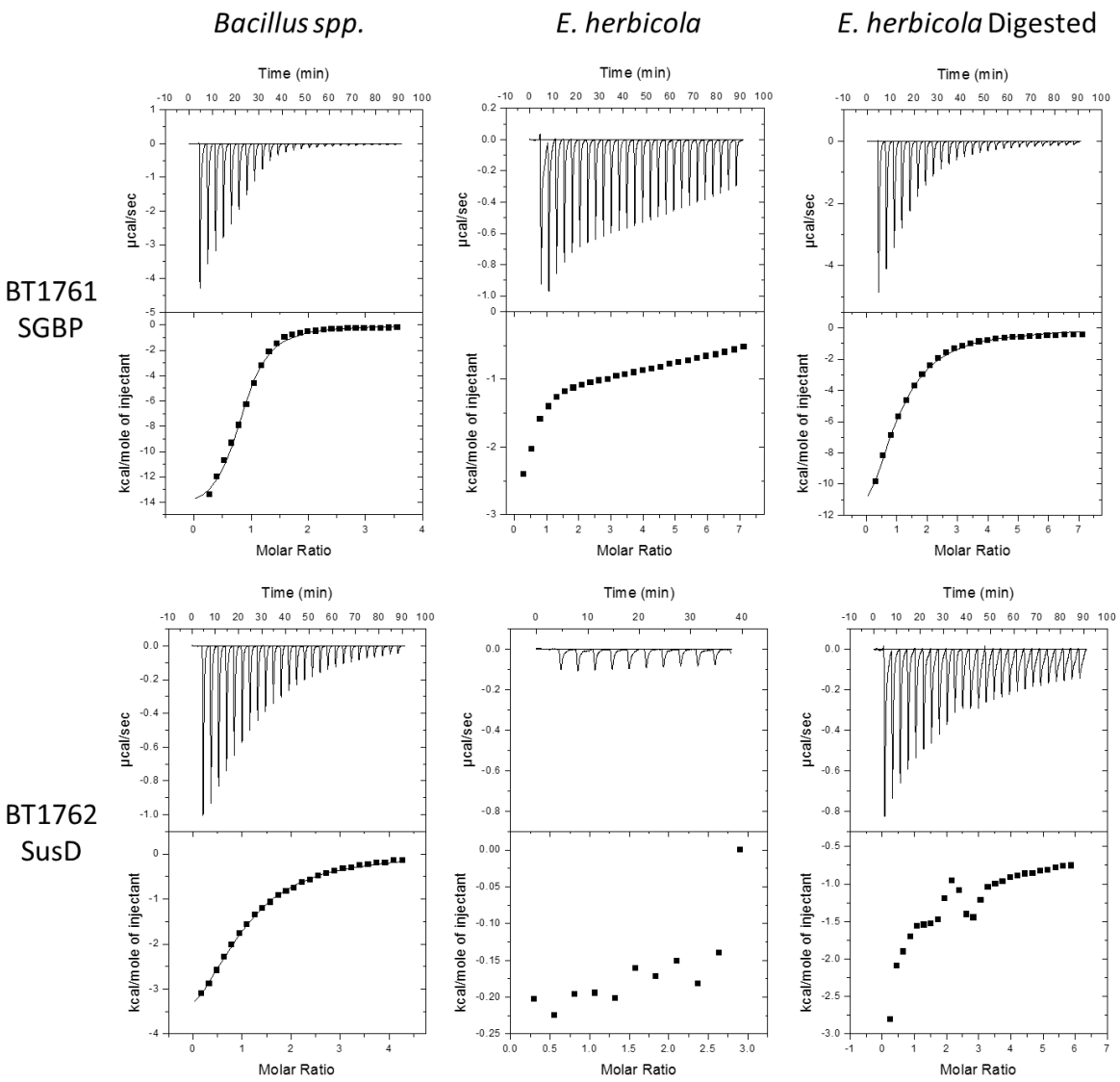


Figure 4.16 Purification of BT1761^{SGBP} and BT1762^{SusD} for ITC
Example SDS-PAGE of BT1761^{SGBP} and BT1762^{SusD} samples following overexpression in *E. coli* BL21 and IMAC purification. Cell free extract (CFE), flow through (FT), wash (W), elution with 5 mM imidazole (E5) and elution with 100 mM imidazole (E100). BT1761^{SGBP}is expected to be ~50 kDa and BT1762SusD is expected to be ~65 kDa Molecular weight (MW) marker shows approximate protein sizes in kDa.



	Bacillus spp.	*E. herbicola*	*E. herbicola* Digested
BT1761 SGBP	Ka = 1.7 (±0.5) x 10^5 Kd = 5.9 x 10^{-6}	No binding	Ka = 1.9 (±1.0) x 10^4 Kd = 5.26 x 10^{-5}
BT1762 SusD	Ka = 2.1 (±0.8) x 10^4 Kd = 4.76 x 10^{-5}	No binding	Affinity too low to fit

Figure 4.17 Binding of BT1761^{SGBP} and BT1762^{SusD} to levan analysed by ITC.
5 mg/ml of ligand was titrated against 55 µM of protein. Microbial levan polysaccharide was used from two different sources; *Bacillus spp.* and *E. herbicola*. The partial digest of *E. herbicola* levan was produced using BT1760 (Figure 4.8A). All ITC was carried out in 20 mM HEPES pH7.5 at 25 °C. Where possible, data was fit using MicroCal Origin 7 software and the Ka was averaged (levan polysaccharides – 3 independent experiments, levan digest – 2 independent experiments). The top half of each panel shows the raw injection heats and the bottom panel shows the integrated peak areas.

4.3.4.2 Cloning, expression, purification and crystallisation

The surface glycan binding protein (SGBP) BT1761 is the only protein from the levan PUL outer membrane apparatus (Figure 4.1) which does not have a solved structure. BT1761^{SGBP} was cloned without its Type II signal sequence into the *E. coli* expression vector pET28a. The construct was overexpressed in BL21 with an N-terminal His_{x6}-tag and purified using IMAC (Figure 4.18A) followed by SEC. Eight 96-condition screens for crystal hits (Molecular Dimensions Structure, Index, PACT, Morpheus, JCSG 1/2/3/4) produced diffracting crystals in one condition; 0.1 M MES pH 5, 1.6 M ammonium sulphate (Figure 4.18B). A dataset was collected at DLS however, there was no suitable complete model for MR. A partial solution for ~25 % of BT1761^{SGBP}, to a resolution of 2.5 Å, was found using an edited model of a predicted glycan binding protein BT2081 (PDB ID: 3HBZ) produced by MrBUMP (Keegan and Winn, 2007). BT1761^{SGBP} with selenomethionine incorporated produced crystals which did not diffract. A C-terminal His_{x6}-tagged BT1761^{SGBP} construct was also produced but did not crystallise.

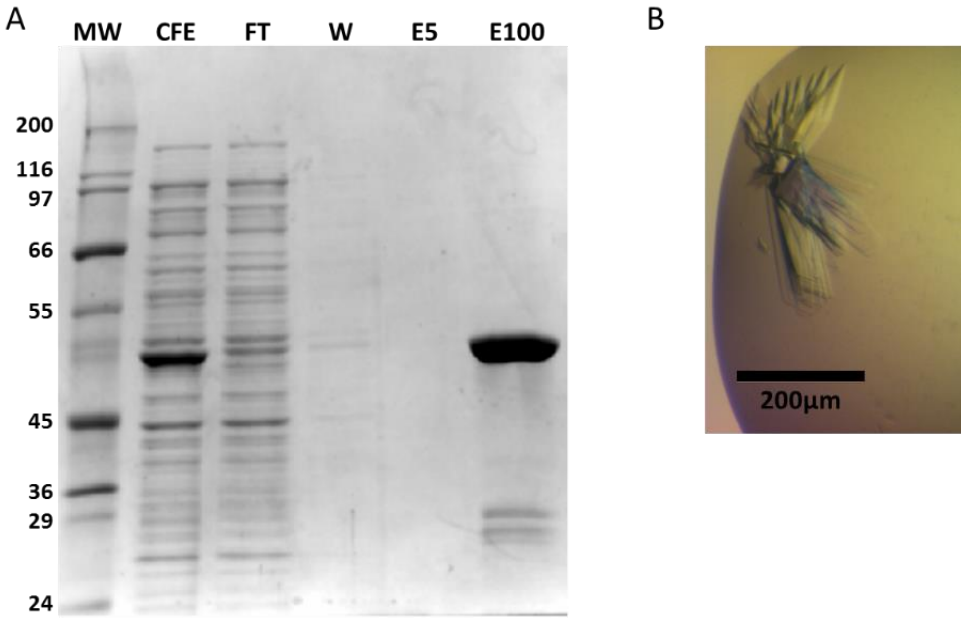


Figure 4.18 Purification and crystallisation of BT1761 surface glycan binding protein.
A: Example SDS-PAGE of BT1761^{SGBP} samples following overexpression in *E. coli* BL21 and IMAC purification. Cell free extract (CFE), flow through (FT), wash (W), elution with 5 mM imidazole (E5) and elution with 100 mM imidazole (E100). The major band at ~50kDa in CFE and E100 lanes is the expected MW of BT1761^{SGBP}. Molecular weight (MW) marker shows approximate protein sizes in kDa. **B:** Crystal hit produced in 0.1 M MES pH 5, 1.6 M ammonium sulphate (Molecular Dimensions screen JCSG 4 condition G5).

Lipoproteins such as SGBPs have N-terminal linker sequences between the lipid anchor and the globular protein domains (Cameron *et al.*, 2012). All existing structures of recombinant lipoproteins from *Bacteroides spp.*, expressed without the signal peptide, are lacking the first 15 – 20 residues suggesting the linker region is disordered when the protein is not membrane associated and could interfere with crystal formation (Koropatkin *et al.*, 2008; Cameron *et al.*, 2012). Linkers from BT2261^{LP}, BT2262^{LP} and BT2263^{SusD} are clear in the BT2261-4 complex structure and are up to 22 residues long (Chapter 3, Figure 3.8). Therefore a truncated construct without the first 22 N-terminal residues was cloned into pB22 with a TEV-cleavable N-terminal His_{x7}-tag. The protein was overexpressed in the periplasm of TUNER *E. coli*, purified using IMAC and the His_{x7}-tag was cleaved before SEC. This construct only produced very small, irregular crystals. Addition of levan oligosaccharides (Figure 4.7) did not improve crystallisation of BT1761^{SGBP}.

4.3.5 Levan Acquisition Outer Membrane Protein Complex BT1760-3

Previous studies have suggested the lipoproteins expressed in *Bacteroides* PULs interact at the cell surface to form a multiprotein complex (Shipman *et al.*, 2000; Foley *et al.*, 2016; Cartmell *et al.*, 2017). Certainly in the BT2261-4 purification the two downstream lipoproteins from the locus ($BT2261^{LP}$ and $BT2262^{LP}$) formed a tight complex with the SusC, supporting this hypothesis. Depending on the strength and nature of these interactions, it may be possible to co-crystallise the entire outer-membrane complex of the levan utilisation system. This would give vital insights into the location of interaction sites on each protein component and level of co-operation between the binding, catalytic and import components of the system.

