HETEROLOGOUS EXPRESSION AND SECRETION OF NANOBODIES TARGETING CAMPYLOBACTER JEJUNI FOR INTESTINAL HEALTH APPLICATIONS

RICHARD MCLEAN
Bachelor of Science, University of Lethbridge, 2013

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Chemistry and Biochemistry
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

HETEROLOGOUS EXPRESSION AND SECRETION OF NANOBODIES TARGETING CAMPYLOBACTER JEJUNI FOR INTESTINAL HEALTH APPLICATIONS

RICHARD MCLEAN

Date of Defence: August 22nd, 2016

Dr. D. Wade Abbott Co-Supervisor	Research Scientist, Adjunct Professor	Ph.D.
Dr. Steven Mosimann Co-Supervisor	Associate Professor	Ph.D.
Dr. G. Douglas Inglis Thesis Examination Committee Member	Research Scientist, Adjunct Professor	Ph.D.
Dr. L. Brent Selinger Thesis Examination Committee Member	Professor	Ph.D.
Dr. Cory L. Brooks External Examiner California State University Fresno, California	Assistant Professor	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

Dedication

To my family, for convincing me to return to school and for your support throughout. You may not have always understood what I was talking about, but your interest and excitement was a constant motivator.

To Sutherland, your endless patience while I stared blankly at these chapters for hours without producing a line of text is something I won't forget. I know it doesn't always seem like I appreciate when you push me to keep going, but I still wouldn't be finished if it wasn't for you. Thank you.

Abstract

As strategies for engineering the enteric microflora continue to advance, the vision of a future with a secondary, artificial immune system increasingly comes into focus. There are numerous challenges that will need to be overcome before this is a reality including the construction of a chassis capable of producing functional antimicrobial compounds at adequate concentrations and the construction of libraries of antimicrobial compounds capable of targeting a range of pathogens including those that develop resistance.

Towards these ends, I have elected to engineer *Bacteroides thetaiotaomicron* (*B. theta*), one of the most prevalent and stable organisms in the human distal gut, to heterologously express and secrete nanobodies that bind the flagella of *Campylobacter jejuni*. Nanobody genes were inserted behind native *B. theta* promoters and integrated into the genome allowing for induction following the introduction of a specific inducing compound. Signalling peptides were fused to the nanobodies allowing for targeting the nanobodies out of *Escherichia coli* (*E. coli*) cells. Also as part of this project, novel signal peptides have also been characterized allowing for the targeting of protein to any subcellular compartment within a gram-negative bacterium.

Table of Contents

List of Tables

List of Figures

Acronyms and Abbreviations

AHL - acyl homoserine lactone

AMAN - acute motor axonal neuropathy

ANOVA - analysis of variance

Asn - asparagine

Asp - aspartic acid

ATCC - American Type Culture Collection

B. fragilis - Bacteroides fragilis

B. ovatus - Bacteroides ovatus

B. theta - Bacteroides thetaiotaomicron

BglII - *Bacillus globigii* restriction endonuclease II

BHIS - supplemented brain heart infusion

BspHI - *Bacillus* species *H* (D. Hall) restriction endonuclease I

Bt-# - *Bacteroides thetaiotaomicron* gene #

C. jejuni - Campylobacter jejuni

CAD - Canadian dollar

CadF - adhesin protein

CAZY - Carbohydrate Active Enzyme Database

CAZyme - carbohydrate active enzyme

CD - Crohn's disease

cDNA - complementary DNA

CheZ - chemotaxis regulator Z

cT - cycle threshold

DNA - deoxyribonucleic acid

DnaG - DNA primase

dUMP - deoxyuridine monophosphate

dTMP - deoxyuridine triphosphate

E. coli - Escherichia coli

ECF-σ - extracytoplasmic function σ factor

Fc - fragment conserved/crystallisable/constant

FlaA - flagella filament protein A

FlaB - flagella filament protein B

FlgM - flagellar secreted protein

Fuc - fucose

FUdR - 5-Fluoro-2'-deoxyuridine

Fv - fragment variable

Gal - galactose

GalNAc - N-acetylgalactosamine

gapdh - glyceraldehyde-3-phosphate dehydrogenase

GBS - Guillain-Barré Syndrome

GC content - guanidine/cytosine content

GD1 - disialotetrahexosylganglioside

gDNA - genomic DNA

Glu - Glutamine

GluS - glutamate-tRNA ligase

Gly - glycine

GM1 - monosialotetrahexosylganglioside

gmk - guanylate kinase

GMO - genetically modified organism

GQ1 - tetrasialoganglioside

GT1 - trisialotetrahexosylganglioside

gyrA - DNA gyrase subunit A

HCAb - heavy chain antibody

HG - homogalacturonan

hIL10 - human interleukin 10

HKGF2 - human keratinocyte growth factor-2

HRP - horse radish peroxidase

Hsp90α - heat shock protein 90α

HTCS - hybrid two component system

IBD - inflamatory bowel disease

IBS - irritable bowel syndrome

ICE - integrative and conjugative element

IgA - immunoglobulin α

IgD - immunoglobulin δ

IgE - immunoglobulin ε

IgG - immunoglobulin γ

IgM - immunoglobulin μ

lpxC - UDP-3-O-acyl N-acetylglucosamine deacetylase

JlpA - adhesin protein

kDa - kilodalton

KpnI - *Klebsiella pneumoniae* restriction endonuclease I

L. lactis - Lactococcus lactis

LB - lysogeny broth

LOS - lipooligosaccharide

maf4 – glycosyl transferase

MAP Kinase - mitogen-activated protein kinase

mcat - malonyl-CoA-acyl carrier protein transacylase

MccS - microcin S

MCS - multiple cloning site

mIL2 - mouse interleukin 2

mIL6 - mouse interleukin 6

mIL10 - mouse interleukin10
mRNA - messenger RNA
MUCLIN - mucin-like glycoprotein
NcoI - *Nocardia coralline* restriction endonuclease I
NeuNAc - N-acetylneuraminic acid
NF-κB - nuclear factor κB
NheI - *Neisseria mucosa heidelbergensis* restriction endonuclease I
OD_{600} - optical density at 600 nm
P. aeruginosa - *Pseudomonas aeruginosa*
P75 - polysaccharide utilization locus 75
PG - pectic galactan
pgk - phosphoglycerate kinase
pH - power of hydrogen scale
PHAC - Public Health Agency of Canada
PstI - *Providencia stuartii* restriction endonuclease I
PUL - polysaccharide utilization locus
PVDF – polyvinylidene fluoride
RCF - relative centrifugal force
ReA - reactive arthritis
recA - recombinase A
RGI - rhamnogalacturonan I
RGII - rhamnogalacturonan II
rho - transcriptional terminator
RIN - RNA integrity number
RNA - ribonucleic acid
RNAP - RNA polymerase
rplI - 50S ribosomal subunit protein L9
rplQ - 50S ribosomal subunit protein L17
rpm - rotations per minute
rpoD - RNA polymerase, σ 70 factor
rpoS - RNA polymerase, σ 38 factor
rRNA - ribosomal RNA
rrsA - 16S rRNA
RT-qPCR - real time quantitative polymerase chain reaction
S. gordonii - *Streptococcus gordonii*
SCFA - short chain fatty acids
scFv - single-chain variable fragment
SEC - general secretory pathway
Ser - serine
spp. - *Species pluralis*
srp - signal recognition particle
SUS - starch utilization system

TAT - twin-arginine translocation
tdk - thymidine kinase
TGF-β1 - transforming growth factor β1
Thr - threonine
thyA - thymidylate synthase
TonB - transport associated protein
TTFC - tetanus toxin fragment C
TYG - tryptone yeast extract glucose
UC - ulcerative colitis
Usp45 - ubiquitin specific protease 45
VPI - Virginia Polytechnic Institute
wlaN - a β-1,3 galactosyltransferase
XbaI - *Xanthomonas badrii* restriction endonuclease I
XhoI - *Xanthomonas holicola* restriction endonuclease I
YebF - small lipoprotein

Chapter 1. Introduction

Humankind's desire to manipulate life is an ancient one. Early humans domesticated fierce, wild wolves more than 30,000 years ago (Skoglund, 2011; Ratnakumar, 2013), the result of which would become "man's best friend." In an effort to establish an agrarian lifestyle, wild plants were first domesticated more than 12,000 years ago, which was followed by the Neolithic founder crops over 9,000 years ago (Lev-Yadun, 2000). Most modern livestock was domesticated between 8,000-10,000 years ago (Bruford, 2003), companion cats (Vigne, 2004) 9,000 years ago, chickens 7,000 years ago (Liu, 2006), and horses 5,000 years ago (Outram, 2009). Wild plants and animals were improved upon by selective breeding to enrich for traits that were appealing, such as physical strength or taste. Although a fundamental understanding of genetics was still several millennia away, artificial selection of gene pools was a common practice that transformed civilization with the goal of improving our quality of life.

As the understanding of genetics and the ability to manipulate genetics expanded through advancing technologies, the desire to manipulate life grew accordingly. In 1973, Cohen *et al.* constructed and replicated recombinant plasmids for the first time and Itakura *et al.* expressed the first heterologous proteins in 1977. Molecular cloning techniques improved to the point that allowed for the design of the first transgenic animal in 1974 (Jaenisch, 1974) with the first transgenic animal capable of passing its transgenes to its offspring in 1981 (Gordon, 1981). The first transgenic plant came shortly after in 1983 (Bevan, 1983). In the mid 1990's, as the first complete genomes of bacteria (Fleischmann, 1995), archaea (Bult, 1996), and eukaryotes (Goffeau, 1996) were sequenced, a plethora of genetic information became available allowing for a better understanding of how genetics functioned on a molecular level.

Although many advances had been made, genetic engineering – or the design of a genetic system for a biotechnological purpose – was still relatively primitive due to a lack of information about the complexity of genetic networks. Early genetic engineering was typically simple, with transgenic organisms producing one or a couple of proteins that were regulated by basic responsive systems. Crops have been produced with herbicide resistance or the ability to secrete natural pesticides (Duke, 2015), the first transgenic plant brought to market was the Flavr Savr tomato which produced an antisense transcript to reduce the level of a polygalacturonase expressed resulting in a slow ripening tomato (Kramer, 1994). Currently, the great majority of transgenic animals are mice and the nematode *Caenorhabditis elegans* used in the laboratory setting.

As the field of transcriptomics and metabolomics has matured, the influx of "meta"-information of how genes are regulated and function has fostered more contemporary engineered systems with increasing levels of sophistication. By the early 2000's, synthetic biology began to grow in popularity due to its multidisciplinary nature. This field collectively engages biologists, chemists, and engineers under a single vision of science to address world problems as diverse as hunger, global warming, and disease. Synthetic biology is a burgeoning subfield of molecular biology which couples the genetic manipulation of traditional genetic engineering with established engineering principles such as incremental rational design, standardization, modularity, and flexibility along with rigorous testing and anticipation of change.

Advances in the field have been used to tackle a diverse assortment of problems from targeting cancer, producing novel organic compounds, biofuel synthesis, carbon capture, and the emulation of digital circuits. Carbon capture techniques have been coupled to biofuel production in the hope of simultaneously reducing our carbon footprint and producing an alternative to fossil

fuels (Atsumi, 2008; Huo, 2011). Engineering complex networks has allowed for the production of novel compounds such as artemisinin which has been used to treat malaria and certain cancers (Martin, 2003). The construction of digital circuit emulators has allowed for the production of synthetic edge detectors (Tabor, 2009), Boolean logic gates (Tamsir, 2011), and "RNA computing (Win, 2008)."

Perhaps one of the most exciting prospects for synthetic biology is preventing and curing disease. Direct engineering of the human genome or engineering of microorganisms associated with the human body has the potential to give these cells the ability to sense bacteria, viruses, or cancerous cells and produce compounds specific for affecting those targets. In this project, I will investigate the potential for engineering the human enteric mutualist *Bacteroides thetaiotaomicron* (*B. theta*) for the production and secretion of compounds that will clear the enteric pathogen *Campylobacter jejuni* from the primary production host (i.e. chickens), in the hope of reducing the burden of gastroenteritis on the general population and the economy by producing safer poultry products. The cost of an average case of gastroenteritis in Canada including health care and loss of productivity was 1,089-1,342 CAD as of 2004 (Majowicz, 2006; Henson, 2008). In Canada, there were an average of 10,147 reported cases of campylobacteriosis per year between 2000 and 2013 (PHAC Notifiable Diseases On-Line, 2016). The burden of campylobacteriosis is believed to be much more extensive; however, as the disease is chronically under reported and it is estimated that for every reported case, there are an additional 22-48 cases that go undetected (Thomas, 2006). These estimates amount to campylobacteriosis costing the Canadian economy and health care system several hundred million CAD per year prior to the inclusion of hospitalization and long-term care expenses for the more severe secondary sequela.

1.1. *Previous Enteric Engineering Approaches* (Figure 1.1)

There have been a number of organisms engineered for enteric applications, most of which were focused on generating vaccines in the gut. Organisms such as *Lactococcus lactis* and *Escherichia coli* have been engineered for a broader range of applications, but there are currently limitations that will need to be overcome if they are to be implemented for long-term, large scale applications.

1.1.1 *Lactococcus lactis* (Figure 1.1A)

The first and most commonly engineered microorganism for the manipulation of the enteric microflora is *L. lactis*, a lactic acid bacterium commonly found in the human diet due to its use in manufacturing numerous dairy products including cheese, yogurt, buttermilk, and sour cream. Heterologous expression and secretion of proteins were demonstrated in the late 1980's (Van de Guchte, 1989, 1990) and *L. lactis'* potential as an enteric delivery organism was recognised shortly thereafter.

In 1990, Iwaki *et al.* used *L. lactis* subsp. *lactis* (formerly *Streptococcus lactis*) to heterologously express a 190 kDa surface protein of *Streptococcus mutans*. After expression, cell death was induced with formalin and the resulting solution was successfully used to immunize mice against *S. mutans*. Although immunization was successful, expression levels were low (0.2% dry weight) and the protein was expressed exclusively in the cytoplasm making it ineffective as an autonomous delivery system. The system was improved by incorporating the T7-RNA polymerase based expression system and a signal peptide from an unidentified secreted 45 kDa protein (Usp45).
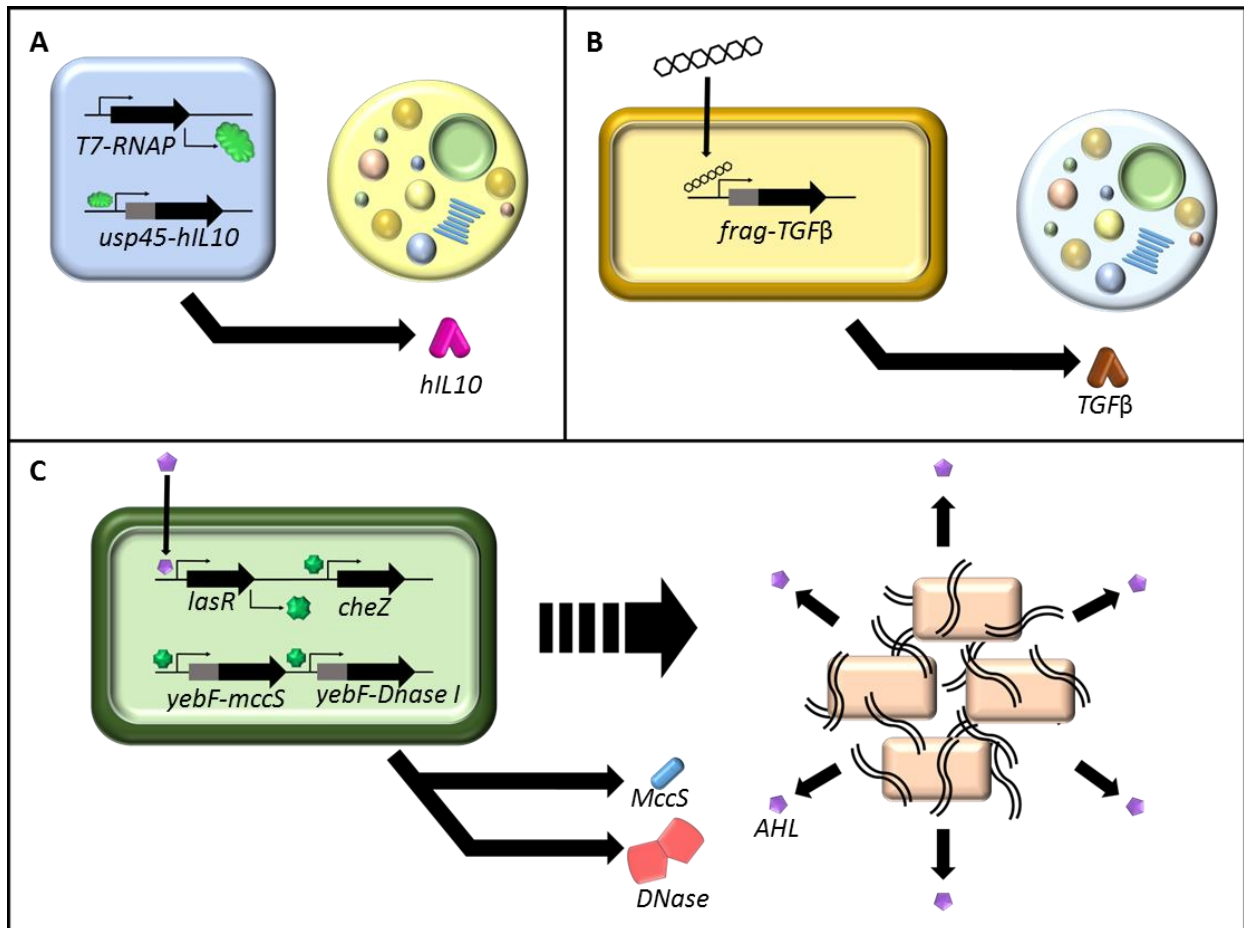
Figure 1.1. Previous enteric engineering approaches. **A.** *L. lactis* was engineered to produce hIL10 with a Usp45 signal peptide which resulted in secretion of the protein. Protein production was regulated by the expression of T7-RNAP. **B.** *E. coli* was engineered to produce MccS, a broad spectrum bacteriocin as well as DNaseI with a yebF signal peptide that resulted in secretion. cheZ was expressed allowing for chemotaxis towards *P. aeruginosa*. Upon arrival, the DNase allowed for degradation of biofilms and the MccS caused cell death. **C.** *B. ovatus* was engineered to produce TGFβ with an enterotoxin secretion tag which resulted in the protein being secreted. Protein production was regulated by exposure to xylan.

The first use of live cells to express biologically active, heterologous protein in the enteric microbial ecosystem was by Steidler *et al.* in 1998. Usp45 was fused to either murine interleukin-2 or -6 (mIL2 or mIL6) cytokines. Tetanus toxin fragment C (TTFC) was co-expressed to elicit a specific immune response and it was determined that anti-TTFC IgA antibody titres were 10-15 fold greater when mIL2 or mIL6 were co-expressed with TTFC rather than TTFC alone suggesting that the cytokines were expressed on site and they were biologically active. Since then, *L. lactis* has been used to produce mIL10 and trefoil factors in attempts to prevent and ameliorate colitis in mice with positive results (Steidler, 2000; Vandenbroucke, 2004).

In an attempt to test this model in human beings, *L. lactis* was engineered to produce human interleukin-10 (hIL10) (Braat, 2006). hIL10 was successfully produced and was demonstrated to be biologically active using bioassays; however, the authors felt that for use in people, additional biosecurity measures should be put in place. To achieve this goal, *L. lactis* thymidylate synthase gene (thyA) was replaced by hIL10; thyA is responsible for the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). In the absence of this enzyme, cells must be supplemented with dTMP, in order to replicate DNA. It was demonstrated that the strain was uninhibited in the presence of dTMP but cell density was reduced by several orders of magnitude in its absence *in vitro*, and the bacterium was undetectable in the patients.

Human trials began in 2006 with a test group of ten subjects with varying severities of Crohn's Disease. Subjects were given ~10^{11} CFU lyophilized *L. lactis* with hIL10 genomically integrated in place of thyA twice daily for seven days. Of the ten subjects, eight showed signs of improvement with five going into complete remission. It was concluded that the method was

both safe, and that transgenic strains were incapable of surviving in the absence of supplemented dTMP.

L. lactis shows considerable promise for future engineering due to the relative ease of growing and manipulating the organism's genetics. Many tools are available including vectors for genomic integration and plasmid-based expression as well as signal peptides targeting proteins to the membrane and the extracellular environment. The greatest drawback to the use of *L. lactis* for these applications is its transient nature in the gut (Zhang, 2016). A result of this is the need to regularly reapply the strain orally. The aforementioned human trial was successful in reducing the symptoms of IBD, but required the subjects to ingest 20 *L. lactis* pills daily and after the trial was complete, half of the subjects positively affected had a significant relapse. This suggests that for applications in chronic disease, treatment would need to be continued indefinitely. *L. lactis* may still be useful for the short-term delivery of a compound, for example a vaccine or proof of concept applications, but long-term therapy is impractical.

1.1.2 *Escherichia coli* (Figure 1.1C)

An appealing bacterium for the manipulation of the enteric flora is *E. coli*. *E coli* was the first organism used to build and replicate a recombinant plasmid (Cohen, 1973) and the first used to express heterologous protein (Itakura, 1977). Since then, it has been a cornerstone of molecular biology research and is one of the best characterized organisms. This coupled with the fact that it is easy to grow and manipulate in the laboratory and that it is a natural commensal resident in many mammalian colons makes *E. coli* appear to be an ideal organism for long term treatment of colitis.

E. coli was engineered by Saeidi *et al.* (2011) with the goal of eliminating enteric *Pseudomonas aeruginosa*. Expression of pyocin S5, a narrow-spectrum antimicrobial peptide, or "bacteriocin", naturally produced by *P. aeruginosa* was chosen for this purpose. In order to hone the system, an acyl homoserine lactone (AHL), the quorum sensing molecule produced by *P. aeruginosa*, was used to detect the presence of the pathogen. Upon binding of the AHL, the cells were induced to express Pyocin S5 as well as the E7 lysis protein. Once Pyocin S5 was produced, the cells would lyse releasing their payload.

This system is effective but results in cell death of the engineered strain and not all *P. aeruginosa* strains are sensitive to Pyocin S5. Gupta *et al.* (2013) improved the system by replacing Pyocin S5 with an engineered bacteriocin hybrid composed of Pyocin S3 and Colicin E3 and fusing it with the FlgM signal peptide allowing it to be exported from the cell.

In order to make the system more robust, Saeidi *et al.* replaced Pyocin S5 with MccS, a more broad-spectrum bacteriocin and fused it to the YebF signal peptide for secretion (Hwang, 2013). In addition to the bacteriocin, a DNase I was added with the same signal peptide allowing the degradation of biofilms. Finally, the cells were equipped with CheZ, a receptor that allowed for chemotaxis toward its target in response to AHL. Taken together, this strain of *E. coli* can recognise the presence of *P. aeruginosa* and move towards it allowing for a more targeted attack, then degrade any protective biofilms present, and secrete an antimicrobial peptide causing cell death to its target.

E. coli is a versatile bacterium that has proven to be extremely amenable to genetic manipulation. The massive toolbox of genetic elements in the literature previously tested in *E. coli* make it a natural choice for engineering applications. The greatest detriment to using *E. coli* for applications such as this is its low cell density in the gut. In order to produce enough protein to

induce an effect, many cells must produce a small amount of protein each, few cells must produce a massive amount of protein each, or an intermediate therein. If the metabolism of these few cells has been fine-tuned to maximize expression of the protein of choice, it may prove difficult to maintain competitiveness in this highly complex environment. In addition, *E. coli* typically colonizes the intestinal lumen (Li, 2015). As a result, if the target is the epithelium (i.e. the production of immunomodulators) or epithelia associated pathogens (i.e. *Campylobacter jejuni*) a large portion of the protein produced will be lost to diffusion prior to reaching the target. Further engineering may by-pass these restrictions but for the time being, it is a legitimate road block for *in vivo* applications.

1.1.3 Bacteroides ovatus (Figure 1.1B)

An increasingly popular view to enteric engineering is to approach the problem from the other direction. Rather than finding an organism that is easy to grow and modify and then try to make it competitive in the gut, groups have been taking the most successful organisms from the gut microbiome and learning to engineer them. Although the tools and techniques are still in their infancy, the advantage to this method is that organisms can be selected for traits that suit the project including physical location, resource niches, and prevalence.

The *Bacteroides* genus is a group of anaerobic, gram-negative, rod-shaped bacteria which are resistant to host bile and do not form endospores. They are among the most common genera in the human gut commonly comprising ~25 % of the healthy human colonic microbiome (Duncan, 2008; De Filippo, 2010; Human Microbiome Project Consortium, 2012).

Farrar *et al*. (2005) engineered a strain of *B. ovatus* to express mIL2 in the presence of xylan. mIL2 was fused to the signal peptide from the *Bacteroides fragilis* enterotoxin fragilysin which

allowed for secretion from the cell. The same system was used to produce and secrete human keratinocyte growth factor-2 (HKGF2) into the gut of intestinally compromised mice (Hamady, 2010). Bioactive HKGF2 was detected and had a positive effect on gut health including reduction of weight loss, reduction of rectal bleeding, and improved stool consistency. The system was subsequently used to express TGF-β_1 with similar positive effects including accelerated epithelial healing and reduction of proinflammatory cytokines (Hamady, 2011). Long term persistence was not investigated so the stability of these strains is not known.

1.2. B. theta Represents a Promising Alternative System for Enteric Engineering

Like other Bacteroides species, *B. theta* is a glycophile, or sugar-loving bacterium. Its genome encodes an impressive 306 catabolic carbohydrate active enzymes (CAZymes), more than three times as many as the human genome (CAZyme database, 2016a,b) allowing it to catabolise many complex carbohydrates refractory to human digestion. As a result, *B. theta* is able to survive and thrive in the gut of an organism with a diet as diverse as human beings.

1.2.1 Carbohydrate Utilization

In spite of the massive array of CAZymes encoded by the *B. theta* genome, its genetics are highly organized. *Bacteroides* contain Polysaccharide Utilization Locus (PUL) systems, which are independently regulated pathways dedicated to the metabolism of discreet polysaccharides (Bjursell, 2006). Archetypical proteins, involved in the transport and modification of glycans are known as "Sus-like", referring to their homology with the *B. theta* starch utilization system (Sus) proteins from PUL66 (Shipman, 2000).
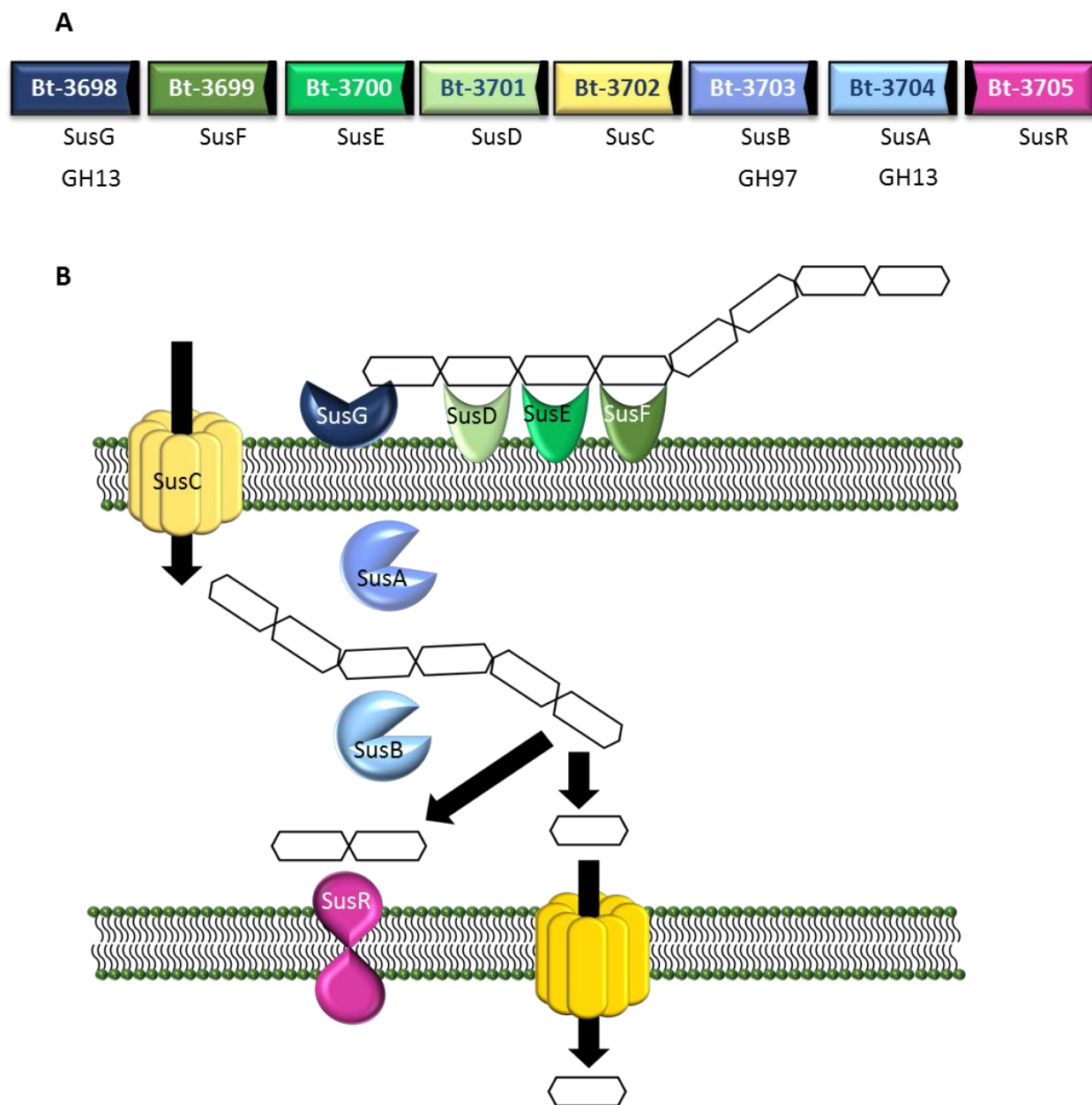
Figure 1.2. A. A genetic schematic and **B.** a model of the eight genes encoded by PUL66. SusD, SusE, and SusF allow for binding of starch, SusG cleaves, and SusC transports oligomers into the periplasm. In the periplasm, SusA and SusB further process the oligomers. SusR binds maltose, upregulating the PUL and monomers are transported into the cytoplasm by proteins not coded by the PUL. The SUS paradigm is used to detect "sus-like" genes during identification of new PULs.

