ENGINEERING DUAL-GLYCAN RESPONSIVE EXPRESSION SYSTEMS FOR TUNABLE PRODUCTION OF HETEROLOGOUS PROTEINS IN *BACTEROIDES THETAIOTAOMICRON*

MARSHALL SMITH
Bachelor of Science, University of Lethbridge, 2016

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 Biological Sciences
University of Lethbridge
LETHBRIDGE ALBERTA CANADA

ENGINEERING DUAL-GLYCAN RESPONSIVE EXPRESSION SYSTEMS FOR TUNABLE PRODUCTION OF HETEROLOGOUS PROTEINS IN *BACTEROIDES THETAIOTAOMICRON*

MARSHALL SMITH

Date of Defence: December 12th 2018

Dr. D. Wade Abbott Co-Supervisor	Research Scientist, Adjunct Professor	Ph.D.
Dr. L. Brent Selinger Co-Supervisor	Professor	Ph.D.
Dr. G. Douglas Inglis Thesis Examination Committee Member	Research Scientist, Adjunct Professor	Ph.D.
Dr. Tony Russell Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Igor Kovalchuk Thesis Examination Committee Member Chair, Thesis Examination Committee	Professor	Ph.D.

Dedication

To my family, for the constant support and encouragement.

To my lab mates. This project would have been no fun without you, and I am proud to call you my friends.

Thank you.

Abstract

Genetically engineering symbiotic bacteria remains an underexploited opportunity to improve host-health and create new classes of biological devices, such as diagnostics or intestinal delivery systems for therapeutics. *Bacteroides thetaiotamicron* (*B. theta*) is a Gram-negative intestinal anaerobe with potential for the capability to produce functional heterologous proteins within a host intestine. To improve the strength and regulatory fidelity of transgene expression in *B. theta*, I have developed platform expression strains with engineered regulatory proteins under control of promoter elements that respond to dextran and arabinogalactan, two chemically distinct glycans. In addition to single glycan induction, I have also developed a novel "dual-glycan" expression system that requires the addition of both dextran and arabinogalactan for induction. Additionally my engineered strains are compatible with a series of chromosomal integration and episomal vectors that improve the throughput of gene cloning, integration, and expression. Together this expression system provides a new collection of glycan-responsive tools to improve transgene expression in *B. theta* and provides the proof-of-concept for engineering more complex dual-glycan expression systems.

Acknowledgements

I would like to thank Dr. Wade Abbott for being my supervisor and giving me the opportunity to work in his lab throughout both my undergraduate and graduate careers, and for helping to power through failed experiments and celebrate successful ones with me. I learned a great deal of valuable skills and lessons during my time in the lab.

To my co-supervisor Dr. Brent Selinger, and to committee member Dr. Douglas Inglis for important discussions and advice on methodology and statistics. To committee member Dr. Anthony Russell for insightful suggestions and perspective.

A special thank-you to past and present members of the Abbott Lab. This project would not have been possible if it were not for previous work done by Darryl Jones, Richard McLean, Julie Grondin, Kaitlyn Shearer, Carolyn Amudsen, Benjamin Farnell, and Anu Anele. I would also like to thank Darryl Jones, Kristin Low, Julie Grondin, and Brent Selinger for providing tremendously useful comments and feedback on various editions of this thesis and the paper manuscript. My writing is still awful, but it is less awful than it was when I started.

Thanks are also due to Jeff Tingley, Stephanie Monteith, Leeann Klassen, Jaclyn MacMillan, Marissa King, Mike Xing, Adam Smith, Muhammed Salah Uddin, Paul Moote, Hannah Dyer, Chris Coté, and Ben Vuong.

Table of Contents

Title Page……………………………………………………………………………………….i
Signature Page. ……………………………………………………………………………ii
Dedication iii
Abstract iv
Acknowledgements v
List of Tables viii
List of Figures ix
List of Abbreviations.. x
Chapter 1 Introduction and Literature Review 1
1.1 Influencing the gut microbiota 2
1.1.1 Probiotics 3
1.1.2 Prebiotics 3
1.1.3 Synbiotics 4
1.2 Genetic engineering and the gut microbiota 5
1.2.1 Escherichia coli 6
1.2.2 Lactococcus lactis 7
1.3 *Bacteroides* spp. are well suited for probiotic engineering 8
1.3.1 *B. theta* as an engineering candidate 10
1.3.2 Genetic tools for use with *B. theta* 10
1.4 Polysaccharide utilization loci regulatory proteins 11
1.4.1 Extracytoplasmic function σ factors 13
1.4.2 Hybrid two component systems 13
1.4.3 SusR-like regulators 16
1.5 B. theta promoters and PULs 16
1.6 Hypothesis..19
1.7 Objectives...19
Chapter 2 Engineering Dual-Glycan Responsive Expression Systems for Tunable Production of Heterologous Proteins in *Bacteroides thetaiotomicron*.. 20
2.1 Introduction..21
2.2 Results..24
2.2.1 Construction of a pExchange based platform for targeted insertion and expression of transgenes 24
2.2.2 Validation of *B. theta* promoters for regulation of transgenes. 27
2.2.3 Engineering tunable expression of glycan-responsive PUL regulator proteins in *B. theta* 29
2.2.4 Augmenting transgene expression using a *B. theta* episomal expression vector ...32
2.2.5 Heterologous production of an agarase in modified *B. theta* strains 35
2.3 Discussion 37
2.4 Conclusion 41
2.5 Methods..42
2.5.1 Vector construction 42
2.5.2 Regulatory gene promoter engineering 45
2.5.3 Conjugations..46
2.5.4 Growth curves 47
2.5.5 *B. theta* wild-type and ΔPUL75 co-culture and qPCR...48

2.5.6 NanoLuc assays 48
2.5.7 Statistical analysis 49
2.5.8 BuGH16 digestion assays 50
2.5.9 Western blotting 50
Chapter 3 Conclusions and future directions 54
3.1 Future Directions 56
3.2 Public opinion on genetically modified organisms 57
3.3 Biocontainment of genetically modified organisms 60
References 61
Appendix: The AppA Phytase 79
Background 79
Results and Discussion 81
Methods 89
AppA activity and protein detection in *E. coli* 89
Attempts to produce AppA in *B. theta* 90
qPCR in B. theta 91
Codon usage analysis 92
Appendix References 93