Purification of BT1762-3 from cells grown on MM-Frc did not yield the auxiliary lipoproteins from the levan PUL ($BT1760^{GH32}$ and $BT1761^{SGBP}$). This is most likely to be because the cells were not actively degrading and importing levan so perhaps the proteins were not interacting. However, it was not possible to purify BT1762-3 from cells incubated with levan (section 4.3.2.1) therefore to gain insight into the overall complex co-crystalliation was utilised.

4.3.5.1 Crystallisation

The mature proteins BT1760^{GH32} and BT1761^{SGBP}, without signal peptides, were cloned into pET28a and overexpressed in BL21. The soluble recombinant proteins were purified using IMAC and size exclusion chromatography. These proteins (60 µM each) were mixed with freshly purified BT1762-3 (45 µM) and 2.5 mM levan oligosaccharides before incubation at room temperature for 1 hour. Molecular Dimensions screens MemGold 1 and MemGold 2 were set up in sitting drop using a Mosquito robot. After 48 hours approximately 25 % of the 192 conditions screened had produced crystals (Figure 4.19). A dataset has not been collected from any of these crystals because none of those tested in-house diffracted sufficiently. Therefore it is unknown how many of the proteins in the mixture are present in the crystals ie. if the crystals are multi-protein complexes or just single proteins or BT1762-3 etc. Crystals were harvested, for analysis of the component proteins by SDS-PAGE, but the crystals were difficult to handle and disintegrated when washed.

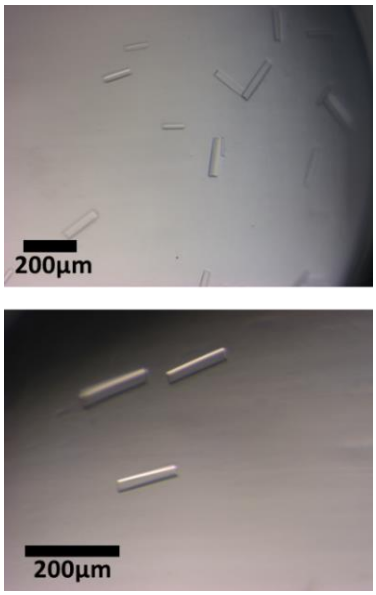


Figure 4.19 Crystallisation screen hits of *B. theta* levan acquisition BT1760-3 complex Crystals produced from a mixture containing approximately 60 µM BT1760^{GH32}, 60 µM BT1761^{SGBP}, 45 µM BT1762-3^{SusCD} and 2.5 mM levan oligosaccharides. Crystallisation conditions; 0.2 M NaCl, 0.05 M Calcium acetate pH 5, 29 % (w/v) PEG 400 (top panel) and 0.04 M Magnesium chloride hexahydrate, 0.05 M NaCl, 0.1 M HEPES pH 7.5, 32 % (w/v) PEG 400 (bottom panel).

4.3.5.2 Native Gel Analysis of the BT1760-3 Complex

A blue native gel (Figure 4.20) was used to investigate the possible oligomerisation and interactions between the outer membrane proteins using the BT1760-3 mixture crystallised in Figure 4.19. The individual components BT1760^{GH32}, BT1761^{SGBP} and BT1762-3^{SusCD} were all prepared with detergent (0.5 % DDM) and used for comparison. The full BT1760-3 complex (lane C) showed four extra bands of higher MW than BT1762-3^{SusCD} alone (Figure 4.20). These bands must be several proteins from the mixture interacting and forming complexes but the exact configuration is unknown. The size of each 'complex' band cannot be estimated because the highest MW protein standard is 320 kDa. Mass spectrometry could be used to identify the very high molecular weight bands. SEC was used by Bert van den Berg to analyse the mixture of proteins but did not indicate that complex formation was occurring as UV peaks were observed for the individual proteins.

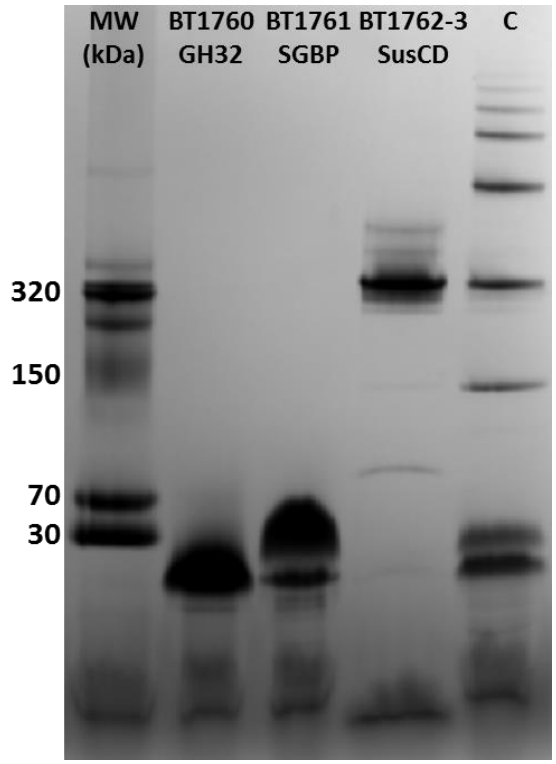


Figure 4.20 Native gel of BT1760-3 levan acquisition complex.
Native-PAGE Bis-Tris 4-16 % polyacrylamide gel (ThermoFisher) showing the soluble components of the complex individually BT1760^{GH32} and BT1761^{SGBP}, the SusCD pair BT1762-3 and the full BT1760-3 complex (C). All samples were prepared with detergent for comparison. MW shows approximate sizes in kDa for four membrane proteins. The 150 kDa protein standard Mep2 appears as a diffuse band.

4.4 Discussion

4.4.1 BT1762-3 Structures

The overall BT1762-3 structures (Figure 4.4 and 4.9) show the same general configuration as BT2263-4; dimeric SusC homologues each covered with a SusD 'lid', with the interface between $BT1762^{SusD}$ and $BT1763^{SusC}$ involving the ligand binding face of the SusD (Figure 4.4D). This suggests the structure is conserved across SusCD homologue complexes. The conformation exposes the conserved TPR domains of $BT1762^{SusD}$ to the cell surface. TPRs domains are often involved in protein-protein interactions and the assembly of multi-protein complexes therefore may form contacts with the levan PUL encoded lipoproteins $BT1760^{GH32}$ and $BT1761^{SGBP}$during glycan import.

The degradation of 30 kDa from the N-terminus of $BT1763^{SusC}$ was completely unexpected (Figure 4.4 and 4.5). The remainder of the barrel remained unaffected by the removal of the plug domain, which suggests the plug could come out of the barrel in its entirety to facilitate the transport of large substrates, a theory which has been the subject of much debate among the TBDT research field (Hickman *et al.*, 2017).

The BT1762-3 structure with levan oligosaccharides bound (Figure 4.9) is highly significant as the first SusCD homologue structure with glycan bound. The substrate is comprised of at least five fructose units with a possible sixth fructose forming a branch (Figure 4.21). the ligand is bound by thirteen residues; seven from $BT1762^{SusD}$ and six from $BT1763^{SusC}$, Figure 4.11. The complex completely encloses the levan oligo suggesting there is a maximum size limit for import (Figure 4.11A).

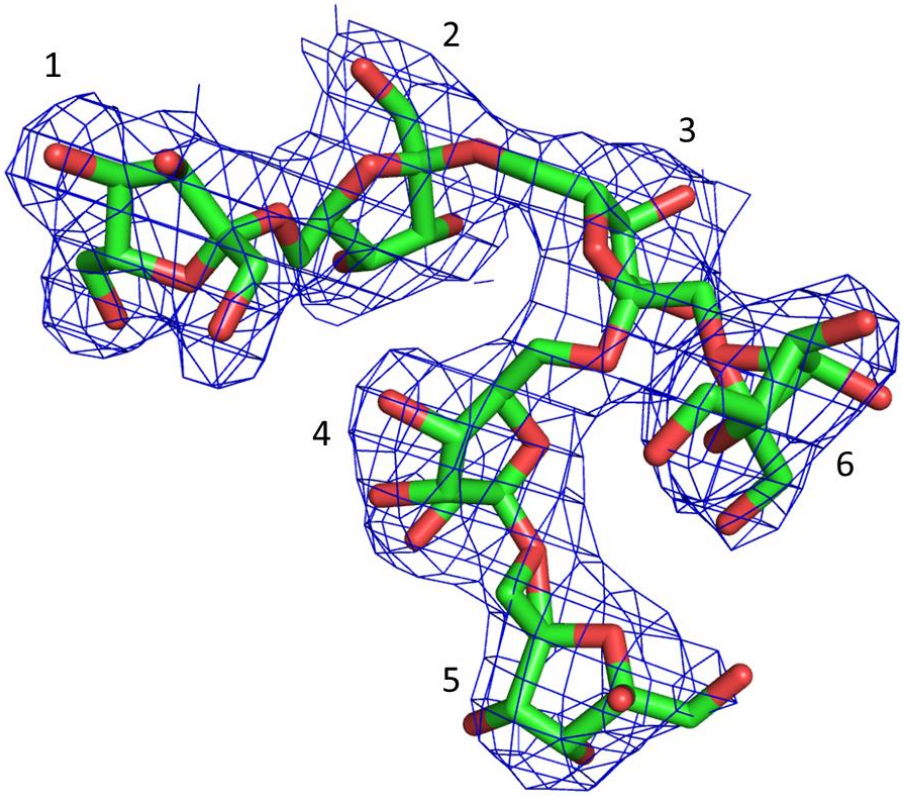


Figure 4.21 Structure and electron density of the levan oligogosaccharide bound by BT1762-3
The BT1762-3 bound levan oligosaccharide consisting of a main chain of five fructose units (1-5) and a possible branch with one additional fructose molecule (6).

4.4.2 Single-channel electrophysiology of BT1762-3

To investigate *in vitro* the findings of the MD simulations of BT2261-4, the spontaneous degradation of the BT1763^{SusC} plug domain was exploited for single-channel electrophysiology. Purified protein was sent to collaborators at Jacobs University (Bremen, Germany) and the experiments were carried out by Satya Bhamidimarri.

The proteins were all re-folded *in vitro* following SDS-PAGE which allowed BT1763^{SusC} to be separated from BT1762^{SusD}. Reconstitution of the full length BT1762-3 SusCD complex and full length BT1763^{SusC} alone in a lipid bilayer generated very small, relatively stable currents (Figure 4.22). This is consistent with a relatively small channel and would be expected with a TBDT due to the plug domain occluding the channel. The complex lacking a plug domain, Δplug-BT1762-3^{SusCD}, produced large channels with average conductance values approximately 1.5 nS. The noisy, unstable trace showed a wide range of conductance values, which indicated dynamic opening and closing of the channel. The truncated complex was separated using SDS-PAGE, Δplug-BT1763^{SusC} was isolated from the gel and refolded before reconstitution in the bilayer. The resulting trace showed a large channel with conductance value around 3 nS with a stable current. This suggested that the extracellular loops of BT1763^{SusC} were not responsible for the gating channel observed in the Δplug-BT1762-3^{SusCD} protein.