PUL66 is comprised of eight genes, *susR,A,B,C,D,E,F,G* associated with the recognition, transport, and catabolism of starch (Figure 1.2); all genes are constitutively expressed at low levels. In the presence of starch, the outer membrane lipidated proteins SusD,E,F bind the polymer allowing for hydrolysis by SusG, which converts α-glucans into smaller chain maltooligosaccharides (D'Elia, 1996b). These products are then transported into the periplasm by SusC, a TonB-dependent receptor (Shipman, 2000). After entering the periplasm, oligosaccharides are further hydrolyzed by SusA and SusB (Shipman, 2000). Maltose, which is generated as a product during this cascade, binds to the inner membrane bound SusR, a transmembrane transcriptional regulator that upregulates the expression of PUL66 (D'Elia, 1996a). Finally, monomeric glucose is transported into the cytoplasm by constitutive MFS transporters, which are not regulated by SusR (Shipman, 2000).

Most PULs share a similar architecture to PUL66 (Martens, 2009; Rogowski, 2015). The core defining feature of a PUL is the presence of at least one *susC/SusD* pair, which have evolved to bind their respective substrates. Transport of more complex glycans have been shown to require multiple SusC/D pairs (Martens, 2011; Cuskin, 2015). *susA, susB*, and *susG* genes are more diversified and are typically replaced by analogous enzyme genes [i.e. glycoside hydrolases, polysaccharide lyases, and other carbohydrate active enzymes (CAZymes)]. The function of PUL enzymes are to remove the side chains and modifications, and depolymerize the targeted glycan. In most cases, *susR* is replaced with functionally analogous ECF-σ/anti-σ pairs (Braun, 2005) or hybrid two-component systems (HTCSs) (Sonnenburg, 2006). HTCSs are structurally and functionally similar to traditional two-component systems with the exception that the two proteins exist as a single polypeptide.

In total, the genome of *B. theta* contains 267 identified glycoside hydrolases from 47 different families, 15 polysaccharide lyases from 8 families, and 18 carbohydrate esterases from 9 families, the majority of which are associated with its 88 PULs (CAZy Database, 2016a). In comparison, human beings only have 98 glycoside hydrolases from 30 different families and a single carbohydrate esterase, the majority of which are associated with glycoprotein processing (Carbohydrate-Active enZYmes Database, 2016b). PULs make up nearly 1/5 of the *B. theta* genome, highlighting the importance of carbohydrate metabolism for this organism.

1.2.2 Promise for Future Engineering

B. theta is an intriguing bacterium for engineering applications in the gut. It is present at the highest concentration of any intestinal *Bacteroides* ($\sim 10^{10}$ per gram dry weight of feces) (Salyers, 1984; Bervoets, 2013) and is generally considered a commensal, if not a mutualist. The massive variety of carbohydrate inducible promotors present in its genome will allow for the construction of highly complex regulatory networks and it has been shown that promoters can be used interchangeably between *Bacteroides* species. It is amongst the most stable species in the gut over 5 year periods (Faith, 2013) which would allow for long term therapeutic strategies. In addition, although *Bacteroides* spp. are typically thought to have a relatively narrow host specificity, *B. theta* is common in humans and mice, it has been identified in chickens, cows, pigs, goats (Atherly, 2014) and camels (Bhatt, 2013) suggesting that it can thrive in a wide variety of animal species, possibly as a result of strain specificity.

Some *B. theta* engineering resources are currently available. These include protocols for conjugation (Bandoh, 1992), transfection (Burt, 1977), and chemical (Smith, 1985) and electrocompetent transformation (Thomson, 1989; Ichimura, 2010). Multiple strategies for conjugation and genomic integration of genes into *B. theta* have been tested and a library of 140

ribosomal binding sites have been published (Mimee, 2015). Although few promoters have been tested for induction of heterologous genes, extensive transcriptomics have been performed in this bacterium and information is available on gene induction profiles (Sonnenburg, 2005; Bjursell, 2006; Martens, 2008; Benjdia, 2011; Marcobal, 2011; Martens, 2011; Lynch, 2012; TerAvest, 2014). This information can be harnessed to develop promoters that respond to distinct carbohydrates and have contrasting levels of expression.

1.2.3 Nutritional Benefits

The colonic microbiome, the collective of microorganisms in the colon, is well-equipped to catabolize the majority of complex carbohydrates it encounters. This metabolic bioreactor provides a core symbiosis with the host. The by-products of glycan catabolism by *B. theta*, and other distal gut members, is the production of short chain fatty acids (SCFA) such as acetate and propionate (Tremaroli, 2012). These SCFAs represent an essential nutrient for the intestinal epithelium and represents as much as 10% of the daily caloric intake of human beings (Bergman, 1990). As a result, germ-free animals require 30% greater caloric intake than wild type animals (Wostmann, 1983).

In addition to unlocking a potential energy source from otherwise indigestible dietary fibre, *B. theta* induces a number of host genes to actively improve nutrient uptake by the host. For example, Na^+/glucose transporters have been shown to be upregulated after colonization with *B. theta* as well as fatty acid-binding protein, pancreatic lipase-related protein-2, apolipoprotein A-IV, and colipase, all of which are involved in the absorption of lipids by the host (Hooper, 2001).

1.2.4 Intestinal Health

The relationship between *B. theta* and human beings extends beyond ensuring that both are well nourished. The presence of *B. theta* in the gut has been shown to positively affect the health of intestinal epithelial cells and to prevent infection by dangerous pathogens through colonization resistance or competitive exclusion, whereby the few cells of an invading pathogen are forced to compete with the established population of *B. theta* for resources, and against the host immune system primed by *B. theta* (Buffie, 2013).

The intestinal mucosa is the first line of defence against pathogens. In the absence of bacteria, there is a dramatic reduction in the surface area of the epithelia due to shorter villi (McCracken, 2001). Epithelial turnover is also considerably slower in germ-free animals which may account for the abnormal morphology (McCracken, 2001). Decorating the intestinal epithelia is a layer of mucin, a densely packed carbohydrate/protein matrix that functions to physically prevent penetration to the epithelia by pathogens. The composition of mucin is dependent on the expression of glycosyl transferases, CAZymes responsible for the polymerization of carbohydrates. Which transferases are expressed is determined in part by the presence of modulins, soluble factors produced by commensal enteric bacteria (Freitas, 2003). Inoculation with *B. theta* causes an increase in Fucα 1-2 Galβ 1- 4 GlcNAc, Galβ 1-4 GalNAc, and a decrease in NeuNAcα 2-6 Gal (Freitas, 2005). It is unknown what the net result of these changes is, but it has been demonstrated that different enteric commensals can elicit different host glycosylation patterns.

1.2.5 Immunomodulation

B. theta has been implicated in the activation of the innate and adaptive immunity. It has been shown to stimulate the production of angiogenin-4, an antimicrobial protein expressed by Paneth cells in intestinal crypts to protect intestinal stem cells (Hooper, 2003), MUCLIN, a component of the mucus layer that can bind gram negative and gram positive bacteria (Hooper, 2001), and matrilysin, an enzyme involved in defensin maturation (López-Boado, 2000). In comparison to germ-free animals, the Peyer's patches and lymph nodes of animals colonized by *B. theta* are larger, IgG and IgA titres are higher, and even extra-intestinal organs such as the thymus and spleen are better developed (Shroff, 1995). In addition to priming effects such as these, *B. theta* actively helps to repair inflammatory damage to the intestine caused by the immune system. *B. theta* can activate the phosphatidylinositol-3-kinase pathway and reverse damage to the intestinal epithelia caused by proinflammatory cytokines induced by pathogenic organisms such as *Salmonella enterica, E. coli* O157:H7, and *Campylobacter jejuni* (Resta–Lenert, 2006).

1.3. Campylobacter jejuni

The *Campylobacter* genus is a group of microaerophilic, gram negative, spiral shaped bacteria which are catalase and oxidase positive and incapable of fermenting glucose. They are the most common cause of bacterial enteritis in the developed world (Bresee, 2012; Tam, 2012, PHAC Notifiable Diseases On-Line, 2016). Although campylobacters can be isolated from water and soil (Ross, 2006; Khan, 2014), they are typically associated with the gut of animals such as livestock, wild game, pets, and importantly, birds. Although there are several avenues for the contraction of campylobacteriosis, as much as 70% can be traced back to the consumption of contaminated birds (Epps, 2013) and as many as 95% of cases are caused by *Campylobacter jejuni* specifically (Gillespie, 2002; Doorduyn, 2010).

1.3.1 Pathology

Campylobacteriosis is generally sporadic and self-limiting with infections typically lasting 7-14 days (Young, 2007) and is equivalent to infections caused by *Salmonella* and *Shigella* species (Acheson, 2001). Symptoms include gastroenteritis causing fever, abdominal pain, nausea, and diarrhea. Although most cases are resolved without antibiotics, some patients have continued diarrheal disease that can last for several weeks while other patients are completely asymptomatic (Acheson, 2001). Infection with *C. jejuni* can trigger a number of secondary sequela with symptoms lasting long after the bacterium is eliminated from the body. In immunocompromised individuals and neonates, campylobacteriosis has been indicated as an antecedent of meningitis, osteomyelitis, cholecystitis, carditis, and pancreatitis. In healthy adults, infection has been implicated in the triggering of numerous debilitating diseases, such as Reactive Arthritis (ReA), Inflammatory Bowel Disease (IBD), and Guillain-Barré Syndrome (GBS).

1.3.1.1 Reactive Arthritis

Reactive Arthritis is the inflammation of joints caused by an infection at a disparate site on the body (Morris, 2012). It is unclear specifically how this disorder develops, but hypotheses include molecular mimicry and the production of Arthritogenic peptides particularly in HLA-B$_{27}$+ individuals. Symptoms include joint pain and stiffness, frequent, painful urination, and eyeball inflammation. Full recovery is common after several months to a year, however severe, chronic arthritis can develop. Numerous studies have investigated Campylobacteriosis as an antecedent of ReA with variable results; between 0.7-24 percent of cases of infection resulted in ReA (Gumpel, 1981; Bremell, 1991).

1.3.1.2 Inflammatory Bowel Disease

Inflammatory Bowel Disease is the chronic inflammation of the gastrointestinal tract (Hanauer, 2006). The principle manifestations of IBD are Ulcerative Colitis (UC) and Crohn's Disease (CD). UC is limited to the most superficial layer of the colon and symptoms include fatigue and weight loss, abdominal pain and cramping, bloody diarrhea, and rectal bleeding. CD is typically limited to the colon and ileum, however it can spread further into the intestine and generally affects the surface and underlying layers. Symptoms are generally similar to that of UC but also include oral ulcers and inflammation of joints, eyes, and the liver. Long-term treatment for both UC and CD are available and UC can effectively be cured with a complete proctocolectomy, there is no cure for CD. IBD has been correlated with a number of cancers including small intestinal cancer and colorectal cancer.

1.3.1.3 Guillain-Barré Syndrome

Guillain-Barré Syndrome is a disorder characterized by acute peripheral neuropathy (Hughes, 2005). The neuropathy is caused by an autoimmune response whereby antibodies bind to the myelin sheath of neurons of the peripheral nervous system resulting in demyelination. Symptoms usually start with pain and weakness in the hands and feet and, as the demyelination progresses, muscle weakness often spreads to the diaphragm causing difficulty breathing and may require a ventilator. In severe cases, demyelination can spread to the autonomic nervous system as well causing issues with blood pressure and heart rate. If it is allowed to progress, paralysis can spread through the entire body, occasionally causing death.

Twenty to fifty percent of cases of GBS are preceded by campylobacteriosis (Baker, 2012). *C. jejuni* lipooligosaccharides (LOS) are structurally similar to human gangliosides GM1, GM1b,

GD1a, and GalNAc-GD1a which can trigger a subtype of GBS called acute motor axonal neuropathy (AMAN) as well as GQ1b and GT1a which can trigger Fisher's Syndrome, a variant of Acute Pandysautonomia, another variant of GBS.

1.3.2 C. jejuni Infection

In order for *C. jejuni* to establish an infection in a new host, an infectious dose of between 500 and 800 cells must be introduced to the system (Epps, 2013). Once in the gut, it must evade the immune system and penetrate mucus to gain access to the epithelium. *C. jejuni* has evolved a number of traits all of which are important in helping it achieve this goal (Young, 2007). From an engineering perspective, any and all of these traits could be used as targets for the delivery of antimicrobial compounds (Figure 1.3). A serendipitous benefit of this approach is that as the organism inevitably mutates the proteins associated with these traits to develop resistance to the antimicrobial compounds, pathogenicity and the ability to colonize will likely be reduced.

1.3.2.1 Lipooligosaccharides

Lipooligosaccharides (LOS) are truncated versions of lipopolysaccharides (LPS) containing both the Lipid A anchor and core polysaccharide like an LPS, but lacking the O-antigen. They are integral portions of the outer membrane of many gram-negative bacteria and are important for pathogens due to their immunomodulatory properties. *C. jejuni* has highly variable LOS with at least 19 unique classes defined by the genotyping of their LOS biosynthetic loci (Houliston, 2011). Some classes of LOS mimic the structure of human antigens such as DU-PAN-2, a pancreatic tumour-associated antigen, lacto-N-neotetraose, a P blood group antigen, and a
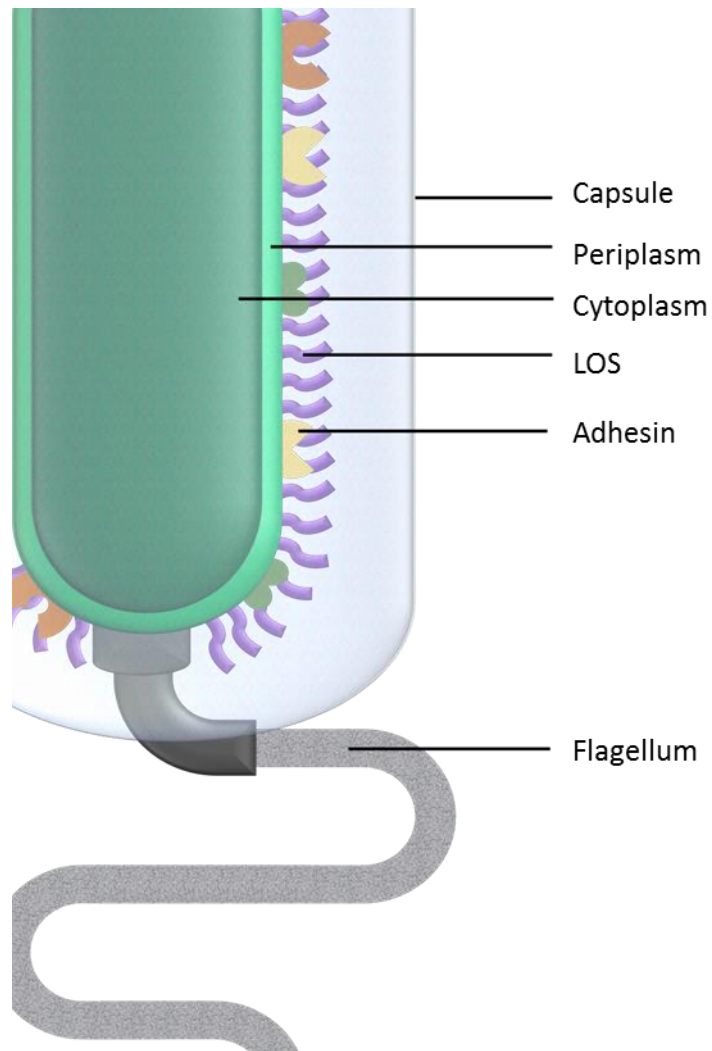
Figure 1.3. Potential nanobody targets on the surface of *C. jejuni. C. jejuni* contains a number of extracellular structures that have been shown to be important for colonization and pathogenicity. Targets such as flagella and adhesins may be targeted to reduce motility or epithelial retention respectively which will result in clearing the bacterium from the gut. Targets such as the LOS and capsule may be used as points of attachment for the delivery of antibacterials or recruitment of the immune system.

number of human gangliosides such as GM1, GM1b, GD1a, GalNAc-GD1a, GQ1b and GT1a linked to GBS.

1.3.2.2 Bacterial Capsule

The capsule is a complex network of carbohydrates found outside of the outer membrane (Karlyshev, 2000). The composition is highly variable between strains and is covalently linked to the membrane. The capsule has been shown to be important for evasion of the immune system and serum resistance and is likely involved in resistance to some viruses and the human compliment system. Knockout mutants have been used to show that strains lacking a capsule have reduced cell adherence and are less successful at colonizing the chicken intestine (Jones, 2004).

1.3.2.3 Protein Glycosylation

Protein glycosylation is common to all kingdoms of life and serves a multitude of purposes including facilitating protein folding, increasing protein stability, reducing proteolytic degradation, and aiding cell adhesion. *C. jejuni* has pathways for both O-linked and N-linked glycosylation and currently has the best characterized bacterial N-linked glycosylation pathway (Szymanski, 1999). For N-linked glycosylation, a heptasaccharide is translocated from the cytoplasm into the periplasm where glycosyl transferases attach it to the common motif: Asp/Glu-X-Asn-X-Ser/Thr (Kowarik, 2006). More than 50 glycoproteins have been identified in *C. jejuni*, however it is expected that as many as 150 glycoproteins are possible. The functions of many glycoproteins are still unknown; however, they have been implicated in cellular adherence, colonization, and the reduced productions of proinflammatory cytokines such as IL6 by host cells (Nothaft, 2016).

1.3.2.4 Adhesins

Cellular adherence, or the ability to bind to and persist on the surface of target cells, is important for colonization. Many pathogens produce fimbriae to bind extracellular glycans and assist in cell adhesion. *C. jejuni* is unusual in that does not produce fimbriae, instead it relies on a number of adhesin proteins to fulfill this roll. JlpA, an adhesin that binds Hsp90α, has been shown to both aid in cell adhesion and induce a proinflammatory response with NF-κB and MAP kinase (Jin, 2001). CadF, another adhesin binds fibronectin, a glycoprotein important for human-human cell adhesion (Konkel, 2005); loss of CadF has a dramatic negative effect on colonization.

1.3.2.5 Flagella

Flagella serve a number of important functions in bacteria, principally locomotion, but also sensing, adhesion, and secretion (Kazmierczak, 2013). *C. jejuni* is either monotrichous or amphitrichous depending on the strain (Kazmierczak, 2013). A complex of proteins in the inner and outer membranes function as a motor and a filament composed of flagellar protein A and B (FlaA and FlaB) is attached to the motor by a hook protein complex (Kazmierczak, 2013).

C. jejuni movement has been described as "rapid" and "darting" and is optimally suited for movement through viscous solutions such as the mucus layer of the intestinal epithelia (Ferrero, 1988). *C. jejuni* has been shown to be motile in higher viscosity solutions that restrict the motility of other enteric bacteria such as *E. coli, Salmonella enteritidis*, or *Vibrio cholera* (Ferrero, 1988). In addition, *C. jejuni* displays greater directional flexibility as viscosity increases (Ferrero, 1988). Immobilized, flagellated *C. jejuni* have been used to demonstrate that functional, locomotive flagella are required for virulence, colonization, and epithelial invasion (Grant, 1993).

Flagella are important for more than just locomotion, they are essential for cell adherence, biofilm formation, and secretion of proteins (Kazmierczak, 2013). Because *C. jejuni* lacks a traditional Type III secretion system, the flagellum is used as a transporter to secrete *Campylobacter* invasion antigens (cia) (Konkel, 2004). The function of these proteins is not understood, but knockout mutants are internalized into epithelia at a lower rate than wild type strains, suggesting they are important for pathogenicity (Buelow, 2011).

1.3.2.6 Genetic Manipulation

C. jejuni is uniquely suited to modify itself genetically in a rapidly changing environment. It is known possess several hyper-variable regions within its genome (Parkhill, 2000) and to go through a process known as phase variation, rapid, reversible, stochastic mutations that can cause changes in a proteins structure or expression, often in an on/off manner. Phase variation in *C. jejuni* is mediated by poly G/C tracts within open reading frames due to slipped-strand mispairing and can influence the epitopes present on the cell surface (Bayliss, 2012). For example wlaN, a β-1,3 galactosyltransferase responsible for galactosylating the LOS (Linton, 2000) and maf4, an enzyme responsible for glycosylation of the flagella (Van Alphen, 2008), in some strains are both prone to being turned on or off under different conditions resulting in different glycosylation patterns on the cell surface. Some studies have suggested that recombination events in *C. jejuni* result in variation at twice the rate of *de novo* mutations and this coupled with the regular exchange of genetic material with other species leads to relatively rapid evolution within this genus. In this light, it was predicted that *C. jejuni* may experience speciation on the timescale of thousands of years rather than millions of years as predicted by traditional molecular clocks (Wilson, 2009). Because of this advanced rate of modification, *C. jejuni* will likely develop resistance to any antimicrobial compound it encounters for extended

periods of time (Luangtongkum, 2009). As a result, an effective strategy must include the ability not only to target a myriad of strains, but to be modified to continue targeting those strains as they develop resistance.

1.4. Nanobodies: a novel in vivo mitigation strategy for targeting C. jejuni

When considering a system that will target an enteric pathogen within the intestine, the choice of antimicrobial compound must be carefully considered. If the compound is toxic, it will cause unwanted harm to the host, if it is too broad range it will cause unwanted harm to the commensal enteric flora, and it it's too narrow range, resistance and immunity will quickly be developed by the pathogen and render the system ineffective. Ideal characteristics include extreme specificity coupled with adaptability, non-toxicity, non-immunogenicity, activity at low concentrations, and ease of engineering. Along these lines, antimicrobial peptides and bacteriophages have both been used *in vitro* with great success (Svetoch, 2010; Connerton, 2011); however, bacteriophages can be difficult to engineer, and resistance to both are common (Svetoch, 2010; Connerton, 2011). An appealing alternative is the use of antibodies due to the virtually limitless possibilities of ligand specificity coupled to the ultra-tight specificity associated with antibody-ligand interactions. In order for antibodies to be a viable option for production in engineered enteric bacteria, however, they must be modified to promote expression in a foreign system and increase yields.

1.4.1 Antibody Classes

Antibodies, or immunoglobulin (Ig), are produced in all animals after the evolutionary division between jawless fish and jawed vertebrates and are the fundamental unit of the adaptive immune system (Pancer, 2008). There are five isotypes of antibodies in most antibody producing animals.

Immunoglobulin-α (IgA) is associated largely with mucus, saliva, tears, and breastmilk. Immunoglobulin-δ (IgD) is still poorly understood. Immunoglobulin-ε (IgE) is involved with allergens and large parasites. Immunoglobulin-μ (IgM) is the first antibody produced in response to exposure to an epitope. Finally, Immunoglobulin-γ (IgG) is the most common antibody and the only antibody capable of crossing the placenta (Brezski, 2016).

1.4.2 Conventional Antibody Structure (Figure 1.4A)

IgG antibodies are comprised of a dimer of heterodimers (Muyldermans, 2013). The subunits of each heterodimer are commonly referred to as the heavy chain and light chain (Muyldermans, 2013). Light chains are composed of two ~12 kDa domains and the heavy chains are composed of four ~12 kDa domains (Ramsland, 2002). The first three domains of the heavy chain and first domain of the light chain on the carboxy-terminus are referred to as the conserved, constant, or crystallisable fragment (F_C) and are responsible for all interactions with the immune system including complement activation and recruitment of lymphocytes (Muyldermans, 2013). The variable fragment (F_V) consists of the terminal domain of both the light and heavy chains (Muyldermans, 2013). These domains contain hypervariable regions that are among the most rapidly mutated sequences found in nature (Li, 2004), and are responsible for the binding to antigen. The variability contributes to specificity and affinity of the antibody (Muyldermans, 2013).

1.4.3 Heavy Chain Antibody Structure (Figure 1.4B)

Camels and llamas have an additional, unusual IgG subtype referred to as heavy chain antibodies (HCAbs) (Muyldermans, 2013). HCAbs are structurally similar to conventional IgG antibodies except they lack a domain in the conserved region of the heavy chain and completely lack a light
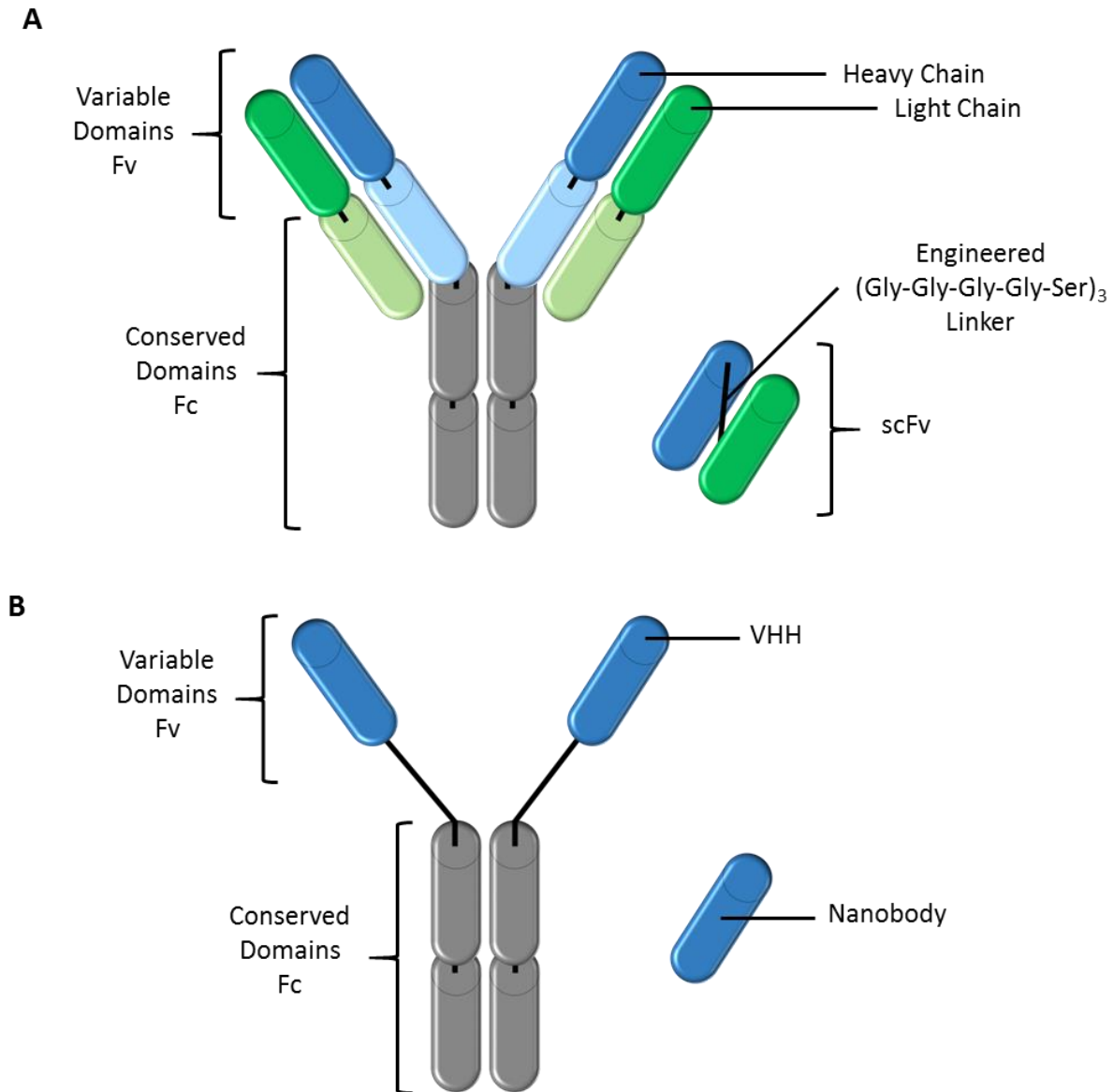
Figure 1.4. Cartoon depiction of the structure of **A.** conventional antibodies and scFv and **B.** heavy-chain antibodies, and nanobodies structure. Conventional antibodies are composed of two heavy and two light chains each of which contains one variable domain (Fv) and 3-4 conserved domains (Fc). Heavy chain antibodies are composed of a pair of heavy chains lacking one conserved domain. scFv are generated by cloning the variable domains of both the heavy and light chain and engineering a linker. Nanobodies are generated by cloning the variable domain of heavy-chain antibodies.

chain (Hamers-Casterman, 1993). The loss of the light chain requires a number of modifications to the HCAb in order for it to be fully functional. First, a number of well conserved residues, typically hydrophobic, are involved in the interaction between light and heavy chain (Muyldermans, 2013). In order to prevent nonspecific interactions and aggregations, many of these residues have been mutated to smaller, more hydrophilic residues (Muyldermans, 2013). Second, the loss of the domain in the conserved region means that the F_V domains are physically closer to each other which may compromise their ability to crosslink antigens and reduce their agglutination properties (Muyldermans, 2013). At least some HCAbs have extended Pro-Gln repeats in this portion of the protein which are believed to act analogously to extend the reach of the HCAb (Muyldermans, 2013). Finally, the F_V of heavy and light chains are positioned next to each other in conventional antibodies and their hypervariable loops work in concert to form an antigen binding platform 600–900 Å^2 (Muyldermans, 2013). Because HCAbs lack one of the domains involved in forming this platform, a dramatic loss of affinity and range of potential antigens would be expected. To prevent this, HCAbs have evolved significantly longer hypervariable loops allowing for a platform 600–800 Å^2 (De Genst, 2006). To minimize the entropy of the unfolded paratope associated with the enlarged loops and, by extension, minimize the entropy loss upon epitope binding, many HCAbs contain an additional disulfide is typically present to restrict movement of the loops (Muyldermans, 2013).

1.4.4 Single Chain F_V (Figure 1.4A)

The greatest detriments to using IgG and HCAbs for engineering and heterologous expression are their size and oligomeric nature making them difficult to express, secrete, properly fold, and organise the subunits. In an attempt to by-pass these limitations, Huston *et al.* (1988) constructed single chain F_V (scFv) against digoxin, a toxic steroid glycoside produced in a number of plants.

27

The sequence of both the light chain Fv and heavy chain Fv were amplified and stitched together by a short sequence encoding (Gly-Gly-Gly-Gly-Ser)$_3$. The result was a protein composed of two domains ~20% of the size of a conventional IgG that was able to bind to its antigen with affinity similar to the unmodified antibody.