List of Tables

Table 2.1: Primers used in this study. 4343
Table 2.2: Plasmids and Strains used in this study 45
Table S1: Novel engineered regulatory systems ……..……52
Table A.1: Signal peptides used to direct localization of the AppA phytase…….……82
Table A.2: Codon usage preference for each amino acid compared between *E. coli* and *B. theta.*……...87

List of Figures

Figure 1.1: General model for polysaccharide acquisition and degradation by Bacteroides PULs. ...12
Figure 1.2: B. theta PULs specific for DX and AG 18
Figure 2.1: Target genomic locus and integrative expression vectors.. 25
Figure 2.2: Inducible expression cassettes produce product in a dose dependent manner under wild-type regulation in Bacteroides thetaiotaomicron.....28
Figure 2.3: Regulatory engineering changes growth kinetics, levels of gene activity, and gene induction requirements.....31
Figure 2.4: Episomal transgene expression increases gene expression compared to chromosomal equivalent.. 34
Figure 2.5: The heterologous BuGH16 agarase can be produced in engineered regulatory strains....36
Figure S1: Comparison of NanoLuc activity from two different genomic loci……...53
Figure A.1: AppA phytase activity and localization patterns when produced with different signal peptides in E. coli. ein was detected from fractions using Western blotting.....82
Figure A.2: AppA protein localized to the *E. coli* outer membrane by signal peptides……83
Figure A.3: Lack of meaningful change in phytase activity detected from transgenic *B. theta* harbouring the *appA* phytase under control of different promoters.. 85
Figure A.4: Detection of *appA* transcript from transgenic *B. theta* by qPCR.. 8686
Figure A.5: Codon frequency in *E. coli* and *B. theta* applied to the *appA* gene……....88

List of Abbreviations

AG	arabinogalactan
AGP	antimicrobial growth promoter
B. theta	*Bacteroides thetaiotaomicron*
BHIS	supplemented brain heart infusion
CAZyme	carbohydrate-active enzyme
CUTD	Codon Usage Table Database
DGM	distal gut microbiota
DNA	deoxyribose nucleic acid
DX	dextran
ECF	extra-cytoplasmic function
FOS	fructooligosaccharides
FTU	phytase unit
FUDR	floxuridine
GalA	galacturonic acid
GH	glycoside hydrolase
GLC	glucose
GM	genetic modification
GMO	genetically modified organism
HG	homogalacturonan
HTCS	hybrid two-component system
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	lysogeny broth
MFS	major facilitator superfamily
MM	minimal medium
NBU	non-replicating *Bacteroides* unit
PBP	periplasmic binding protein
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PL	polysaccharide lyase
PUL	polysaccharide utilization loci
PVDF	polyvinlyidine fluoride
qPCR	quantitative polymerase chain reaction
RBS	ribosome binding site
RNA	ribonucleic acid
RNAP	RNA polymerase
SCFA	short chain fatty acid
SUS	starch utilization system
TA	toxin- antitoxin
TCS	two-component system
TYG	tryptone yeast glucose
WT	wild-type

Chapter 1 Introduction and Literature Review

The digestive tract of animals is populated by hundreds of species of microorganisms, collectively referred to as the microbiota. These microorganisms have been found to colonize the entire tract, with relatively low densities of 10^3 microorganisms per mL in the stomach to the much denser 10^{11} microorganisms per mL in the colon [1]. In recent years, both scientific and public interest in the microbiota, especially the distal gut microbiota (DGM), has increased as knowledge of this community's role in host health has expanded. The total collection of genes encoded within the DGM, or distal gut microbiome, greatly outnumbers the host genome [2] and provides functionality for key physiological roles, such as synthesis of vitamins and other essential molecules, and digestion of dietary glycans [3]. In the colon, low motility, neutral pH range, and the absence of oxygen create the perfect environment for fermentation of glycans that pass undigested from the small intestine [4]. The mammalian DGM creates short chain fatty acids (SCFAs) from these glycans. In humans, SCFAs supply an estimated 10% of daily caloric intake and are the main energy source utilized by enterocytes [5]. In ruminant livestock, such as cattle, SCFAs from microbial fermentation can supply up to 70% of the animals' daily calories [6]. The DGM has also been shown to be vital in proper intestinal development in neonates by guiding angiogenesis [7], and does additionally contribute to immune system development and priming [8, 9]. Indeed, germ-free animals display impaired intestinal and immune system development and function, yet these defects can be largely reversed by colonization with a conventional microbiota [7, 10]. A healthy and diverse microbiota has been shown to

benefit hosts by competitively excluding pathogen colonization [11]. However, reduced species diversity and imbalances in the DGM, termed dysbiosis [12], have also been linked with chronic health conditions in humans such as diabetes [13], obesity [14], and inflammatory bowel disease [15]. Reduced DGM diversity is implicated in recurrent *Clostridium difficile* associated diarrhea [16], a current treatment for which is fecal microbiota transplant from a healthy donor [17]. Clinical and experimental evidence suggests that the DGM can interact with the central nervous system to influence the host's mental state through interactions with intestinal cells and the enteric nervous system, or by secreting neuroactive compounds that reach their targets through the circulatory system [18]. Correspondingly, dysbiosis in the DGM has been associated with nervous disorders such as depression [19, 20], anxiety [21], and autism [22]. Dysbiosis is problematic in animals as well. In livestock, dysbiosis in the DGM has been implicated in Johne's disease (an ailment similar to Crohn's disease in humans) and hemorrhagic diarrhea in cattle [23, 24], intestinal irritation and inflammation in postweaning pigs [25], and inflammation and shortened villi in chickens [26]. A healthy microbiota is therefore critical for human and animal health, and tools and treatments that influence the DGM are important topics of research.

1.1 Influencing the DGM

Probiotics and prebiotics are products designed to improve health by interacting with the microbiota post-ingestion. These are commonly marketed towards consumers, but a number exist for commercial livestock production as well. Commercial probiotic products usually take the form of fermented foods such as yogurt, but capsules, health drinks, and other innovative forms, such as ice cream [27], are on the market. Pro- and

prebiotics that are designed to function in concert are termed “synbiotics” for their expected synergistic effects.