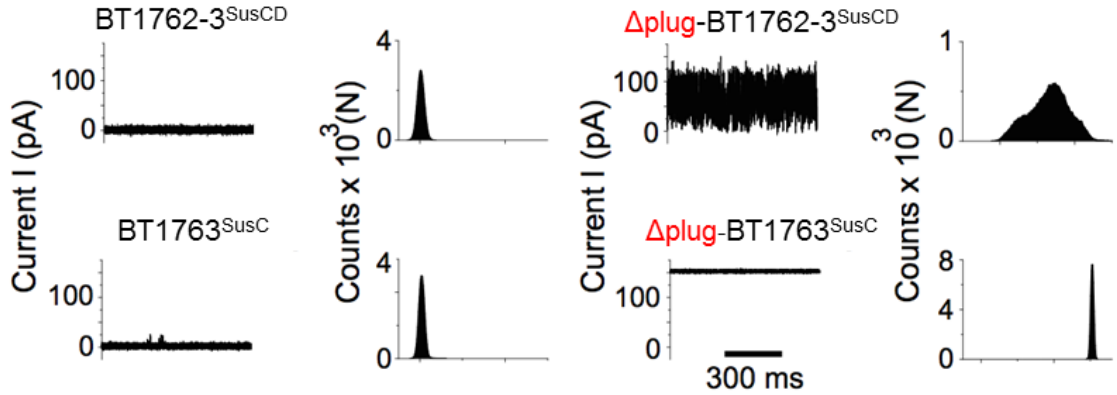


Figure 4.22 Single-channel electrophysiology of BT1762-3 complexes.
BT1762-3^{SusCD} complex and BT1763^{SusC} alone tested with and without the plug domain (*Δplug*). Traces and histograms are representative of ten experiments. *N* represents the number of events, defined as the current value in a 200 µs window. Data obtained at +50 mV. Figure adapted from Glenwright *et al.* 2017.

4.4.3 Investigating the function of BT1762-3

Single base mutagenesis of BT1762^{SusD} indicates six of the seven binding residues are essential for binding of BT1762^{SusD} in isolation to levan (Figure 4.12). Six residues from BT1763^{SusC} complete the binding site. BT1762-3 cannot be purified in high enough quantities for ITC (20 mg per test) therefore a different method would be required to probe binding affinity of the whole complex for levan.

Based on the ITC data of recombinant BT1762^{SusD} W103A which showed that this residue was essential for levan binding, the W103A mutant was made in the genomic copy of BT1762 to produce the *W103A-BT1762* strain. The lack of growth defect for this mutant strain on levan was somewhat unexpected (Figure 4.14) and may be due to the binding site of the complex being composed of multiple residues from both the SusD and the SusC, such that loss of a single residue does not prevent glycan import. In addition it may be that the BT1762-3 SusCD in the *W103A-BT1762* strain functions less efficiently than the WT, but that this reduced efficiency is masked/unable to be observed in the assay used.

MD simulations of BT2261-4 showed the hinge loop (L7) moved with BT2263^{SusD} providing most of the remaining BT2264^{SusC}-BT2263^{SusD} interactions in the open state, and underwent the biggest conformational change (Chapter 3, Figure 3.15B). The conserved loop was deleted in the genomic copy of BT1763^{SusC} but surprisingly only caused a slightly increased lag-phase on MM-Levan (Figure 4.14). Therefore the BT1762^{SusD} 'lid' must retain the ability to open/close without the hinge loop. The dynamic movement of BT1762^{SusD} 'lid' may also involve a second loop L8 of BT1763^{SusC} which is closely associated to the binding protein (Figure 4.9 and 4.23). Further mutations or deletions of the extracellular loops of BT1763^{SusC} could be used to identify the essential contacts between the two proteins which allow the dynamic movement of BT1762^{SusD}.

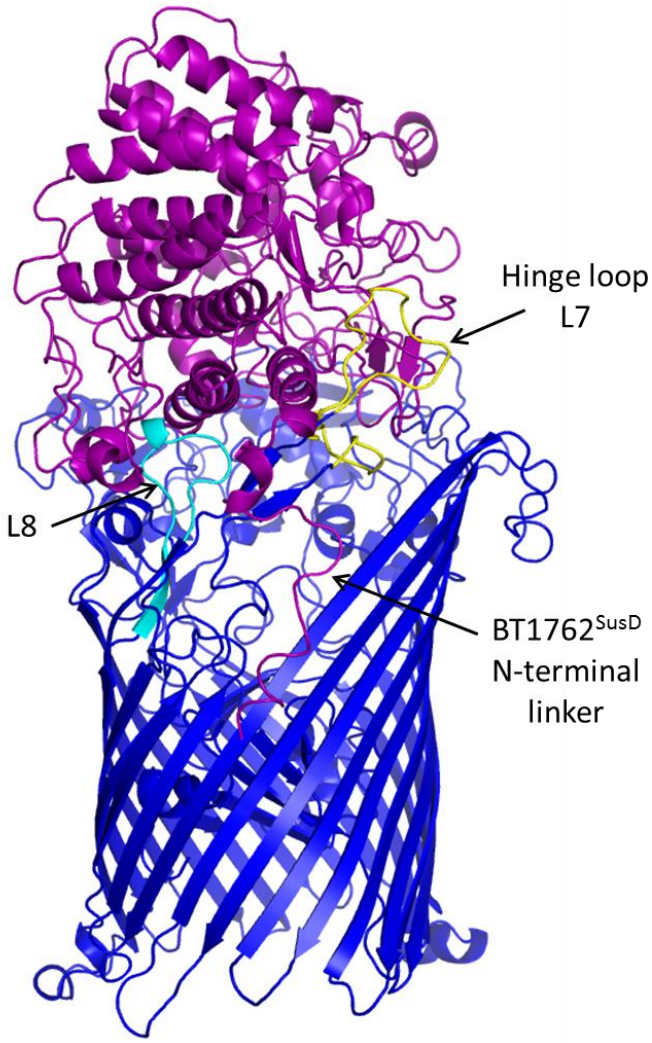


Figure 4.23 Structure of BT1762-3 highlighting extracellular loops 7 and 8
Cartoon representation of BT1762-3 showing BT1762^{SusD} in purple and BT1763^{SusC} in dark blue. The hinge loop (L7) highlighted is yellow and the L8 loop is cyan. The N-terminal linker sequence of BT1762^{SusD} which attaches to the lipid anchor is also highlighted.

A knockout of the DUF4480 domain (*ΔDUF-BT1763*) cannot grow on MM-Levan after 48 hours which suggests the domain is essential for BT1763^{SusC} to transport levan (Figure 4.14). Expression tests of BT1762^{SusD} and BT1763^{SusC} show both proteins in the OM suggesting the phenotype is due to disrupted functioning of BT1763^{SusC} rather than problems with expression (Figure 4.15). DUF4480 is present in ~80 % of SusC homologues from *B. theta*. The domain immediately precedes the TonB box (Figure 4.13B) which interacts with part of the inner membrane

TonB apparatus in the periplasm, providing energy for transport (Hickman *et al.*, 2017). Consequently DUF4480 may also be involved in energy procurement.

4.4.4 BT1761 surface glycan binding protein

Binding of levan to BT1761^{SGBP} measured by ITC confirms that it is a glycan binding protein with higher affinity for levan polysaccharides and oligosaccharides than BT1762^{SusD} (Figure 4.17). SGBPs are thought to interact with and sequester the target glycan at the cell surface for degradation and subsequent import by other proteins from the PUL. This may be why the binding affinity of BT1761^{SGBP} is higher than that of BT1762^{SusD}. Producing well diffracting crystals of BT1761^{SGBP} was difficult; only one dataset has been collected, for which there is only a partial MR solution. To produce a structure anomalous x-ray scattering data will be required as SGBPs do not have sequence homology and finding structural homology is more challenging. Existing structures of SGBPs have several domains and the higher flexibility of multi domain proteins often reduces the chance of crystallisation (Cameron *et al.*, 2012; Tauzin *et al.*, 2016; Cartmell *et al.*, 2017). This may explain the problems producing diffracting crystals of BT1761^{SGBP}. The heparin binding SGBP BT4661 consists of six domains (Figure 4.24) (Cartmell *et al.*, 2017).

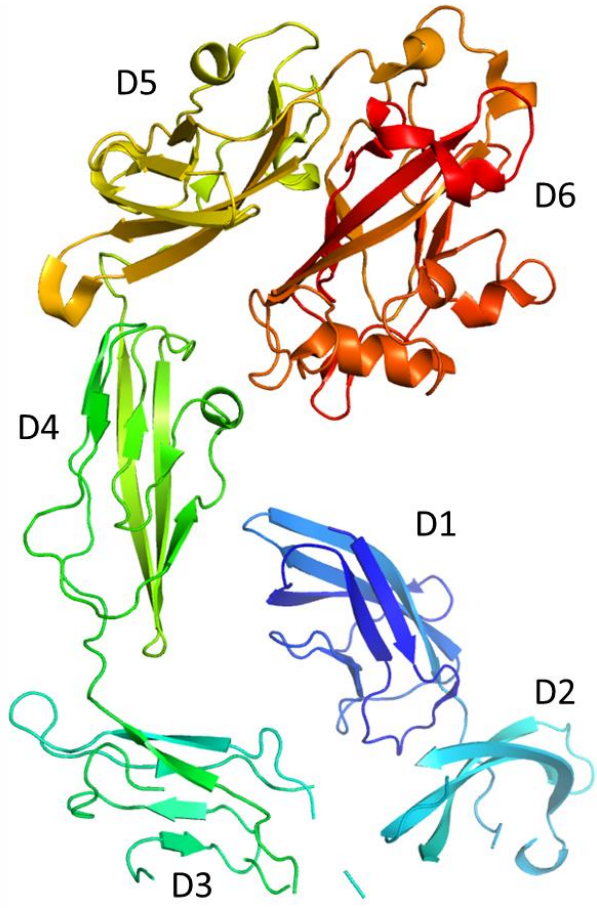


Figure 4.24 Structure of BT4461 SGBP from the *B. theta* heparin PUL
Cartoon representation of BT4661^{SGBP} (PDB ID: 4AK1) coloured blue to red from the N-terminus to C-terminus. The six discrete domains are labelled (D1-6).

4.4.5 Levan Acquisition Outer Membrane Protein Complex

A recent study used live-cell super-resolution microscopy to show dynamic recruitment and assembly of a complex of proteins from the *B. theta* Starch Utilisation System (Karunatilaka *et al.*, 2014). Therefore, in this study, BT1760^{GH32}, BT1761^{SGBP} and BT1762-3^{SusCD} were purified separately and mixed together to produce the OM levan acquisition complex. The native gel of all four components (BT1760-3) shows very large bands over 320 kDa which suggests the proteins were interacting in various combinations (Figure 4.20). Crystals produced from the mixture of proteins (Figure 4.19) did not diffract and SDS-PAGE analysis of the crystal components was unfortunately unsuccessful.

Chapter 5 – Final Discussion

5.1 Structure of SusCD homologues

This study presents the X-ray crystal structures of two functionally distinct SusCD complexes. The relative orientation of SusC and SusD homologues is the same in the two structures (Figure 5.1). BT1762-3 is a classical glycan importing SusC-like protein whereas BT2261-4 likely transports peptides suggesting this configuration could be conserved across SusCD homologues. SusC-like transporters only have ~20 % sequence homology with each other but this study suggests high structural homology. The dimeric conformation observed in both structures is unique among characterised TBDTS and was not a consequence of crystallisation (Noinaj *et al.*, 2010). The transporters were shown to exist as dimers using mass spectrometry, analytical gel filtration and size exclusion chromatography multi-angle light scattering (SEC-MALS) (Glenwright *et al.*, 2017).