Over the past few decades, this procedure has been replicated by many groups targeting a wide variety of antigens, boundaries have been better defined, and variable linkers have been investigated to manipulate flexibility and reduce protease sensitivity (Ahmad, 2012). Dimers and trimers of scFvs have been engineered to create high avidity cross-linkers (Hudson, 1999), and a great variety of fusions have been created for everything from cellular imaging (Casey, 2000), to cancer detection (Todorovska, 2001), to delivery of antimicrobial agents (Peschen, 2004). There are two significant drawbacks to this approach, first, the linker region between domains is exposed to mechanical and protease degradation which may cause the domains to dissociate; and second, because the domains are portions of different genes and antibody maturation is a highly complex process, cloning of these domains can be challenging (Ahmad, 2012).

1.4.5 Nanobodies (Figure 1.4B)

To by-pass these issues, HCAbs have been used in the same manner to develop independent binding domains resulting in a single domain antibody. Significantly these ~12 kDa proteins can bind with the same affinity and specificity as their parental HCAbs (Muyldermans, 2013). This group of proteins have been named nanobodies (Gibbs, 2005). There are a number of benefits to working with nanobodies apart from their size and binding properties, including the lack of labile linker sequences and the quick and easy cloning due to its monomeric nature (Muyldermans, 2013).

Nanobodies possess a unique shape that is significantly different from F_V or scFv. Antigen binding platforms of conventional antibodies typically fall into one of three categories, *cavities* for binding of small molecules, *groove* for binding of peptides, or *flat* for binding of larger antigens (Muyldermans, 2013). Because of the additional size of the hypervariable loops, the antigen binding platform of nanobodies is typically convex. This shape, coupled with their smaller size results in a protein that is ideally suited for binding to the inside of clefts such as enzyme active sites or entrances to transporters (Lauwereys, 1998). This property not only allows nanobodies to act as excellent inhibitors but unlocks potential antigens that conventional antibodies physically cannot reach. This may enable recognition and binding of compounds with low immunogenicity (Muyldermans, 2013). Traditional approaches such as reduction in motility due to nanobody binding of the flagella and cell lysis inducing nanobody fusions are also possible (Riazi, 2013; Muyldermans, 2013).

The small size of the nanobody provides a number of benefits for engineering applications in addition to the relative ease of cloning and expression. Because of the limited number of residues present, exhaustive studies have been done to mutate nanobodies to become more heat and pH stable as well as resistant to proteases (Hussack, 2014). Because nanobodies are typically cleared from the blood quickly and are generally not immunogenic, they can be used in the gut of an animal without eliciting an immune response (Coppieters, 2006).

1.5. Hypotheses

1. *Bacteroides* promoters are not gene specific and will function to induce the expression of heterologous genes in vivo.

2. *Bacteroides* signal peptides will be functional across *Bacteroides* species allowing for borrowing of signal peptides from other *Bacteroides* species.

3. Recombinant techniques will allow for the integration of exogenous genes into the genome of *B. theta*.

4. Nanobodies can be expressed and secreted in concentrations adequate for the reduction of motility of clinically relevant strains of *C. jejuni*.

1.6. Project Objectives

The objectives for this project were:

(1) To identify multiple *Bacteroides* signal peptides and characterize them in vivo.

(2) To identify inducible *B. theta* promoters and test them in vivo.

(3) To identify novel reference genes for RT-qPCR in *B. theta*.

(4) To construct vectors that will allow for plasmid based expression of nanobodies in *E. coli, L. lactis*, and *B. theta* as well as integration and genomic expression in *B. theta*.

(5) To characterize the specific nanobody binding of *C. jejuni* flagella.

Chapter 2. Genome Integration, Gene Induction, and RT-qPCR of Nanobody genes in *Bacteroides thetaiotaomicron*

Abstract

Currently in the literature, there are few validated promoters to evaluate the heterologous expression of genes in *Bacteroides* species. RT-qPCR experiments involving *B. theta* often only use ribosomal RNA as a reference gene despite previous evidence that rRNA is not optimal due to its high concentrations relative to mRNA, slower degradation rates, and extensive secondary and tertiary structure. In contrast, incorporating multiple reference genes tends to provide more consistent data than a single reference gene. In this chapter, 15 putative reference genes were selected and tested for expression stability along with the 16s rRNA when grown on eight carbon sources and under ten stress conditions. Two carbohydrate-inducible promoters, the fructan inducible *susC* Bt-1763 and dextran inducible *susC* Bt-3090 were selected, and expression levels of nanobody genes integrated into the genome or expressed from a novel vector pPPRAMA were compared to native Bt-1763 and Bt-3090 genes. When wild type *B. theta* was grown on fructans, Bt-1763 was upregulated as much as 3,500-fold. When *B. theta* was grown on dextran, Bt-3090 was upregulated as much as 1,000-fold. However, when these strains contained genomically integrated heterologous genes behind either of these promoters at an alternative genomic locus they were not upregulated in the presence of their inducing carbohydrate and plasmid-based expression was lower than anticipated. Unexpectedly, under these conditions native Bt-1763 and Bt-3090 genes were repressed. These data suggest that gene induction has been affected in an unanticipated manner. Further progress will require a re-evaluation of the genome integration strategy and selection of additional inducible promoters.

2.1. Introduction

2.1.1 Genetic Engineering in B. theta

Knowledge of *Bacteroides* genetic elements has been exploited to construct a number of useful tools. A multitude of vectors have been constructed using cryptic plasmids, insertion sequences, and fragments of ICEs. Shuttle vectors have been generated for transferring genetic elements between *E. coli* and *Bacteroides* species (Smith, 1985; Shoemaker, 1986; Thompson, 1990; Wong, 2003), and among *Bacteroides* species (Shoemaker, 1991; Gardner, 1996), and cosmids have been generated for large-scale DNA transfer (Shoemaker, 1989). Options for inducible expression of heterologous proteins; however, are less developed. To date, heterologous expression systems exist for three *Bacteroides* species, *B. theta* (Mimee, 2015), *B. fragilis* (Parker, 2012), and *B. ovatus* (Farrar, 2005; Hamady, 2010, 2011). Part of the challenge for these systems is that *Bacteroides* spp. have an unusual promoter structure compared to other Gram negative species, such as *E. coli*. While the *E. coli* promoter, and many others, are defined by the -10 TATAAT and -35 TTGACA sequence elements, the *Bacteroides* promoter is characterized by a -7 TAnnTTTG and -33 TTTG motif (Bayley, 2000; Mastropaolo, 2009). As a result, *E. coli* promoters that function in a wide variety of bacteria are non-functional in *Bacteroides* species (Mastropaolo, 2009). To address this challenge, further understanding of *Bacteroides* promoter function is required. This can be accomplished by monitoring gene expression under different environmental conditions.

2.1.2 Transcriptomics

An abundance of papers have been published which describe the transcriptome of *Bacteroides* spp. during growth on different plant and host glycans (Sonnenburg, 2005; Martens, 2011;

Benjdia, 2011; Lynch, 2012), periods of host development (Bjursell, 2006; Martens, 2008; Marcobal, 2011), and biofilm or planktonic growth (TerAvest, 2014). These studies typically make use of Affymetrix Microarray Genechips allowing for the accurate measurement of mRNA for thousands of genes at the same time. Although this information will be an invaluable tool for advanced engineering projects in the future, a relative few promoters have been tested in non-native genetic loci for the expression of heterologous proteins. Of those tested, xylan (Whitehead, 1990; Farrar, 2005; Hamady, 2010, 2011), rhamnose, chondroitin-sulfate, arabinogalactan, allolactose (Mimee, 2015), maltose, starch and oxygen inducible promoters (Lobo, 2011), as well as 16S (Mastropaolo, 2009), insertion sequence IS4351 (Smith, 1992), and synthetic constitutive (Mimmee, 2015) promoters have been successful. These studies have used various reporter systems, including Flavin mononucleotide fluorescence (Lobo, 2011), luciferase luminescence (Mastropaolo, 2009; Mimee, 2015), anti-inflammatory compounds (Farrar, 2005; Hamady, 2010, 2011), and enzyme reporters such as thioredoxin (Parker, 2012), xylanase (Whitehead, 1990), and chloramphenicol acetyltransferase (Smith, 1992).

2.1.2.1 RT-qPCR

In addition to microarray techniques, RT-qPCR has been employed to characterize transcript production. Although the majority of literature making use of RT-qPCR with *Bacteroides spp.* is aimed at detection, for example of fecal contamination of water (Yampara-Iquise, 2008; Converse, 2009), some studies have used RT-qPCR to quantify mRNA levels under different culture conditions (Martens, 2011; Hehemann, 2012; Rogers, 2013). Historically, most of these papers have made use of the 16s rRNA as a sole reference gene for normalization. It has been previously reported (Kozera, 2013) that rRNA is a poor control for two reasons; Firstly, rRNA is present in high concentrations within the cell amounting to more than 80% of all RNA while

mRNA represents less than 5%. This means that reactions must be run at either very high reference gene transcript concentrations, very low gene of interest transcript concentrations, or on different dilutions of template, all of which introduce unnecessary error. Secondly, the use of a single reference gene rather than multiple genes has been shown to cause error from 3.0 to 6.4-fold due to the inherent fluctuation of non-inducible genes and it is typically recommended that two to five reference genes be used to reduce this error (Vandesompele, 2002).

In this study I selected 15 putative reference genes and along with the 16s rRNA, determined the most stably expressed genes while *B. theta* was grown on eight different carbon sources, as well as under ten different stress conditions including temperature, oxygen concentration, variable pH, and osmotic shock. The three most commonly referenced programs for the determination of relative mRNA concentrations, or the stability of gene expression, geNorm, BestKeeper, and NormFinder were used and the four genes with least variable expression were selected for further RT-qPCR experiments.

To compare the expression levels of heterologous genes to wild-type *susC*, two pExchange-*tdk* derivative vectors were designed and constructed for the genomic integration or plasmid-based expression of genes in *B. theta* and two novel carbohydrate inducible promoter/terminator combinations (i.e. Bt-1763 and Bt-3090) were explored, using nanobody-*DSB* as a reporter gene. Bt-1763, from PUL22, has been shown to be upregulated in the presence of fructose containing carbohydrates, including in ascending order, levan, fructose, fructooligosacharides, sucrose, and inulin (Sonnenburg, 2010) for a maximum induction of ~10,000-fold. Bt-3090, from PUL48, is upregulated in the presence of dextran for a maximum induction of ~1,000-fold (Rogers, 2013). RT-qPCR was performed on strains containing genomically integrated heterologous genes or plasmid-based genes allowing for a comparison between dextran and fructan-inducible

promoters, the effect a terminator had on transcript levels, and the difference between genes expressed from a plasmid relative to integration into the genome.

2.22. Materials and Methods

2.2.1. Bacterial strains

2.2.1.1 E. coli

pPPRAMA and pINT based constructs (Figure 2.1) were propagated in *E. coli* S17-1 λpir (*TpR SmR recA, thi, pro, hsdR⁻M⁺ RP4: 2-Tc:Mu: Km Tn7*) as well as all biparental mating with *B. theta* strains.

2.2.1.2 B. theta

The thymidine kinase knockout mutant *B. theta* ATCC 29148 ΔBt-2275 (Koropatkin, 2008) was used as a recipient strain both for the vector pPPRAMA and to generate a strain lacking the homogalacturonan PUL, PUL75, *B. theta* ATCC 29148 *Δtdk ΔBt4108-4124* (*B. theta Δtdk ΔP75*). This strain was used as a recipient strain for all pINT conjugations. Strains generated in this study are outlined in Table 2.1. These strains along with *B. theta* VPI-5482 were used in growth studies and to harvest mRNA for RT-qPCR.

2.2.2. Culture conditions

2.2.2.1 Lysogeny Broth (Bertani, 1951)

E. coli were grown in 1.0% (w/v) Bacto-tryptone, 0.5% (w/v) yeast extract, and 171 mM NaCl aerobic atmosphere with 180 rpm agitation. Cell cultures were grown at 37°C until an optical density (OD_{600}) of 0.8.
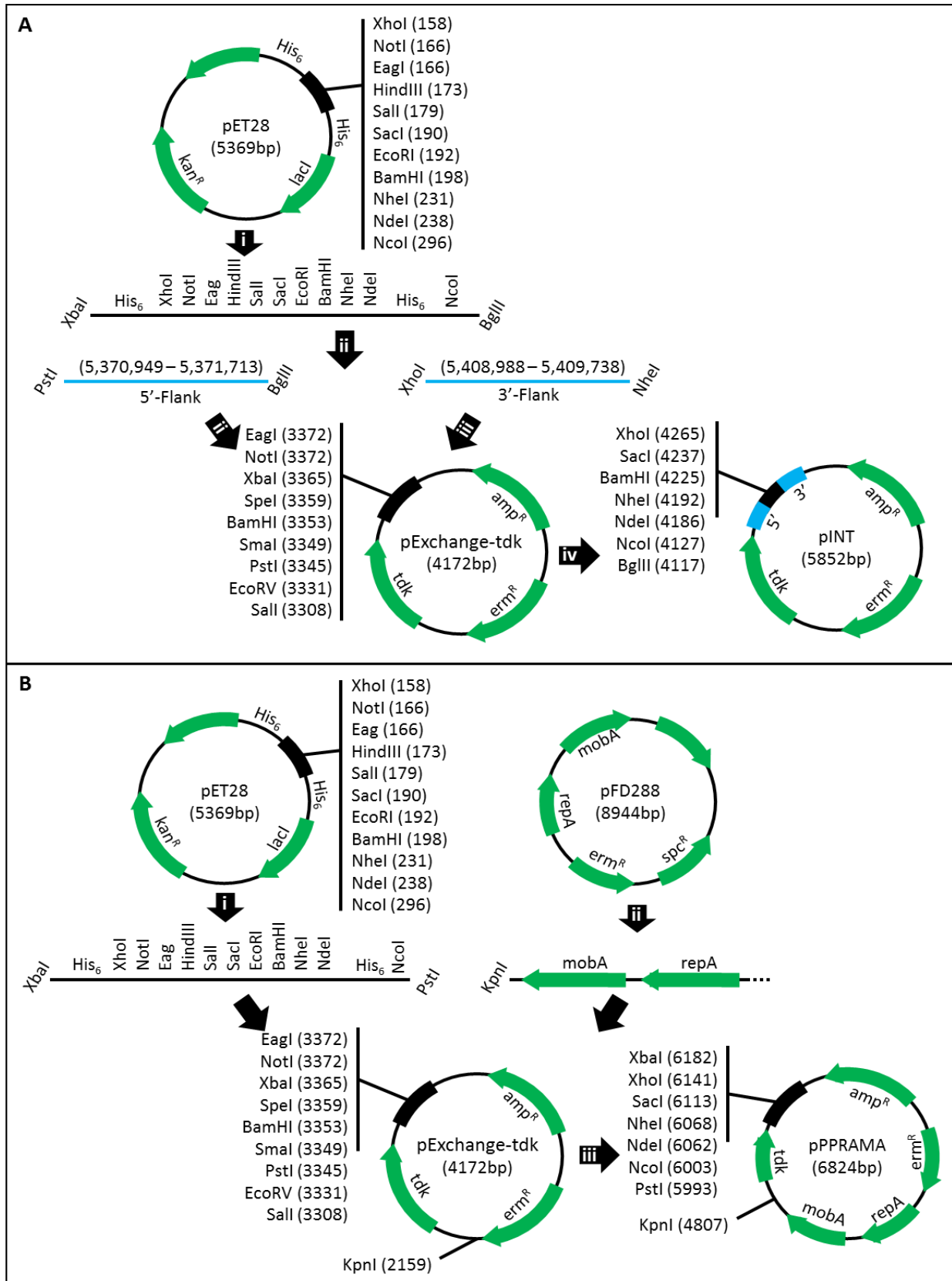
Figure 2.1. Cloning of parental vectors for the introduction of foreign genetics into *B. theta*. **A.** Generation of the pINT vector, used for genomic integration into the PUL75 knockout. **i.** The multiple cloning site of pET28a including both N-terminal and C-terminal hexahistidine-tags was amplified with primers containing a BglII recognition sequence on the 5'-end and an XbaI recognition sequence on the 3'-end. **ii.** A 764 bp sequence 5' of Bt-4108 was amplified with primers containing a PstI recognition sequence on the 5'-end and BglII recognition sequence on the 3'-end. Sequences from i. and ii. Were ligated at the BglII recognition sequence. **iii.** A 750 bp sequence 3' of Bt-4123 was amplified with primers containing an XhoI recognition sequence on the 5'-end and an NheI recognition sequence on the 3'-end. iv. These sequences were ligated into a pExchange-tdk vector between the PstI recognition sequence and XbaI recognition sequence generating the pINT vector. **B.** Generation of the pPPRAMA vector, used for plasmid-based expression. **i.** The pET28a MCS was amplified as in A. except with a 5' primer containing a PstI recognition sequence. This sequence was ligated into pExchange-tdk vector digested similarly. **ii.** The RepA and MobA genes were amplified with a 5' primer containing partial complementarity to the pExchange-tdk vector and a 3' primer containing a KpnI recognition sequence. **iii.** The sequence was stitched to the vector generated in i. by PCR.

2.2.2.2 Supplemented Brain Heart Infusion (BHIS)

Strains of *B. theta* were refreshed from glycerol stocks by streaking ~2 μL aliquots on an agar medium containing, 1.25% (w/v) calf brain infusion, 1.0% (w/v) peptone, 0.5% (w/v) beef heart infusion, 85.6 mM NaCl, 23.8 mM $NaHCO_3$, 17.6 mM $NaHPO_4$, 11.1 mM glucose, 8.25 mM cysteine, 7.67 μM hemin solution, 4.36 μM resazurin solution.

2.2.2.3 Tryptone Yeast Extract Glucose (TYG) Medium (Holdeman, 1977)

Strains of *B. theta* were cultured by inoculating ~5 μL aliquots of cells obtained from BHIS cultures in 2% (w/v) Tryptone, 1% (w/v) yeast extract, 0.5% (w/v) glucose, 23.8 mM $NaHCO_3$, 8.25 mM cysteine, 1.4 mM NaCl, 230 μM K_2HPO_4, 75 μM $MgSO_4$, 70 μM $CaCl_2$, 29.4 μM KH_2PO_4, 7.67 μM hemin solution, and 4.36 μM resazurin solution. Cultures were incubated at 35°C, anaerobic atmosphere (20% CO_2, 80% N_2) throughout without agitation (standard conditions) for 12 h.

2.2.2.4 Bacteroides Minimal Medium (Martens, 2008)

B. theta, was grown to a cell density OD_{600} of 0.6 in 0.5% (w/v) glucose, 100 mM KH_2PO_4 pH 7.2, 15 mM NaCl, 4.7 mM Na_2CO_3, 3.75 mM $(NH_4)_2SO_4$, 2.0 mM cysteine, 0.2 mM histidine, 100 μM $MgCl_2$, 72 μM $CaCl_2$, 5.25 μM iron (II) sulfate, 3.0 μM vitamin K3 (2-methylnaphthalene-1,4-dione), 1.9 μM hematin, 0.25 μM resazurin (7-hydroxy-10-oxidophenoxazin-10-ium-3-one), and 18.5 pM vitamin B12 (α-(5,6-dimethylbenzimidazolyl) cobamidcyanide). To harvest cells, cultures were then centrifuged at 5,000 RCF for 10 min in an anaerobic atmosphere. Cell pellets were resuspended in minimal medium and target genes were induced with 0.5% (w/v) of glucose, galactose, fructose, homogalacturonan, rhamnogalacturonan I, pectic galactan, α-mannan, or dextran. Cultures were incubated for 1-48 h at standard

conditions. Alternatively, wild type *B. theta* was inoculated directly into minimal medium with a final concentration of 0.5% (w/v) inducing carbohydrate. Cultures were incubated for 48 h at standard conditions and samples were taken at 1, 2, 3, 6, 12, 24, and 48 h.

2.2.3. Stress Conditions

Wild type *B. theta* was grown in minimal medium with 0.5 % (w/v) glucose to an OD_{600} of 0.4. Cultures were then centrifuged at 5,000 RCF for 2 min in an anaerobic atmosphere. Cells were washed once with anaerobic minimal medium lacking a carbohydrate and resuspended in fresh minimal medium with 0.5 % (w/v) glucose prior to stress treatment.

2.2.3.1. Heat/Cold Stress

To induce a heat or cold stress response, cultures were rapidly equilibrated to either 42°C in a water bath or 4°C in an ice bath. Cells were incubated for 1 h prior to harvesting by centrifugation at 5,000 RCF for 5 min.

2.2.3.2. Oxidative Stress

To induce an oxidative stress response, cultures were either supplemented with a final concentration of 500 μM paraquat (N, N'-Dimethyl-4, 4′-bipyridinium dichloride), or aerated with atmospheric oxygen every 5 min with a 1 mL pipet. Cultures were maintained at 37°C for 1 h prior to harvesting by centrifugation at 5,000 RCF for 5 min.

2.2.3.3. pH Stress

To induce a pH shock response, cultures were resuspended in minimal media with 0.5 % (w/v) glucose adjusted to a pH of 6.0, 6.5, 7.0, 7.5, or 8.0 with KH_2PO_4. Cultures were maintained at 37°C for 1 h prior to harvesting by centrifugation at 5,000 RCF for 5 min.

2.2.3.4. Osmotic Stress

To induce hyperosmotic stress response, cell pellets were resuspended in minimal media with 0.5 % (w/v) glucose supplemented with 300 mM NaCl, 600 mM NaCl 600 mM sucrose, or 1.2 M sucrose (Cheung, 2009). Cultures were maintained at 37°C for 1 h prior to harvesting by centrifugation at 5,000 RCF for 5 min.

2.2.4. Nucleic acid extraction and cDNA synthesis

2.2.4.1. RNA Purification

2 mL of culture grown as described above was centrifuged at 5,000 RCF for 5 min. Cell pellets were resuspended in 300 μL RNAprotect Bacteria Reagent (Qiagen Cat# 76506) for 5 min prior to centrifugation at 15,000 RCF for 5 min. Total cell RNA was the purified from the cell pellets with a GeneJET RNA Purification kit (Thermo Cat# K0731) according to the Bacteria Total RNA Purification Protocol.

2.2.4.2. RNA Quality

Cell metabolism of cultures was suspended using RNAprotect Bacteria Reagent to ensure additional gene expression would not be induced during the purification procedure. RNA was purified using a GeneJET RNA Purification kit according to the Bacteria Total RNA Purification Protocol. A_{260}/A_{280} ratios were used to determine protein contamination and values less than 1.9 or above 2.1 were discarded and reproduced. Purified RNA samples were checked for degradation using an Agilent 2100 Bioanalyzer. The RNA 6000 Nano Kit (Aligent Cat# 5067-1511) was used according to manufacturer's instructions under the Prokaryote Total RNA Nano settings. Samples with a RNA Integrity Number (RIN) of less than 9.7 were discarded and reproduced. RNA samples were reverse transcribed immediately after purification.

2.2.4.3. Reverse Transcription

cDNA was reverse transcribed using the QuantiTect Reverse Transcription Kit (Qiagen Cat# 205310) according to manufacturer's instructions. 250-500 ng of RNA was incubated in gDNA Wipeout Buffer in a final volume of 14 μL for 2 min at 42°C to eliminate intact genomic DNA. Nucleotides, buffer, and Quantiscript Reverse Transcriptase were then added to a final volume of 20 μL and incubated at 42°C for 30 min. Newly generated cDNA samples were diluted to a final concentration of 50 ng μL^{-1}. cDNA samples were stored at -80°C for no longer than one week prior to use.

2.2.5. Primer design

Primers were designed using OligoCalc (Kibbe, 2007). Sequences were selected based on salt-adjusted melting points nearest 58°C, Nearest Neighbour melting points nearest 52°C, GC content nearest 50% and amplicons nearest 200 base pairs. Preference was given for sequences nearest 20 bps. Primers are outlined in Table 2.

2.2.6. Optimization of real-time qPCR reaction conditions

To optimize reaction conditions, a cDNA pool containing cDNA from all samples was diluted to 100, 10, 1, 0.1, and 0.01 ng/μL. Samples were run with primer concentrations of 1, 1.5, or 2 μM with annealing temperatures ranging from 54-60°C. Cycling conditions were 15 min at 95°C followed by 35 cycles of 15 sec at 95°C, 15 seconds at the annealing temperature, and 30 sec at 72°C. This was followed by a dissociation stage from 60-95°C.

Representative samples from each primer pair were electrophoresed through a 2% (w/v) agarose gel to ensure a single product was produced and the melting point of these samples as determined through the dissociation curve was used in future runs to ensure the correct product was

produced. Once it was established that the correct product was produced and that there were no contaminating products, cycle thresholds (cT values) were plotted against template concentration for each primer pair under each condition and the efficiency of reaction was determined using Equation 1. Reaction conditions for further RT-qPCR were chosen such that all primer pairs resulted in efficiency ranging from 90-110% with a single product.

Equation 1.

$$Reaction\ Efficiency = (10^{\left(-\frac{1}{slope}\right)} - 1) \times 100$$

2. 2.7. geNorm, BestKeeper, and NormFinder analysis

To determine the stability of gene expression, three programs were used which employ independent methods for measuring expression stability.

2.2.7.1 geNorm

In order to determine the most stably expressed genes in a sample, the geNorm software calculates a parameter based on raw, non-normalized transcript levels. Ratios of transcript levels of stable genes should be consistent between all samples. Based on this assumption, pairwise variation between a gene and all other genes is calculated, and the average of these variations is called the reference gene stability value, or M value. The software removes the gene with the highest M value in a step-wise fashion and recalculates this value until only two genes remain. In a final step, the software calculates the minimal number of reference genes required out of the set to attain consistent results. This is achieved by calculating the pairwise variation of a normalization factor while introducing genes in a step-wise manner.

2.2.7.2 BestKeeper

Rather than transcript values, the BestKeeper software uses raw cT values to calculate the

Pearson correlation coefficient for each gene with all other genes and calculates whether the

correlation is statistically significant. The geometric mean of the most highly correlated genes is

calculated to generate the BestKeeper index for a set of genes. Finally, the correlation coefficient

and R^2 value between the Bestkeeper index and each reference gene is calculated as well as

whether the correlation is statistically significant.

2.2.7.3 NormFinder

The NormFinder software uses log-transformed expression values and performs a two-way

ANOVA to determine intra- and inter-group variance and combines them to obtain a stability

value. Unbiased estimates of intra-group variation are determined using moment equations, and

inter-group variation is calculated based on the assumption that none of the genes experience

systematic variation. The two variances are then combined in equal weighting to generate a

stability value.

2.2.8. Vector construction

Two parent vectors were generated from the pExchange-*tdk* vector for introducing genetics into

B. theta. All products were sequence confirmed (Figure 2.1).

2.2.8.1 Generation of pINT

The pINT vector was generated for the integration of genes into the PUL75 locus. A 750 bp

sequence of genomic DNA 3' of Bt-4123 (RefSeq NC_004663.1, base pairs 5,408,988 –

5,409,738) was amplified with Primer-1 (Table. 2.3) and Primer-2. This fragment was digested

with XbaI and NheI and ligated into a pExchange vector digested with XbaI resulting in an XbaI cut site 5' of the flank, and the nullification of the 3' cut site.

A 764bp sequence of genomic DNA 5' of Bt-4108 (base pairs 5,370,949 – 5,371,713) were amplified with Primer-3 and Primer-4. A 179bp sequence containing the multiple cloning site of pET28a (Novagen Cat.69864-3 base pairs 301-123) was amplified with Primer-5 and Primer-6. These sequences were digested with PstI and BglII and BglII and XbaI respectively and ligated into the above vector digested with PstI and XbaI.

2.2.8.2 Generation of pPPRAMA (pETpExchange-RepA_MobA)

A 2,648bp sequence containing replication (RepA) and mobilization (MobA) genes was amplified from pFD288 (gi:1016679 base pairs 6,267 – 8,915) with Primer-7 which contains partial complimentary sequence to pExchange, and Primer-8. This sequence was stitched onto pExchange with Primer-9 and Primer-10. The final product was digested with KpnI and recircularized by ligation.

The MCS from pET28a was amplified as with pINT except with Primer-5 and Primer-11. This product was digested with PstI and XbaI and ligated into the above vector digested similarly.

2.2.8.3 Regulation

For both pINT and pPPRAMA vectors, promoters and terminators were introduced flanking the MCS. A promoter/terminator pair was taken from Bt-1763, a fructan-inducible susC-like carbohydrate transporter, as well as Bt-3090, a dextran-inducible susC-like carbohydrate transporter.

The Bt-1763 promoter was amplified from the *B. theta* genome (base pairs 2,183,023 – 2,182,827) with Primer-12 and Primer-13. This sequence was digested with BglII and NcoI and ligated into either pINT or pPPRAMA. The Bt-1763 terminator (base pairs 2,174,961 – 2,174,833) was amplified with Primer-14 and Primer-15. This sequence was digested with XbaI and NheI and ligated into the above vector digested with XbaI.

Bt-3090 promoter/terminator vectors were prepared similarly except the Bt-3090 promoter (base pairs 3,924,009 – 3,923,822) was amplified with Primer-16 and Primer-17 and the terminator (base pairs 3,913,436 – 3,913,237) was amplified with Primer-18 and Primer-19.

2.2.8.4 Expression vectors

Because the above vectors have had the pET28 MCS introduced in frame, all further cloning was performed identically for all vectors. Nanobody *DSB* was amplified with Primer-20 and Primer-21. In each case, the sequence was then digested with BspHI and XhoI and ligated into a vector digested with NcoI and XhoI.

2.2.9. Construction of B. theta Mutants

pINT constructs were transformed into chemically competent *E. coli* S17-1 λpir cells and positive transformants were selected by growth on 200 μg mL^{-1} ampicillin. Cultures were grown to an OD_{600} of ~0.8 in LB broth at 37°C with 200 rpm agitation followed by centrifugation at 4,000 RCF for 10 min and the spent medium was aspirated.

Recipient strains of *B. theta* were grown in TYG broth at standard conditions to an OD_{600} of ~0.8. Cultures were then centrifuged anaerobically for 10 min at 5,000 RCF and the spent medium was aspirated. Fresh, 37°C, aerobic TYG broth was used to resuspend *E. coli* and *B.*

theta cell pellets and the conjugation mixture was plated on BHIS agar medium with no antibiotic. Cultures were incubated aerobically at 37°C for 20 h to allow matings to occur.