1.1.1 Probiotics

Probiotics are defined by the International Scientific Association for Probiotics and Prebiotics as “live microorganisms that, when administered in adequate amounts, confer a health benefit on the host” [28]. Probiotics have been shown to reduce symptoms of irritable bowel disease [29], reduce anxious behavior in laboratory animals [30], and ease symptoms in human schizophrenia sufferers [31]. Additionally, probiotics have demonstrated anti-tumourigenic activities both indirectly by suppressing carcinogen-producing microbial species and directly via altering expression of genes involved in cell cycle and apoptosis [32]. In livestock production, probiotics are reported to increase average daily gains [33, 34] and reduce the amount of antimicrobial supplementation required by replacing antimicrobial growth promoters (AGPs) [35], which has important implications for industry as AGPs are phased out after Dec. 1, 2018. The mode of action will differ between probiotic species, but interactions with host receptors [30], stimulation of host immune system [29], and production of certain metabolites such as hydrogen peroxide and short chain fatty acids have all been implicated [36].

1.1.2 Prebiotics

A prebiotic is a substance, typically an indigestible carbohydrate, that stimulates the growth and/or activities of a select group of microorganisms that confer a health benefit to the host [37]. Fructans and galactooligosaccharides were among the first sugars shown to have a prebiotic effect and most prebiotic research involves these sugars [38]. These sugars are selectively fermented by and promote the growth of *Bifidobacterium* and

Lactobacillus species. Prebiotics have been shown to have a variety of effects including: reducing the severity of irritable bowel disease [39], decreasing the incidence of cancer [40, 41], improving the uptake of calcium and other minerals [42, 43], improving bone mineralization [44], decreasing levels of hunger hormones, increasing levels of satiety hormones [45], and reducing weight gain [46]. The exact mechanisms by which prebiotics are effective remain unknown, despite the compelling evidence linking the function of prebiotics to the modulation of the microbiota [47].

1.1.3 Synbiotics

Synbiotics consist of pro- and prebiotics functioning concertedly to impart a synergistic effect. For example, feeding *Lactobacillus rhamnosus* and *Bifidobacterium lactis* probiotics when combined with fructans comprises one example of a synbiotic [48]. A number of studies have examined synbiotics and demonstrated positive outcomes when compared to negative controls. A blend of bacterial species in combination with fructo-oligosaccharides (FOS) can shorten the duration of diarrhea in children [49, 50], and administering *B. longum* with inulin and FOS can reduce histological disease scores in patients with Crohn's disease [51]. Like probiotics and prebiotics, synbiotics are also under consideration for livestock applications. A bacterial multispecies mix and inulin-type fructans can increase average daily gains in broiler chickens [33], and a year-long study found that *Lactobacillus casei* and dextran can significantly improve milk quantity and quality in Holstein cows [52].

Unfortunately, most studies involving synbiotics have neglected to test the probiotic and prebiotic component separately to determine whether or not the synbiotic truly function synergistically [53]. Investigations of synergy have been varied, with some

studies reporting synergistic effects [54], and others finding no differences between administering prebiotics and probiotics individually or together [48, 55, 56]. To summarize, though synbiotic formulations can be effective, current evidence for being more effective than either probiotics or prebiotics individually is weak at best.

1.2 Genetic engineering and the DGM

Genetically-engineered bacteria provide a potentially powerful tool for medicine. There are already pre-clinical and clinical trials of live modified bacteria being administered to hosts. For example, *Listeria monocytogenes* is being investigated as a potential cancer treatment when engineered to produce cell-surface tumor antigens, triggering a response from the host immune system [57]. Additionally, within native microbial communities, engineered *Lactobacillus jensenii* have been shown to produce anti-viral peptides to inhibit HIV [58], and *Lactococcus lactis* can be manipulated to produce anti-inflammatory cytokines to alleviate symptoms of Crohn's disease [59] and proinsulin for the treatment of diabetes [60]. The treatment of human disease is only one potential application for genetically-engineered bacteria, and possibilities exist outside the realm of human health. In agricultural settings for example, *Butyrivibrio fibrisolvens* carrying an engineered plasmid can protect sheep from the toxic plant secondary metabolite monofluoroacetate [61]. Potentially, the addition of digestive enzymes may benefit livestock production by improving the digestibility of feed and could potentially open up new feed options such as marine algae [62, 63]. Engineered bacteria could increase animal production by reducing complications such as the antinutritive effects of viscous feed [64] and reduced feed intake due to slowed transit [65].

1.2.1 Escherichia coli

Escherichia coli is a natural first choice for engineering probiotics. Well-studied since 1961 when the *lac* operon prompted Jacob and Monod to theorize the existence of regulatory circuits [66], *E. coli* has been used for construction of the first recombinant plasmid [67], the first transgenic protein production [68], and the first use of a phage promoter system [69]. *E. coli* is simple to culture and there is a wealth of tools available for manipulation of the organism itself. In addition to its historical and practical value, strains of *E. coli* are commensal residents in the mammalian colon and have been investigated for probiotic benefits [70]. *E. coli* strain NGF-1 has been engineered to sense tetrathionate, a marker of inflammation produced by certain pathogens, and produce β-galactosidase in response, allowing the strain to function as a non-invasive diagnostic tool [71]. Another study has reprogrammed *E. coli* via the introduction of novel genetic circuits, allowing for chemotaxis towards the pathogenic bacterium *Pseudomonas aeruginosa* and release of nuclease DNase I and the antimicrobial peptide microcin S to both degrade the *P. aeruginosa* protective biofilm and cause cell death [72].

Though these remarkable advancements have been made, there are disadvantages towards using *E. coli*. Compared to some other species, *E. coli* is not very abundant in the colon [73] which can be problematic when considering an engineered activity to have clinical relevance for colonic treatments. Producing the threshold amount of protein or other product required to have a detectable effect will likely require each of the few *E. coli* cells to produce a large amount. Dedicating resources to processes that are non-essential for survival lowers growth rate *in vitro* [74], and would likely impact the fitness of *E. coli* in the highly competitive gut environment and hinder its ability to survive long-

term. For these reasons, *E. coli* may be better suited for applications requiring either low levels of production or short-term colonization. Another potential issue arises when one considers that *E. coli* is known to preferentially colonize the enteric lumen rather than the mucosal surface [75], limiting the bacterium's utility in epithelial-specific operations such as targeting epithelium-associated pathogens or delivering signals to epithelial cells. Thus, while *E. coli* can be well-suited to a limited set of applications, many opportunities exist for alternate bacterial species in other situations.