The ligand in each complex is bound at the SusCD interface and involves the highly variable SusD binding site (Figure 5.1). This conformation exposes the conserved TPR domains which were previously predicted to be the site of SusC-SusD interactions (Bolam and Koropatkin, 2012). TPRs have been linked with protein-protein interactions, therefore it is feasible that they could mediate interactions between the SusD-likes and the auxiliary surface lipoproteins expressed from the same PUL. The peptide and levan oligosaccharide are completely enclosed within their respective complexes, which suggests a size limit for import by SusC homologues. This may explain the requirement of most PULs to have a functional surface-exposed endo-acting enzyme to degrade the target glycan into small enough oligosaccharides for import (Bolam and Sonnenburg, 2011).

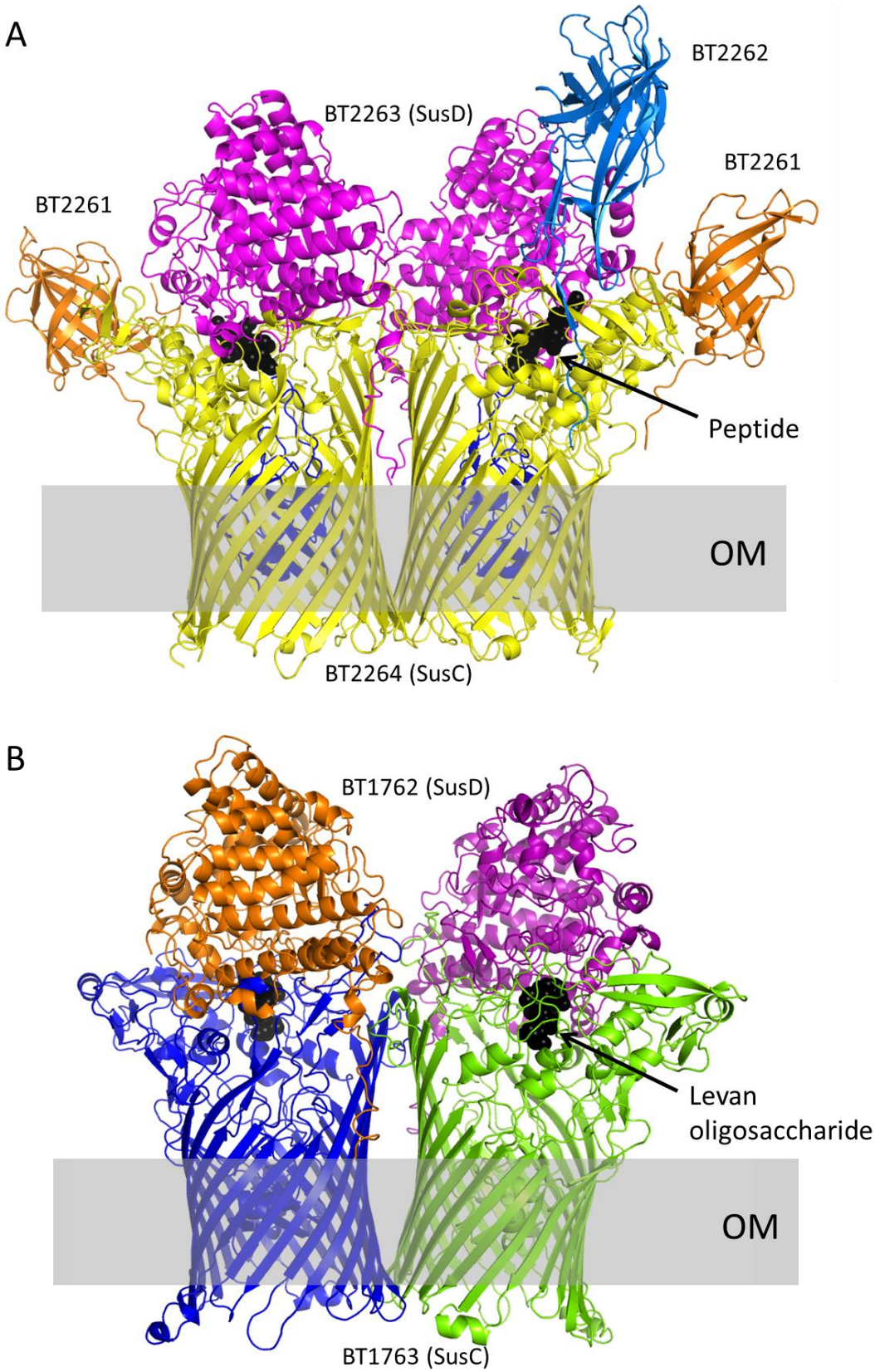


Figure 5.1 Overall architecture of SusCD complexes
Cartoon representation of the SusCD complexes from the side within the OM. **A:** BT2261-4 with the bound peptide shown as black spheres. **B:** BT1762-3 with the bound levan oligosaccharide shown as black spheres.

Purifying proteins directly from the *B. theta* OM was very advantageous and allowed the complexes to remain intact. The association of SusC and SusD homologues within the OM had been predicted previously, but the interactions observed here were robust enough to withstand several rounds of purification with strong detergent. BT2261-4 included two auxiliary lipoproteins which also remained closely associated with the SusCD pair (Figure 5.1). In comparison, BT1762-3 did not co-purify with the other OM lipoproteins from the PUL (BT1760^{GH32} and/or BT1761^{SGBP}), this could be because the complex was not actively processing and importing levan therefore the OM components were not interacting when the cells were harvested (Figure 5.1). Perhaps the interactions between BT1762-3 and the other proteins were not strong enough to maintain the complex during purification because the complex formation may be more dynamic than those observed in BT2261-4.

5.2 Mechanism of Import

The ligand binding site is excluded from solvent, indicating the SusD-like binding protein must move away from the transporter to bind substrate from the environment. MD simulations showed the complex remained closed when the peptide ligand was present (Figure 3.15). However, during most of the apo-complex simulations BT2263^{SusD} moved in a hinge-like motion away from BT2264^{SusC} to expose the ligand binding site (Figure 3.15). This opening/closing dynamic was supported *in vitro* by the single-channel electrophysiology of BT1762-3 which exploited the spontaneous degradation of the plug domain (Figure 4.22).

BT2264^{SusC} loop L7 moves with BT2263^{SusD} in the MD simulations and was required for most of the remaining BT2264^{SusC}-BT2263^{SusD} interactions in the open state. The subsequently named 'hinge loop' is conserved in the BT1763^{SusC} structure, which suggests it could be important for SusC function (Figure 5.2A and B). Surprisingly though, a mutant strain lacking the hinge loop of BT1763^{SusC} (*ΔHinge-BT1763*) showed very little growth defect on levan (Figure 4.15). These data indicate that the function of the complex was mostly unaffected by the loss of the hinge loop and retained the ability to open/close. Therefore a different loop or region could be important in

BT1762-63 and possibly BT2263-64 function. Inspection of the SusC structures reveals a second conserved loop (L8) that could act as a hinge point and it may be that this loop is actually the key critical hinge point between the SusC and SusD, or is able to compensate for the loss of the L7 loop (Figure 5.2A and B). This loop could also be removed and function of the complex could be verified by growing the strain on levan (as with hinge loop L7 in Figure 4.15).

In both complete SusCD structures the N-terminal TonB box is not visible which is consistent with the theory that the domain is flexible, when substrate is bound, to interact with TonB across the periplasm (Hickman *et al.*, 2017). The N-terminal plug domain, which is conserved between TBDTs, is very similar in size, structure and position in the two SusC homologues. However, the long plug loop of $BT2264^{SusC}$ which interacts with the peptide (via Leu120) is not present in $BT1763^{SusC}$ (Figure 5.2C). The two ligands are bound in the same relative location but the equivalent loop is not present in the $BT1763^{SusC}$ structure. This contradicts the idea, explored in Chapter 3 (Figure 3.16), that the interaction between the ligand and plug loop provides the signal to TonB that the binding site is occupied thus instigating the transport cycle.

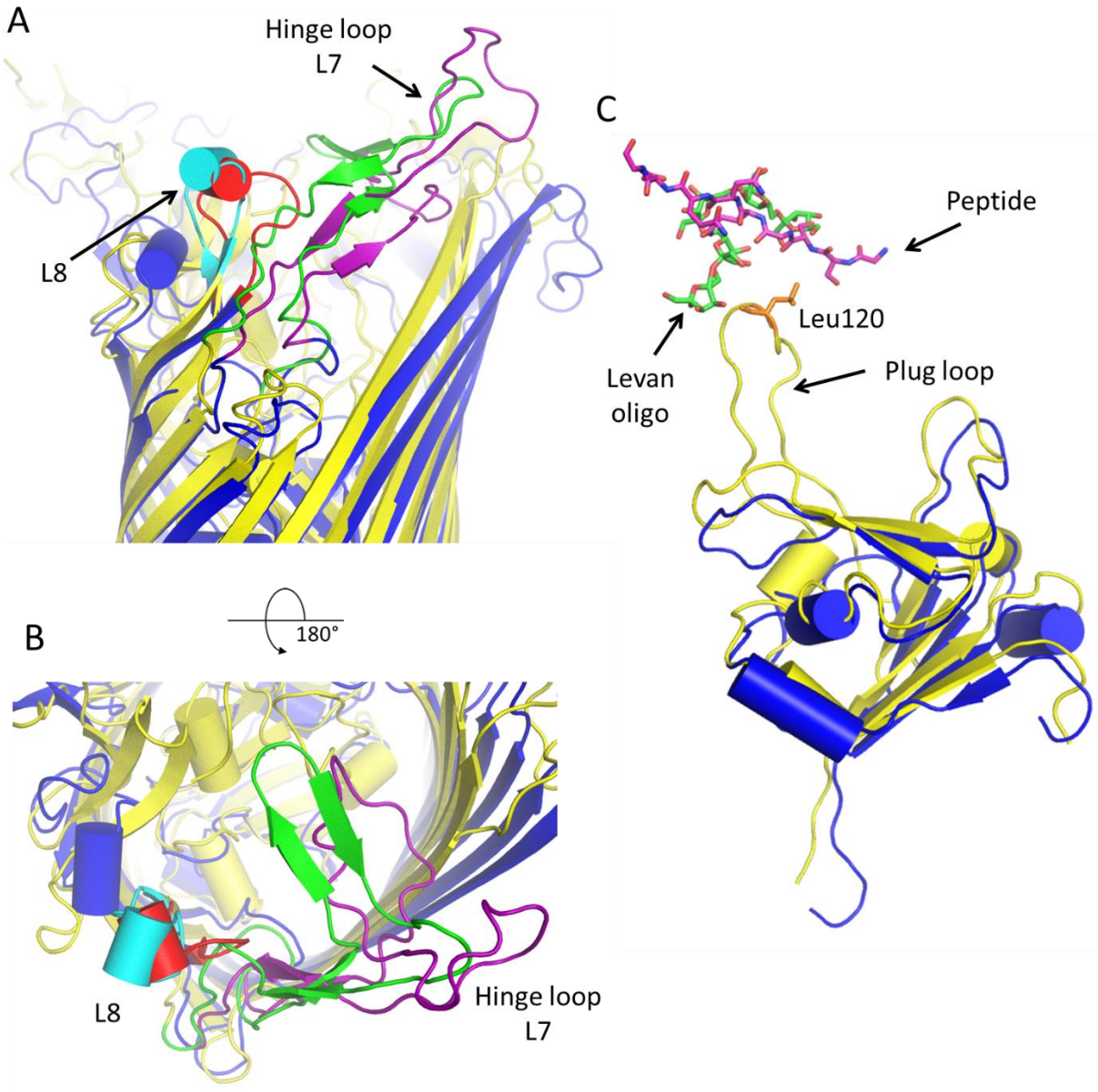


Figure 5.2 Close-up overlays of BT2264^{SusC} and BT1763^{SusC}
A: Overlay of BT2264^{SusC} (yellow) and BT1763^{SusC} (blue) from the side showing the conserved hinge loop (L7) in green (BT2264) and purple (BT1763). Loop 8 is shown in cyan (BT2264) and red (BT1763). Helices are shown as cylinders. **B:** Overlay of BT2264^{SusC} and BT1763^{SusC} (blue) from the outside of the cell (coloured as in **A**). **C:** Overlay of the plug domains of of BT2264^{SusC} (yellow) and BT1763^{SusC} (blue). The bound peptide ligand is shown as magenta sticks and the levan oligosaccharide is shown as green sticks. The plug loop of BT2264^{SusC} is highlighted and Leucine 120 is shown in orange.