Cell masses from these cultures were resuspended in 37°C TYG broth and dilutions were plated on BHIS-agar containing 25 μg/mL erythromycin and 200 μg/mL gentamycin. Cultures were incubated under standard conditions for 30 h and colonies were streaked on fresh agar medium and incubated as above.

Cells from randomized colonies were harvested and grown to an OD_{600} of ~0.8 in TYG broth at standard conditions. Dilutions of these cultures were plated on BHIS-agar plates containing 200 μg/mL 5-fluoro-2'-deoxyuridine (FUdR) and plates were incubated under standard conditions. Colonies were streaked on fresh plates and incubated equivalently.

Cells were screened by colony PCR using P75 flanking primers Primer-22 and Primer-23 and products were sequenced.

2.2.10 RT-qPCR of native genes

To determine the expression levels of native Bt-1763 and Bt-3090 genes, wild type *B. theta* VPI-5482 was grown with one of five carbohydrates as the sole carbon source: glucose as a negative induction control, fructose and inulin to induce Bt-1763, dextran to induce Bt-3090, and HG to determine to what extent the presence or absence of PUL75 affected the expression of Bt-1763 and Bt-3090. Samples were taken at 1, 2, 3, 6, 12, and 24 h after induction. The growth was repeated on three separate occasions and five reactions containing each primer pair was performed in each replicate. To ensure that the removal of Bt-2275 and PUL75 were not detrimental to the expression of these genes the experiment was replicated with *B. theta* Δ*tdk*, ΔP75.

2.2.11. RT-qPCR of nanobody constructs

Strains of *B. theta* containing the pPPRAMA vector, including the Bt-3090 promoter and a nanobody gene, or with a genomically integrated nanobody gene flanked by the Bt-3090 promoter, Bt-3090 promoter/terminator pair, or Bt-1763 promoter/terminator were investigated. All strains with chromosomal insertions of nanobodies were grown on glucose as a reference condition. Heterologous strains containing the Bt-3090 promoter were grown on dextran and strains containing the Bt-1763 promoter were grown on fructose and inulin. The induction of the nanobody gene as well as the native susC gene, were monitored for changes in the level of transcripts.

2.2.12. RT-qPCR data analysis

RT-qPCR outputs were imported into Biogazelle, Qbase^{+} where genes were normalized to reference genes. Fold induction was determined by comparing normalized values from treatment conditions to values during growth on glucose. Statistical significance was determined using two way repeated measures ANOVA with P≤0.05.

2.3. Results

2.3.1 Vector generation

Two new derivatives of the pExchange-*tdk* vector were generated allowing for genomic integration into the PUL75 locus (pINT) or plasmid-based expression (pPPRAMA). The Bt-1763 (fructan inducible) and Bt-3090 (dextran inducible) promoter/terminator pairs as well as Bt-3090 promoter lacking terminator were cloned into pINT to investigate specific induction of genes in the presence of these carbon sources and to what extent the terminator affected expression levels. The Bt-3090 promoter/terminator pair was cloned into the pPPRAMA vector to test whether

plasmid-based expression was significantly different from genome-based expression. Nanobody-*DSB* was cloned into these vectors, and the pINT constructs were conjugated into the *B. theta* genome. Constructs generated can be found in Table 1.

2.3.2. Selection of Putative RT-qPCR Reference Genes

Putative reference genes were chosen from unique, pathways with diverse functionalities in order to minimize co-regulation, as well as for their location in the *B. theta* VPI-5482 genome with the intention of minimizing transcriptional coupling (Takle, 2007). Finally, genes were chosen based on their stable expression in studies of other species (Table. 2.2). In total, 15 genes were chosen from essential pathways such as glycolysis (*pgk*, *gapdh*), amino acid (*gluS*), nucleic acid (*gmk*), and fatty acid synthesis (*mcat*), protein translation (*rplI*, *rplQ*, *rrsA*) and transport (*srp*), transcription (*rpoD*, *rpoS*, *rho*), DNA repair (*recA*), DNA replication (*gyrA*, *dnaG*) and membrane stability (*lpxC*). 16s rRNA was used to compare with results reported in the literature.

2.3.2.2. RT-qPCR Reaction Efficiency and Specificity

cDNA was generated from purified RNA from each of the experimental conditions and a pool of all cDNAs was produced. Concentrations ranging from 0.01 - 100 ng/μL were used as a template and reactions were run in triplicate. Log-transformed sample concentration was plotted against cT value and the slope of the curve was determined using linear regression. Reaction efficiency (Table 2.2) was determined with Equation 1. Dissociation curves were generated by gradually increasing the temperature of the sample from 60-95°C while measuring fluorescence to ensure a single product was produced. Representative samples of these reactions were electrophoresed through a 2% agarose gel to ensure products were of the anticipated size.

2.3.2.3. Determination of Stable Reference Gene Expression

To determine the stability of putative reference gene expression, data was analysed by linear mixed-effects modeling using the Normfinder software, repeated pairwise correlation of mRNA with other putative reference genes and regression analysis using the Bestkeeper software, and stepwise elimination of genes with the greatest pairwise variation with other putative reference genes using the geNorm software of Biogazelle's qbase$^+$ (Figure 2.2, Table 2.4).

A primary round of screening was performed testing the expression of the 16 putative reference genes one h after induction on 0.5% (w/v) galactose, fructose, HG, PG, RGI, yeast mannan, and dextran. Based on these data, *gluS, gyrA, rho, srp, gmk, rplQ, ipxC, gapdh* were eliminated due to instability. In order to further reduce the number of putative reference genes under potential experimental conditions, a secondary round of screening incorporated heat and cold shock, acid and alkaline shock, and oxygen shock conditions and investigated the induction of the remaining eight genes one h after induction (Figure 2.2). Based on these data, *rpoD*, *pgk*, *dnaG*, and *rpoS* were determined to be the most stable reference genes under these conditions for *B. theta* and were used as reference genes for the remaining RT-qPCR experiments.

Although *rrsA* was the second most stable gene as determined by NormFinder, it was the fifth and sixth most stable gene as determined by BestKeeper and geNorm respectively. This suggests that rRNA is stably expressed under these conditions; however, more stable alternatives are available.
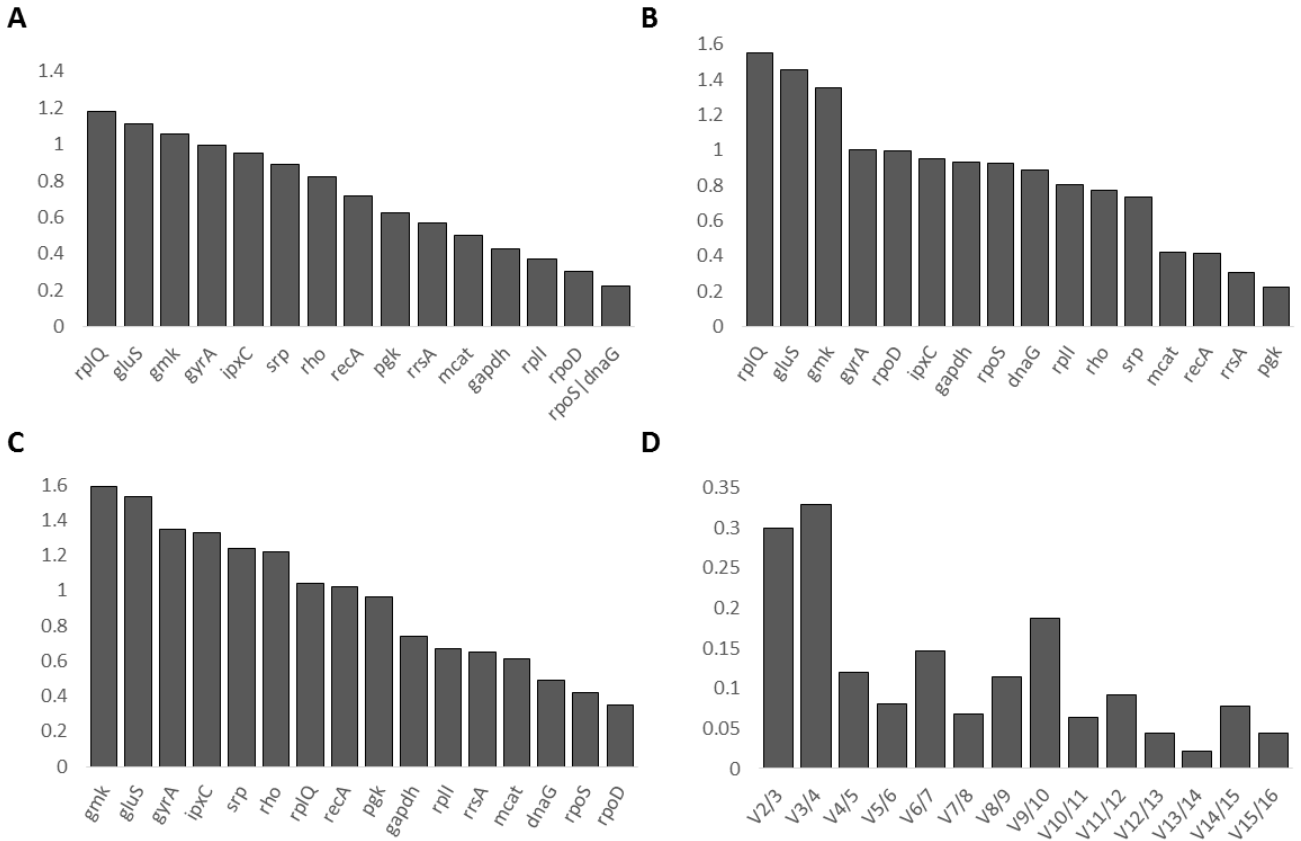
Figure 2.2. Determination of stability of putative reference genes for RT-qPCR. Gene expression was monitored during growth on eight carbon sources and ten stress conditions. (A-C from least stable gene on the left, to most stable on the right) **A.** geNorm "M-values" or mean gene stability during stepwise removal of the least stable gene. Higher values indicate less stable genes. **B.** NormFinder stability values base on intra- and inter-group variance. **C.** Standard deviation of C_T values as calculated by BestKeeper. Higher values indicate less stable genes. **D.** geNorm "V-values" or the pairwise variation of normalization factors of increasing numbers of genes. Indicative of the minimum number of reference genes required for consistent normalization.

2.3.3. RT-qPCR

2.3.3.1 Native susC expression

In all cases, fold increase of transcript had returned to one by the six h time point and time points between 6-24 h are not shown. There was no difference between the level of induction between wild type and *B. theta* Δtdk, $\Delta P75$ for either of the genes on any carbon source.

In the presence of fructose, Bt-1763 increased to ~400-fold and decreased to ~150-fold by 3 h. In the presence of inulin, Bt-1763 increased from ~1,000-fold to over 3,000-fold by 3 h. In both cases, the level of expression of Bt-3090 did not change. In the presence of dextran, Bt-3090 increased from ~500-fold to ~1,000-fold over 3 h. Bt-1763 was significantly upregulated in the presence of dextran to a maximum of 16-fold; however, consistent increase or decrease over time was not observed and this amounts to less than half of one percent of the maximal upregulation of this gene in the presence of inulin and is likely a negligible fluctuation.

In the presence of HG, Bt-1763 was upregulated to ~2,000-fold, peaking at 2 h for both strains before beginning to decrease. Bt-3090 was not significantly induced in the presence of HG.

2.3.3.2 Heterologous gene expression

In all cases, the genomically integrated nanobody gene was not upregulated (Figure 2.3B). Interestingly, unlike the wild type or *B. theta* Δtdk, $\Delta P75$ strains, the native Bt-1763 and Bt-3090 susC genes were also not upregulated in any of these strains at any time point. In the case of pPPRAMA plasmid-based expression, both nanobody and native Bt-3090 susC were upregulated; however, to much lower levels that the wild type counterpart.
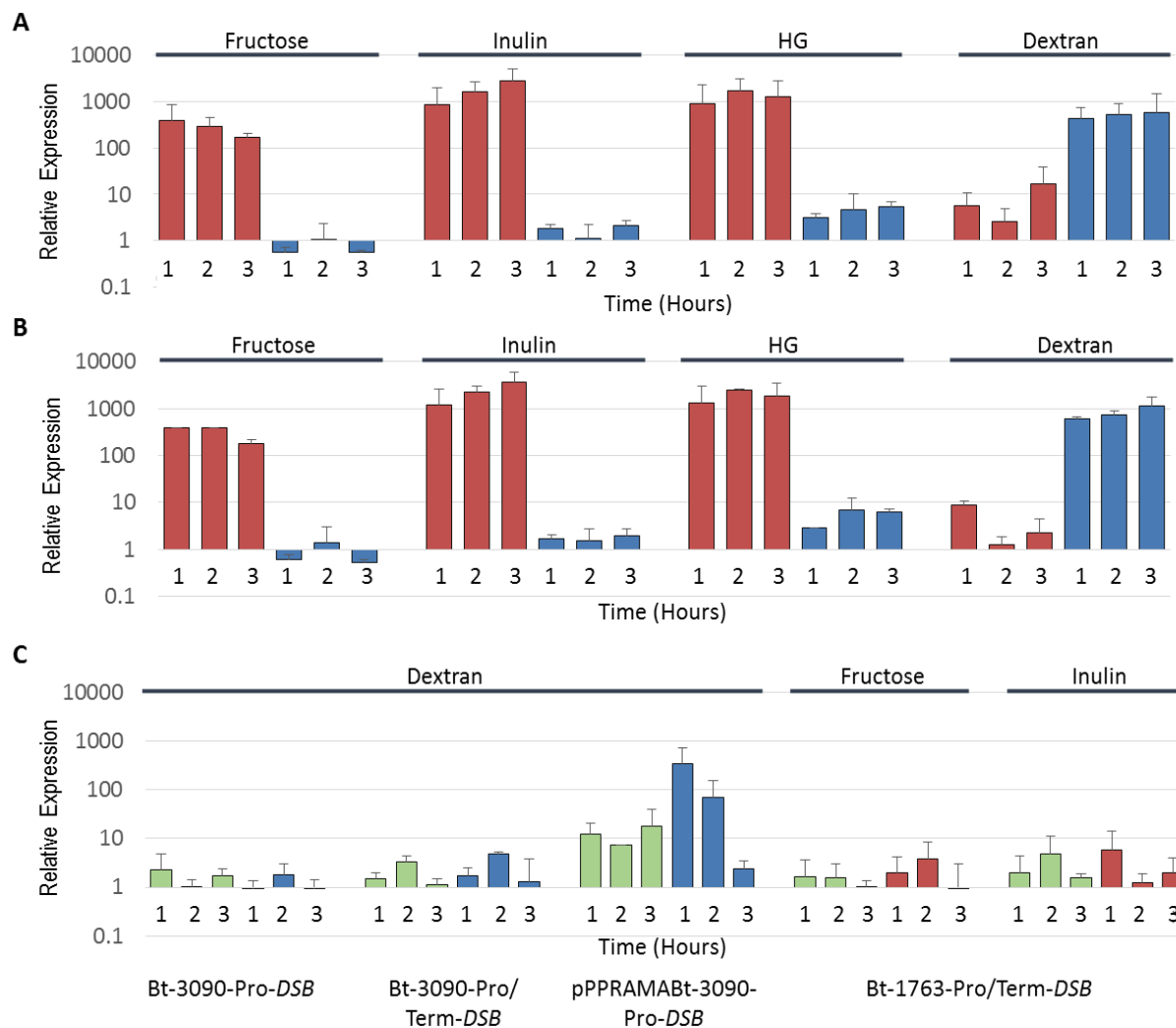
Figure 2.3. Relative gene expression as determined by RT-qPCR. All values are fold difference in transcript concentration relative to the equivalent sample grown on glucose over three h after induction. Error bars represent standard deviation from the mean. In all panels, red bars represent native Bt-1763 susC, blue bars represent native Bt-3090, and green bars represent nanobody-*DSB*. **A.** Wild type *B. theta* VPI-5482 grown on fructose, inulin, HG, and dextran. **B.** *B. theta Δtdk, ΔP75*. **C.** Heterologous constructs. Constructs containing the Bt-3090 promoter are grown on dextran and shown along with the native Bt-3090 susC transcript level. Constructs containing the Bt-1763 promoter are grown on fructose and inulin and shown along with the native Bt-1763 susC transcript level.

2.4. Discussion

2.4.1 RT-qPCR reference genes

In this project, 16 putative reference genes from *B. theta* were chosen and their expression levels were analyzed after growth on eight carbon sources and under ten stress conditions. Expression stability was determined using three commonly cited programs geNorm, BestKeeper, and NormFinder. geNorm and BestKeeper ranked the most stable genes similarly with geNorm ranking *rpoS/dnaG* followed by *rpoD* and BestKeeper ranking *rpoD*, *rpoS*, and *dnaG* as the top three most stable genes (Figure 2.2; Table. 2.4). Interestingly, although NormFinder and geNorm were largely in agreement as to which genes were least stable, ranking *gyrA*, *Gmk*, *gluS*, and *rplQ* in the bottom four, NormFinder largely disagreed with both geNorm and BestKeeper about the most stably expressed genes ranking *pgk*, *rrsA*, and *recA* in the top three.

The variance in gene expression as calculated by geNorm suggests that four reference genes is the minimal number required to achieve consistent normalisation. Because there is no established method for combining these rankings or weighting ranking by the different programs in order of significance, *rpoD*, *rpoS*, and *dnaG* were chosen based on the agreement of both geNorm and BestKeeper, and *pgk* was chosen as it was the most stable gene as calculated by NormFinder.

rrsA was ranked the second, fifth, and sixth most stable gene as determined by NormFinder, BestKeeper, and geNorm respectively. This suggests that *rrsA* was stably expressed under these conditions, but it is not the ideal choice. In spite of relative stability, issues with consistent measurement as a result of high concentrations relative to mRNA, slower degradation rates, and extensive secondary and tertiary structure reported elsewhere make rRNA an undesirable choice and these alternatives are recommended.

2.4.2 RT-qPCR of native susC genes

Promoter/terminator pairs from two susC genes were inserted into vectors for the heterologous expression of genes under defined conditions (Figure 2.2). Bt-1763 is inducible by a variety of fructans whereas Bt-3090 is inducible by dextran. To determine native levels of expression of these genes in the presence of their inducing carbohydrates, wild type *B. theta* VPI-5482 was grown on a number of carbohydrates including glucose, fructose, inulin, HG, and dextran (Figure 2.3). To ensure that the Bt-2275 tdk and Bt-4108-4123 PUL 75 knockouts did not have an effect on expression levels, identical experiments were run with this strain. As predicted, Bt-1763 was upregulated in response to growth on both fructose and inulin. Levels of upregulation were less than those reported in the literature (Sonnenburg, 2010) with ~400-fold rather than ~1,000-fold, and ~3,500-fold rather than ~10,000-fold for fructose and inulin respectively. Because of the exponential nature of the data, these are relatively minor differences in expression and could be the result of negligible differences in culture conditions or experimental error. In response to growth on dextran, Bt-3090 was upregulated ~1,000-fold which is very similar to previously reported values (Rogers, 2013). Also reported in the literature, the level of expression fell from a consistently high level to baseline between six and eight h after induction. Although the time course presented here does not contain as many time points, expression reached a peak at three h and had fallen to baseline by six h.

The expression of both Bt-1763 and Bt-3090 was monitored during growth on HG as a control as the pINT constructs are inserted into the PUL75, HG catabolic PUL knockout locus. It was unclear whether any residual genetic elements would affect the expression of genes integrated here. Unexpectedly, the native Bt-1763 susC gene was upregulated in the presence of HG.

Levels of induction were roughly half of that grown on inulin and 3-5 times higher than when grown on fructose.

It is not clear what caused the upregulation; however, it is possible that there are trace quantities of fructans present in the commercially produced HG. This seems unlikely; however, as trace quantities would not result in equivalent induction levels (Rogers, 2013). Alternatively, it has been previously shown that some PULs are coregulated. For example, PUL75 is upregulated to its highest levels when grown on HG, but it is also upregulated when grown on PG, RGI, and RGII (Martens, 2011). This may not be surprising as all four polysaccharides are common components of the pectic network (Martens, 2011), and therefore, the metabolic response of *B. theta* to a pectic polysaccharide may be linked to the metabolism of the complete network. In nature, the components of pectin are not pure, and therefore, the enzymes that target multiple branches may be required for pruning and efficient metabolism. Although a co-regulated response has not been demonstrated for PUL22, the locus containing Bt-1763, inulins are found in pectin rich plants such as chicory, onions, and asparagus. This is a plausible explanation as the ΔPUL75 mutant is no longer capable of expressing HG receptors contained on PUL75 suggesting that another, more promiscuous receptor may be involved.

Importantly, there was no statistically significant difference in levels of induction between wild type *B. theta* and *B. theta* Δ*tdk*, ΔP75. This suggests both that the experiments were consistent and that the removal of these genes does not negatively affect the expression of Bt-1763 or Bt-3090.

2.4.3 RT-qPCR of heterologous genes

To determine whether heterologous genes integrated at a defined genomic location were induced to levels comparable to their native counterparts, strains with a *DSB* nanobody flanked by the Bt-3090 promoter, Bt-3090 promoter/terminator pair, or the Bt-1763 promoter/terminator pair were inserted into the vacated PUL75 space knockout. Additionally, the *DSB*-pPPRAMA vector under regulation by the Bt-3090 promoter was analyzed as an episomal control. All strains were grown on glucose as a reference condition and constructs containing the Bt-1763 promoter were grown on fructose or inulin, while constructs containing the Bt-3090 promoter were grown on dextran. All other conditions were identical to those that the wild type and *B. theta* Δ*tdk*, ΔP75 strains were exposed to.

In all cases of genomically integrated *DSB*, there was no significant increase in transcript level over the 3 h period observed and although plasmid-based expression was comparatively higher, it was significantly lower that native susC levels. Surprisingly, the native Bt-1763 and Bt-3090 susC genes which were upregulated in both wild type and *B. theta* Δ*tdk*, ΔP75 were also not upregulated when the strains contained genomically integrated *DSB*, and were not only upregulated to significantly lower levels in strains containing plasmids, the transcript levels decreased at an increased rate. It is unclear why this is the case; however, a number of potential explanations will be discussed below.

First, two identical promoters in the genome (i.e. Bt-3090 or Bt-1763) may compete for binding of regulatory proteins resulting in fewer interactions per promoter. If this is the case, it may take longer to reach the same levels of induction. To investigate this, additional later time points could be observed to monitor transcript levels. This would be problematic in the long term as it would result in slower reaction time to carbon sources relative to wild type strains and a loss of

56

fitness; however, this problem could be overcome by choosing promoters from the knocked out PUL75, promoters with minor modifications to their sequences or from other closely related *Bacteroides* species.

Second, mechanisms of cross-talk between PULs are still poorly understood. It may be the case that introducing a dextran inducible promoter into PUL75 has resulted in an unanticipated interaction between this gene and PUL48, the dextran catabolic PUL by an unknown mechanism. If this is the case, it is again possible to make use of additional promoters or to integrate constructs with the Bt-1763 and Bt-3090 promoters in to PUL22 and PUL48 respectively.

Third, because all genome integrated constructs stem from the same stock of *B. theta* Δtdk, ΔP75, it is possible that a point mutation was unknowingly introduced to the genome prior to integration in a gene involved in transcription. Because transcript levels for reference genes were not significantly different between wild type and nanobody containing strains, it would likely not be a mutation to RNA polymerase or associated genes, but possibly in a gene involved in regulation. This is not a likely scenario, however, as the genes involved in recognition and regulation of PULs are largely independent and the effect was seen for both Bt-1763 and Bt-3090. If this was the case, it would be a simple matter to repeat the genomic integration procedure on a different stock of *B. theta*.

Fourth, because nanobodies are binding domains, it may be the case that once produced they are binding to an unknown target within the *B. theta* cell. Because *B. theta* does not have a flagella it was predicted that this would not be an issue; however, nonspecific binding can occur with antibodies. As a result, if nanobodies are being produced at low levels, they may be binding to a protein important for the upregulation of these genes. Although it is not known what this protein might be, this would explain why the native Bt-1763 and Bt-3090 genes are not upregulated as

they were in the absence of nanobody genes. If this was the case, crude lysates of *B. theta* could be visualized with a western blot using the nanobodies as the primary antibody. If any signal was detected, this would suggest an unexpected binding event.

Finally, *B. theta* genetics are highly organised and possibly take advantage of avidity effects of regulator binding as well as transcriptional coupling as a result of several genes induced in the same condition being present in the same loci (Terrapon, 2015). Integrating a gene into the PUL75 knockout loci would result in a loss of both of these effects. Although this would only affect the expression of the heterologous gene and not the native susC, it may be responsible for a portion of the observed effect. If this were the case, the most straightforward solution is to integrate constructs into their respective PULs as described above.

2.5. Conclusion

Although induction of genomically integrated heterologous genes was not observed and plasmid-based expression was lower than desired, a foundation of genetic parts has been established for further experiments. Any promoter/terminator pair from the *B. theta* genome or other *Bacteroides* species can be inserted into either the pINT or pPPRAMA vector to test for expression from the genome or plasmid and reference genes have been validated for commonly used laboratory conditions. If the observed effect is consistent when new promoters are introduced, it is a simple matter to switch the PUL75 flanking region in these vectors for regions flanking any locus within the genome which will help avoid any issues associated with this locus.

Data presented here demonstrate that *rssA* is stably expressed. This is consistent with what is known about *rssA* espression; however, this sequence is generally not recommended due to

difficulties maintaining consistent measurements. $rpoD$, pgk, $dnaG$, and $rpoS$ have been shown to be the most stably expressed genes under these conditions and are recommended as alternatives to the commonly used $rrsA$.

Table. 2.1. Bacterial strains and plasmids used in this study. Relevant features and sources are listed.

Bacterial Strains	Features/Use	Source/Reference
B. theta VPI-5482	Wild type strain	(ATCC® 29148™)
B. theta VPI-5482 Δ*tdk*	Wild type with Bt-2275 knockout, FUdR resistant	(Koropatkin, 2008)
B. theta VPI-5482 Δ*tdk, ΔPUL75*	*B. theta* VPI-5482 Δtdk with Bt-4108 - 4123 knockout, recipient of pINT vectors	This study
B. theta VPI-5482 Δ*tdk, ΔPUL75::Pro1763-DSB-Term1763*	*B. theta* VPI-5482 Δ*tdk* with the *DSB* gene flanked by the Bt-1763 promoter and terminator inserted into the PUL75 knockout	This study
B. theta VPI-5482 Δ*tdk, ΔPUL75::Pro3090-DSB*	*B. theta* VPI-5482 Δ*tdk* with the *DSB* gene flanked by the Bt-3090 promoter inserted into the PUL75 knockout	This study
B. theta VPI-5482 Δ*tdk, ΔPUL75::Pro3090-DSB-Term3090*	*B. theta* VPI-5482 Δ*tdk* with the *DSB* gene flanked by the Bt-3090 promoter and terminator inserted into the PUL75 knockout	This study
B. theta VPI-5482 Δ*tdk/pPPRAMA, Pro3090-DSB*	*B. theta* VPI-5482 Δ*tdk* propagating the pPPRAMA, Pro3090 vector containing the *DSB* gene	This study
Plasmid	Features/Use	Source/Reference
pExchange-*tdk*	Parent vector for pINT and pPPRAMA, returns FUdR sensitivity	(Koropatkin, 2008)
pET28a	Source of pINT and pPPRAMA MCS	(Novagen, Cat.69864-3)
pFD288	Source of pPPRAMA RepA/MobA	(Smith, 1995)
pINT, Pro/Term1773	Integration vector for the insertion of a gene flanked by Bt-1763 Promoter/Terminator into the PUL75 knockout	This study
pINT, Pro3090	Integration vector for the insertion of a gene flanked by Bt-3090 Promoter into the PUL75 knockout	This study
pINT, Pro/Term3090	Integration vector for the insertion of a gene flanked by Bt-3090 Promoter/Terminator into the PUL75 knockout	This study
pPPRAMA, Pro3090	Expression vector for plasmid-based expression of a gene regulated by the Bt-3090 promoter	This study

Table. 2.2. Primers used for RT-qPCR. *B. theta* gene number and the gene symbol used in this chapter are listed. References listed for putative reference genes are studies that used analogous genes in other species. References listed for *susC* genes are studies that used these genes previously and are the sources of these sequences.