1.2.2 Lactococcus lactis

Lactococcus lactis has several advantages over *E. coli*. *Lactococcus* species have been consumed in milk and other dairy products for over ten thousand years [76]. Whereas some strains of *E. coli* are known pathogens [77], *L. lactis* is recognized by Health Canada to "have a long history of safe use in food" [78], is designated "Generally Recognized As Safe" by the American Food and Drug Administration [79], and has a Qualified Presumption of Safety from the European Food Safety Authority [80]. *L. lactis* is the first genetically-modified bacterium to be administered live to human beings in a clinical trial, for which it was engineered to produce human interleukin-10, an anti-inflammatory cytokine, as a treatment for Crohn's disease [59]. In a clinical trial, eight of the ten participants showed improvement and four of the participants experienced relapse after the probiotic was withdrawn. Unfortunately, the study lacked both a placebo control group, and a conventional treatment positive control, and so did not allow for the determination of efficacy. Mouse studies performed prior to the human trials showed that non-transgenic *L. lactis* did not beneficially affect inflammation [81], suggesting that the benefit observed in humans was a result of the genetically-engineered organism. The

most common adverse effect of the treatment was excessive flatulence in three out of ten patients. *L. lactis* can be effectively used outside of the gut in some circumstances, as evidenced by its successful use as a live intranasal tetanus vaccine in mice [82]. *L. lactis* is seemingly best suited for short term applications with respect to gastro-intestinal use; although *L. lactis* has been shown to survive passage through the alimentary tract, it does not appear to colonize, becoming undetectable by selective plating or PCR 4 days after administration ceases [83].

1.3 *Bacteroides* spp. are well suited for probiotic engineering

The Gram-negative bacterial genus *Bacteroides*, a member of the phylum Bacteroidetes, is another a promising candidate for engineered probiotics. Bacteroidetes is one of the most prominent and prevalent phyla in the DGM. Unlike non-colonizing transient species such as *L. lactis*, Bacteroidetes are long-term colonizers of the gut and comprise approximately 25% of microorganisms in the human DGM [73]. The Bacteroidetes phylum is prominent in livestock as well, comprising roughly 25% of the bacteria both the cow rumen [23] and the chicken ceca [84]. Though *Bacteroides spp.* are able to cause significant pathology and sepsis if they enter the blood stream [85], e.g. through an open wound, they are generally considered to be commensals or mutualists in the colon [86]. *Bacteroides spp.* are generally known for their ability to metabolize a wide variety of structural complex polysaccharides that originate from diverse biological sources, including terrestrial plants [3], seaweed [87], bacteria [88], fungi [89], and animals [90]. The genes that encode for the transporters, enzymes, and regulatory proteins responsible for these metabolic activities are organized into discrete clusters called polysaccharide utilization loci (PULs) and are abundant in *Bacteroides*. The archetypical

PUL is the starch utilization system (SUS) discovered in *Bacteroides thetaiotaomicron* (*B. theta*) [91]. The hallmark components of a PUL include a SusC-like Ton-B dependent transporter and a SusD-like outer membrane binding protein pair [92]. The SusC homolog is an outer membrane protein with a beta-barrel structure that relies on the proton-motive force generated by the plasma membrane TonB complex in order to transport sugars from the extracellular environment to the periplasm. SusD-like outer membrane proteins bind the target substrate and function to enhance substrate binding for transport [93]. The SusD interacts with its associated SusC at the opening of the pore and is thought to swing open in a 'pedal bin'-like mechanism to expose the substrate binding site on the SusC-like transporter [94]. PULs typically contain several carbohydrate-active enzymes (CAZymes) specific for the targeted degradation of glycosidic bonds and modifications within the identified polysaccharide, and a regulatory protein that activates the PUL upon sensing the appropriate substrate. The capabilities of PULs and the signaling networks needed to coordinate them are intricate and could provide useful tools for engineered probiotics where engineered activities are tuned by carbohydrates in the host diet.

Indeed, *Bacteroides* species have begun to be investigated as potential engineered bacterial species for use in the mammalian colon. Using a promoter responsive to the plant cell wall polysaccharide xylan [95], *B. ovatus* has been engineered to produce a variety of compounds intended to be bioactive in the distal gut including interleukin-2 [96], trefoil factor 3 [95], keratinocyte growth factor-2 [97], and transforming growth factor β [98]. These products are immunomodulatory or otherwise involved in healing and the goal of their production is to speed recovery from colonic inflammation. *B. ovatus* cells producing heterologous protein with this xylan system have been tested on colonic epithelial cell lines

[95, 96] and in mice [97]. When mice were fed xylan and gavaged with *B. ovatus* engineered to secrete keratinocyte growth factor in the presence of xylan, the severity of dextran sodium sulfate-induced colitis was significantly reduced. No human studies have been carried out so far.

1.3.1 *B. theta* as an engineering candidate

First described by Distasso in 1912 [99], *B. theta* was briefly designated a subspecies of *B. fragilis* [100] from 1973 to 1976 until full species designation was restored due to genetic distinctiveness [101]. *B. theta* is one of the most studied members of Bacteroidetes. It can be detected in roughly half of human fecal microbiotas [73] and has been shown to lessen intestinal inflammation by promoting nuclear export of pro-inflammatory transcription factor NF-κB subunit RelA [102]. The ability to interact with host transcription factors combined with its prevalence indicate that *B. theta* is very well adapted to its hosts.

The genome of *B. theta* is estimated to contain at least 88 PULs based upon the identification of SusC-like/D-like pairs [103], corresponding to approximately 18% of the genome, suggesting that this microorganism is well-endowed as a "generalist" to adapt to fluctuating levels of dietary- and host-derived glycans present within the DGM [89]. In this regard, selective metabolic strategies, and unique promoter responses are under-exploited opportunities for engineering glycan-responsive bacteria that function within the colon.

1.3.2 Genetic tools for use with *B. theta*

Bacteroidetes diverged from other bacteria relatively early in their evolutionary history, with many differences observed within the organization of genetic elements. Rather than the canonical -10 and -35 consensus sequences in promoters, *Bacteroides*

instead have -7 and -33 sequences [104] and unique RNA polymerase subunits adapted to recognize them [105]. The Bacteroidetes also appear to have unique translation initiation signals. Their ribosome binding sites (RBS) are enriched in adenosine and thymidine and ribosomes are more selective for RBS sequence and spacing [106, 107]. For these reasons, molecular biology techniques and tools commonly used in *E. coli* are often not adaptable for use in *Bacteroides* species.