Research into SusC-like transporters is just beginning. *B. theta* encodes 103 predicted SusCD homologues of which approximately 20 are expressed as part of characterised PULs. The successful native expression and subsequent purification of two SusCD homologues from *B. theta* has opened up a new avenue for studying highly expressed outer membrane proteins from the *Bacteroides spp.* Questions which remain include, what is the absolute size limit for import by a SusC TBDT? Can heavily branched oligosaccharides be imported? PULs which target highly complex glycans often encode multiple SusCD homologues, so does each transporter target different oligosaccharides from the glycan? Crystal structures of further ligand-bound SusCD homologues would be invaluable for answering these questions.

Regarding the function of the complexes; what is the purpose of the essential DUF4480? Why is the domain only present in ~80% of SusC-like transporters? The open/close mechanism of the SusD 'lid' does not appear to require the Hinge loop as predicted, therefore how much contact is required between the SusC and SusD homologues? Would removing further loops disrupt the system? Which binding residues are required for the complex recognise the substrate? Mutation of the BT1762-3 complex, which must be functional for growth of *B. theta* on levan, could possibly answer these questions and more.

Finally, what is the rate of transport by SusC-like proteins? Some PULs have been shown to lose oligosaccharides at the cell surface which are then scavenged by other members of the HGM (Rogowski *et al.*, 2015). This would suggest the rate of oligosaccharide production by the endo-acting enzymes is significantly faster than the rate of import in such systems. . The rate of transport by TBDTs remains unknown because of the inherent difficulty of studying a system which requires two membranes.

Glycan use by the HGM is a key factor driving community composition so studies which reveal the mechanism used by the HGM to access their major nutrient sources are important for understanding HGM-host symbiosis and the role it plays in human health. This work has addressed

some of the questions regarding the mechanism of nutrient import by Sus-like systems but many questions remain.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. and Zwart, P.H. (2010) 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallographica Section D: Biological Crystallography*, 66(Pt 2), pp. 213-221.

Allan, R.K. and Ratajczak, T. (2011) 'Versatile TPR domains accommodate different modes of target protein recognition and function', *Cell Stress Chaperones*, 16(4), pp. 353-67.

Anderson, K.L. and Salyers, A.A. (1989a) 'Biochemical evidence that starch breakdown by Bacteroides thetaiotaomicron involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes', *Journal of Bacteriology*, 171(6), pp. 3192-3198.

Anderson, K.L. and Salyers, A.A. (1989b) 'Genetic evidence that outer membrane binding of starch is required for starch utilization by Bacteroides thetaiotaomicron', *Journal of Bacteriology*, 171(6), pp. 3199-3204.

Arora, T., Sharma, R. and Frost, G. (2011) 'Propionate. Anti-obesity and satiety enhancing factor?', *Appetite*, 56(2), pp. 511-5.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Dore, J., Antolin, M., Artiguenave, F., Blottiere, H.M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K.U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'Rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S.D. and Bork, P. (2011) 'Enterotypes of the human gut microbiome', *Nature*, 473(7346), pp. 174-80.

Bakolitsa, C., Xu, Q., Rife, C.L., Abdubek, P., Astakhova, T., Axelrod, H.L., Carlton, D., Chen, C., Chiu, H.J., Clayton, T., Das, D., Deller, M.C., Duan, L., Ellrott, K., Farr, C.L., Feuerhelm, J., Grant, J.C., Grzechnik, A., Han, G.W., Jaroszewski, L., Jin, K.K., Klock, H.E., Knuth, M.W., Kozbial, P., Krishna, S.S., Kumar, A., Lam, W.W., Marciano, D., McMullan, D., Miller, M.D., Morse, A.T., Nigoghossian, E., Nopakun, A., Okach, L., Puckett, C., Reyes, R., Tien, H.J., Trame, C.B., van den Bedem, H., Weekes, D., Hodgson, K.O., Wooley, J., Elsliger, M.A., Deacon, A.M., Godzik, A., Lesley, S.A. and Wilson, I.A. (2010) 'Structure of BT_3984, a member of the SusD/RagB family of nutrient-binding molecules', *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 66(Pt 10), pp. 1274-80.

Baslé, A., Rummel, G., Storici, P., Rosenbusch, J.P. and Schirmer, T. (2006) 'Crystal structure of osmoporin OmpC from E. coli at 2.0 A', *J Mol Biol*, 362(5), pp. 933-42.

Beher, M.G., Schnaitman, C.A. and Pugsley, A.P. (1980) 'Major heat-modifiable outer membrane protein in gram-negative bacteria: comparison with the ompA protein of Escherichia coli', *Journal of Bacteriology*, 143(2), pp. 906-913.

Biagi, E., Nylund, L., Candela, M., Ostan, R., Bucci, L., Pini, E., Nikkila, J., Monti, D., Satokari, R., Franceschi, C., Brigidi, P. and De Vos, W. (2010) 'Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians', *PLoS One*, 5(5), p. e10667.

Bolam, D.N. and Koropatkin, N.M. (2012) 'Glycan recognition by the Bacteroidetes Sus-like systems', *Curr Opin Struct Biol*, 22(5), pp. 563-9.

Bolam, D.N. and Sonnenburg, J.L. (2011) 'Mechanistic insight into polysaccharide use within the intestinal microbiota', *Gut Microbes*, 2(2), pp. 86-90.

Brownlee, I.A., Havler, M.E., Dettmar, P.W., Allen, A. and Pearson, J.P. (2003) 'Colonic mucus: Secretion and turnover in relation to dietary fibre intake', *Proceedings of the Nutrition Society*, 62(1), pp. 245-249.

Buchanan, S.K., Smith, B.S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., van der Helm, D. and Deisenhofer, J. (1999) 'Crystal structure of the outer membrane active transporter FepA from Escherichia coli', *Nat Struct Mol Biol*, 6(1), pp. 56-63.

Byndloss, M.X., Olsan, E.E., Rivera-Chávez, F., Tiffany, C.R., Cevallos, S.A., Lokken, K.L., Torres, T.P., Byndloss, A.J., Faber, F., Gao, Y., Litvak, Y., Lopez, C.A., Xu, G., Napoli, E., Giulivi, C., Tsolis, R.M., Revzin, A., Lebrilla, C.B. and Bäumler, A.J. (2017) 'Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion', *Science*, 357(6351), p. 570.

Cameron, E.A., Kwiatkowski, K.J., Lee, B.-H., Hamaker, B.R., Koropatkin, N.M. and Martens, E.C. (2014) 'Multifunctional Nutrient-Binding Proteins Adapt Human Symbiotic Bacteria for Glycan Competition in the Gut by Separately Promoting Enhanced Sensing and Catalysis', *mBio*, 5(5).

Cameron, E.A., Maynard, M.A., Smith, C.J., Smith, T.J., Koropatkin, N.M. and Martens, E.C. (2012) 'Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism', *Journal of Biological Chemistry*, 287(41), pp. 34614-34625.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) 'The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics', *Nucleic Acids Res*, 37(Database issue), pp. D233-8.

Cartmell, A., Lowe, E.C., Basle, A., Firbank, S.J., Ndeh, D.A., Murray, H., Terrapon, N., Lombard, V., Henrissat, B., Turnbull, J.E., Czjzek, M., Gilbert, H.J. and Bolam, D.N. (2017) 'How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans', *Proc Natl Acad Sci U S A*, 114(27), pp. 7037-7042.

Cascales, E., Buchanan, S.K., Duche, D., Kleanthous, C., Lloubes, R., Postle, K., Riley, M., Slatin, S. and Cavard, D. (2007) 'Colicin biology', *Microbiol Mol Biol Rev*, 71(1), pp. 158-229.

Celia, H., Noinaj, N., Zakharov, S.D., Bordignon, E., Botos, I., Santamaria, M., Barnard, T.J., Cramer, W.A., Lloubes, R. and Buchanan, S.K. (2016) 'Structural insight into the role of the Ton complex in energy transduction', *Nature*, 538(7623), pp. 60-65.

Chapla, D., Pandit, P. and Shah, A. (2012) 'Production of xylooligosaccharides from corncob xylan by fungal xylanase and their utilization by probiotics', *Bioresour Technol*, 115, pp. 215-21.

Chen, J., Domingue, J.C. and Sears, C.L. (2017) 'Microbiota dysbiosis in select human cancers: Evidence of association and causality', *Semin Immunol*.

Cho, K.H. and Salyers, A.A. (2001) 'Biochemical Analysis of Interactions between Outer Membrane Proteins That Contribute to Starch Utilization byBacteroides thetaiotaomicron', *Journal of Bacteriology*, 183(24), pp. 7224-7230.

Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) 'Nonchromosomal Antibiotic Resistance in Bacteria: Genetic Transformation of Escherichia coli by R-Factor DNA', *Proceedings of the National Academy of Sciences*, 69(8), pp. 2110-2114.

Cowan, S.W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R.A., Jansonius, J.N. and Rosenbusch, J.P. (1992) 'Crystal structures explain functional properties of two E. coli porins', *Nature*, 358(6389), pp. 727-33.

Cummings, J.H. and Macfarlane, G.T. (1991) 'The control and consequences of bacterial fermentation in the human colon', *J Appl Bacteriol*, 70(6), pp. 443-59.

Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E.A., Pudlo, N.A., Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., Rogowski, A., Hamilton, B.S., Chen, R., Tolbert, T.J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munoz-Munoz, J.L., Day, A., Pena, M.J., McLean, R., Suits, M.D., Boraston, A.B., Atherly, T., Ziemer, C.J., Williams, S.J., Davies, G.J., Abbott, D.W., Martens, E.C. and Gilbert, H.J. (2015) 'Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism', *Nature*, 517(7533), pp. 165-169.

D'Elia, J.N. and Salyers, A.A. (1996) 'Effect of regulatory protein levels on utilization of starch by Bacteroides thetaiotaomicron', *J Bacteriol*, 178(24), pp. 7180-6.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J. and Turnbaugh, P.J. (2014) 'Diet rapidly and reproducibly alters the human gut microbiome', *Nature*, 505(7484), pp. 559-63.