Gene Symbol	Gene Number	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Gene ID	Reference	Efficiency (%)
gluS	Bt-0785	CGACACTTTCGATGCTTTGT	GATTCATGATAACCGCGTTG	>gi|880966697|	(Takle, 2007; Hesami, 2011)	96.2
mcat	Bt-0789	GTTTGCGTACCTGCCAACTA	ATGGATTTCTGTTGCGTTGA	>gi|939728750|	(Theis, 2007)	99.7
gyrA	Bt-0899	ATTGAAACCGGACAGACACA	GCTTGCATTAGCATCACGTT	>gi|880967236|	(Rocha, 2015)	100.0
rpoD	Bt-1311	CCCGTTGAATCAGGTAGGTT	GGCTGTTGTCTTCTCCTTCC	>gi|29346721|	(Savli, 2003; Desroche, 2005, Theis, 2007)	96.9
rho	Bt-1595	GATTCTTGATGAGCAAGCCA	TAGGCTGTTGCTGAGGTTTG	>gi|499420420|	(Rocha, 2015)	99.7
srp	B-1601	GCTATCGAACAGTTGCGTGT	CGGTGATCTCGTTCATCATC	>gi|496043411|	(Takle, 2007)	96.3
pgk	Bt-1672	TCATCGCAGGTGGTATGACT	GCGCAGAACTGAGTGTTAGC	>gi|29347082|	(Vafadarnejad, 2015)	99.6
rpoS	Bt-1907	TGGGCGTACAGATATCAATCA	TGTCGGCATGGAGTTATAGG	>gi|499420585|	(Savli, 2003)	95.2
gmk	Bt-2009	GATGTAGTAGGCGGATGCAA	CATTCACGATGACACGATCA	>gi|499420622|	(Theis, 2007)	99.7
rplI	Bt-2161	AAAGAAGACATCGTAAACTTGGG	GAAACGCCTTCCAACTTAGC	>gi|29347571|	(Hesami, 2011)	97.5
rplQ	Bt-2700	GTCCGGGTGGTTATACTCGT	TGGTTCTTCAGTTGCAGGAG	>gi|29348109|	(Hesami, 2011)	97.2
dnaG	Bt-3932	ACTCCATCAAGTCCCTACGC	CTTCCTGCGGAATAAAGGAG	>gi|499421561|	(Desroche, 2005)	97.8
ipxC	Bt-4206	CCAGCCGTACTTTCGTCTTT	TGGTTTGTGATTGATGTACCC	>gi|29341524|	(Hommais, 2011)	100.0
gapdh	Bt-4263	AGGCTCATATCGAAGCTGGT	AACCGTCAAGGATACCGAAC	>gi|29349671|	(Rocha, 2015)	96.5
recA	Bt-4610	ACAGGTGGTAACGCATTGAA	CCCAAATCGATGATTTCTCC	>gi|29350018|	(Rocha, 2015)	100.0
rrsA	Bt-r03 Bt-r06 Bt-r09	GGTAGTCCACACAGTAAACGATGAA	CCCGTCAATTCCTTTGAGTTTC	>gi|444303855|	(Converse, 2009; Nava, 2011; Bloom, 2011; Martens, 2011; Hehemann,	96.1

61

	Bt-r12 Bt-r13			2012; Rogers, 2013; van Bueren, 2015)	
sc22	Bt-1763	ATGCCTGGTCACCTACGAAC	CAAGCGGTCCATTCTCATTT	>gi|29339072|	(Sonnenburg, 2010)	97.4
sc48	Bt-3090	ATGCTGAATGCCGCCCAATA	CGAGAAAACCGCCGGATACATA	>gi|29348499|	(Rogers, 2013)	98.1
DSB	N/A	CTGTGCAACCTCTGGTCTCA	TGTTCTTGGCGTTGTCTCTG	>gi|576090030	(Riazi, 2013)	96.7

Table. 2.3. Putative reference gene ranking from most to least stable.

	geNorm	geNorm Stability Value	NormFinder	NormFinder Stability Value	BestKeeper	BestKeeper Stability Value
1	rpoS/dnaG	0.233	pgk	0.226	rpoD	0.35
2	-	-	rrsA	0.308	rpoS	0.42
3	rpoD	0.306	recA	0.418	dnaG	0.49
4	rplI	0.373	Mcat	0.42	mcat	0.61
5	Gapdh	0.428	Srp	0.735	rrsA	0.65
6	Mcat	0.504	Rho	0.771	rplI	0.67
7	rrsA	0.571	rplI	0.805	gapdh	0.74
8	Pgk	0.628	dnaG	0.888	pgk	0.96
9	recA	0.718	rpoS/gapdh	0.928	recA	1.02
10	Rho	0.825	-	0.934	rplQ	1.04
11	Srp	0.891	ipxC	0.952	rho	1.22
12	ipxC	0.955	gyrA/rpoD	0.998	srp	1.24
13	gyrA	0.996	-	1	ipxC	1.33
14	Gmk	1.059	gmk	1.352	gyrA	1.35
15	gluS	1.112	gluS	1.459	gluS	1.53
16	rplQ	1.18	rplQ	1.551	gmk	1.59

Table. 2.4. Primers used in vector generation. Lower case letters represent restriction endonuclease recognition sequences used.

Primer Name	Sequence (5'-3')
Primer-1	TATTtctagaAATATCGGTGTTTTTTGTTAGAAC
Primer-2	AATAgctagcTTTCAGTCCTTCCAGTTCTG
Primer-3	TATTctgcagAATGTAGCAAATGATAGCGG
Primer-4	TTAAagatctGGTTTATTGTTTTAAGGATTATAC
Primer-5	TTAAagatctTATACCATGGGCAGCAGC
Primer-6	ATATATtctagaTTAGCAGCCGGATCTCAGTGGTGGTGGTGG
Primer-7	TGTTTACGCGTCCTCGGTACTGATTTTGTGCTGTTTTGGGGG
Primer-8	AAAGGTACCCACAACCGTTTTCCGAACCC
Primer-9	TTTggtaccGCTTCTTCCACAACAGTCTGCGG
Primer-10	AAAGGTACCCACAACCGTTTTCCGAACCC
Primer-11	TATTctgcagTTAAAGATCTTATACCATGGGCAGCAGC
Primer-12	TATATAagatctCTATGATTGAGAGAGAATATAATCC
Primer-13	ATATATccatggTAGTTTAATGTTATTAATTTAAAAGTACG
Primer-14	TATATAtctagaGAAAACTACACCGGCAAG
Primer-15	TATATAgctagcGTTCTATGGTATTAAGGTTGTTATAC
Primer-16	TATATAAGATCTGAAATCCACTACTTTTTTAGCACG
Primer-17	ATATATCCATGGGTACATCAATTTAAAGTTAATATTAGG
Primer-18	TATATATCTAGAGACTTGAACTCAACGCCTAC
Primer-19	TATATAGCTAGCGTTTCATACCTTAAAACAAACTCG
Primer-20	TATAtcatgaCAGGTGCAGCTGGTGG
Primer-21	TTAActcgagCCACCAGCTGCACCTGA
Primer-22	CCAATCACGATACCCAGGAATAA
Primer-23	CCACGAATCAACTGTAACTGATG

Chapter 3. Nanobody Expression, Targeting, and Functionality

Abstract

In order to engineer complex pathways in *B. theta*, it would be beneficial to have the capacity to target proteins to different subcellular locations within the cell. This can be achieved by fusing signal peptides to protein cargo, allowing native secretion pathways to deliver proteins in a regulated manner. Although signal peptides are well studied in a number of other bacteria, it is currently unknown whether *B. theta* secretion systems and signal peptides function similarly to *E. coli* systems. Presented here, seven novel signal peptides from a number of *Bacteroides* spp. and two previously characterized signal peptides that are predicted to target the periplasm, outer membrane, or extracellular environment, have been fused to two single domain antibodies (nanobody) for expression in *E. coli*. In order to reproducibly determine the localization of the nanobodies, a method has been developed for the systematic fractionation of *E. coli* cells. Using this approach, four signal peptides that target proteins to the extracellular environment and one that targets the outer membrane have been demonstrated to be functional in *E. coli*. Notably, secreted proteins displayed a statistically significant retention in epitope binding. This study represents a proof of concept for the validation of *Bacteroides* signal peptides in a model *E. coli* expression system.

3.1. Introduction

3.1.1 Signal Peptides

Signal peptides are short amino acid sequences appended to the N-terminus, C-terminus, or internal structure of a protein (Pfeffer, 1987). They are employed across all domains of life and serve to signal the subcellular localization of secreted proteins (Bernstein, 1989; 1993). In Gram-

negative cells, signal peptides traffic proteins out of the cell into the environment, into the periplasm, or onto the surface of or into cellular membranes (Ivankov, 2013). Conventionally, proteins that lack signal peptides are not secreted and are destined to function within the cytoplasm (Ivankov, 2013) (Figure 3.1A).

Accurate subcellular localization of proteins is essential for the function of many different classes of proteins, including virulence factors, transporters, and catabolic enzymes. Virulence factors must be secreted from the cell to reach their target, transporters must be correctly positioned within membranes to facilitate the passage of metabolites, and many catabolic pathways require targeting to multiple locations within the cell.

3.1.2 Secretion Systems

There are a number of pathways used by prokaryotes to translocate proteins across plasma membranes and into the environment typically referred to as Type I to Type IX secretion systems (T1SS – T9SS) (Desvaux, 2009). *Bacteroides thetaiotaomicron* (*B. theta*) has several putative T1SS including Bt-4288 and Bt-4289, homologous to *HlyB* and *HlyD* respectively, the minimal units of the *E. coli* hemolysin secretion pathway and Bt-0304 - Bt-0306, homologous to *AcrA*, *AcrB*, and *TolC* of the *E. coli* Acridine efflux pathway (Xu, 2003). *B. theta* also contains Bt-0303, which is homologous to the recently discovered Type Vd Patatin-like pathway of *Pseudomonas aeruginosa* (Salacha, 2010). Interestingly, *B. theta* lacks homologous structures to the commonly found T2SS-T4SS, subtypes a-c of T5SS, and the less common T6SS-T9SS. This raises questions as to how similar *B. theta* secretion systems are to the well-studied systems of other species and whether or not signal peptide prediction algorithms will be accurate.
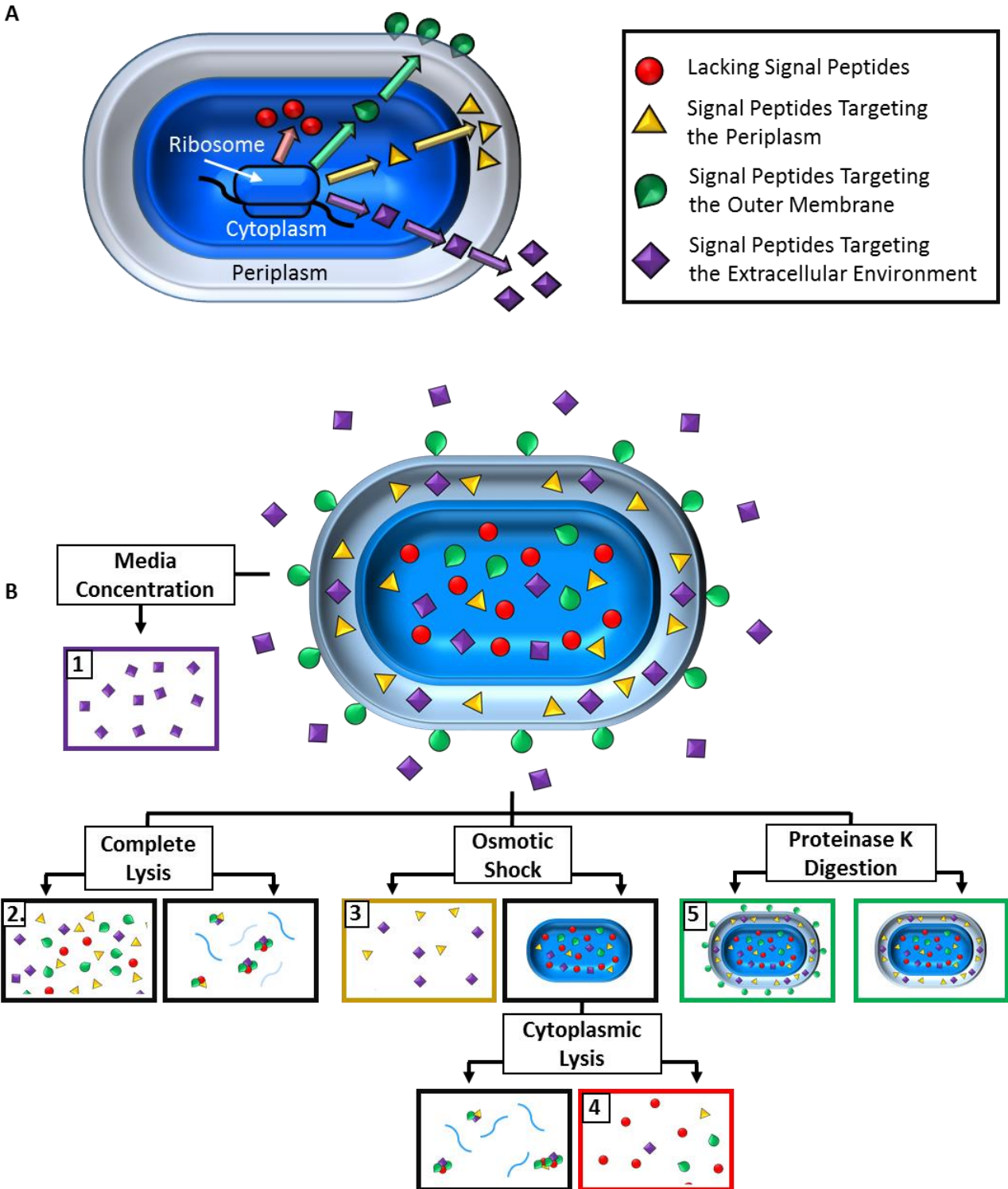
Figure 3.1: (A) General structure of a Gram-negative cell. Proteins are synthesized in the cytoplasm, where they can remain (red circles) or be targeted into the periplasm (yellow triangles), outer membrane surface (green teardrop), or released into the environment (purple diamonds). **(B)** Cell fractionation protocol: 1. Spent media is removed from the culture and concentrated allowing for the identification of secreted proteins, 2. Cells are lysed into soluble

and insoluble fractions allowing for visualization of all cellular proteins, 3. Osmotic shock fractionates the periplasmic contents and leaves the inner membrane intact, 4. Cytoplasmic compartment remaining after osmotic shock is lysed into soluble and insoluble fractions, 5. Proteinase K digestion followed by full cell dot blot allows for the identification of proteins on the cell surface. The purple panel represents the fraction containing secreted proteins, yellow panel represents periplasmic proteins, red panel represents cytoplasmic proteins, and green panels were used to verify surface exposed proteins.

Typically, bacteria use T1SS, T3SS, and T4SS to secrete proteins directly from the cytoplasm into the environment in a one-step process; and T2SS and T5SS to transport proteins from the periplasm across the outer membrane into the environment in a two-step process (Desvaux, 2009). Two-step processes typically use the general secretory pathway (SEC-dependent secretion), which translocates the unfolded polypeptide; or the Twin-Arginine Translocation pathway (TAT-dependent secretion), identified by the tandem pair of arginine residues in its signal peptide, which translocates the protein in a fully-folded form (Natale, 2008). Once in the periplasm, proteins destined for extracellular secretion are transported through the T2SS or T4SS machinery (Johnson, 2006). T5SS proteins contain a transmembrane pore and are able to auto-translocate across the outer membrane (Henderson, 2004).

3.1.3 SEC-dependent Secretion

The SEC-dependent secretion pathway is typified by the *E. coli SecABDEFGY* translocon (Beckwith, 2013). *SecY* (homolog: Bt-2707), *SecG*, (homolog: Bt-4378), and *SecE* (no predicted *B. theta* homolog) form the membrane complex SecYEG, which serves as the inner membrane transporter. Interestingly, although it has been shown that the complex is minimally functional in the absence of *SecG*, it is commonly thought that the complex is non-functional in the absence of *SecE* which is responsible for scaffolding and maintaining the structure of the complex (Akimaru, 1991). *SecD* and *SecF* (homolog: Bt-2835) form a membrane-bound complex on the periplasmic side of the inner membrane and are thought to aid in translocation of polypeptides making use of the proton motive force (Tsukazaki, 2011). *SecA* (homolog: Bt-4362), is an ATPase that reversibly binds to the *SecYEG* complex and hydrolyses ATP, using the energy to translocate the nascent polypeptide through the *SecYEG* complex (Tsukazaki, 2011). Importantly, *SecB*, a protein that binds to nascent polypeptide strands preventing their folding prior to arriving

at the translocon, does not have a detectable homologue in *B. theta* (Xu, 2003). This suggests that secreted proteins in *B. theta* interact with an analogous protein, interact with a homologous protein with low sequence identity, unfold prior to transport, or are synthesised at the site of translocation.

3.1.4 TAT-dependent Secretion

The TAT-dependent secretion pathway is a relatively uncomplicated system. In *E. coli*, three proteins are required, *TatABC* (Natale, 2008). *TatC* (homolog: Bt-4103) is an integral membrane protein that binds to the TAT motif of the signal peptide (Natale, 2008). Upon binding, *TatA* (homolog: Bt-4102) associates with *TatC* and multimerises with other TatA monomers forming a pore that the fully folded protein can pass through (Natale, 2008). Following secretion the signal peptide is cleaved and the *TatA* complex dissociates. *TatB* (no known homolog) serves a function analogous to *TatA* (Natale, 2008). Although it is unusual in Gram-negative bacteria some Gram-positives such as *Bacillus subtilis* also lack a *TatB* homologue (Berks, 2015).

Because *B. theta* lacks detectable homologs of T2SS and T4SS, it is unclear how proteins lacking their own trans-membrane pore, as in the case of T5SS, manage to cross the outer membrane after being transported into the periplasm. As a result of the non-typical secretion and poorly understood strategy for secretion in *B. theta*, it was unclear whether signal peptides from other species would be functional in *B. theta*. To circumvent this potential issue, I opted for investigating the function of novel *Bacteroides* signal peptides rather than adopting the well-established peptides of other Gram-negative species.

Defining a library of signal peptides from *Bacteroides* species will be important for bioengineering new catabolic pathways as well as the targeted secretion of heterologous genes

such as immunomodulators and bio-active compounds. Bioinformatic algorithms exist for predicting subcellular localization of proteins in Gram-negative bacteria such as SignalP (Petersen, 2011), which predicts signal peptides and cleavage sites; LipoP (Juncker, 2003; Rahman, 2008), which predicts signal peptide type; and PSORTb (Nancy, 2010), which predicts subcellular localization based on probability scores. These programs however, rely on published data for select peptides in Gram-negative bacteria and only have limited coverage of functional secretion signals in nature (Petersen, 2011). As a result, these algorithms have limited accuracy in their predictions and it is unknown whether they can accurately predict signal peptides from *B. theta*. Consequently, peptides require validation for applications in protein trafficking.

3.1.5 Previous Bacteroides Signal Peptide Validation

Previously, the signal peptide from the secreted *B. fragilis* metalloprotease Fragilysin was fused to murine interleukin-2 (mIL-2) (Farrar, 2005), human keratinocyte growth factor-2 (HKGF2) (Hamady, 2010), and TGF-β_1 (Hamady, 2011). This signal peptide is predicted to target the extracellular environment by PSORTb, and is predicted to be a lipoprotein by LipoP. When expressed in *Bacteroides ovatus*, this sequence successfully targeted the protein to the extracellular environment, albeit at low levels (Farrar, 2005; Hamady, 2011). A maximum concentration of 849.9 pg ml^{-1} of mIL-2 was detected in the medium of constructs containing the signal peptide, increased from 44.2 pg ml^{-1} found in the media of a construct lacking the signal peptide. Nearly a three-fold increase (539.5 pg ml^{-1}) of mIL-2 was found in the cell lysate of the construct lacking the signal peptide compared to the cell lysate of the construct with the signal peptide. This suggests that not only was mIL-2 successfully targeted out of the cell but a greater gross concentration of protein was produced with the signal peptide. Cytoplasmic concentrations were not reported for HKGF2 or TGF-β_1; however both proteins were detected in the media

(Hamady, 2010; 2011). Concentrations of HKGF2 similar to mIL-2 were detected with a maximum concentration of greater than 700 pg ml^{-1} detected after 24 h however lower concentrations of TGF-β_1 were reported with a maximum concentration of greater than 300 pg ml^{-1} (Hamady, 2010).

Presented here, I have selected seven signal peptides from several *Bacteroides* species using LipoP and PSORTb for evaluation (Table. 3.1). These sequences where selected for specific outcomes: *Theta-OM* for outer membrane-bound, *Theta-SEC1, Theta-SEC2, Sala-SEC, Vulg-SEC,* and *Frag-SEC* for SEC-dependent secretion, and *Theta-TAT1* and *Theta-TAT2* for TAT-dependent secretion. Additionally, the *B. fragilis* Fragilysin peptide and *E. coli* OmpA (Takahara, 1985) were selected as positive controls for trafficking to the extracellular environment and periplasm, respectively. For this study, signal peptides were appended to the N-terminus of two single domain antibodies (nanobody): FlagV1-F23-DSB (*DSB*) (Riazi, 2013) and FlagV6 (*V6*) (Hussack, 2014). These proteins were selected for proof-of-concept studies because they express well in recombinant systems, are detectable with anti-hexahistidine antibodies, and can be functionally assayed for retention of binding. Both nanobodies bind epitopes on the flagella of *Campylobacter jejuni* (*C. jejuni*), which leads to a reduction in motility and loss of virulence (Riazi, 2013).

In order to determine the final destination of a secreted protein, a protocol was developed to systematically fractionate the cell and detect both periplasmic and the outer membrane surface localization with greater accuracy (Figure 3.1).

3.1.6 Periplasmic trafficking

Although, several protocols exist for the complete cell lysis of Gram-negative bacteria including sonication (Harvey, 1929), French-pressure cells (French, 1955), and the application of lysozyme with detergents; methods to separate periplasmic from cytoplasmic contents without contamination from cytoplasmic proteins are uncommon. The standard technique for this purpose is 'osmotic shock', which ruptures the cell by introducing a rapid change in osmolarity (Neu, 1965; Nossal, 1966). Cells are typically incubated in a hyperosmotic solution such as buffered sucrose, as the cell imports sucrose into the periplasm, the osmolarity of the periplasm increases. The cells are then rapidly removed from the hyperosmotic solution and resuspended in a hypotonic solution such as dH_2O. Because the inner plasma membrane is supported by peptidoglycan and the outer plasma membrane is not, the outer membrane ruptures allowing the periplasmic contents to be removed from intact cytoplasmic compartments. Although this protocol is effective for releasing periplasmic proteins from the cell, it often results in minor contamination of the periplasmic proteins with proteins from the cytoplasm (French, 1996; Chen, 2004). Contamination with cellular proteins can lead to the erroneous conclusion that they have been biologically secreted across the inner membrane. To improve upon the method of Neu and Nossal (Neu, 1965; Nossal, 1966), concentrations of hypo and hyperosmotic solutions have been optimized to ensure that the inner membrane remains intact throughout the process. Although, complete lysis of the outer membranes may not occur using this method, which would result in an underestimation of the amount of protein in the periplasm, it provides more confidence that the periplasmic fraction represents periplasmic proteins.

3.1.7 Outer membrane surface trafficking

To determine whether protein is efficiently secreted to the outer membrane surface, a protocol was adapted from (Sharp, 2012). Cells expressing *V6* and *DSB* nanobodies fused to an outer membrane-targeting peptide at their N-termini were adsorbed to a PVDF membrane. Blots were visualized using immunochemistry before and after Proteinase K treatment to determine whether proteins were detected on the cell surface. Assuming the cells remained intact, only nanobodies exposed on the outer membrane should be visualized, and this signal should be ablated after proteolytic digestion. To ensure Proteinase K was exclusively digesting outer membrane proteins and not all cellular proteins resulting in false positives, treated and untreated cells were lysed and analyzed by western blots.

3.1.8 Validation of Nanobodies

To determine the spectrum of interaction for *V6* and *DSB* which were raised against *C. jejuni* 81-176, 79 strains of clinically relevant *C. jejuni* which had previously been isolated from poultry were selected based on genetic and environmental diversity discussed below (Section 3.2.1.2). To determine whether the nanobodies bound the flagella of these strains, whole cell ELISA were performed (Elder, 1982; Prieto, 2003). Live cells were adsorbed to polystyrene plates, nanobodies were allowed to bind the flagella, and antibodies specific for *V6* and *DSB* were detected colourimetrically. To ensure that nanobody binding resulted in reduced motility of these strains after binding, semi-solid agar motility assays were performed (Guerry, 1991). Cells were incubated with *DSB* or *V6*, or left untreated prior to being injected into the agar plates. Unhindered bacteria remain motile and translocate through the agar matrix while nanobody-immobilized bacteria do not. The area of growth disks was compared between treated and untreated samples to determine the reduction in motility for the full spectrum of *C. jejuni* strains

74

tested. To ensure that secreted nanobodies retain their binding function, assays were repeated on strains that had been detected with ELISA using secreted nanobodies purified from the media.

3.2. Materials and Methods

3.2.1. Bacterial Strains

3.2.1.1 E. coli

Protein expression and cell fractionation was performed on *E. coli* BL21 Star (DE3): F⁻ *ompT hsdS$_B$* (r$_B$⁻m$_B$⁻) *gal dcm lon rne131* λ (DE3 [*lacI lacUV5-T7* gene 1 *ind1 sam7 nin5*]).

3.2.1.2 C. jejuni

Strains were selected using comparative genomic fingerprinting (CGF) as described previously (Taboada, 2012). The 5,535 chicken isolates previously analyzed by the National Comparative Genomic Fingerprinting Database (NCGFDb) were screened for the following characteristics: (1) clusters of at least 25 isolates based on CGF data, (2) isolations from at least two geographically distant sources, (3) at least 15% of isolates in a cluster were sourced from poultry to ensure that poultry is a relevant reservoir for the cluster, and (4) at least 5% of isolates in a cluster were sourced from humans to ensure that the cluster is associated with diarrheic disease and is therefore clinically relevant. This screen established 38 unique clusters from which 79 representative strains were selected for further experiments. Because both nanobodies were raised against the common laboratory strain *C. jejuni* 81-176, this strain was used as a control in further experiments. Motility assays and whole-cell ELISAs were performed on the 80 strains identified in (Table. 3.2).

3.2.2. Culture conditions

3.2.2.1 Lysogeny Broth (Bertani, 1951)

E. coli were grown in 1.0% (w/v) Bacto-tryptone, 0.5% (w/v) yeast extract, 171 mM NaCl under aerobic conditions with 180 rpm agitation. Cells were grown at 37°C until an optical cell density (OD_{600nm}) of 0.8. Cultures were then cooled to 16°C and induced to express protein with a final concentration of 250 mM IPTG. Cultures were subsequently grown for an additional 16 h prior to protein harvest.

3.2.2.2 Karmali Agar (Karmali, 1986)

Strains of *C. jejuni* were refreshed from glycerol stocks by streaking ~2 μL portions on agar containing 1.5% (w/v) agar, 1.0% (w/v) yeast enriched peptone, 0.5% (w/v) enzyme hydrolyzed casein, 0.5% enzyme hydrolyzed animal tissue, 0.4% (w/v) activated charcoal, 0.3% (w/v) enzyme hydrolyzed cardiac tissue, 25 mM Tris pH 7.3, 13.9 mM glucose, 8.5 mM NaCl, 5.6 mM Na_2CO_3, 830 μM $MgSO_4$, 400 μM L-cysteine, 131 μM $FeSO_4$, and 50 μM haemin. Cultures were incubated at 37°C in a microaerobic atmosphere (5% O_2, 10% CO_2, and 85% N_2) for 36-48 h.

3.2.2.3 Mueller-Hinton Broth (Mueller, 1941)

Strains of *C. jejuni* were cultured by transferring cells from Karmali agar plates into Mueller-Hinton broth (1.75% (w/v) acid hydrolysed casein, 0.3% (w/v) beef extract, 0.15% (w/v) starch). Cultures were incubated at 37°C in a microaerobic atmosphere (5% O_2, 10% CO_2, and 85% N_2) for 36-48 h without agitation.

3.2.2.4 Semi-soft Mueller-Hinton Agar

Motility assays of *C. jejuni* were performed on semi-soft Mueller-Hinton agar agar plates containing 1.75% (w/v) acid hydrolysed casein, 0.4% (w/v) agar, 0.3% (w/v) beef extract, 0.15% (w/v) starch. Plates were cured at room temperature for at least one h prior to inoculation. Inoculated plates were incubated at 37°C and microaerobic conditions (5% O_2, 10% CO_2, and 85% N_2) for 36-72 h.

3.2.3. Protein Expression and Cell Fractionation

3.2.3.1 Signal Peptides

Proteins from *B. theta* species were first identified based on predictions of PSORTb (Nancy, 2010) in the PSORTdb database by selecting the highest probability signal peptides in the genome. Proteins predicted to be targeted to either the cytoplasm, cytoplasmic membrane, periplasm, outer membrane, extracellular and probabilities were given as a value out of 10. Selected protein predictions were analyzed further with LipoP (Juncker, 2003; Rahman, 2008) which predicts whether a protein lacks a signal peptide, has a signal peptide type I (SPI), typically secreted, or signal peptide type II (SPII), typically a lipoprotein. Finally, the specific cleavage site of the signal peptide which determines the full length of the peptide was predicted with SignalP (Petersen, 2011). All peptides were used as they were found on the native protein with the exception of *Theta-SEC2* which had a cysteine as its C-terminal residue. This residue was mutated to a serine because this is often associated with palmitoylation (Kovacs-Simon, 2011). Sequences and predictions are presented in Table 3.1.

3.2.3.2 Protein Purification

2 L of *E. coli* expressing nanobody lacking a signal peptide were centrifuged at 8 000 RCF for 10 min and medium was discarded. Cell pellets were lysed by sonication and lysates were centrifuged for 1 h at 13 000 RCF. Lysates were passed through a gravity-fed Ni^{2+}-NTA sepharose affinity chromatography column. The column was washed with 20 mM Tris pH 8.0 with 500 mM NaCl and protein was eluded in a stepwise fashion by increasing imidazole concentration from 10-500 mM. Protein containing fractions were identified by electrophoresis and dialyzed against 20 mM Tris pH 8.0 with 50 mM NaCl.

To purify secreted nanobodies, 2 L of *E. coli* expressing either nanobody fused to *Theta-SEC1, Theta-SEC2, Sala-SEC, Vulg-SEC, Frag-SEC* signal peptides were centrifuged at 8 000 RCF for 10 min. Spent medium was then passed through a 0.45 μm cut-off filter to remove remaining cells. Filtered media was then concentrated with an EMD Millipore Amicon™ Bioseparations Stirred Cell to a volume of 100 mL and dialyzed against 20 mM Tris pH 8.0 with 50 mM NaCl. Samples were purified with a Nickel-NTA column as above.

3.2.3.3 Cell Lysis (Figure 3.1B.2)

Culture equivalents of ~10^{10} cells were centrifuged for 10 min at 13 000 RCF at 4°C . Cell pellets were resuspended in 500 μL of room temperature BugBuster Primary Amine-Free Solution (EMBMillipore #70923), and incubated at room temperature for 10 min. Samples were centrifuged for 10 min at 13 000 RCF at 4°C. The supernatant was used as total cell lysate.

3.2.3.4 Optimization of Osmotic Shock

Culture equivalents of ~10^{10} cells were centrifuged for 10 min at 13 000 RCF at 4°C. Cell pellets were resuspended in 1 mL of 5-20% (w/v) sucrose, 20 mM Tris pH 8.0, 1 mM EDTA at 4°C.

In addition, cell pellets were resuspended at each step by either gentle inversion within a tube, gentle resuspension with a wide-bore tip, vigorous resuspension with a standard-bore pipet tip, or by vortex to determine the effect of mechanical shearing of the membranes during the procedure.The resulting supernatant from the osmotic shock step was used as the periplasmic fraction (Figure 3.1B.3). The cell pellet, comprised of intact cytoplasmic compartments, was lysed as in section 2.3.1. The supernatant from this step was used as the cytoplasmic fraction (Figure 3.1B.4).

3.2.3.5 Media Concentration

In order to achieve comparable concentration of cellular protein to extra-cellular protein, samples of media were concentrated with a Savant™ SPD111V P1 SpeedVac™. 12.5 mL of culture was centrifuged 10 min at 13 000 RCF while refrigerated. Media was aspirated and centrifuged for an additional 10 min. Samples were then passed through a 0.4 μm membrane to remove any remaining cells. The remaining solution was dialyzed against 20 mM Tris pH 8.0 and concentrated to 500 μL (25X concentrate).