A number of genetic tools have been created for use in *B. theta.* The pExchange plasmid for homologous recombination has existed since 1991 [108]. Using this counter-selectable system, unmarked chromosomal deletions and insertions can be made with single base pair resolution [109, 110]. Alternatively, non-replicating *Bacteroides* unit 2 (NBU2), a mobilizable transposon element [111] incorporated into a plasmid (pNBU2) [110], can be used to quickly integrate the entire heterologous plasmid sequence into the chromosome of *B. theta*. More recently, constitutive promoter elements and RBSs have been screened and ranked to provide a wide-range of expression levels for heterologous proteins [109, 112]. Episomal and chromosomal inducible gene expression systems also exist, and can be regulated by anhydrotetracycline, a synthetic chemical inducer [109], or the natural polysaccharide α-mannan [113].

1.4 Polysaccharide utilization loci regulatory proteins

PULs are typically controlled by one of three possible regulatory mechanisms: an extra-cytoplasmic function (ECF) σ factor, a hybrid two component system (HTCS), or a SusR-like regulator (Fig. 1.1). HTCSs and SusR are unique to Bacteroidetes, as homologs have not been found in any non-Bacteroidetes organisms [114].

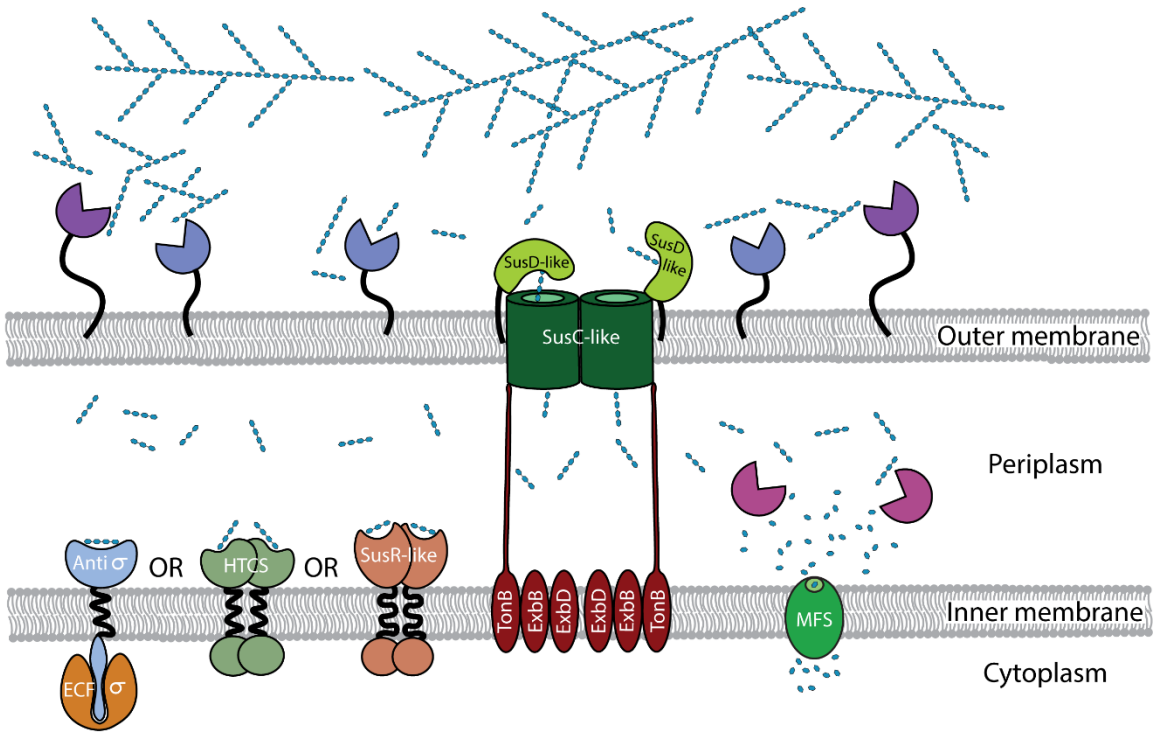


Figure 1.1: General model for polysaccharide acquisition and degradation by *Bacteroides* PULs. Large polysaccharides are partially degraded by outer membrane CAZymes. The resulting products are imported by SusC-like TonB dependent transporters with the aid of SusD-like binding proteins. In the periplasm, oligosaccharides are typically detected by one of three regulatory apparatus; ECF-σ/Anti-σ factors, HTCSs, or SusR homologs, which upregulate their respective PUL. Periplasmic oligosaccharides are further digested into monosaccharides for transport to the cytoplasm by major facilitator superfamily (MFS) transporters.

1.4.1 Extracytoplasmic function σ factors

σ-factors are subunits of RNA polymerase (RNAP) complexes and can be broken into two families based on structure: σ^{54} and σ^{70}, both identified first from *E. coli* and named for their molecular weight. The σ^{70} family is the largest and most bacteria possess at least one member [115]. There are four conserved regions in σ^{70}-factors, accordingly numbered 1 through 4. ECF type σ-factors are a diverse group within the σ^{70} family [116]. They contain only regions 2 and 4 and are typically smaller than other family members. The archetypical pathway of regulation with ECF-type σ-factors involves an anti σ-factor. The anti σ-factor is a transmembrane protein with periplasmic and cytoplasmic domains. The cytoplasmic domain binds between regions 2 and 4 of the σ-factor and prevents the σ-factor from interacting with DNA or RNAP. When the periplasmic region senses the presence of its target sugar, the σ-factor is released from the cytoplasmic domain of the anti σ-factor to recruit RNAP and trigger gene expression. The genome of *B. theta* is highly enriched for ECF type σ-factors with 50 out of 54 total σ-factors being ECF type. Furthermore, out of the 25 ECF type σ-factors located in a PUL, 22 are adjacent to an anti σ-factor, indicating that this regulatory mechanism is an important part of *B. theta*'s nutrient acquisition strategy.