Delcour, A.H. (2003) 'Solute uptake through general porins', *Front Biosci*, 8, pp. d1055-71.

Delcour, A.H. (2009) 'Outer Membrane Permeability and Antibiotic Resistance', *Biochimica et biophysica acta*, 1794(5), pp. 808-816.

Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., Pudlo, N.A., Kitamoto, S., Terrapon, N., Muller, A., Young, V.B., Henrissat, B., Wilmes, P., Stappenbeck, T.S.,

Nunez, G. and Martens, E.C. (2016) 'A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility', *Cell*, 167(5), pp. 1339-1353.e21.

Earle, K.A., Billings, G., Sigal, M., Lichtman, J.S., Hansson, G.C., Elias, J.E., Amieva, M.R., Huang, K.C. and Sonnenburg, J.L. (2015) 'Quantitative Imaging of Gut Microbiota Spatial Organization', *Cell Host and Microbe*, 18(4), pp. 478-488.

Eckburg, P.B., Lepp, P.W. and Relman, D.A. (2003) 'Archaea and their potential role in human disease', *Infect Immun*, 71(2), pp. 591-6.

Eren, E., Vijayaraghavan, J., Liu, J., Cheneke, B.R., Touw, D.S., Lepore, B.W., Indic, M., Movileanu, L. and van den Berg, B. (2012) 'Substrate Specificity within a Family of Outer Membrane Carboxylate Channels', *PLoS Biology*, 10(1), p. e1001242.

Ferguson, A.D., Hofmann, E., Coulton, J.W., Diederichs, K. and Welte, W. (1998) 'Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide', *Science*, 282(5397), pp. 2215-20.

Filip, C., Fletcher, G., Wulff, J.L. and Earhart, C.F. (1973) 'Solubilization of the cytoplasmic membrane of Escherichia coli by the ionic detergent sodium-lauryl sarcosinate', *J Bacteriol*, 115(3), pp. 717-22.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J. and Punta, M. (2014) 'Pfam: the protein families database', *Nucleic Acids Research*, 42(D1), pp. D222-D230.

Foley, M.H., Cockburn, D.W. and Koropatkin, N.M. (2016) 'The Sus operon: a model system for starch uptake by the human gut Bacteroidetes', *Cell Mol Life Sci*, 73(14), pp. 2603-17.

Frost, G., Sleeth, M.L., Sahuri-Arisoylu, M., Lizarbe, B., Cerdan, S., Brody, L., Anastasovska, J., Ghourab, S., Hankir, M., Zhang, S., Carling, D., Swann, J.R., Gibson, G., Viardot, A., Morrison, D., Louise Thomas, E. and Bell, J.D. (2014) 'The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism', *Nat Commun*, 5, p. 3611.

Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K., Tobe, T., Clarke, J.M., Topping, D.L., Suzuki, T., Taylor, T.D., Itoh, K., Kikuchi, J., Morita, H., Hattori, M. and Ohno, H. (2011) 'Bifidobacteria can protect from enteropathogenic infection through production of acetate', *Nature*, 469(7331), pp. 543-7.

Garcia, E.C., Brumbaugh, A.R. and Mobley, H.L. (2011) 'Redundancy and specificity of Escherichia coli iron acquisition systems during urinary tract infection', *Infect Immun*, 79(3), pp. 1225-35.

Gibson, G.R., Probert, H.M., Loo, J.V., Rastall, R.A. and Roberfroid, M.B. (2004) 'Dietary modulation of the human colonic microbiota: updating the concept of prebiotics', *Nutr Res Rev*, 17(2), pp. 259-75.

Gibson, G.R. and Roberfroid, M.B. (1995) 'Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics', *J Nutr*, 125(6), pp. 1401-12.

Glenwright, A.J., Pothula, K.R., Bhamidimarri, S.P., Chorev, D.S., Baslé, A., Firbank, S.J., Zheng, H., Robinson, C.V., Winterhalter, M., Kleinekathofer, U., Bolam, D.N. and van den Berg, B. (2017) 'Structural basis for nutrient acquisition by dominant members of the human gut microbiota', *Nature*.

Goodell, E.W. and Higgins, C.F. (1987) 'Uptake of cell wall peptides by Salmonella typhimurium and Escherichia coli', *J Bacteriol*, 169(8), pp. 3861-5.

Grondin, J.M., Tamura, K., Dejean, G., Abbott, D.W. and Brumer, H. (2017a) 'Polysaccharide Utilization Loci: Fueling Microbial Communities', *J Bacteriol*, 199(15).

Grondin, J.M., Tamura, K., Déjean, G., Abbott, D.W. and Brumer, H. (2017b) 'Polysaccharide Utilization Loci: Fueling Microbial Communities', *Journal of Bacteriology*, 199(15).

Gu, S., Chen, D., Zhang, J.N., Lv, X., Wang, K., Duan, L.P., Nie, Y. and Wu, X.L. (2013) 'Bacterial community mapping of the mouse gastrointestinal tract', *PLoS One*, 8(10), p. e74957.

Guyer, C.A., Morgan, D.G. and Staros, J.V. (1986) 'Binding specificity of the periplasmic oligopeptide-binding protein from Escherichia coli', *J Bacteriol*, 168(2), pp. 775-9.

Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) 'Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter', *J Bacteriol*, 177(14), pp. 4121-30.

Hantke, K. (1976) 'Phage T6--colicin K receptor and nucleoside transport in Escherichia coli', *FEBS Lett*, 70(1), pp. 109-12.

Hedemann, M.S., Theil, P.K. and Bach Knudsen, K.E. (2009) 'The thickness of the intestinal mucous layer in the colon of rats fed various sources of non-digestible carbohydrates is positively correlated with the pool of SCFA but negatively correlated with the proportion of butyric acid in digesta', *British Journal of Nutrition*, 102(1), pp. 117-125.

Hickman, S.J., Cooper, R.E.M., Bellucci, L., Paci, E. and Brockwell, D.J. (2017) 'Gating of TonB-dependent transporters by substrate-specific forced remodelling', *Nat Commun*, 8, p. 14804.

Hill, C., Guarner, F., Reid, G., Gibson, G.R., Merenstein, D.J., Pot, B., Morelli, L., Canani, R.B., Flint, H.J., Salminen, S., Calder, P.C. and Sanders, M.E. (2014) 'Expert consensus document: The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic', *Nat Rev Gastroenterol Hepatol*, 11(8), pp. 506-514.

Jathore, N.R., Bule, M.V., Tilay, A.V. and Annapure, U.S. (2012) 'Microbial levan from Pseudomonas fluorescens: Characterization and medium optimization for enhanced production', *Food Science and Biotechnology*, 21(4), pp. 1045-1053.

Jeong, J.Y., Yim, H.S., Ryu, J.Y., Lee, H.S., Lee, J.H., Seen, D.S. and Kang, S.G. (2012) 'One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies', *Appl Environ Microbiol*, 78(15), pp. 5440-3.

Kabsch, W. (2010) 'XDS', *Acta Crystallographica Section D: Biological Crystallography*, 66(Pt 2), pp. 125-132.

Karunatilaka, K.S., Cameron, E.A., Martens, E.C., Koropatkin, N.M. and Biteen, J.S. (2014) 'Superresolution Imaging Captures Carbohydrate Utilization Dynamics in Human Gut Symbionts', *mBio*, 5(6).

Keegan, R.M. and Winn, M.D. (2007) 'Automated search-model discovery and preparation for structure solution by molecular replacement', *Acta Crystallogr D Biol Crystallogr*, 63(Pt 4), pp. 447-57.

Keith, J., Wiley, B., Ball, D., Arcidiacono, S., Zorfass, D., Mayer, J. and Kaplan, D. (1991) *Continuous culture system for production of biopolymer levan using Erwinia herbicola*.

Kekez, B.D., Gojgic-Cvijovic, G.D., Jakovljevic, D.M., Stefanovic Kojic, J.R., Markovic, M.D., Beskoski, V.P. and Vrvic, M.M. (2015) 'High Levan Production by Bacillus licheniformis NS032 Using Ammonium Chloride as the Sole Nitrogen Source', *Applied Biochemistry and Biotechnology*, 175(6), pp. 3068-3083.

Kien, C.L. (1996) 'Digestion, absorption, and fermentation of carbohydrates in the newborn', *Clin Perinatol*, 23(2), pp. 211-28.

Knights, D., Ward, T.L., McKinlay, C.E., Miller, H., Gonzalez, A., McDonald, D. and Knight, R. (2014) 'Rethinking "enterotypes"', *Cell Host Microbe*, 16(4), pp. 433-7.

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T. and Ley, R.E. (2011) 'Succession of microbial consortia in the developing infant gut microbiome', *Proc Natl Acad Sci U S A*, 108 Suppl 1, pp. 4578-85.

Koropatkin, N.M., Cameron, E.A. and Martens, E.C. (2012) 'How glycan metabolism shapes the human gut microbiota', *Nat Rev Micro*, 10(5), pp. 323-335.

Koropatkin, N.M., Martens, E.C., Gordon, J.I. and Smith, T.J. (2008) 'Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices', *Structure*, 16(7), pp. 1105-15.

Koropatkin, N.M., Martens, E.C., Gordon, J.I. and Smith, T.J. (2009) 'Structure of a SusD Homologue, BT1043, Involved in Mucin O-Glycan Utilization in a Prominent Human Gut Symbiont†,‡', *Biochemistry*, 48(7), pp. 1532-1542.

Laemmli, U.K. (1970) 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature*, 227(5259), pp. 680-5.

Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., Klinter, S., Pudlo, N.A., Urs, K., Koropatkin, N.M., Creagh, A.L., Haynes, C.A., Kelly, A.G., Cederholm, S.N., Davies, G.J., Martens, E.C. and Brumer, H. (2014) 'A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes', *Nature*, 506(7489), pp. 498-502.

Lau, K., Srivatsav, V., Rizwan, A., Nashed, A., Liu, R., Shen, R. and Akhtar, M. (2017) 'Bridging the Gap between Gut Microbial Dysbiosis and Cardiovascular Diseases', *Nutrients*, 9(8).

Lefranc-Millot, C., Guerin-Deremaux, L., Wils, D., Neut, C., Miller, L.E. and Saniez-Degrave, M.H. (2012) 'Impact of a resistant dextrin on intestinal ecology: how altering the digestive ecosystem with NUTRIOSE(R), a soluble fibre with prebiotic properties, may be beneficial for health', *J Int Med Res*, 40(1), pp. 211-24.

Letunic, I., Doerks, T. and Bork, P. (2012) 'SMART 7: recent updates to the protein domain annotation resource', *Nucleic Acids Res*, 40(Database issue), pp. D302-5.

Lichtman, J.S., Ferreyra, J.A., Ng, K.M., Smits, S.A., Sonnenburg, J.L. and Elias, J.E. (2016) 'Host-Microbiota Interactions in the Pathogenesis of Antibiotic-Associated Diseases', *Cell Rep*, 14(5), pp. 1049-61.

Louis, P., Hold, G.L. and Flint, H.J. (2014) 'The gut microbiota, bacterial metabolites and colorectal cancer', *Nat Rev Microbiol*, 12(10), pp. 661-72.