3.2.3.6 Western Blot

10 μL of cell lysate, periplasmic fraction, cytoplasmic fraction, and concentrated media each were added to 10 μL of 2 M glycerol, 1.28 M β-mercaptoethanol, 125 mM Tris pH 6.8, 140 mM SDS, 60 μM bromophenol blue. Samples were heated to 95°C for 10 minute to denature proteins and loaded into a 15% (w/v) polyacrylamide gel with a Precision Plus Protein™ All Blue Prestained Protein Standard (Bio-Rad #1610373). Gels were run at 200 volts in 52 mM glycine, 25 mM Tris pH 8.3, 3.5 mM SDS running buffer for ~1 h until the bromophenol blue load dye reached the bottom of the gel.

PVDF membranes were activated by soaking in methanol for 1 minute prior to sandwiching with a previously electrophoresed SDS-PAGE gel. Blots were electrophoresed for 2 h at 30 volts with 190 mM glycine, 25 mM Tris pH 8.3, 20% methanol transfer buffer.

Membranes were blocked by submersion in 5% (w/v) casein in TBS buffer and incubated for 2 h at room temperature. Membranes were then transferred into 5% (w/v) casein in TBS buffer containing 400 ng mL^{-1} anti-hexahistidine rabbit polyclonal IgG-HRP conjugate against hexahistidine (Bethyl Laboratories #A190-114P) and incubated at 4°C for 16 h.

Membranes were washed thrice in TBS buffer containing 0.1% (w/v) Tween-20. Finally, blots were visualized colourimetrically using an Opti-4CN™ Substrate Kit (BioRad #1708235). A solution containing 9 mL dH$_2$O, 1 mL Opti-4CN™ Diluent, and 200 μL Opti-4CN™ substrate was added to the blots and incubated at room temperature. Reactions were stopped by thoroughly rinsing the blots with dH$_2$O.

3.2.3.7 ELISA

Samples of cell lysate, periplasmic fraction, cytoplasmic fraction, and concentrated media were each diluted 1:10, 1:20, 1:40, and 1:80. One hundred μL of each dilution of each construct was added to a well of a Nunc Maxisorp 96-well plate. Plates were incubated at 4°C overnight and subsequently dried at room temperature overnight. Plates were washed twice with TBS buffer and 100 μL of 5% (w/v) casein in TBS buffer was added to each well followed by an incubation overnight at 4°C. Plates were washed with TBS and 100 μL of 5% (w/v) casein in TBS buffer containing 400 ng mL^{-1} rabbit polyclonal anti-hexahistidine IgG-HRP conjugate against hexahistidine was added to each well followed by incubation for 16 h at 4°C.

Plates were washed twice with TBS buffer and 100 μL of 1-Step™ Turbo TMB-ELISA Substrate Solution (Thermo Fisher Scientific #34022) was added to each well. Reactions were incubated for 10 min at room temperature and 100 μL of 2 M H_2SO_4 was added to each well to stop the reaction. Absorbances were measured at 450 nm with a Synergy™ HT Multi-Detection Microplate Reader. Samples were viewed in duplicate and average absorbances were normalized to a standard curve of purified, quantified *DSB* or *V6* expressed without a signal peptide.

3.2.4 Whole-Cell Dot Blot

To determine whether the signal peptide targeting the outer membrane was functional, constructs containing nanobody with the *Theta-OM* signal peptide were compared to nanobody lacking any signal peptide.

3.2.4.1 Proteinase K Digestion

Culture equivalents of ~10^{10} cells were centrifuged for 10 min at 13 000 RCF at 4°C. Cells were washed thrice with TBS by resuspending the pellet in TBS buffer, centrifuging and discarding the TBS buffer. The cultures were split evenly in two and centrifuged to remove any remaining TBS buffer. One pellet was resuspended in TBS buffer, the second was resuspended in TBS buffer containing 100 μM Proteinase K. Samples were incubated for one h at 37°C and cells were washed again to remove all Proteinase K. Samples were prepared in triplicate.

3.2.4.2 Dot Blot

PVDF membrane was activated with methanol and Proteinase K treated and untreated samples were spotted onto the membrane. PVDF membranes were allowed to dry completely at 30°C. Membranes were activated again with methanol prior to submersion in 5% (w/v) casein in TBS buffer. Blots were incubated for 2 h at room temperature, then transferred into 5% (w/v) casein in

TBS buffer containing 400 ng mL^{-1} rabbit anti-hexahistidine polyclonal IgG-HRP conjugate, and incubated with antibody for 16 h at 4°C.Blots were washed thrice in TBS buffer containing 0.1% (w/v) Tween-20 and visualized colourimetrically using an Opti-4CN™ Substrate Kit as in 3.2.3.4.

3.2.4.3 Confirmation of Intact Cells

To ensure that loss of signal due to Proteinase K was the result of membrane proteins being digested and not that the Proteinase K caused cell lysis and a loss of signal due to all cellular proteins being digested, Proteinase K treated and untreated samples were lysed as in 3.2.3.1 and analyzed by Western blot.

3.2.5. Nanobody Functionality

3.2.5.1 Whole-Cell ELISA

C. jejuni was grown in Mueller-Hinton broth to a cell density OD$_{600}$ of 0.3-0.6. Cultures were then washed twice with TBS buffer by centrifuging at 13 000 RCF for ten min followed by resuspension in TBS buffer. Cells were resuspended to a final cell density OD$_{600}$ of 0.1, 0.01, and 0.001. 100 μL of each dilution per strain was added to a well of a Nunc Maxisorp 96-well plate. Plates were incubated at 4°C overnight and subsequently dried at 37°C aerobically. Plates were washed twice with TBS buffer and 100 μL of 5% (w/v) casein in TBS buffer was added to each well followed by incubation overnight at 4°C.

Samples of *V6* and *DSB* without signal peptide were added to wells in 100 μL of 5% (w/v) casein in TBS buffer at a final concentration of 1 mg mL^{-1}. Plates were washed with TBS and 100 μL of 5% (w/v) casein in TBS buffer containing 400 ng mL^{-1} rabbit anti-hexahistidine polyclonal IgG-HRP conjugate was added to each well. Plates were incubated with antibody for 16 h at 4°C.

Plates were visualized with 1-Step™ Turbo TMB-ELISA Substrate Solution as in 3.2.3.5. Samples were viewed in duplicate and average absorbances were normalized to the maximally bound strain, values are reported as a percentage of maximal absorbance.

The protocol was then repeated on the 15 strains shown to bind at 50% or greater total absorbance with nanobodies isolated from the media of *E. coli* expressing either nanobody fused to *Theta-SEC2 or Vulg-SEC* signal peptide. Constructs containing nanobodies fused to either *Theta-SEC1, Sala-SEC,* or *Frag-SEC* signal peptide were not tested due to low yield. All concentrations and incubation times were maintained.

3.2.5.2 Motility Assay

C. jejuni was prepared as in 3.2.5.1 except final cell suspensions were diluted to an OD_{600nm} of 0.2. 10 μL of cell suspension was added to either 10 μL of TBS buffer for a negative control or 10 μL of 1 mg mL^{-1} of either *V6* (Riazi, 2013) or *DSB* (Hussack, 2014) resulting in a final concentration of cells at OD_{600} of 0.1 with or without a nanobody at 0.5 mg mL^{-1}. Solutions were incubated at room temperature for 30 min prior to being injected by pipet into the center of a 60 x 15 mm semi-solid agar plate. Plates were allowed to dry at room temperature prior to being incubated face-down at 37°C in microaerobic atmosphere for 36-72 h.

After negative controls had grown to ~3 cm in diameter, plates were photographed on a 0.5 cm^2 grid. Area of *C. jejuni* growth was calculated using ImageJ software (Schneider, 2012). Mean areas of nanobody treated cells were compared to mean areas of untreated cells per strain to calculate a percent reduction in area, reported as the percent reduction in motility. To determine which strains had significant effects, one-way ANOVA were performed followed by a protected Tukey-Kramer test. A value of $p < 0.05$ was deemed significant.

3.3. Results

3.3.1. Osmotic Shock Optimization

To determine the stringency of the osmotic shock protocol, cultures expressing *DSB* or *V6*, which should remain in the cytoplasm, were compared with OmpA-tagged *DSB* and *V6*, which should be targeted to the periplasm.

First, the effect of mechanical shearing of plasma membranes was examined by performing a stock procedure of 20% sucrose, 20 mM Tris pH 8.0, 1 mM EDTA followed by 5 mM $MgCl_2$ solution where at each step, cell pellets were resuspended by inversion, with a standard-bore pipet, a wide-bore pipet, or vortex. Samples were viewed in a 15% (w/v) polyacrylamide gel with Coomassie Brilliant Blue (Figure 3.2A). There was no apparent difference in results between resuspension methods. All subsequent protocols were performed by gentle resuspension with a standard-bore pipet.

Next, the length of incubation was investigated (Figure 3.2B, C). Again, a standard protocol was used except the incubation in sucrose solution was shortened to 2.5, 5.0, 7.5, or 10.0 min followed by a 10.0 minute incubation in $MgCl_2$ solution or a 10.0 minute incubation in sucrose solution followed by a 2.5, 5.0, 7.5, or 10.0 minute incubation in $MgCl_2$ solution. Length of incubation had no apparent effect on results. All subsequent protocols were performed with 10.0 minute sucrose solution incubation followed by 10.0 minute $MgCl_2$ incubation.

Finally, the concentration of hypertonic and hypotonic solutions were modified (Figure 3.2D, E). Samples were incubated in 5.0, 10.0, 15.0, or 20.0% (w/v) sucrose, 20 mM Tris pH 8.0, 1 mM EDTA followed by incubation in 5 mM $MgCl_2$ solution or 20.0% (w/v) sucrose, 20 mM Tris pH 8.0, 1 mM EDTA followed by incubation in 5, 25, 100, or 500 mM $MgCl_2$ solution. These

solutions were further optimized to either 5.0 or 10.0% (w/v) sucrose, 20 mM Tris pH 8.0, 1 mM EDTA followed by incubation in 100 or 200 mM $MgCl_2$ solution (Figure 3.2F).

Different resuspension techniques and incubation times had little effect on what proteins were released during the osmotic shock; however, variable hypertonic and hypotonic solutions showed greater variability. Determination of subcellular localization of proteins with other signal peptides was performed with standard-bore pipets, with 10 minute incubation in 10.0% (w/v) sucrose, 20 mM Tris pH 8.0, 1 mM EDTA, a 10 minute centrifugation at 13 000 RCF while refrigerated, followed by incubation in 100 mM $MgCl_2$ solution and finally, a 10 minute centrifugation at 13 000 RCF while refrigerated as this protocol released detectable concentrations of periplasmic nanobody without releasing detectable concentrations of cytoplasmic nanobody.

3.3.2. Cell Fractionation and Subcellular Targeting

DSB or *V6* fused to one of eight Bacteroides signal peptides, the *E. coli* OmpA periplasmic signal peptide, or lacking any signal peptide were expressed in *E. coli*. These cultures were fractionated by complete cell lysis for the release or all cellular protein, osmotic shock releasing periplasmic proteins, the lysis of cytoplasmic compartments releasing cytoplasmic proteins, and the removal and concentration of spent media for the identification of secreted proteins (Figure 3.1). Samples were also viewed before and after digestion of outer membrane proteins with Proteinase K for the identification of outer membrane-bound proteins.

3.3.2.1 Western Blot

Constructs lacking signal peptides were used as controls to demonstrate that strictly cytoplasmic proteins were not indiscriminately leaked during the osmotic shock protocol and OmpA-tagged
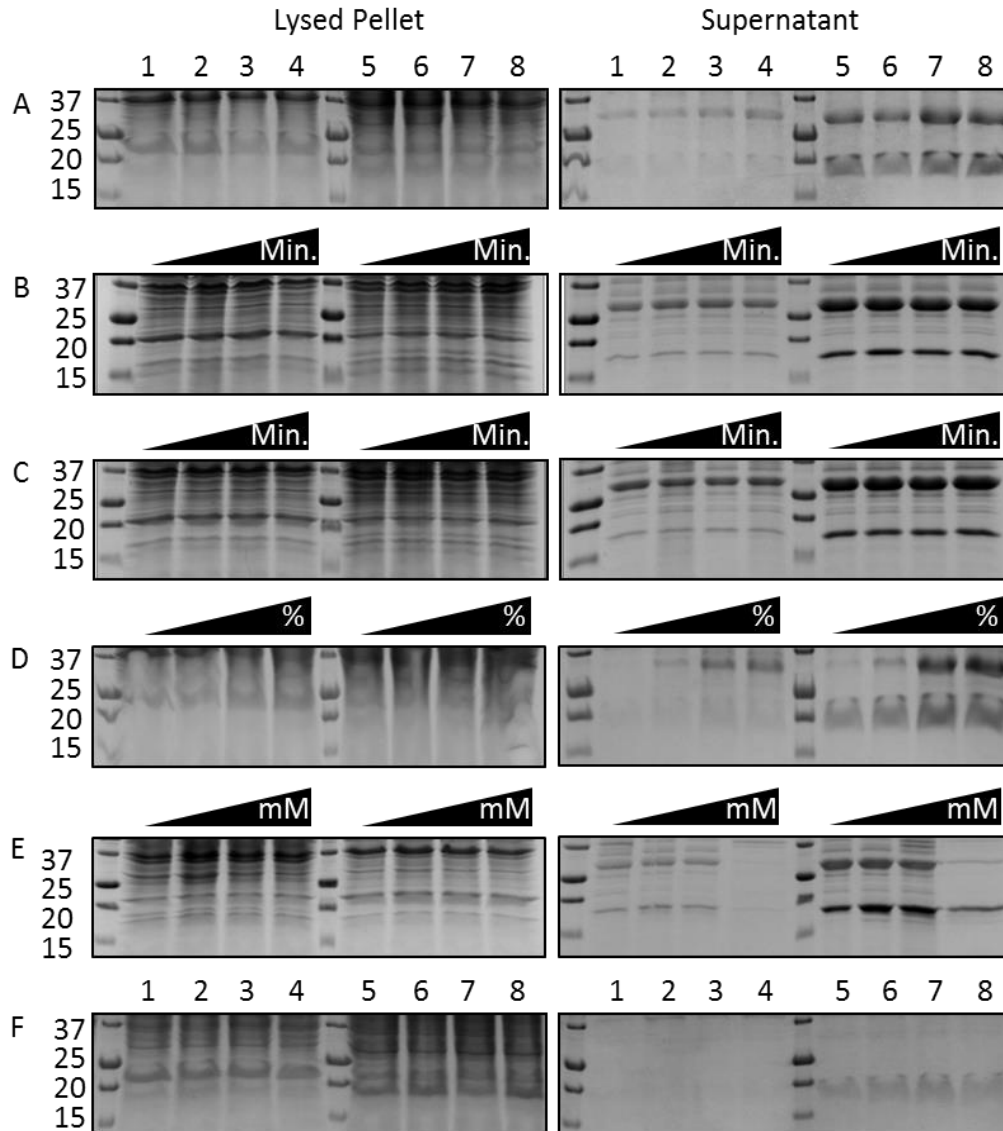
Figure 3.2: Optimization of osmotic shock protocol in *E. coli* for the determination of periplasm-targeted proteins. The left column contains the lysed cell pellets after the osmotic shock and the right column contains soluble protein released by the osmotic shock. Lanes 1-4 are constructs lacking a signal peptide (i.e. cytoplasmic), lanes 5-8 are constructs containing the OmpA signal peptide, which targets protein to the periplasm. **A.)** Standard protocol performed with 1.) inversion, 2.) Wide-bore pipet tip, 3.) Standard-bore pipet tip and 4.) Vortex. **B.)** Increasing length of incubation in hypotonic solution from 2.5 to 10 min. **C.)** Increasing length of incubation in hypertonic solution from 2.5 to 10 min. **D.)** Increasing concentration of hypertonic solution from 5 to 20% sucrose. **E.)** Increasing concentration of hypotonic solution from 5 to 500mM NaCl. **F.)** Further optimization of hyperosmotic and hypoosmotic solutions 1.) 5% sucrose, 100mM NaCl, 2.) 5% sucrose, 200mM NaCl, 3.) 10% sucrose, 100mM NaCl, 4.) 10% sucrose, 200mM NaCl. Solutions were incubated on ice for 2.5-10 min. Samples were then centrifuged for 10 min at 13 000 RCF at 4°C. Cell pellets were resuspended in 500 μL of room

temperature 5-500 mM MgCl$_2$ solution. Samples were incubated at room temperature for 2.5-10 min. Solutions were then centrifuged for 10 min at 13 000 RCF while refrigerated.

constructs were used as controls to ensure that periplasmic proteins were released. The protocol was repeated on a culture of *E. coli* lacking a nanobody construct to ensure other cellular proteins were not non-specifically bound by the antibody.

Nanobodies were detected in the periplasmic fraction as well as the concentrated media from highest to lowest concentration for constructs containing the *Vulg-SEC > Theta-SEC2 > Theta-SEC1 > Frag-SEC >* and *Sala-SEC* signal peptide suggesting a two-step process whereby proteins are targeted to the periplasm prior to being secreted. Neither TAT-dependent secretion tag appeared to have any effect on protein targeting as no protein was detected outside of the cytoplasm (Figure 3.3).

Whole cell dot blots reveal that protein is present outside of the cell when it is fused to the *Theta-OM* signal peptide and absent when the construct lacks signal peptide (Figure 3.4). The signal was lost after the cells were digested with Proteinase K, and is still present in the cell lysates when analyzed by Western blots of cell lysates. Interestingly, although protein was detected on the cell surface, it was not detected in the periplasm for these tagged constructs (Figure 3.3). This suggests that either there is a one-step mechanism for directing the protein from the cytoplasm to the membrane or that the mechanism for membrane insertion occurs rapidly.

3.3.2.2 ELISA

Samples were loaded into a Nunc Maxisorp 96-well plate, dried, and labelled with rabbit polyclonal anti-hexahistidine IgG-HRP conjugate. Blots were visualized with 1-Step™ Turbo TMB-ELISA Substrate Solution. Samples were diluted 1:10, 1:20, 1:40, and 1:80 to ensure that wells were not loaded above their binding capacity and that reactions would not surpass the maximum absorbance limit. It was found that often, inconsistent values were recorded for the
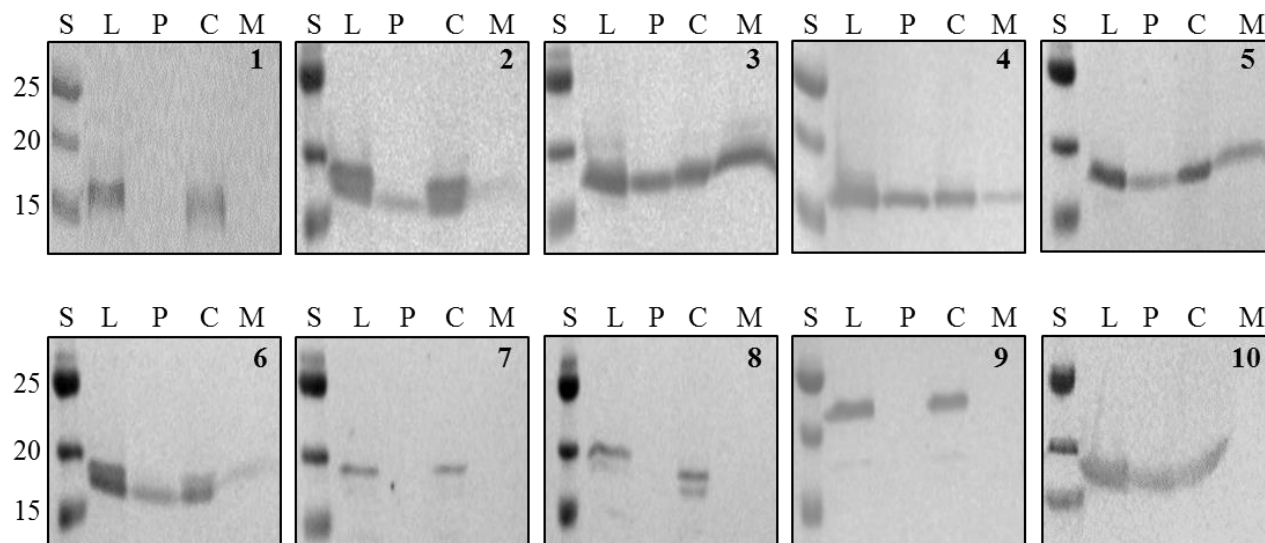
Figure 3.3 Western blots of fractionated *E. coli* expressing nanobody fused to **1.)** no signal peptide, **2.)** *Sala-SEC, 3.)* *Vulg-SEC,* **4.)** *Theta-SEC1*, **5.)** *Theta-SEC2*, **6.)** *Frag-SEC*, **7.)** *Theta-TAT1*, **8.)** *Theta-TAT2,* **9.)** *Theta-OM* and **10.)** *OmpA*. Lanes are a protein marker, cell lysate (L), periplasmic fraction (P), cytoplasmic fraction (C), and concentrated media (M).

concentrated media samples, likely due to the abundance of yeast extract and tryptone in solution. Dialysis to remove as much media concentrate from the secreted protein as possible allowed for binding of higher concentrations of sample that could be detected in a consistent manner.

3.3.3. Nanobody Functionality

To determine the variety of *C. jejuni* strains that the recombinantly produced *DSB* and *V6* are capable of binding, nanobodies were expressed in *E. coli* and purified. These purified proteins were then used in both whole-cell ELISA and motility assays.

3.3.3.1. Whole-Cell ELISA

79 strains of *C. jejuni* were grown in Mueller-Hinton broth at 37°C under microaerobic conditions. Cultures were washed in TBS buffer and diluted to a cell density OD_{600} of 0.1, 0.01, and 0.001. 100 μL of each solution was added to a well of a Nunc Maxisorp 96-well plate, samples were viewed in duplicate. The strain resulting in the highest absorbance and therefore the greatest nanobody binding potential was Cry_548 bound by the *V6* nanobody. This absorbance associated with this strain was used as the maximal absorbance and all other absorbances were expressed as percentages of this value in order to compare absorbances. Values calculated in this manner are presented in Table. 3.2 and represented in Figure 3.5B.

Of the strains examined, 66 (83.5%) were not bound by *DSB* and 54 (68.4%) were not bound by *V6* or were below the level of detection. In addition to V6 binding to an increased variety of strains, a greater percentage of strains were bound by higher concentrations of nanobody. 13 strains (16.4%) were bound at greater that 80% of maximal binding by *V6* compared to only 2 (2.5%) bound by *DSB*.
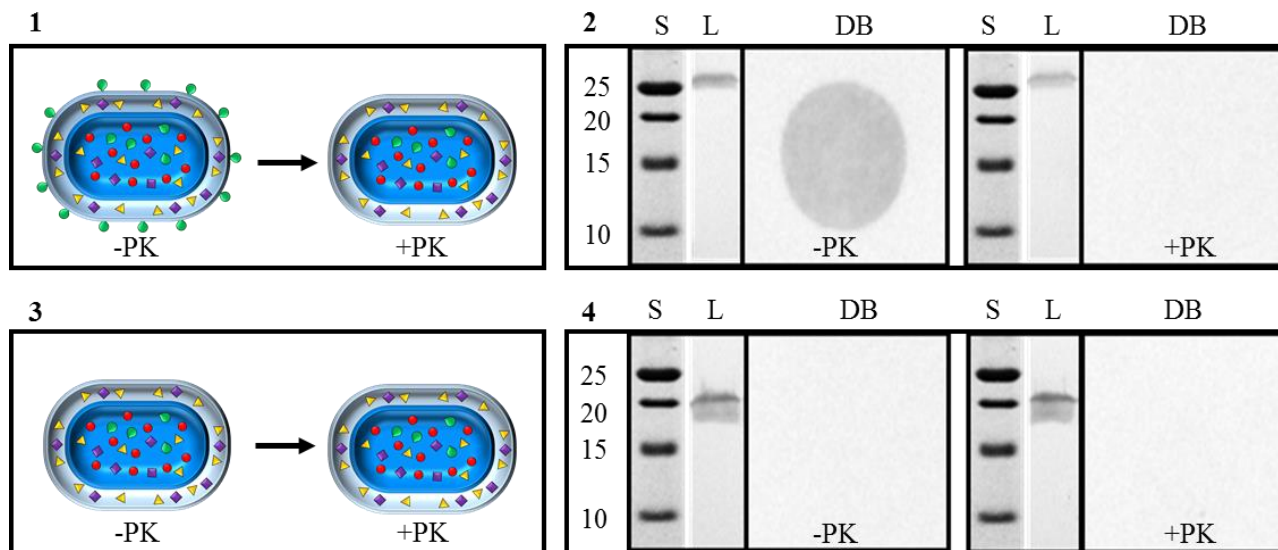
Figure 3.4: Full cell dot blot. The blot on the left (L) is a Western blot of the cell lysate of the spotted whole cell sample (DB). The right panel is a dot blot. **1.)** *E. coli* expressing nanobody fused to *Theta-OM.* **2.)** *E. coli* expressing nanobody fused to *Theta-OM* after digestion with Proteinase K. **3.)** *E. coli* expressing nanobody lacking signal peptide. **4.)** *E. coli* expressing nanobody lacking signal peptide after digestion with Proteinase K.

The protocol was repeated using nanobodies purified from the media of *E. coli* expressing nanobodies fused to *Theta-SEC2 or Vulg-SEC* signal peptide with the 15 strains that exhibited more than 50% of maximum binding by the nanobodies lacking signal peptide. Absorbances were not different ($p \leq 0.01$). This suggests that some nanobodies targeted out of the cell with these two signal peptides were capable of folding properly and retaining full functionality.

3.3.3.2 Motility Assay

Cells were diluted to OD_{600nm} of 0.2 and diluted 1:1 with either TBS buffer, or TBS buffer with a final concentration of 0.5 mg mL^{-1} of either nanobody. After incubation, the cells were injected into the center of a semi-solid agar plate and incubated at 37°C under microaerobic conditions. The samples lacking nanobody were allowed to grow cell disks ~3 cm in diameter, because of different growth rates and levels of natural motility, this required 36-72 h of incubation. Once this diameter was achieved, nanobody-incubated cell disks were compared to nanobody-free cell disks and the reduction of size was reported as a percentage in Table. 3.2 and represented in Figure 3.5B.

Of the strains examined, 67 (84.8%) did not have a reduction in motility by *DSB* and 55 (69.6%) were unaffected by *V6*. In spite of greater numbers of *V6* binding to a greater number of strains as determined by ELISA, this did not translate into a greater reduction in motility in many cases. Only two strains (2.5%) were reduced by greater than 80% by either nanobody (09:MS1956 by *DSB*, Wa10_732 by *V6*, and CHSK-I_158 by both), and only four strains (5.6%) were reduced by greater than 60% by *DSB* (06:MS2284, 09:MS1956, AMR_8257, and CHSK-I_158) and 5
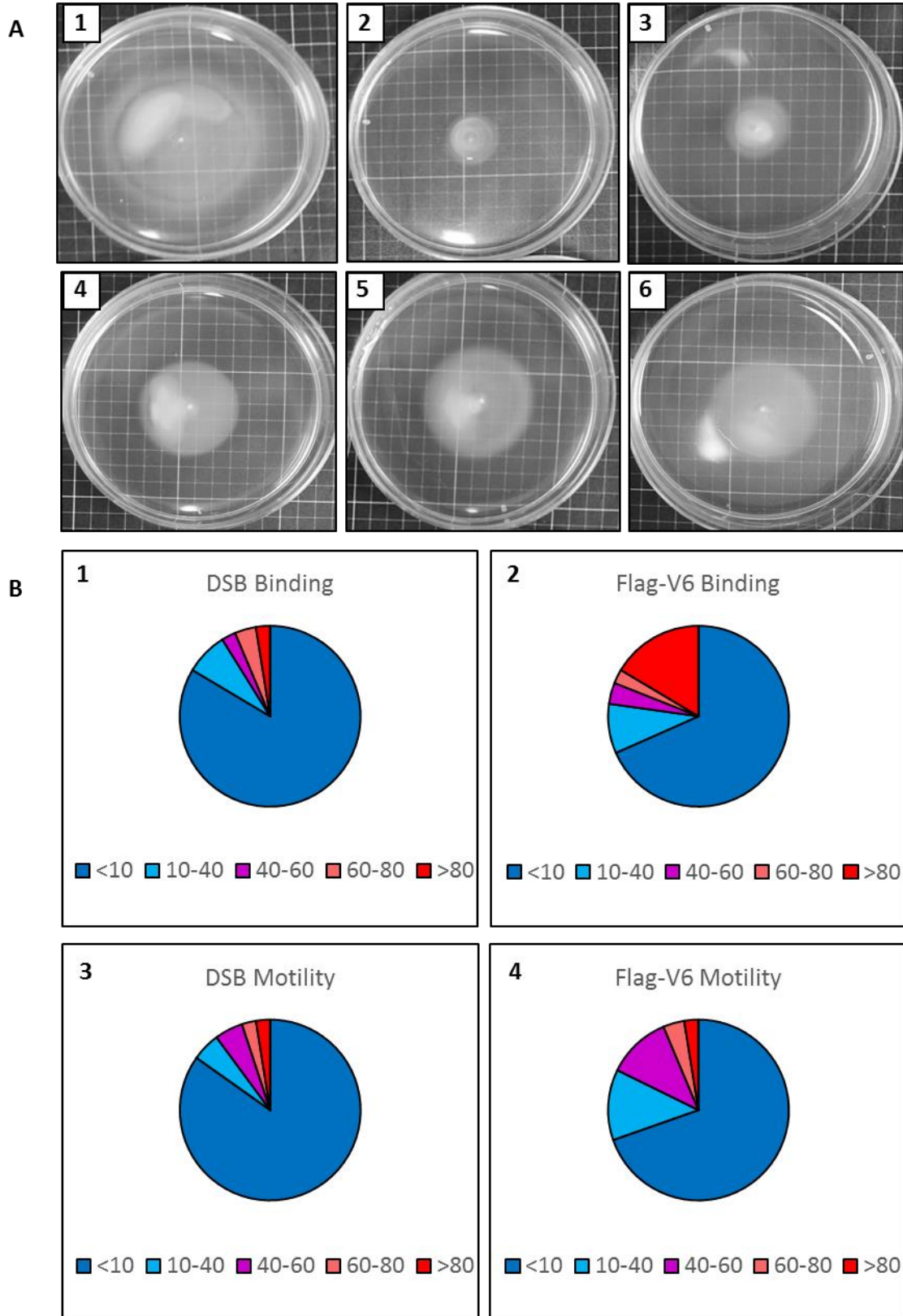
Figure 3.5: Nanobody binding and functionality. Nanobodies were assayed for binding using full cell ELISA as well as motility reduction assays. In all cases *C. jejuni* 81-176, the positive

control, fell within the red slice of the chart amongst the greatest reduction in motility and the greatest number of nanobodies bound. **A.)** Motility assay of *C. jejuni* through semi-solid agar after exposure to nanobodies. **1-3.)** Strain bound by nanobody resulting in reduced motility. **1.)** negative control. **2.)** Bound by 0.5mgmL^{-1} *DSB*. **3.)** Bound by 0.5mg mL^{-1} *V6*. **4-6.)** Strain unaffected by either nanobody. **4.)** negative control. **5.)** Bound by 0.5mg mL^{-1} *DSB*. **6.)** Bound by 0.5mg mL^{-1} *V6*. **B.)** Pie charts representing percent of maximal binding by **1.)** *DSB*. Or **2.)** *V6* as well as reduction in motility by **3.)** *DSB*. Or **4.)** *V6*.

strains (6.3%) by *V6* (LH 07:MS1180b, 10:MS2577, CHRB_1537, CHSK-I_158, and Wa10_732).