1.4.2 Hybrid two component systems

Two-component systems (TCS) are common in bacteria and are used for many sensing and signaling applications. The first component is a response regulator frequently involved in DNA-binding [117]. The second component, a membrane-bound sensor, consists of a histidine kinase and a phosphoacceptor domain that, upon sensing an

environmental cue, transfers a phosphoryl group from a histidine residue in the sensor to an aspartic acid residue in the cognate response regulator [118], causing the output domain of the response regulator to become active and allowing the cell to respond to the signal.

A hybrid two-component system (HTCS) has all the hallmarks of a conventional TCS combined into a single polypeptide chain. HTCSs are embedded in the inner membrane with periplasmic sensors and cytoplasmic effector domains. The response regulators of HTCSs are AraC type DNA binding domains [114]. The histidine and aspartic acid residues involved in phosphotransfer in classical TCSs are conserved in HTCSs, suggesting that the basic mechanism is also conserved. *B. theta* has 32 HTCS proteins, more than any other sequenced prokaryote [119], and 30 of those HTCSs are located in PULs. It is thought that HTCSs provide *B. theta* with advantages for glycan foraging. With the sensor domain and the effector domain incorporated into a single protein, there is no detectable cross-talk between non-cognate sensors and effectors *in vivo*, even though sensors may exhibit promiscuity *in vitro* [120]. This single peptide structure limits signal amplification as the sensors cannot easily activate multiple response regulators, but, in the context of the competitive DGM, it may allow *B. theta* to more accurately and rapidly sense and respond to specific sugars [121]. Being a single peptide may also allow novel HTCSs to evolve more rapidly because the sensor and the effector do not need to coevolve the unique interfaces required to prevent cross-talk. PUL regulators are usually constitutively expressed, but at least one HTCS is inducible [122].

B. theta has two classes of HTCS proteins [123]. The more common Reg_prop class is named for the large β-propeller structures in the N-terminal domain. The periplasmic domain of *bt4663*, the Reg_prop HTCS for heparan sulfate utilization, has been structurally

characterized by X-ray crystallography in both apo and ligand-bound forms [124]. The protein exists as a dimer, and upon substrate binding the two monomers engage in a "scissor blade" like movement which is thought to bring the kinases close enough to autophosphorylate and begin signal transduction through phosphotransfer [124]. The other class of HTCS is the periplasmic binding protein (PBP) class, of which *B. theta* has only one member, *bt1754*. This HTCS is unique in another capacity as it is the only known HTCS to recognize a monosaccharide, fructose, as other HTCSs recognize disaccharides or larger [125].

Computational analysis of HTCS regulons in *B. theta* and other *Bacteroides* has identified DNA motifs predicted to be bound by HTCSs following substrate recognition by the sensor region. The DNA motifs consist of two conserved sequences separated by an average of 21 bp, suggesting that HTCSs bind to DNA as a dimer where the monomers are separated by two helix turns [114] albeit the exact mechanism remains unclear. Though it is known that HTCSs relay information through phosphotransfer, it is not known how the active response regulator actually comes into contact with its DNA target. A conventional TCS has a response regulator that can diffuse through the cytoplasm while the sensor kinase stays membrane bound, but these two entities are fused in an HTCS. However, a chimeric HTCS created by joining the sensor domain of one with the regulatory domain of another was found to bind directly to DNA sequences that it could either positively and negatively regulate [126]. The HTCS was probed with an antibody specific to the sensor domain, indicating that it is not proteolytically cleaved before binding DNA. Thus, either the HTCS must leave the membrane and diffuse towards the chromosome, or the chromosome must come into contact with the membrane-bound HTCS.

1.4.3 SusR-like regulators

The final and smallest group is the SusR-like regulators, which are identified by homology to the regulatory protein from the SUS. *B. theta* has four known SusR-like proteins in addition to the canonical SusR [114]; one is from PUL48 and responds to dextran, two other regulators have unknown specificities, and the final sensor is not associated with a PUL. SusR-like proteins are similar to HTCS proteins in that a periplasmic sensing domain is physically linked to a cytoplasmic DNA binding effector domain. The mechanism of SusR-like activation is unknown, but the C-terminus of SusR has a helix-turn-helix motif similar to those found in some other transcriptional regulators [127]. Analysis of SusR-like regulons has revealed DNA motifs similar to those predicted for HTCSs in that there are two conserved sequences separated by a non-conserved linker. The linkers for SusR homologs are longer than those of HTCSs at 67 or 77 bp long indicating that SusR homologs bind DNA as a dimer where the monomers are separated by six or seven helix turns [114].

1.5 B. theta promoters and PULs

Gene expression profiling studies on *B. theta* have revealed genes with high transcription levels whose promoters could potentially be used for bioengineering. A study analyzing constitutive promoters showed that the *B. theta* σ-factor *rpoD* (*bt1311*) resulted in the highest level of activity for the luminescent reporter gene NanoLuc® [112]. Exchanging the ribosome binding site (RBS) naturally associated with the *bt1311* promoter region for the RBS from a gene encoding a ribosomal protein lead to a slight increase in NanoLuc activity.

Inducible promoters have been found in the *B. theta* PULs for dextran (DX, Fig. 1.2a) and arabinogalactan (AG, Fig. 1.2b). DX is a polysaccharide composed of glucose monomers. The main chain has α-1,6 linkages and side chains branch out from α-1,3 linkages. Molecular weights can vary from 3 kDa to 2000 kDa [128]. DX is created from sucrose by lactic acid bacteria such as *Leuconostoc mesenteroides* and can naturally be found in wine, some fermented meats, and dental plaque [128, 129]. It has medical applications as an anti-clotting agent and a blood volume expander [130]. DX utilization in *B. theta* is controlled by PUL48 [131]. PUL48 contains a *susC/D* pair, a *susR*, two glycoside hydrolases (GHs) and a gene of unknown function. Every gene save the *susR* is part of the same operon. The promoter of this operon, P_{bt3090} (P_{DX}), which is regulated by *bt3091* (*SusR*^{DX}) has been shown to be capable of maintaining transcript levels close to 1000-fold higher on DX than on glucose (GLC) [131].