Lupp, C., Robertson, M.L., Wickham, M.E., Sekirov, I., Champion, O.L., Gaynor, E.C. and Finlay, B.B. (2007) 'Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae', *Cell Host Microbe*, 2(3), p. 204.

Lupton, J.R. (2004) 'Microbial degradation products influence colon cancer risk: the butyrate controversy', *J Nutr*, 134(2), pp. 479-82.

Mardo, K., Visnapuu, T., Vija, H., Aasamets, A., Viigand, K. and Alamae, T. (2017) 'A Highly Active Endo-Levanase BT1760 of a Dominant Mammalian Gut Commensal Bacteroides thetaiotaomicron Cleaves Not Only Various Bacterial Levans, but Also Levan of Timothy Grass', *PLoS One*, 12(1), p. e0169989.

Marques, F.Z., Mackay, C.R. and Kaye, D.M. (2017) 'Beyond gut feelings: how the gut microbiota regulates blood pressure', *Nat Rev Cardiol*, advance online publication.

Martens, E.C., Koropatkin, N.M., Smith, T.J. and Gordon, J.I. (2009) 'Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm', *Journal of Biological Chemistry*, 284(37), pp. 24673-24677.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N. and Gordon, J.I. (2011) 'Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts', *PLoS Biol*, 9(12), p. e1001221.

Martinez, I., Stegen, J.C., Maldonado-Gomez, M.X., Eren, A.M., Siba, P.M., Greenhill, A.R. and Walter, J. (2015) 'The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes', *Cell Rep*, 11(4), pp. 527-38.

Martinez, K.B., Leone, V. and Chang, E.B. (2017) 'Western diets, gut dysbiosis, and metabolic diseases: Are they linked?', *Gut Microbes*, 8(2), pp. 130-142.

Metzker, M.L. (2010) 'Sequencing technologies - the next generation', *Nat Rev Genet*, 11(1), pp. 31-46.

Mulder, N.J. and Apweiler, R. (2008) 'The InterPro database and tools for protein domain analysis', *Curr Protoc Bioinformatics*, Chapter 2, p. Unit 2.7.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986) 'Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction', *Cold Spring Harb Symp Quant Biol*, 51 Pt 1, pp. 263-73.

Mysak, J., Podzimek, S., Sommerova, P., Lyuya-Mi, Y., Bartova, J., Janatova, T., Prochazkova, J. and Duskova, J. (2014) 'Porphyromonas gingivalis: major periodontopathic pathogen overview', *J Immunol Res*, 2014, p. 476068.

Ndeh, D., Rogowski, A., Cartmell, A., Luis, A.S., Baslé, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R.A., Zhu, Y., O'Neill, M.A., Urbanowicz, B.R., York, W.S., Davies, G.J., Abbott, D.W., Ralet, M.-C., Martens, E.C., Henrissat, B. and Gilbert, H.J. (2017) 'Complex pectin metabolism by gut bacteria reveals novel catalytic functions', *Nature*, advance online publication.

Newgard, C.B., An, J., Bain, J.R., Muehlbauer, M.J., Stevens, R.D., Lien, L.F., Haqq, A.M., Shah, S.H., Arlotto, M., Slentz, C.A., Rochon, J., Gallup, D., Ilkayeva, O., Wenner, B.R., Yancy, W.S., Jr., Eisenson, H., Musante, G., Surwit, R.S., Millington, D.S., Butler, M.D. and Svetkey, L.P. (2009) 'A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance', *Cell Metab*, 9(4), pp. 311-26.

Ni, J., Wu, G.D., Albenberg, L. and Tomov, V.T. (2017) 'Gut microbiota and IBD: causation or correlation?', *Nat Rev Gastroenterol Hepatol*.

Nikaido, H. (2003) 'Molecular Basis of Bacterial Outer Membrane Permeability Revisited', *Microbiology and Molecular Biology Reviews*, 67(4), pp. 593-656.

Noinaj, N., Buchanan, S.K. and Cornelissen, C.N. (2012) 'The transferrin–iron import system from pathogenic Neisseria species', *Molecular Microbiology*, 86(2), pp. 246-257.

Noinaj, N., Guillier, M., Barnard, T.J. and Buchanan, S.K. (2010) 'TonB-dependent transporters: regulation, structure, and function', *Annu Rev Microbiol*, 64, pp. 43-60.

Ottman, N., Smidt, H., de Vos, W.M. and Belzer, C. (2012) 'The function of our microbiota: who is out there and what do they do?', *Front Cell Infect Microbiol*, 2, p. 104.

Pages, J.M., James, C.E. and Winterhalter, M. (2008) 'The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria', *Nat Rev Microbiol*, 6(12), pp. 893-903.

Panigrahi, P., Parida, S., Nanda, N.C., Satpathy, R., Pradhan, L., Chandel, D.S., Baccaglini, L., Mohapatra, A., Mohapatra, S.S., Misra, P.R., Chaudhry, R., Chen, H.H., Johnson, J.A., Morris, J.G.,

Paneth, N. and Gewolb, I.H. (2017) 'A randomized synbiotic trial to prevent sepsis among infants in rural India', *Nature*, 548(7668), pp. 407-412.

Pao, S.S., Paulsen, I.T. and Saier, M.H. (1998) 'Major Facilitator Superfamily', *Microbiology and Molecular Biology Reviews*, 62(1), pp. 1-34.

Park, J., Kim, M., Kang, S.G., Jannasch, A.H., Cooper, B., Patterson, J. and Kim, C.H. (2015) 'Short-chain fatty acids induce both effector and regulatory T cells by suppression of histone deacetylases and regulation of the mTOR-S6K pathway', *Mucosal Immunol*, 8(1), pp. 80-93.

Perkins-Balding, D., Ratliff-Griffin, M. and Stojiljkovic, I. (2004) 'Iron transport systems in Neisseria meningitidis', *Microbiol Mol Biol Rev*, 68(1), pp. 154-71.

Porter, N.T. and Martens, E.C. (2017) 'The Critical Roles of Polysaccharides in Gut Microbial Ecology and Physiology', *Annu Rev Microbiol*, 71, pp. 349-369.

Postle, K. (2007) 'TonB system, in vivo assays and characterization', *Methods Enzymol*, 422, pp. 245-69.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S.D. and Wang, J. (2010) 'A human gut microbial gene catalogue established by metagenomic sequencing', *Nature*, 464(7285), pp. 59-65.

Raetz, C.R. and Whitfield, C. (2002) 'Lipopolysaccharide endotoxins', *Annu Rev Biochem*, 71, pp. 635-700.

Rajilic-Stojanovic, M., Heilig, H.G., Molenaar, D., Kajander, K., Surakka, A., Smidt, H. and de Vos, W.M. (2009) 'Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults', *Environ Microbiol*, 11(7), pp. 1736-51.

Reeves, A.R., D'Elia, J.N., Frias, J. and Salyers, A.A. (1996) 'A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch', *J Bacteriol*, 178(3), pp. 823-30.

Reichardt, N., Duncan, S.H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K.P., Flint, H.J. and Louis, P. (2014) 'Phylogenetic distribution of three pathways for propionate production within the human gut microbiota', *Isme j*, 8(6), pp. 1323-35.

Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Basle, A., Morland, C., Day, A.M., Zheng, H., Rogers, T.E., Thompson, P., Hawkins, A.R., Yadav, M.P., Henrissat, B., Martens, E.C., Dupree, P., Gilbert, H.J. and Bolam, D.N. (2015) 'Glycan complexity dictates microbial resource allocation in the large intestine', *Nat Commun*, 6.

Roy, C.C., Kien, C.L., Bouthillier, L. and Levy, E. (2006) 'Short-chain fatty acids: ready for prime time?', *Nutr Clin Pract*, 21(4), pp. 351-66.

Salyers, A.A. (1984) 'Bacteroides of the human lower intestinal tract', *Annu Rev Microbiol*, 38, pp. 293-313.

Salyers, A.A., Vercellotti, J.R., West, S.E. and Wilkins, T.D. (1977) 'Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon', *Appl Environ Microbiol*, 33(2), pp. 319-22.

Savage, D.C. (1977) 'Microbial Ecology of the Gastrointestinal Tract', *Annual Review of Microbiology*, 31(1), pp. 107-133.

Schauer, K., Rodionov, D.A. and de Reuse, H. (2008) 'New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'?', *Trends Biochem Sci*, 33(7), pp. 330-8.

Schirmer, T., Keller, T.A., Wang, Y.F. and Rosenbusch, J.P. (1995) 'Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution', *Science*, 267(5197), pp. 512-4.

Scholtens, P.A., Oozeer, R., Martin, R., Amor, K.B. and Knol, J. (2012) 'The early settlers: intestinal microbiology in early life', *Annu Rev Food Sci Technol*, 3, pp. 425-47.

Schulz, G.E. (2002) 'The structure of bacterial outer membrane proteins', *Biochim Biophys Acta*, 1565(2), pp. 308-17.

Schwartz, S., Friedberg, I., Ivanov, I.V., Davidson, L.A., Goldsby, J.S., Dahl, D.B., Herman, D., Wang, M., Donovan, S.M. and Chapkin, R.S. (2012) 'A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response', *Genome Biol*, 13(4), p. r32.

Sean Peacock, R., Weljie, A.M., Peter Howard, S., Price, F.D. and Vogel, H.J. (2005) 'The Solution Structure of the C-terminal Domain of TonB and Interaction Studies with TonB Box Peptides', *Journal of Molecular Biology*, 345(5), pp. 1185-1197.

Sekirov, I., Russell, S.L., Antunes, L.C. and Finlay, B.B. (2010) 'Gut microbiota in health and disease', *Physiol Rev*, 90(3), pp. 859-904.

Shipman, J.A., Berleman, J.E. and Salyers, A.A. (2000) 'Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron', *J Bacteriol*, 182(19), pp. 5365-72.

Shipman, J.A., Cho, K.H., Siegel, H.A. and Salyers, A.A. (1999) 'Physiological Characterization of SusG, an Outer Membrane Protein Essential for Starch Utilization by Bacteroides thetaiotaomicron', *Journal of Bacteriology*, 181(23), pp. 7206-7211.

Silbir, S., Dagbagli, S., Yegin, S., Baysal, T. and Goksungur, Y. (2014) 'Levan production by Zymomonas mobilis in batch and continuous fermentation systems', *Carbohydrate Polymers*, 99, pp. 454-461.

Smits, S.A., Leach, J., Sonnenburg, E.D., Gonzalez, C.G., Lichtman, J.S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J.E., Dominguez-Bello, M.G. and Sonnenburg, J.L. (2017) 'Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania', *Science*, 357(6353), pp. 802-806.

Sonnenburg, E.D., Smits, S.A., Tikhonov, M., Higginbottom, S.K., Wingreen, N.S. and Sonnenburg, J.L. (2016) 'Diet-induced extinctions in the gut microbiota compound over generations', *Nature*, 529(7585), pp. 212-5.

Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N. and Sonnenburg, J.L. (2010) 'Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations', *Cell*, 141(7), pp. 1241-52.

Sonnenburg, J.L. and Backhed, F. (2016) 'Diet-microbiota interactions as moderators of human metabolism', *Nature*, 535(7610), pp. 56-64.

Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.-H., Westover, B.P., Weatherford, J., Buhler, J.D. and Gordon, J.I. (2005) 'Glycan Foraging in Vivo by an Intestine-Adapted Bacterial Symbiont', *Science*, 307(5717), pp. 1955-1959.