3.4. Discussion

In the pursuit of developing tools for engineering in *B. theta*, a series of vectors were constructed containing *V6* and *DSB* nanobodies fused to *Bacteroides* spp. signal peptides for expression in Gram-negative model bacterium *E. coli*, (Table 3.1); and a protocol was established to determine the subcellular localization of proteins after targeting by the signal peptides (Figure 3.1). Although it is not yet known whether the *B. theta* pathways behave analogously to those in *E. coli*, it was hypothesized that if signal peptides from *Bacteroides* spp. (Phylum *Bacteroidetes*) behave predictably in *E. coli* (Phylum *Proteobacteria*), they would be interchangeable between *Bacteroides* species as well. *E. coli* is amongst the most thoroughly studied organisms at the genetic level with a multitude of tools available for vector-based genetic manipulation and protein over-expression and it is a model Gram-negative enteric resident. Taken together, this makes *E. coli* a good system to screen the function of *Bacteroides* signal peptides. These results would still need to be confirmed in *B. theta*.

3.4.1 Cell Fractionation

By manipulating the concentration of sucrose and $MgCl_2$ in the hypertonic and hypotonic solutions, the optimized fractionation protocol developed here allows for the specific release of periplasmic protein without any discernible contamination by cytoplasmic protein. This is an improvement over previous protocols as they were focused on complete periplasmic release which resulted in minor contamination of the periplasmic fraction with cytoplasmic proteins. The

protocol presented here provides confidence in determining the level of protein trafficking to the periplasm; however, calculations of protein levels may be understated.

3.4.2 Targeted secretion of proteins

The subcellular localization of *V6 and DSB* proteins fused to the *E. coli* OmpA, the *Frag-SEC*, and one of seven *Bacteroides* signal peptides characterized in this study, were compared to untagged protein controls by Western blot and ELISA. As expected, the negative control constructs (i.e. lacking signal peptides) were found only in whole cell and cytoplasmic lysates but were not identified in the periplasm, on the outer membrane, or outside of the cell (Figure 3.3). Constructs containing the OmpA signal peptide were detected in the cytoplasm, which is the site of their synthesis, as well as in the periplasm but were not found outside of the cell. Interestingly, neither TAT-dependent signal peptide had any detectable effect on protein targeting in spite of being well conserved across several *Bacteroides* spp., they did, however, result in a lower concentration of protein being expressed relative to non-secreted controls. This result may reflect a phyla specific secretion mechanism as *B. theta* TAT-dependent translocation proteins TatA and TatC have 29.7 and 29.0% identity respectively with *E. coli* homologues and *B. theta* completely lacks a homolog to the *E. coli* TatB. It is also possible that inappropriate cleavage sites were chosen for these signal peptides, and that several key residues were omitted. In this regard extended signal peptides could also be investigated.

As predicted, constructs containing the *Frag-SEC* signal peptide were found in the cytoplasm, periplasm, and outside of the cell suggesting a two-step process for secretion (Figure 3.5). This is consistent with the T2SS following SEC-dependent translocation across the inner membrane. This is a promising finding; however, it will remain unclear to what extent *B. theta* would transport these constructs across the outer membrane until *B. theta* expression is investigated.

The other SEC-dependent signal peptides trafficked proteins to the cytoplasm, periplasm, and outside of the cell as well; however, different ratios of protein were detected across the fractions (Figure 3.3.2-5). The signal peptides that resulted in extracellular and periplasmic protein from highest to lowest levels are *Vulg-SEC* > *Theta-SEC2* > *Theta-SEC1* > *Frag-SEC* > and *Sala-SEC*. In all cases a higher level of periplasmic protein trafficking results in a higher concentration of extracellular protein. *Theta-SEC1, Sala-SEC,* and *Frag-SEC* resulted in a greater amount of protein persisting in the periplasm relative to extracellular secretion levels; *Vulg-SEC* and *Theta*-SEC2 targets a higher relative level of protein outside of the cell. These peptides, therefore, may represent tools for subcellular pathway engineering by enabling fine-tuned compartmentalization of protein cargo using specific SEC-dependent signal peptides.

3.4.3 Outer membrane targeting of proteins

A protocol for detecting outer membrane-bound protein was adopted from Sharp *et al* (2012) and validated in *E. coli*. Cells expressing proteins lacking signal peptides were compared to cells expressing proteins containing signal peptides predicted to target the outer membrane. When cells were immobilized on a PVDF membrane and labeled using anti-hexahistidine IgG-HRP conjugates, strains with *V6* and *DSB* fused to *Theta-OM* signal peptide were detected whereas untagged proteins were not. Significantly, when these cells were exposed to Proteinase K to digest all outer membrane protein the signal was ablated, suggesting that the protein was exposed on the cell surface. To ensure that this observation was not an artefact of total cellular protein being digested, cell lysates of Proteinase K digested cells were visualized with Western blot and the presence of intracellular protein was confirmed. Finally, to ensure that proteins were attached to the outer membrane and not secreted out of the cell, concentrated media from these expression trials were visualized with Western blot and no signal was detected.

3.4.4 Secreted Nanobody Assays

In order to investigate whether the secreted nanobodies from *E. coli* were functional, *V6* and *DSB* nanobodies secreted into the media were collected and compared against non-secreted controls. *V6* and *DSB* were raised against flagella of *C. jejuni* 81-176; however, *C. jejuni*, is a highly heterogeneous species (Young, 2007). As such, it was important to determine the binding range of *C. jejuni* strains that these nanobodies interact with to ensure that pathogenic strains will be affected. In order to evaluate the interaction spectrum for *V6* and *DSB*, 79 clinically relevant test strains previously isolated from poultry were selected. These strains were cultured and a whole cell ELISA was used to detect and quantify the amount of nanobody binding to each strain, while motility assays were used to determine the extent of motility reduction.

First, non-secreted nanobodies were used to determine the variety of strains that were bound by *V6* and *DSB*. Of the strains selected, 67 (84.8%) were not bound by *DSB* and 55 (69.6%) were not bound by *V6*. Of the strains that were bound by nanobodies, there was a gradient of motility reduction from 16.8% to 87.4%. This highlights both the variability of the strains and the necessity of a more complex approach in the future, likely using different nanobodies raised against a variety of highly heterogeneous strains of *C. jejuni*. As both of the nanobodies in this study were raised against *C. jejuni* 81–176 flagella, is may be a simple matter of raising new nanobodies against one or more of the unaffected strains and using them in conjunction with *DSB* and *V6*. It is currently unknown: however, how variable *C. jejuni* flagellar protein epitopes are, how extensive *C. jejuni*'s ability to develop new glycosylation patterns is, or how quickly different strains will adapt resistance. As a result, this approach may require the development of many nanobodies initially, as well as a continued effort as different strains evolve.

Of the two nanobodies tested, more copies of *V6* appeared to bind per cell than *DSB* as determined by absorbance in ELISA. In spite of this, *V6* binding did not seem to induce a greater reduction in motility, and in some cases, strains that bound relatively few nanobodies per cell resulted in the largest reduction in motility suggesting that the strength of each binding event or other flagellar factors are more important for motility reduction than the number of binding events. Although lower concentrations of nanobodies were not investigated, this evidence suggests that lower concentrations of nanobodies may have similar effects. This is significant as it has been shown that producing nanobodies at this concentration *in vivo* will likely be difficult (Hamady, 2010; 2011). Importantly, of the strains tested, only 13 (16.5%) were reduced in motility by 50% or more by one or both nanobodies. This further emphasizes the need for additional nanobody variety if this approach is to be effective *in vivo*.

To determine whether secretion of the nanobodies had an effect on their ability to bind *C. jejuni* flagella, nanobodies secreted with either *Vulg-SEC* or *Theta-SEC2* were concentrated from culture media. Each nanobody construct was used to bind the 15 strains exhibiting greater than 50% of maximum binding in the previous ELISA. The absorbances detected were not significantly different from those of the previous ELISA suggesting that the nanobodies were capable of being secreted and were properly folded. This is a success of this project particularly in the light of difficulties expressing appreciable quantities of protein in *B. theta*. If a significant quantity of protein was lost due to the inability to fold in the presence of the signal peptide or while exiting a transporter, an even greater net concentration of protein would need to be produced to compensate for this loss, which may not be practical.

Table. 3.1. Signal peptides used in this study. The LipoP is a prediction of either no signal peptide (NA) signal peptide type I (SPI), typically secreted, or signal peptide type II (SPII), typically lipoproteins. PSORTb is a prediction of the likelihood of the tag to target the protein to the cytoplasm (C), cytoplasmic membrane (CM), periplasm (P), outer membrane (OM), and extracellular (EC) out of 10.0.

Peptide	Species	Gene	Gene ID	Sequence	LipoP	PSORTb
Theta-OM	*B. theta*	α-amylase	gi:499421473	MNKHLHFLSLLWLSMLMAFM TACSDDKNITDPAPEPEPPVEG	SPII	CM 0.17 P 2.99 OM 6.58 EC 0.26
Theta-SEC1	*B. theta*	Pectate Lyase family 9	gi:29349578	MKDITKITYLLLGLMLSVPLAA	SPI	CM 0.01 P 0.20 OM 0.07 EC 9.72
Theta-SEC2	*B. theta*	Endonuclease	gi:29348967	MKYKFIQTLLLVLLPTLFAA**S***	SPI	EC 10.00
Theta-TAT1	*B. theta*	Peptidase U62	gi:496042545	MDRRNFLKTGGIALLGSLA	SPI	CM 3.24 P 6.49 OM 0.14 EC 0.14
Theta-TAT2	*B. theta*	Twin Arginine Translocation Pathway Signal Sequence	gi:511013107	MDRRDFLKTVAITGAAMTIQ	NA	C 8.96 CM 0.51 P 0.26 OM 0.01 EC 0.26
Sala-SEC	*Bacteroides salanitronis*	Glycoside Hydrolase family 28	gi:325299708	MKTLLISLLALAGLSLQAS	SPI	CM 0.02 P 0.20 EC 9.78
Vulg-SEC	*Bacteroides vulgatus*	Pectin Esterase family 8	gi:696379322	MRKVLGLLLLLSVVSAAWAQ	SPI	CM 0.02 P 0.20 EC 9.78
Frag-SEC	*Bacteroides fragilis*	Enterotoxin Fragilysin	gi:3046920	MKNVKLLLMLGTAALLAA	SPII	EC 10.00

OmpA	E. coli	Outer Membrane Protein	gi:585319598	MKKTAIAIAVALAGFATVAQA	SPI	OM 10.00

*The terminal serine was a cysteine in the native protein but was mutated to prevent lipidation.

Table. 3.2. *C. jejuni* strains used in this study. Strains sourced from Dr. Doug Inglis at the AAFC in Lethbridge, AB. Representative strains were selected using CGF data from the National Comparative Genomic Fingerprint Database curated by Eduardo Taboada at the Lethbridge National Microbiology Laboratory. The prevalence of each strain is given for cattle, chickens, humans, and environmental samples. Percent of maximal nanobody binding as determined by ELISA (ELISA Normalized) and the percent reduction in motility (MR) are given for comparison.

Strain Name	% Cattle	% Chicken	% Human	% Enviro	*DSB* ELISA	*DSB* MR	*V6* ELISA	*V6* MR
06:MS1594	0.00	68.71	24.10	1.80	0.00	0.00	0.21	0.00
06:MS1838	17.14	68.57	11.43	2.86	62.38	57.31	93.40	46.53
06:MS2087	5.41	37.84	13.51	24.32	4.27	9.81	0.00	3.66
06:MS2169	8.70	44.66	8.89	20.75	8.75	5.15	39.92	39.04
06:MS2284	53.23	20.97	16.13	4.84	76.19	69.73	94.14	54.19
07:MS459b	0.00	91.07	6.55	1.79	0.00	0.00	0.00	3.63
07:MS940	5.26	67.37	23.16	0.00	0.00	0.00	0.49	7.56
LH 07:MS1180	2.33	20.93	6.98	58.14	0.00	1.22	68.65	79.35
08:MS939	31.13	28.30	27.36	13.21	1.53	0.00	0.00	0.00
08:MS1246	0.00	91.18	8.82	0.00	0.00	3.89	0.00	0.22
08:MS1589	50.38	16.03	33.59	0.00	2.32	6.02	0.92	1.48
08:MS1659	39.97	15.35	34.74	8.94	0.00	3.98	4.36	0.00
08:MS2105	12.99	29.87	16.88	19.48	0.01	0.00	4.46	3.08
08:MS3402	8.11	59.46	29.73	0.00	0.00	2.00	10.06	26.58
09:MS444a	0.00	72.97	27.03	0.00	1.25	0.00	0.00	0.00
09:MS1956	0.00	45.37	7.32	12.20	13.07	85.71	96.09	51.03
09:MS1976	9.39	33.70	7.73	33.15	0.00	0.53	0.00	6.26
09:MS2715	3.75	76.25	8.75	10.00	2.17	4.23	0.55	0.00
10:MS2577	62.19	17.91	13.93	4.98	24.04	56.24	97.70	63.19
10:MS2949	8.03	32.85	6.93	28.47	2.00	1.61	0.07	1.80
10:MS3576	0.00	74.65	25.35	0.00	0.00	0.00	0.00	0.00
11:MS1854	3.75	76.25	8.75	10.00	2.04	2.59	0.00	0.00
11:MS2010	7.81	56.25	21.88	3.13	0.00	5.17	1.21	5.49
LH 11:MS3915	56.10	29.27	9.76	2.44	0.68	0.00	4.19	0.26
LH 12:MS1664	25.93	38.89	33.33	0.00	0.10	8.47	0.00	0.00
LH 12:MS1923	53.23	20.97	16.13	4.84	99.56	57.10	93.38	34.35
LH 12:MS3399	29.58	47.89	22.54	0.00	1.38	0.00	0.18	0.00
LH 13:MS1175	0.00	58.33	8.33	13.89	1.97	4.04	28.45	44.89
LH 13:MS3490	0.00	77.08	9.38	0.00	0.00	0.00	1.06	0.00
LH 14:MS2512	0.00	72.97	27.03	0.00	2.28	2.51	89.52	56.25
LH 14:MS1551	22.22	62.96	14.81	0.00	0.00	1.30	1.12	1.15
LH 14:MS1994	56.10	29.27	9.76	2.44	0.23	0.00	5.02	4.16
LH 14:MS2257	39.97	15.35	34.74	8.94	0.00	0.00	1.70	0.00

Strain Name	% Cattle	% Chicken	% Human	% Enviro	DSB ELISA	DSB MR	V6 ELISA	V6 MR
LH 14:MS2457	0.00	77.08	9.38	0.00	0.00	7.08	3.37	0.00
LH 14:MS2950	0.00	73.41	20.81	4.05	0.31	0.00	0.00	0.00
Ab_12891	14.10	42.31	25.64	6.41	2.17	0.00	0.00	3.01
Ab_13522	17.14	68.57	11.43	2.86	57.27	23.31	96.59	56.61
Ab_15547	8.03	32.85	6.93	28.47	0.25	8.98	0.00	0.00
AMR_1738	29.58	47.89	22.54	0.00	1.10	0.00	0.00	0.00
AMR_3388	50.38	16.03	33.59	0.00	0.09	0.00	0.00	8.42
AMR_3806	5.41	37.84	13.51	24.32	7.28	0.00	0.00	0.00
AMR_3947	25.93	38.89	33.33	0.00	0.00	6.49	0.00	3.96
AMR_4172	8.70	44.66	8.89	20.75	0.00	7.61	5.23	3.42
AMR_4684	5.26	67.37	23.16	0.00	0.49	3.37	11.06	24.98
AMR_7432	39.97	15.35	34.74	8.94	0.00	0.00	3.31	4.54
AMR_7850	12.99	29.87	16.88	19.48	0.00	0.62	2.02	6.22
AMR_8257	62.19	17.91	13.93	4.98	25.26	72.82	94.65	45.64
AMR_8300	31.13	28.30	27.36	13.21	0.00	7.90	0.14	4.33
CHRA 43	0.00	73.41	20.81	4.05	42.21	35.22	0.00	0.00
CHRA 52	14.10	42.31	25.64	6.41	0.00	0.80	0.00	9.22
CHRA 181	0.00	79.41	20.59	0.00	0.16	0.23	36.43	16.82
CHRA 223	0.00	75.00	20.45	0.00	0.00	2.80	23.63	36.35
CHRA 243	7.78	50.90	16.77	7.19	12.66	0.00	0.00	0.00
CHRA 312	54.05	16.22	24.32	5.41	0.00	0.00	41.49	23.82
CHRB_1537	0.00	73.08	23.08	0.00	80.41	35.60	98.39	65.27
CHRB_2146	8.11	59.46	29.73	0.00	0.14	0.00	1.09	0.37
CHRB_2442	0.00	60.56	30.99	0.00	0.18	4.68	0.00	0.00
CHRB 2776	0.55	55.49	15.38	5.49	0.62	0.00	88.79	31.29
CHSK-I_4	9.39	33.70	7.73	33.15	1.65	4.00	3.48	0.00
CHSK-I_28	0.00	74.65	25.35	0.00	0.00	0.00	0.00	7.04
CHSK-I_40	7.81	56.25	21.88	3.13	0.00	0.00	0.42	0.00
CHSK-I_41	2.33	20.93	6.98	58.14	0.00	8.01	66.16	29.40
CHSK-I_148	0.00	68.71	24.10	1.80	0.00	0.00	0.00	0.92
CHSK-I_158	62.19	17.91	13.93	4.98	16.26	87.42	96.61	83.72
CHSK-I_165	0.00	60.56	30.99	0.00	0.00	0.00	34.77	52.61
CHSK-I_178	39.97	15.35	34.74	8.94	0.00	0.00	0.43	5.67
Cry_138	0.55	55.49	15.38	5.49	1.75	0.00	42.43	0.00
Cry_185	25.93	38.89	33.33	0.00	0.00	0.00	0.00	0.00
Cry_319	0.00	75.00	20.45	0.00	0.00	0.00	0.00	0.00
Cry_430	7.78	50.90	16.77	7.19	2.43	0.00	0.00	0.00
Cry_508	12.99	29.87	16.88	19.48	0.29	1.50	0.00	0.00
Cry_548	0.00	60.56	30.99	0.00	24.43	25.93	100.00	38.46
Wa08 4953	0.00	30.30	6.06	42.42	0.00	0.00	1.67	0.00

Strain Name	% Cattle	% Chicken	% Human	% Enviro	DSB ELISA	DSB MR	V6 ELISA	V6 MR
Wa10 181	2.33	20.93	6.98	58.14	0.00	0.00	0.00	7.61
Wa10_732	17.14	68.57	11.43	2.86	65.22	46.10	97.11	85.67
Wa11_62	31.13	28.30	27.36	13.21	0.00	0.00	6.74	0.00
Wa11_289	54.05	16.22	24.32	5.41	0.00	0.00	0.24	0.00
Wa11_693	8.70	44.66	8.89	20.75	6.72	4.64	6.49	0.00
Wa11_1600	39.97	15.35	34.74	8.94	0.00	1.00	49.50	47.73
81-176	NA	NA	NA	NA	94.12	88.20	96.43	81.54

Chapter 4. Conclusion

In this thesis, I have generated several novel molecular tools for bioengineering *B. theta*, which will facilitate the production and secretion of heterologous proteins. Major advances include the generation of constructs for genome integrated and plasmid-based expression of nanobody genes; identification of four novel *Bacteroides* SEC-dependent signal peptides and one outer membrane attachment signal peptide for targeted secretion of proteins in Gram-negative cells; development of a novel cell fractionation protocol for validation of protein trafficking in Gram-negative cells; and the identification of 15 putative reference genes in *B. theta*, which will expand the resources in the field beyond 16s rRNA for RT-qPCR experiments. In addition, I have expressed and purified two nanobodies and characterized their specificity against a library of *C. jejuni* strains to determine their spectrum of binding and effect on motility reduction.

4.1. Improved Induction and Protein Expression

Attempts were made to induce expression of *B. theta* constructs both genomically and plasmid based with two different promoter/terminator pairs, and to detect mRNA with RT-qPCR. Although transcripts were detected, induction was not achieved and protein was not detected with Western blots. There are several potential strategies that could be undertaken in an attempt to overcome the lack of nanobody production in *B. theta*, including investigating additional genetic loci to detect PUL75 deletion effects; additional promoter-RBS combinations could be explored, including alternative carbohydrate induction and strong constitutive promoters ; and mimicking alternative hyper-expression systems, such as the insertion of the T7 RNA polymerase (T7RNAP) into the genome of *B. theta*, coupled with a plasmid containing a nanobody gene flanked by a T7 promoter and terminator.

4.2. Expansion of Nanobody Efficacy

The issue of nanobodies binding a relatively small subset of strains of *C. jejuni* was anticipated due to the heterogeneity of this species. There are a number of potential approaches that may address the problem, including raising antibodies against the flagella of resistant strains; using a combination of nanobodies that bind alternative cellular structures; using nanobody fusions to deliver other compounds such as antigenic proteins or toxic enzymes, which could increase the immunogenicity of *C. jejuni* or directly kill the pathogen; and generating nanobodies that interfere with *C. jejuni* metabolism or block selective transport.

4.3. The Problem of Public Perception

Perhaps the greatest challenge faced by this project is not the inherent difficulty of engineering a strain to safely deliver protein cargo to the chicken intestine, but the public perception of the use of these technologies in general. Because this strain is intended to live in the intestine of chickens, it will, on occasion, be transferred to humans that consume contaminated meat. As a result of the success of *B. theta* in the human gut, it is probable that an engineered strain, in the absence of a potent and selective kill-switch, would survive and even thrive is this environment.

Possible contamination of the human gut is problematic in the current atmosphere of public distrust of genetically modified organisms (GMO). In recent surveys in the United States (Hallman, 2013; CRNRC, 2015), 50.5% of respondents thought that consuming GMOs was dangerous, 65.1% would be upset to learn that they had unknowingly consumed GMOs at a restaurant, and 75% considered avoiding GMOs important. Superficially, these statistics appear to be a death-knell for this and similar projects. This is not necessarily the case, however, and may reflect a lack of understanding on behalf of the general public as GMOs are now a

foundational component of modern agriculture and has permeated many different sectors of food production. This helps to highlight a fundamental lack of understanding of these technologies by the public. The emergence of new GMOs may be facilitated by accurate education programs for public and additional scientific research expanding on the stability and safety of GMOs in agriculture.

4.4. Future Applications

If this technology is implemented, it will likely not be adopted by all poultry producers. This is problematic as it has been shown that the rate of contamination increases after processing in abattoirs and that nearly three quarters of carcasses of *Campylobacter*-free flocks are contaminated with *C. jejuni* if they are slaughtered after a *Campylobacter*-positive flock (Reich, 2008). Because of this, a single farm that does not adopt the technology could result in the contamination of the majority of poultry slaughtered in the abattoir that day. In addition, it has been shown that even after treatment with chloride containing compounds and quaternary-ammonium compounds, *C. jejuni* is still culturable at greater than 90% (Ellerbroek, 2010) suggesting that a single contaminated flock could affect the abattoir for days rendering the entire effort useless. If these strains were used in the human gut instead, the consumption of contaminated poultry would be less of an issue as pathogenic strains are not capable of colonizing the human gut without a functional flagellum.

If this technology is accepted for use in the human gut, it could be adapted for use against a variety of human pathogens. Nanobodies can be raised against any pathogen whether it is bacterial, viral, or eukaryotic, regardless of its preferred habitat. Heterologous protein expression has been achieved in the mouth and nasal mucosa with *Streptococcus gordonii* (Medaglini, 1995; Bonn, 1998) as well as the vagina with *S. gordonii* and *Lactococcus spp.* (Medaglini, 1997;

Mercenier, 2000). The system could be used as prophylaxis *for Streptococcus pneumoniae*, *Haemophilus influenzae*, and others in the nasal cavity or *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, and others in the vaginal canal.

The future of this and similar technologies is bound only by imagination and the discovery of new genetic tools. As the systems continue to progress and become increasingly complex, these organisms may serve as a secondary, pseudo-immune system both in humans and in the animals we consume. The goal of this project was to setup the foundation for the expression and secretion of heterologous proteins in *B. theta* which will be used as a vehicle for delivery of proteins into the gut. Although the project was not successful on all grounds, a great deal of progress was made and numerous potential solutions to its shortcomings have been outlined. Future work will focus heavily on induction and expression of genes in *B. theta*. Once detectable levels of protein have been produced it will be easier to optimize the system to maximize output, as well as ensure that the signal peptides function equivalently in *B. theta* and *E. coli*. From that point, the system could be used to deliver any number of nanobodies as described above, as well as any other proteins that may prove useful.

References Cited

Acheson, D., *et al*. (2001). *Campylobacter jejuni* infections: update on emerging issues and trends. Clinical infectious diseases, 32(8), 1201-1206.

Ahmad, Z., *et al*. (2012). scFv antibody: principles and clinical application. Clinical and Developmental Immunology, 2012.

Akimaru, J., *et al*. (1991). Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from *Escherichia coli*. Proceedings of the National Academy of Sciences, 88(15), 6545-6549.

Atherly, T., *et al*. (2014). *Bacteroides* isolated from four mammalian hosts lack host-specific 16S rRNA gene phylogeny and carbon and nitrogen utilization patterns. Microbiology Open, 3(2), 225-238.

Atsumi, S., *et al*. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature, 451(7174), 86-89.

Baker, M., *et al*. (2012). Declining Guillain-Barré syndrome after campylobacteriosis control, New Zealand, 1988–2010. Emerging infectious diseases, 18(2), 226.

Bandoh, K., *et al*. (1992). Conjugal transfer of imipenem resistance in *Bacteroides fragilis*. The Journal of antibiotics, 45(4), 542-547.

Bayley, D., *et al*. (2000). Analysis of cepA and other *Bacteroides fragilis* genes reveals a unique promoter structure. FEMS microbiology letters, 193(1), 149-154.

Bayliss, C, *et al*. (2012). Phase variable genes of *Campylobacter jejuni* exhibit high mutation rates and specific mutational patterns but mutability is not the major determinant of population structure during host colonization. Nucleic acids research, gks246.

Beckwith, J. (2013). The Sec-dependent pathway. Research in microbiology, 164(6), 497-504.

Berks, B. (2015). The twin-arginine protein translocation pathway. Annual review of biochemistry, 84, 843-864.

Bervoets, L., *et al*. (2013). Differences in gut microbiota composition between obese and lean children: a cross-sectional study. Gut pathogens, 5(1), 1.

Benjdia, A., *et al*. (2011). Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont, *Bacteroides thetaiotaomicron*. Journal of Biological Chemistry, 286(29), 25973-25982.

Bergman, E. (1990). Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. Physiol. Rev, 70(2), 567-590.

Bernstein, H., *et al*. (1989). Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. Nature 340.6233 : 482-486.

Bernstein, H., *et al*. (1993). Functional substitution of the signal recognition particle 54-kDa subunit by its *Escherichia coli* homolog. Proceedings of the National Academy of Sciences 90.11 : 5229-5233.

Bertani, G. (1951). Studies on Lysogenesis I.: The Mode of Phage Liberation by Lysogenic *Escherichia coli*. Journal of bacteriology, 62(3), 293.

Bevan, M., *et al*. (1983). A chimeric antibiotic resistance gene as a selectable marker for plant cell transformation. Nature 304, 184 – 187.

Bhatt, V., *et al*. (2013). Molecular analysis of the bacterial microbiome in the forestomach fluid from the dromedary camel (*Camelus dromedarius*). Molecular biology reports, 40(4), 3363-3371.

Bjursell, M., *et al*. (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. Journal of biological chemistry.

Bloom, S., *et al*. (2011). Commensal *Bacteroides* species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. Cell host & microbe, 9(5), 390-403.

Bonn, D. (1998). Commensal bacteria armed against biological warfare. Molecular medicine today, 4(9), 369-369.

Braat, H., *et al*. (2006). A phase I trial with transgenic bacteria expressing interleukin-10 in Crohn's disease. Clinical gastroenterology and hepatology, 4(6), 754-759.

Braun, V., *et al*. (2005). Transmembrane transcriptional control (surface signaling) of the *Escherichia coli* Fec type. FEMS microbiology reviews, 29(4), 673-684.

Bremell, T., *et al*. (1991). Rheumatic symptoms following an outbreak of *Campylobacter* enteritis: a five year follow up. Annals of the rheumatic diseases, 50(12), 934-938.

Bresee, J., *et al*. (2012). The etiology of severe acute gastroenteritis among adults visiting emergency departments in the United States. Journal of Infectious Diseases, 205(9), 1374-1381.

Brezski, R. J., *et al*. (2016). Immunoglobulin isotype knowledge and application to Fc engineering. Current opinion in immunology, 40, 62-69.

Bruford, M., *et al*. (2003). DNA markers reveal the complexity of livestock domestication. Nature Reviews Genetics, 4(11), 900-910.

Buelow, D., *et al*. (2011). *Campylobacter jejuni* survival within human epithelial cells is enhanced by the secreted protein CiaI. Molecular microbiology, 80(5), 1296-1312.