AG is composed of D-galactose and L-arabinose monomers in a ratio of roughly 6:1. The main chain is composed of galactose linked through β-1,3 bonds. Side chains are linked to the main chain through α-1,6 bonds. AG is a component of plant cell walls [132] and is commercially isolated from larch wood. It can be safely used as a thickening agent for both foods and cosmetics. AG is the substrate of two PULs in *B. theta*, PUL5 and PUL65 [103]. The *susC* of PUL5 (*bt0268*) is more highly expressed than the corresponding *susC* in PUL65 (*bt3680*) [103]. The fold change in expression levels between the induced and non-induced states of *bt0268* range from 29 to 71-fold depending on publication [103, 112]. PUL5 is a relatively large PUL with HTCS type regulatory gene *bt0267* (*HTCs*^{AG}), two *susC/D* pairs, three GHs, a polysaccharide lyase (PL), an integrase and 15 genes of unknown function. The P_{bt0268} (P_{AG}) promoter has

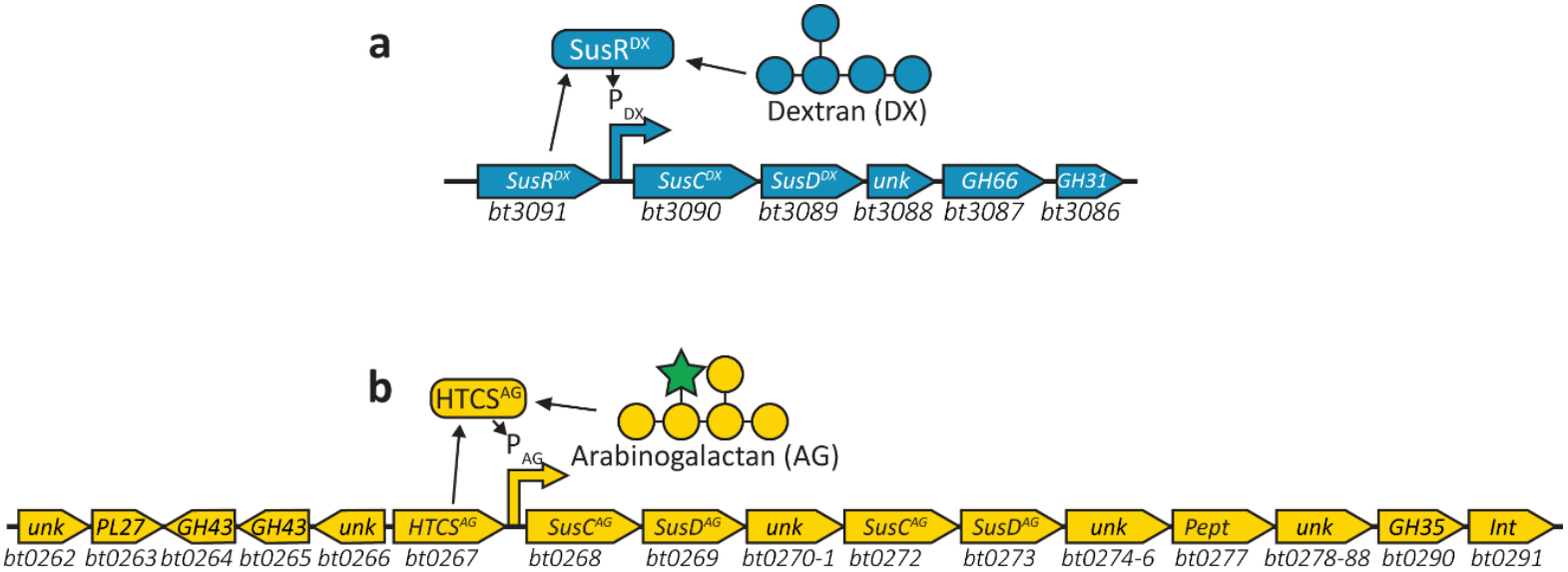


Figure 1.2.2: B. theta PULs specific for DX and AG . (a) Schematic of DX responsive PUL48. The highlighted promoter P_{DX} is activated by the SusR-like protein $SusR^{DX}$ in the presence of DX. DX is represented by linked blue circles. **(b)** Schematic of AG responsive PUL5. The highlighted promoter P_{AG} from PUL5 is activated by the HTCS protein $HTCS^{AG}$ in response to the presence of AG. Galactose monomers in AG are represented by yellow circles and arabinose monomers are represented by green stars. GH = glycoside hydrolase, HTCS = hybrid two-component system, Int = integrase, Pept = peptidase, PL = polysaccharide lyase, Sus= starch utilization system, unk = unknown.

been used to drive expression of NanoLuc and demonstrated to do so at a high level only in the presence of AG [112].

1.6 Hypothesis

Changing the regulation of the dextran regulatory gene *SusR*^{DX} by replacing its promoter with a strong constitutive promoter, or a dextran or arabinogalactan responsive promoter will change the growth kinetics and / or expression patterns that affect gene product patterns in response to the presence of dextran, arabinogalactan, or a mixture of the two sugars.

1.7 Objectives

1. Replace the native promoter sequence of *SusR*^{DX} with the P_{ON} (P_{bt1311}), P_{DX} (P_{bt3090}), and P_{AG} (P_{bt0268}) promoter sequences.
2. Investigate how these modifications change the growth profile of engineered bacteria (see objective 1) when cultured on dextran, arabinogalactan, or a mixture of the two sugars.
3. Investigate how these modifications change the activity patterns of NanoLuc, a luciferase reporter when engineered bacteria (see objective 1) are cultured on dextran, arabinogalactan, or a mixture of the two sugars.
4. Test production of a new reporter protein: endo-β-agarase BuGH16 in select engineered bacteria.
5. Generate "reciprocal strains" by replacing the arabinogalactan responsive promoter of the regulatory gene *HTCS*^{AG} with the P_{ON}, P_{DX}, and P_{AG} promoter sequences.

Chapter 2 Engineering Dual-Glycan Responsive Expression Systems for Tunable Production of Heterologous Proteins in *Bacteroides thetaiotomicron*

The following chapter was prepared as a manuscript for submission to the Journal of Scientific Reports.