Szmelcman, S. and Hofnung, M. (1975) 'Maltose transport in Escherichia coli K-12: involvement of the bacteriophage lambda receptor', *J Bacteriol*, 124(1), pp. 112-8.

Tame, J.R., Murshudov, G.N., Dodson, E.J., Neil, T.K., Dodson, G.G., Higgins, C.F. and Wilkinson, A.J. (1994) 'The structural basis of sequence-independent peptide binding by OppA protein', *Science*, 264(5165), pp. 1578-81.

Tame, J.R.H., Dodson, E.J., Murshudov, G., Higgins, C.F. and Wilkinson, A.J. (1995) 'The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands', *Structure*, 3(12), pp. 1395-1406.

Tancula, E., Feldhaus, M.J., Bedzyk, L.A. and Salyers, A.A. (1992) 'Location and characterization of genes involved in binding of starch to the surface of Bacteroides thetaiotaomicron', *J Bacteriol*, 174(17), pp. 5609-16.

Tauzin, A.S., Kwiatkowski, K.J., Orlovsky, N.I., Smith, C.J., Creagh, A.L., Haynes, C.A., Wawrzak, Z., Brumer, H. and Koropatkin, N.M. (2016) 'Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont', *MBio*, 7(2), pp. e02134-15.

Thaver, D. and Zaidi, A.K. (2009) 'Burden of neonatal infections in developing countries: a review of evidence from community-based studies', *Pediatr Infect Dis J*, 28(1 Suppl), pp. S3-9.

Topping, D.L. and Clifton, P.M. (2001) 'Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides', *Physiol Rev*, 81(3), pp. 1031-64.

Tzianabos, A.O., Pantosti, A., Baumann, H., Brisson, J.R., Jennings, H.J. and Kasper, D.L. (1992) 'The capsular polysaccharide of Bacteroides fragilis comprises two ionically linked polysaccharides', *J Biol Chem*, 267(25), pp. 18230-5.

Van den Ende, W. (2013) 'Multifunctional fructans and raffinose family oligosaccharides', *Frontiers in Plant Science*, 4(247).

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J. (1993) 'Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin', *J Mol Biol*, 229(1), pp. 105-24.

Xu, J. and Gordon, J.I. (2003) 'Honor thy symbionts', *Proc Natl Acad Sci U S A*, 100(18), pp. 10452-9.

Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., Cordum, H., Van Brunt, A., Kim, K., Fulton, R.S., Fulton, L.A., Clifton, S.W., Wilson, R.K., Knight, R.D. and Gordon, J.I. (2007) 'Evolution of Symbiotic Bacteria in the Distal Human Intestine', *PLOS Biology*, 5(7), p. e156.

Yamashita, E., Zhalnina, M.V., Zakharov, S.D., Sharma, O. and Cramer, W.A. (2008) 'Crystal structures of the OmpF porin: function in a colicin translocon', *Embo j*, 27(15), pp. 2171-80.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., Heath, A.C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J.G., Lozupone, C.A., Lauber, C., Clemente, J.C., Knights, D., Knight, R. and Gordon, J.I. (2012) 'Human gut microbiome viewed across age and geography', *Nature*, 486(7402), pp. 222-7.

Zheng, H. (2009) *Nutrient acquisition in a human gut symbiont: molecular analysis of the carbohydrate utilisation apparatus of Bacteroides thetaiotaomicron*. Newcastle University.

Appendix A: Primer List

All primers used in this study are listed in the table below (Table A).

Construct	Fwd/ Rev	Sequence 5'-3'	Restr. site	SLIC
pET28a BT2262	F	CTCCATATGTGCGACAAATCAACAGATGAC	NdeI	No
	R	CTCCTCGAGTTATTATTGTTTTACTATAATGTC	XhoI	No
pET9 BT2262	F	TACTTCCAATCCATGACGTTGGAAAGAGAGGGTGAT	-	Yes
	R	TATCCACCTTTACTGTTATTATTGTTTTACTATAATGTCAAAAGA	-	Yes
pET28a BT2263	F	CTCGCTAGCGGTTGCGACCTGAACATC	NheI	No
	R	CTCCTCGAGTTACTATTTGATATCCCACCA	XhoI	No
pET9 BT2263	F	TACTTCCAATCCATGGTGACGGCGGATTTGATATTCC	-	Yes
	R	TATCCACCTTTACTGTTATTATTTGATATCCCACCAAACCGGAGT	-	Yes
pExchange FLAG-tag BT2263	F1	AAAGAAGATAACATTCGAGTCGACGTGGGATACGGGTGA	SalI	Yes
	R1	CTATTTGTCGTCATCGTCTTTGTAGTCAGCAGCAGCAGCTTTGATATCCCACCA	-	-
	F2	GCTGCTGCTGCTGACTACAAAGACGATGACGACAAATAGTAACCAATAACTTAG	-	-
	R2	CACCGCGGTGGCGGCCGCTCTAGACAGTTTATATTCCCA	XbaI	Yes
FLAG-tag screen	F	GCTGCTGCTGCTGACTACAAAGACGATGACGACAAATAG	-	-
pExchange BT2264 ko	F1	AAAGAAGATAACATTCGAGTCGACCAAAAGCTGCTGGGA	SalI	Yes
	R1	ATGATGCATATAATAATCCGATGTACGAAATTTG	-	-
	F2	CAAATTTCGTACATCGGATTATTATATGCATCAT	-	-
	R2	GGTGGCGGCCGCTCTAGAGATAGAGTAATCAAT	XbaI	Yes
pET28a CT His BT1761	F	TAAGAAGGAGATATACCATGGGCAGTGATGACTTCAAATCC	NcoI	Yes
	R	GTGGTGGTGGTGGTGCTCGAGTTTACACAAGTAGTTGATTG	XhoI	Yes
pET28a NT His BT1761	F	GTGCCGCGCGGCAGCCATATGAGTGATGACTTCAAATCC	NdeI	Yes
	R	GTGGTGGTGGTGGTGCTCGAGTTATTTACACAAGTAGTTGATTG	XhoI	Yes
pB22 NT-22res. trunc. BT1761	F	TTATATTTTCAAGGTCTCGAGGAATACGCAGGAACTGTCGAC	XhoI	Yes
	R	AGCGGGTGTGAATACTCTAGATTATTTACACAAGTAGTTGATTG	XbaI	Yes
pExchange CT 6xHis-tag BT1763	F1	CTCGTCGACGGTAACGACAATATCAAATGG	SalI	No
	R1	TTAGTGGTGGTGGTGGTGGTGAAATCCGATGTTAAGTCCGAA	-	-
	F2	CACCACCACCACCACCACTAACTTACTTCATCAATAAAACGA	-	-
	R2	CTCTCTAGACTGGCTCGGCTTGTGGAAGTC	XbaI	No
pExchange CT 4xAla 6xHis-tag BT1762	F1	AAAGAAGATAACATTCGAGTCGACTGCCCTTACCGGAATCAAC	SalI	Yes

	R1	TTAGTGGTGGTGGTGGTGGTGAGCAGCAGCAGCCCAACCGAAATTCTGTGT	-	Yes
	F2	GCTGCTGCTGCTCACCACCACCACCACCACTAACTAACCATCTTAAAACAA	-	Yes
	R2	CACCGCGGTGGCGGCCGCTCTAGATCTTCGTTTCCTCCCCAG	XbaI	Yes
His-tag screen	R	GTGGTGGTGGTGGTGGTG	-	-
Internal BT1762 sequencing	F	CTGTTCCATTCGAACAGATC	-	-
pExchange Hinge 20res BT1763	F1	AAAGAAGATAACATTCGAGTCGACACAATCTTTCCGTCAGCA	SalI	Yes
	R1	TTTGAAACCACCAGGTGCATAGATAGTATAACGGGC	-	-
	F2	TATGCACCTGGTGGTTTCAAACGCAACCAGATCGGT	-	-
	R2	CACCGCGGTGGCGGCCGCTCTAGAGATCATTCGTTTTATTGA	XbaI	Yes
Hinge mutation sequencing	F	CTCTCCGTCGGGACGGTTC	-	-
pExchange W103A BT1762	F1	AAAGAAGATAACATTCGAGTCGACGAAGAGTGTAGTCGGACATAC	SalI	Yes
	R1	ATTGATATTCGCGTCGGTGGTATTGATACC	-	-
	F2	ACCACCGACGCGAATATCAATGATATATGG	-	-
	R2	CACCGCGGTGGCGGCCGCTCTAGACTGAATCAGTGCTTCGGCAC	XbaI	Yes
W103A sequencing	F	CTACGCCATCTGGGCGAC	-	-
pExchange BT1763 DUF4480 ko	F1	AAAGAAGATAACATTCGAGTCGACCTACGGGAGCATG	SalI	Yes
	R1	AACTTCATCTACCATCTGAATGTTTTGCCCCCACAG	-	-
	F2	CTGTGGGGGCAAAACATTCAGATGGTAGATGAAGTT	-	-
	R2	GGTGGCGGCCGCTCTAGAGAGGGTTACGACGGT	XbaI	Yes
DUF ko sequencing	F	CTGGTATCATGAAGAACA	-	-
BT1762^{SusD} Site Directed Mutagensis				
BT1762 D59A/N61A	F	GCGACAGGTGATGCAATCGCATCTTCCTTCTCT		
	R	AGAGAAGGAAGATGCGATTGCATCACCTGTCGC		
BT1762 D59A	F	GCGACAGGTGATGCAATCAACTCTTCCTTCTCT		
	R	AGAGAAGGAAGAGTTGATTGCATCACCTGTCGC		
BT1762 N61A	F	GCGACAGGTGATGATATCGCATCTTCCTTCTCT		
	R	AGAGAAGGAAGATGCGATATCATCACCTGTCGC		
BT1762 D85A	F	TCGGGAACAGAAGCAGGCGGTGTGTTC		
	R	GAACACACCGCCTGCTTCTGTTCCCGA		
BT1762 C316A	F	CAGATACTGGGCGCATGCGACTTCCAC		
	R	GTGGAAGTCGCATGCGCCCAGTATCTG		
BT1762 R386A	F	GATGACTGGAGCGCAAGCAAGGGACTT		
	R	AAGTCCCTTGCTTGCGCTCCAGTCATC		
BT1762 Y413A	F	AAGAAAGGTTCTGCATGGGCCAGTTCT		
	R	AGAACTGGCCCATGCAGAACCTTTCTT		

Table A Primers used in this study. Sewing PCR requires two set of primers labelled F1/R1 and F2/R2.

Appendix B: Protein Structure PDB codes

All PDB codes of the structures produced in this study are listed in the table below (Table B).

Protein Structure	Expression	PDB code	Resolution
BT2262	Soluble – *E. coli*	5FQ3	3.1 Å
BT2263	Soluble – *E. coli*	5FQ4	1.9 Å
BT2261-4	Native, membrane – *B. theta*	5FQ6	2.8 Å
		5FQ7	3.4 Å
		5FQ8	2.75 Å
BT1762-3 Apo	Native, membrane – *B. theta*	5T3R	3.1 Å
		5T4Y	3.1 Å
BT1762-3 with levan oligosaccharide bound	Native, membrane – *B. theta*	N/A Not deposited/unpublished	2.9 Å

Table B Protein Structure PDB codes. All protein structures produced in this study, listed with expression details and PDB codes.