Buffie, C., *et al*. (2013). Microbiota-mediated colonization resistance against intestinal pathogens. Nature Reviews Immunology, 13(11), 790-801.

Bult, C., *et al*. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science, 273(5278), 1058-1073.

Burt, S., *et al*. (1977). Transfection of the anaerobe *Bacteroides thetaiotaomicron* with phage DNA. Journal of General Microbiology, 103(1), 181-187.

Carbohydrate-Active enZYmes Database. 25-01-2016a. Retrieved from: http://www.cazy.org/b134.html

Carbohydrate-Active enZYmes Database. 25-01-2016b. Retrieved from: http://www.cazy.org/e358.html

Casey, J., *et al*. (2000). Green fluorescent antibodies: novel *in vitro* tools. Protein engineering, 13(6), 445-452.

Chen, Y., *et al*. (2004). A modified osmotic shock for periplasmic release of a recombinant creatinase from *Escherichia coli.* Biochemical engineering journal, 19(3), 211-215.

Cheung, C., *et al*. (2009). The effect of ionic (NaCl) and non-ionic (sucrose) osmotic stress on the expression of β-galactosidase in wild type E. coli BW25993 and in the isogenic BW25993ΔlacI mutant. J. Exp. Microbiol. Immunol, 13, 1-6.

Cohen, S., *et al*. (1973). Construction of biologically functional bacterial plasmids *in vitro*. Proceedings of the National Academy of Sciences, 70(11), 3240-3244.

Connerton, P., *et al*. (2011). *Campylobacter* bacteriophages and bacteriophage therapy. Journal of applied microbiology, 111(2), 255-265.

Consumer Reports National Research Center (CRNRC). (2015). Natural Food Labels Survey. 2015 Nationally-Representative Phone Survey.

Converse, R., *et al*. (2009). Rapid QPCR-based assay for fecal *Bacteroides* spp. as a tool for assessing fecal contamination in recreational waters. Water research, 43(19), 4828-4837.

Coppieters, K., *et al*. (2006). Formatted anti–tumor necrosis factor α VHH proteins derived from camelids show superior potency and targeting to inflamed joints in a murine model of collagen-induced arthritis. Arthritis & Rheumatism, 54(6), 1856-1866.

Cuskin, F., *et al*. (2015). Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. Nature, 517(7533), 165-169.

D'Elia, J., *et al*. (1996a). Effect of regulatory protein levels on utilization of starch by *Bacteroides thetaiotaomicron*. Journal of bacteriology, 178(24), 7180-7186.

D'Elia, J., *et al*. (1996b). Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. Journal of bacteriology, 178(24), 7173-7179.

De Filippo, C., *et al*. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proceedings of the National Academy of Sciences, 107(33), 14691-14696.

De Genst, E., *et al*. (2006). Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proceedings of the National Academy of Sciences of the United States of America, 103(12), 4586-4591.

Desroche, N., *et al*. (2005). Determination of an internal control to apply reverse transcription quantitative PCR to study stress response in the lactic acid bacterium *Oenococcus oeni*. Journal of Microbiological Methods, 60(3), 325-333.

Desvaux, M., *et al*. (2009). Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. Trends in microbiology, 17(4), 139-145.

Doorduyn, Y., *et al*. (2010). Risk factors for indigenous *Campylobacter jejuni* and *Campylobacter coli* infections in The Netherlands: a case-control study. Epidemiology and infection, 138(10), 1391-1404.

Duke, S. (2015). Perspectives on transgenic, herbicide-resistant crops in the United States almost 20 years after introduction. Pest management science, 71(5), 652-657.

Duncan, S., *et al*. (2008). Human colonic microbiota associated with diet, obesity and weight loss. International journal of obesity, 32(11), 1720-1724.

Elder, B., *et al*. (1982). Whole-bacterial cell enzyme-linked immunosorbent assay for *Streptococcus sanguis* fimbrial antigens. Journal of clinical microbiology 16.1 : 141-144.

Ellerbroek, L., *et al*. (2010). *Campylobacter spp*. in Broiler Flocks at Farm Level and the Potential for Cross-Contamination During Slaughter. Zoonoses and public health, 57(7-8), e81-e88.

Epps, S., *et al*. (2013). Foodborne *Campylobacter*: infections, metabolism, pathogenesis and reservoirs. International journal of environmental research and public health, 10(12), 6292-6304.

Faith, J., *et al*. (2013). The long-term stability of the human gut microbiota. Science, 341(6141), 1237439.

Farrar, M., *et al*. (2005). Engineering of the gut commensal bacterium *Bacteroides ovatus* to produce and secrete biologically active murine interleukin-2 in response to xylan. Journal of applied microbiology, 98(5), 1191-1197.

Ferrero, R., *et al*. (1988). Motility of *Campylobacter jejuni* in a viscous environment: comparison with conventional rod-shaped bacteria. Journal of general microbiology, 134(1), 53-59.

Freitas, M., *et al*. (2003). Host-pathogens cross-talk. Indigenous bacteria and probiotics also play the game. Biology of the Cell, 95(8), 503-506.

Freitas, M., *et al*. (2005). Indigenous microbes and their soluble factors differentially modulate intestinal glycosylation steps *in vivo*. Histochemistry and cell biology, 124(5), 423-433.

French, C., *et al*. (1955). Disintegration of bacteria and small particles by high-pressure extrusion. Methods in enzymology 1: 64-67.

French, C., *et al*. (1996). Development of a simple method for the recovery of recombinant proteins from the *Escherichia coli* periplasm. Enzyme and Microbial Technology, 19(5), 332-338.

Gardner, R., *et al*. (1996). Use of a modified *Bacteroides-Prevotella* shuttle vector to transfer a reconstructed beta-1, 4-D-endoglucanase gene into *Bacteroides uniformis* and *Prevotella ruminicola* B (1) 4. Applied and environmental microbiology, 62(1), 196-202.

Gibbs, W. (2005). Nanobodies. Scientific American, 293(2), 78-83.

Gillespie, I., *et al*. (2002). A case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: a tool for generating hypotheses. Emerging infectious diseases, 8(9), 937-42.

Goffeau, A., *et al*. (1996). Life with 6000 genes. Science, 274(5287), 546-567.

Gordon, J., *et al*. (1981). Integration and stable germ line transmission of genes injected into mouse pronuclei. Science, 214(4526), 1244-1246.

Grant, C., *et al*. (1993). Role of flagella in adherence, internalization, and translocation of *Campylobacter jejuni* in nonpolarized and polarized epithelial cell cultures. Infection and immunity, 61(5), 1764-1771.

Guerry, P., *et al*. (1991). Role of two flagellin genes in *Campylobacter* motility. Journal of bacteriology 173.15 : 4757-4764.

Gumpel, J., *et al*. (1981). Reactive arthritis associated with *Campylobacter* enteritis. Annals of the rheumatic diseases, 40(1), 64-65.

Gupta, S., *et al*. (2013). Genetically programmable pathogen sense and destroy. ACS synthetic biology, 2(12), 715-723.

Hallman, W., *et al*. (2013). Public perceptions of labeling genetically modified foods. In Working Paper 2013-01. Rutgers University.

Hamady, Z., *et al*. (2010). Xylan-regulated delivery of human keratinocyte growth factor-2 to the inflamed colon by the human anaerobic commensal bacterium *Bacteroides ovatus*. Gut, 59(4), 461-469.

Hamady, Z., *et al*. (2011). Treatment of colitis with a commensal gut bacterium engineered to secrete human tgf-β1 under the control of dietary xylan 1. Inflammatory bowel diseases, 17(9), 1925-1935.

Hamers-Casterman, C., *et al*. (1993). Naturally occurring antibodies devoid of light chains. Nature, 363(6428), 446-448.

Hanauer, S. (2006). Inflammatory bowel disease: epidemiology, pathogenesis, and therapeutic opportunities. Inflammatory bowel diseases, 12(5), S3-S9.

Harvey, E., *et al*. (1929). The destruction of luminous bacteria by high frequency sound waves. Journal of Bacteriology 17.5 : 373.

Hehemann, J., *et al*. (2012). Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. Proceedings of the National Academy of Sciences, 109(48), 19786-19791.

Henderson, I., *et al*. (2004). Type V protein secretion pathway: the autotransporter story. Microbiology and molecular biology reviews, 68(4), 692-744.

Henson, S., *et al*. (2008). Estimation of the costs of acute gastrointestinal illness in British Columbia, Canada. International journal of food microbiology, 127(1), 43-52.

Hesami, S., *et al*. (2011). Identification of cold-temperature-regulated genes in *Flavobacterium psychrophilum*. Applied and environmental microbiology, 77(5), 1593-1600.

Hommais, F., *et al*. (2011). lpxC and yafS are the most suitable internal controls to normalize real time RT-qPCR expression in the phytopathogenic bacteria *Dickeya dadantii*. PloS one, 6(5), e20269.

Hooper, L., *et al*. (2001). Molecular analysis of commensal host-microbial relationships in the intestine. Science, 291(5505), 881-884.

Hooper, L., *et al*. (2003). Angiogenins: a new class of microbicidal proteins involved in innate immunity. Nature immunology, 4(3), 269-273.

Houliston, R., *et al*. (2011). Lipooligosaccharide of *Campylobacter jejuni* Similarity With Multiple Types Of Mammalian Glycans Beyond Gangliosides. Journal of Biological Chemistry, 286(14), 12361-12370.

Hudson, P., *et al*. (1999). High avidity scFv multimers; diabodies and triabodies. Journal of immunological methods, 231(1), 177-189.

Hughes, R., *et al*. (2005). Guillain-barre syndrome. The Lancet, 366(9497), 1653-1666

Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. Nature, 486(7402), 207-214.

Huo, Y., *et al*. (2011). Conversion of proteins into biofuels by engineering nitrogen flux. Nature biotechnology, 29(4), 346-351.

Hussack, G., *et al*. (2014). Protease-resistant single-domain antibodies inhibit *Campylobacter jejuni* motility. Protein Engineering Design and Selection, gzu011.

Huston, J., *et al*. (1988). Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. Proceedings of the National Academy of Sciences, 85(16), 5879-5883.

Hwang, I., *et al*. (2013). Reprogramming microbes to be pathogen-seeking killers. ACS synthetic biology, 3(4), 228-237.

Ichimura, M., *et al*. (2010). Efficient electrotransformation of *Bacteroides fragilis*. Applied and environmental microbiology, 76(10), 3325-3332.

Itakura, K., *et al*. (1977). Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science, 198(4321), 1056-1063.

Ivankov, D., *et al*. (2013). How many signal peptides are there in bacteria?" Environmental microbiology 15.4 : 983-990.

Jaenisch, R., *et al*. (1974). Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. Proceedings of the national academy of sciences, 71(4), 1250-1254.

Jin, S., *et al*. (2001). JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. Molecular microbiology, 39(5), 1225-1236.

Johnson, T., *et al*. (2006). Type II secretion: from structure to function. FEMS microbiology letters, 255(2), 175-186.

Jones, M., *et al*. (2004). Adaptation of *Campylobacter jejuni* NCTC11168 to high-level colonization of the avian gastrointestinal tract. Infection and immunity, 72(7), 3769-3776.

Juncker, A., *et al*. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria." Protein Science 12.8 : 1652-1662.

Karlyshev, A., *et al*. (2000). Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. Molecular microbiology, 35(3), 529-541.

Karmali, M., *et al*. (1986). Evaluation of a blood-free, charcoal-based, selective medium for the isolation of *Campylobacter* organisms from feces. Journal of clinical microbiology, 23(3), 456-459.

Kazmierczak, B., *et al*. (2013). Spatial and numerical regulation of flagellar biosynthesis in polarly flagellated bacteria. Molecular microbiology, 88(4), 655-663.

Khan, I., *et al*. (2014). A national investigation of the prevalence and diversity of thermophilic *Campylobacter* species in agricultural watersheds in Canada. Water research, 61, 243-252.

Kibbe, W. (2007). OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res. 35(webserver issue): May 25.

Koga, T., *et al*. (1990). Surface hydrophobicity, adherence, and aggregation of cell surface protein antigen mutants of *Streptococcus mutans* serotype c. Infection and immunity, 58(2), 289-296.

Konkel, M., *et al*. (2004). Secretion of virulence proteins from *Campylobacter jejuni* is dependent on a functional flagellar export apparatus. Journal of bacteriology, 186(11), 3296-3303.

Konkel, M., *et al*. (2005). Identification of a fibronectin-binding domain within the *Campylobacter jejuni* CadF protein. Molecular microbiology, 57(4), 1022-1035.

Kovacs-Simon, A., *et al*. (2011). Lipoproteins of bacterial pathogens. Infection and immunity, 79(2), 548-561.

Kowarik, M., *et al*. (2006). Definition of the bacterial N-glycosylation site consensus sequence. The EMBO Journal, 25(9), 1957-1966.

Kozera, B., *et al*. (2013). Reference genes in real-time PCR. Journal of applied genetics, 54(4), 391-406.

Kramer, M., *et al*. (1994). Commercialization of a tomato with an antisense polygalacturonase gene: The FLAVR SAVR™ tomato story. Euphytica, 79(3), 293-297.

Lauwereys, M., *et al*. (1998). Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. The EMBO journal, 17(13), 3512-3520.

Lev-Yadun, S., *et al*. (2000). The cradle of agriculture. Science, 288(5471), 1602-1603.

Li, H., *et al*. (2015). The outer mucus layer hosts a distinct intestinal microbial niche. Nature communications, 6.

Li, Z., *et al*. (2004). The generation of antibody diversity through somatic hypermutation and class switch recombination. Genes & development, 18(1), 1-11.

Linton, D., *et al*. (2000). Phase variation of a β-1, 3 galactosyltransferase involved in generation of the ganglioside GM1-like lipo-oligosaccharide of *Campylobacter jejuni*. Molecular microbiology, 37(3), 501-514.

Liu, Y., *et al*. (2006). Multiple maternal origins of chickens: out of the Asian jungles. Molecular phylogenetics and evolution, 38(1), 12-19.

Lobo, L., *et al*. (2011). Flavin mononucleotide (FMN)-based fluorescent protein (FbFP) as reporter for gene expression in the anaerobe *Bacteroides fragilis*. FEMS microbiology letters, 317(1), 67-74.

López-Boado, Y., *et al*. (2000). Bacterial exposure induces and activates matrilysin in mucosal epithelial cells. The Journal of cell biology, 148(6), 1305-1315.

Luangtongkum, T., *et al*. (2009). Antibiotic resistance in *Campylobacter*: emergence, transmission and persistence.

Lynch, J., *et al*. (2012). Prioritization of a plant polysaccharide over a mucus carbohydrate is enforced by a *Bacteroides* hybrid two-component system. Molecular microbiology, 85(3), 478-491.

Majowicz, S., *et al*. (2006). Burden and cost of gastroenteritis in a Canadian community. Journal of Food Protection®, 69(3), 651-659.

Marcobal, A., *et al*. (2011). *Bacteroides* in the infant gut consume milk oligosaccharides via mucus-utilization pathways. Cell host & microbe, 10(5), 507-514.

Martens, E., *et al*. (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell host & microbe, 4(5), 447-457.

Martens, E., *et al*. (2009). Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. Journal of Biological Chemistry, 284(37), 24673-24677.

Martens, E., *et al*. (2011). Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS-Biology, 9(12), 2492.

Martin, V., *et al*. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nature biotechnology, 21(7), 796-802.

Mastropaolo, M., *et al*. (2009). Comparison of *Bacteroides thetaiotaomicron* and *Escherichia coli* 16S rRNA gene expression signals. Microbiology, 155(8), 2683-2693.

McCracken, V., *et al*. (2001). The gastrointestinal ecosystem: a precarious alliance among epithelium, immunity and microbiota. Cellular microbiology, 3(1), 1-11.

Medaglini, D., *et al*. (1995). Mucosal and systemic immune responses to a recombinant protein expressed on the surface of the oral commensal bacterium *Streptococcus gordonii* after oral colonization. Proceedings of the National Academy of Sciences, 92(15), 6868-6872.

Medaglini, D., *et al*. (1997). Commensal bacteria as vectors for mucosal vaccines against sexually transmitted diseases: vaginal colonization with recombinant *Streptococci* induces local and systemic antibodies in mice. Vaccine, 15(12), 1330-1337.

Mercenier, A., *et al*. (2000). Lactic acid bacteria as live vaccines. Current issues in molecular biology, 2, 17-26.

Mimee, M., *et al*. (2015). Programming a human commensal bacterium, *Bacteroides thetaiotaomicron*, to sense and respond to stimuli in the murine gut microbiota. Cell Systems, 1(1), 62-71.

Morris, D., *et al*. (2012). Reactive arthritis: developments and challenges in diagnosis and treatment. Current rheumatology reports, 14(5), 390-394.

Mueller, J., *et al*. (1941). A Protein-Free Medium for Primary Isolation of the *Gonoeoceus* and *Meningococcus*. Proceedings of the Society for Experimental Biology and Medicine, 48(1), 330-33.

Muyldermans, S. (2013). Nanobodies: natural single-domain antibodies. Annual review of biochemistry, 82, 775-797.

Nancy, Y., *et al*. (2013). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes." Bioinformatics 26.13 : 1608-1615.

Natale, P., *et al*. (2008). Sec-and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. Biochimica et Biophysica Acta (BBA)-Biomembranes, 1778(9), 1735-1756.

Nava, G., *et al*. (2011). Spatial organization of intestinal microbiota in the mouse ascending colon. The ISME journal, 5(4), 627-638.

Neu, H., *et al*. (1965). The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. Journal of Biological Chemistry 240.9 (1965): 3685-3692.

Nossal, N., *et al*. (1966). The release of enzymes by osmotic shock from *Escherichia coli* in exponential phase." Journal of Biological Chemistry 241.13 : 3055-3062.

Nothaft, H., *et al*. (2016). Engineering the *Campylobacter jejuni* N-glycan to create an effective chicken vaccine. Scientific reports, 6.

Pancer, Z., *et al*. (2008). The oldest antibodies newly discovered. Nature biotechnology, 26(4), 402-403.

Parker, A., *et al*. (2012). Development of an IPTG inducible expression vector adapted for *Bacteroides fragilis*. Plasmid, 68(2), 86-92.

Parkhill, J., *et al*. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature, 403(6770), 665-668.

Peschen, D., *et al*. (2004). Fusion proteins comprising a *Fusarium*-specific antibody linked to antifungal peptides protect plants against a fungal pathogen. Nature biotechnology, 22(6), 732-738.

Petersen, T., *et al*. (2011) "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nature methods 8.10 : 785-786.

Pfeffer, S., *et al*. (1987). Biosynthetic protein transport and sorting by the endoplasmic reticulum and Golgi. Annual review of biochemistry, 56(1), 829-852.

PHAC Notifiable Diseases On-Line. 25-01-2016. Retrieved from: http://dsol-smed.phac-aspc.gc.ca/dsol-smed/ndis/charts.php?c=pl

Prieto, C., *et al*. (2003) Whole-bacterial cell enzyme-linked immunosorbent assay for cell-bound *Moraxella bovis* pili. Veterinary microbiology 91.2 : 157-168.

Outram, A., *et al*. (2009). The earliest horse harnessing and milking. Science, 323(5919), 1332-1335.

Rahman, O., *et al*. (2008). Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. World Journal of Microbiology and Biotechnology, 24(11), 2377-2382.

Ramsland, P., *et al*. (2002). Crystal structures of human antibodies: a detailed and unfinished tapestry of immunoglobulin gene products. Journal of Molecular Recognition, 15(5), 248-259.

Ratnakumar, A. (2013). Detecting Signatures of Selection within the Dog Genome.

Reich, F., *et al*. (2008). The effects of *Campylobacter* numbers in caeca on the contamination of broiler carcasses with *Campylobacter*. International journal of food microbiology, 127(1), 116-120.

Resta–Lenert, S., *et al*. (2006). Probiotics and commensals reverse TNF-α–and IFN-γ–induced dysfunction in human intestinal epithelial cells. Gastroenterology, 130(3), 731-746.

Riazi, A., *et al*. (2013). Pentavalent single-domain antibodies reduce *Campylobacter jejuni* motility and colonization in chickens. PloS one, 8(12), e83928.

Rocha, D., *et al*. (2015). Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. Antonie van Leeuwenhoek, 108(3), 685-693.

Rogers, T., *et al*. (2013). Dynamic responses of *Bacteroides thetaiotaomicron* during growth on glycan mixtures. Molecular microbiology, 88(5), 876-890.

Rogowski, A., *et al*. (2015). Glycan complexity dictates microbial resource allocation in the large intestine. Nature communications, 6.

Ross, C., *et al*. (2006). *Campylobacter jejuni* inactivation in New Zealand soils. Journal of applied microbiology, 101(5), 1188-1197.

Saeidi, N., *et al*. (2011). Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, a human pathogen. Molecular systems biology, 7(1), 521.

Salacha, R., *et al*. (2010). The *Pseudomonas aeruginosa* patatin-like protein PlpD is the archetype of a novel Type V secretion system. Environmental microbiology, 12(6), 1498-1512.

Salyers, A. (1984). *Bacteroides* of the human lower intestinal tract. Annual Reviews in Microbiology, 38(1), 293-313.

Savli, H., *et al*. (2003). Expression stability of six housekeeping genes: a proposal for resistance gene quantification studies of *Pseudomonas aeruginosa* by real-time quantitative RT-PCR. Journal of Medical Microbiology, 52(5), 403-408.

Schneider, C., *et al*. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat methods 9.7 (2012): 671-675.

Sharp, Julia A., *et al*. (2012). *Staphylococcus aureus* surface protein SdrE binds complement regulator factor H as an immune evasion tactic. PloS one 7.5 (2012): e38407.

Shoemaker, N., *et al*. (1986). Regions in *Bacteroides* plasmids pBFTM10 and pB8-51 that allow *Escherichia coli-Bacteroides* shuttle vectors to be mobilized by IncP plasmids and by a conjugative *Bacteroides* tetracycline resistance element. Journal of bacteriology, 166(3), 959-965.

Shoemaker, N., *et al*. (1989). Cloning and characterization of a *Bacteroides* conjugal tetracycline-erythromycin resistance element by using a shuttle cosmid vector. Journal of bacteriology, 171(3), 1294-1302.

Shoemaker, N., *et al*. (1991). Conjugal transfer of a shuttle vector from the human colonic anaerobe *Bacteroides uniformis* to the ruminal anaerobe *Prevotella* (*Bacteroides*) *ruminicola* B (1) 4. Applied and environmental microbiology, 57(8), 2114-2120.

Shipman, J., *et al*. (2000). Characterization of Four Outer Membrane Proteins Involved in Binding Starch to the Cell Surface of *Bacteroides thetaiotaomicron*. Journal of bacteriology, 182(19), 5365-5372.

Shroff, K., *et al*. (1995). Commensal enteric bacteria engender a self-limiting humoral mucosal immune response while permanently colonizing the gut. Infection and immunity, 63(10), 3904-3913.

Skoglund, P., *et al*. (2011). Estimation of population divergence times from non-overlapping genomic sequences: examples from dogs and wolves. Molecular biology and evolution, 28(4), 1505-1517.

Smith, C. (1985). Development and use of cloning systems for *Bacteroides fragilis*: cloning of a plasmid-encoded clindamycin resistance determinant. Journal of bacteriology, 164(1), 294-301.

Smith, C. (1992). Heterologous gene expression in *Bacteroides fragilis*. Plasmid, 27(2), 141-154.

Sonnenburg, J., *et al*. (2005). Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. Science, 307(5717), 1955-1959.

Sonnenburg, E., *et al*. (2006). A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing *in vivo* to carbohydrate metabolism. Proceedings of the National Academy of Sciences, 103(23), 8834-8839.

Sonnenburg, E., *et al*. (2010). Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations. Cell, 141(7), 1241-1252.

Steidler, L., *et al*. (1995). Secretion of biologically active murine interleukin-2 by *Lactococcus lactis* subsp. *lactis*. Applied and environmental microbiology, 61(4), 1627-1629.

Steidler, L., *et al*. (1998). Mucosal delivery of murine interleukin-2 (IL-2) and IL-6 by recombinant strains of *Lactococcus lactis* co-expressing antigen and cytokine. Infection and immunity, 66(7), 3183-3189.

Steidler, L., *et al*. (2000). Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. Science, 289(5483), 1352-1355.

Svensson, S., *et al*. (2014). Flagella-mediated adhesion and extracellular DNA release contribute to biofilm formation and stress tolerance of *Campylobacter jejuni*. PloS one, 9(8), e106063.

Svetoch, E., *et al*. (2010). Bacteriocins to control *Campylobacter spp*. in poultry—a review. Poultry science, 89(8), 1763-1768.

Szymanski, C., *et al*. (1999). Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Molecular microbiology, 32(5), 1022-1030.

Taboada, E., *et al*. (2012). Development and validation of a comparative genomic fingerprinting method for high-resolution genotyping of *Campylobacter jejuni*. Journal of clinical microbiology, 50(3), 788-797.

Tabor, J., *et al*. (2009). A synthetic genetic edge detection program. Cell, 137(7), 1272-1281.

Takahara, M., *et al*. (1985). The ompA signal peptide directed secretion of *Staphylococcal* nuclease A by *Escherichia coli*. Journal of Biological Chemistry, 260(5), 2670-2674.

Takle, G., *et al*. (2007). Evaluation of reference genes for real-time RT-PCR expression studies in the plant pathogen *Pectobacterium atrosepticum*. BMC plant biology, 7(1), 1.

Tam, C., *et al*. (2012). Changes in causes of acute gastroenteritis in the United Kingdom over 15 years: microbiologic findings from 2 prospective, population-based studies of infectious intestinal disease. Clinical infectious diseases, cis028.

Tamsir, A., *et al*. (2011). Robust multicellular computing using genetically encoded NOR gates and chemical/wires/'. Nature, 469(7329), 212-215.

TerAvest, M., *et al*. (2014). Regulated expression of polysaccharide utilization and capsular biosynthesis loci in biofilm and planktonic *Bacteroides thetaiotaomicron* during growth in chemostats. Biotechnology and bioengineering, 111(1), 165-173.

Terrapon, N., *et al*. (2015). Automatic prediction of polysaccharide utilization loci in *Bacteroidetes* species. Bioinformatics, 31(5), 647-655.

Theis, T., *et al*. (2007). Identification of suitable internal controls to study expression of a *Staphylococcus aureus* multidrug resistance system by quantitative real-time PCR. Journal of microbiological methods, 70(2), 355-362.

Thomas, M., *et al*. (2006). Estimated numbers of community cases of illness due to *Salmonella*, *Campylobacter* and verotoxigenic *Escherichia coli*: pathogen-specific community rates. The Canadian Journal of Infectious Diseases & Medical Microbiology, 17(4), 229.

Thomson, A., *et al*. (1989). Electroporation induced transformation of *Bacteroides ruminicola* and *Bacteroides uniformis* by plasmid DNA. FEMS microbiology letters, 61(1-2), 101-104.

Thompson, J., *et al*. (1990). Sequencing the gene for an imipenem-cefoxitin-hydrolyzing enzyme (CfiA) from *Bacteroides fragilis* TAL2480 reveals strong similarity between CfiA and *Bacillus cereus* beta-lactamase II. Journal of bacteriology, 172(5), 2584-2593.

Todorovska, A., *et al*. (2001). Design and application of diabodies, triabodies and tetrabodies for cancer targeting. Journal of immunological methods, 248(1), 47-66.

Tremaroli, V., *et al*. (2012). Functional interactions between the gut microbiota and host metabolism. Nature, 489(7415), 242-249.

Tsukazaki, T., *et al*. (2011). Structure and function of a membrane component SecDF that enhances protein export. Nature, 474(7350), 235-238.

Vafadarnejad, E., *et al*. (2015). The rad2 gene of haloarchaeum *Halobacterium salinarum* is functional in the repair of ultraviolet light induced DNA photoproducts.

Van Alphen, L., *et al*. (2008). A functional *Campylobacter jejuni* maf4 gene results in novel glycoforms on flagellin and altered autoagglutination behaviour. Microbiology, 154(11), 3385-3397.

van Bueren, A., *et al*. (2015). Differential metabolism of exopolysaccharides from probiotic *Lactobacilli* by the human gut symbiont *Bacteroides thetaiotaomicron*. Applied and environmental microbiology, 81(12), 3973-3983.

Van de Guchte, M., *et al*. (1989). Construction of a lactococcal expression vector: expression of hen egg white lysozyme in *Lactococcus lactis* subsp. *lactis*. Applied and Environmental Microbiology, 55(1), 224-228.

Van de Guchte, M., *et al*. (1990). Heterologous gene expression in *Lactococcus lactis* subsp. *lactis*: synthesis, secretion, and processing of the *Bacillus subtilis* neutral protease. Applied and environmental microbiology, 56(9), 2606-2611.

Vandenbroucke, K., *et al*. (2004). Active delivery of trefoil factors by genetically modified *Lactococcus lactis* prevents and heals acute colitis in mice. Gastroenterology, 127(2), 502-513.

Vandesompele, J., *et al*. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology, 3(7), 1-12.

Vigne, J., *et al*. (2004). Early taming of the cat in Cyprus. Science, 304(5668), 259-259.

Whitehead, T., *et al*. (1990). Heterologous expression of the *Bacteroides ruminicola* xylanase gene in *Bacteroides fragilis* and *Bacteroides uniformis*. FEMS microbiology letters, 66(1-3), 61-65.

Wilson, D., *et al*. (2009). Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. Molecular biology and evolution, 26(2), 385-397.

Win, M., *et al*. (2008). Higher-order cellular information processing with synthetic RNA devices. Science, 322(5900), 456-460.

Wostmann, B., *et al*. (1983). Dietary intake, energy metabolism, and excretory losses of adult male germfree Wistar rats. Laboratory animal science, 33(1), 46-50.

Wong, C., *et al*. (2003). Family of shuttle vectors for ruminal *Bacteroides*. Journal of molecular microbiology and biotechnology, 5(2), 123-132.

Xu, J., *et al*. (2003). A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science, 299(5615), 2074-2076.

Yampara-Iquise, H., *et al*. (2008). Use of a *Bacteroides thetaiotaomicron*-specific α-1-6, mannanase quantitative PCR to detect human faecal pollution in water. Journal of applied microbiology 105.5 : 1686-1693.

Young, K. T., *et al*. (2007). *Campylobacter jejuni*: molecular biology and pathogenesis. Nature Reviews Microbiology, 5(9), 665-679.

Zhang, C., *et al*. (2016). Ecological robustness of the gut microbiota in response to ingestion of transient food-borne microbes. The ISME journal.