Jones DR[1¶], Smith MB[1,2¶], McLean R[1], Grondin JM[1], Amundsen C[1], G. Douglas Inglis[1], L. Brent Selinger[2], Richard R. E. Uwiera[3], and D. Wade Abbott[1,2]

[1] Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, Alberta, Canada, T1J 4B1

[2]Department of Biological Sciences, University of Lethbridge, Lethbridge, Alberta, Canada, T1K 3M4

[3]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, Canada, T6G 2P5

[¶]These authors contributed equally to this work

2.1 Introduction

The lower digestive tract of mammals contains hundreds of species of microorganisms [133], collectively referred to as the distal gut microbiota (DGM). The total collection of genes encoded within the DGM greatly outnumbers the host genome, and is essential for physiological roles, such as digestion of dietary fibre and the synthesis of vitamins and amino acids [3]. In recent years, public and scientific interest in how the structure of the DGM is established and dynamically changes has increased as mechanistic links between the DGM and host health have been defined [134]. The DGM is required for proper intestinal and immune system development and function [135], and can reduce pathogen colonization by competitive exclusion [11]. Imbalances in the community structure, termed dysbiosis [12], have also been linked with chronic health conditions such as diabetes [13], obesity [14], and inflammatory bowel disease [15]. These issues have been exacerbated by Western diets (low in dietary fiber) [136] and medicine (rich in prophylactic and therapeutic uses of broad spectrum antibiotics) [137], which lead to decreases in microbial diversity. Mitigating dysbiosis and stimulating DGM diversity through diet and limiting use of antibiotics have been intense areas of research for human health [138, 139] and livestock agriculture [35, 140]. Genetically engineering bacteria may provide clinical applications for intestinal health, such as *Listeria monocytogenes* producing tumor antigens to treat cancer [57], vaginal bacteria producing anti-viral peptides to treat HIV [58], and *Lactococcus lactis* producing anti-inflammatory cytokines to treat Crohn's disease [59], and proinsulin to treat diabetes [60].

Augmenting the metabolism of bacteria by genetic engineering is another approach for shaping DGM structure. Introducing enzymes with new functions or that operate at

faster rates could release more fermentable sugars [141] and reduce complications associated with feed viscosity and slow transit of ingesta [64, 65]. Elegant applications have also been developed using these principles. *B. theta* and *B. stercoris* have been engineered to utilize porphyran, a structural polysaccharide found in the cell walls of red algae [142]. The transfer of a functional porphyranolytic pathway provided access to a "privileged" nutrient, and selective feeding on porphyran enabled engraftment of the engineered strain into an established microbiota for at least fifteen days.

Genetic elements for tunable and tightly controlled expression of heterologous proteins in bacterial strains adapted for long-term host colonization, such as *Bacteroides* spp., are still required. Members of phylum Bacteroidetes are known for their ability to saccharify a wide variety of chemically complex polysaccharides from diverse biological sources, including terrestrial plants [3], seaweed [87], bacteria [88], fungi [89], and animals [90]. Glycan metabolism in these microorganisms are encoded within polysaccharide utilization loci (PULs) [143], which are clusters of co-regulated genes encoding transporters, carbohydrate active enzymes (CAZymes), and regulatory proteins specific for the metabolism of discrete substrates. The archetypical PUL is the starch utilization system (SUS), which was discovered in *B. theta* VPI-5482 [91]. The defining feature of a PUL is the presence of at least one sequential pair of *susC* and *susD* homologs, which encode a Ton-B dependent transporter and an outer membrane surface glycan binding protein, respectively [92]. PULs also contain CAZymes specific for dismantling the specific glycosidic linkages and carbohydrate modifications within the target substrate, and a regulatory protein that activates the PUL upon sensing the appropriate substrate. PULs can be activated by one of three regulators, including SusR-like, AraC hybrid two-component

systems (HTCSs), or extracytoplasmic function σ/anti- σ systems [103, 114]. Although these regulatory proteins operate through different mechanisms, they serve analogous functions within the cell, which are to detect signature products released from the substrate and induce PUL expression.

B. theta is an obligate anaerobic bacterium that colonizes the distal gut of mammals. It is detectable in 46% of healthy Western fecal microbiomes [73] and is one of the most studied members of Bacteroidetes [144]. *B. theta* is estimated to contain at least 88 PULs based upon the identification of *susC*/*susD* homologs [103]. Based on the genetic environment of these pairs, a subset of these PULs are predicted to target various sugars including homogalacturonan (HG) [145, 146], dextran (DX) [131], and arabinogalactan (AG) [103], which suggests that this bacterium is well-endowed as a generalist to adapt to fluctuating levels of dietary and host-derived glycans present in the large intestine. In this regard, selective metabolic strategies, and unique promoter responses are under-exploited opportunities for engineering artificial glycan responses in *B. theta*.

Bacteroides spp. have unique regulatory and conjugation elements [104-106], and genetic tools developed for other organisms generally do not function well in *B. theta.* In this regard, several *Bacteroides* spp. specific tools have been previously developed. The pExchange plasmid for homologous recombination has existed since 1991 [108]. Using this counter-selectable system, unmarked chromosomal deletions and insertions can be made with single base pair resolution [109, 110]. Alternatively, NBU2, a mobilizable transposon element [111] incorporated into a plasmid [110], can be used to quickly integrate the entire heterologous plasmid sequence into the chromosome of *B. theta*. More recently, constitutive promoter elements and ribosome binding sites (RBS) have been

screened and ranked to provide a wide-range of expression levels for heterologous proteins [109, 112]. Episomal and chromosomal inducible gene expression systems also exist, and can be regulated by anhydrotetracycline, a synthetic chemical inducer [109], or the natural polysaccharide α-mannan [113]. Each of these systems requires the addition of a single inducer to activate the expression of transgenes.

To increase the fidelity of glycan responsive episomal and chromosomal expression systems, I have altered the natural regulatory mechanisms of PULs by placing them behind DX and AG responsive promoters and expanded the available tool kit for engineering *B. theta* promoter elements. In addition, I have engineered novel dual glycan expression platforms that are dependent upon the addition of two chemically distinct glycans. This dual glycan expression system provides unprecedented regulatory control of transgene expression in *B. theta* and increases the activity of reporter enzymes (e.g. luciferase and an agarase) likely resulting from augmented levels of expression. Importantly, the majority of these genetic modifications are not deleterious to bacterial fitness *in vitro.* This study provides unique insights into promotor engineering for improved control of expression in *B. theta* and presents a platform for the future development of other dual glycan responsive expression systems.

2.2 Results

2.2.1 Construction of a pExchange based platform for targeted insertion and expression of transgenes

To develop a recombinant chromosomal expression system in *B. theta*, I sought to recommission the genome space of a characterized PUL. PUL75 (Fig. 2.1a) is a locus dedicated to the metabolism of HG, a pectic glycan comprised of galacturonic acid

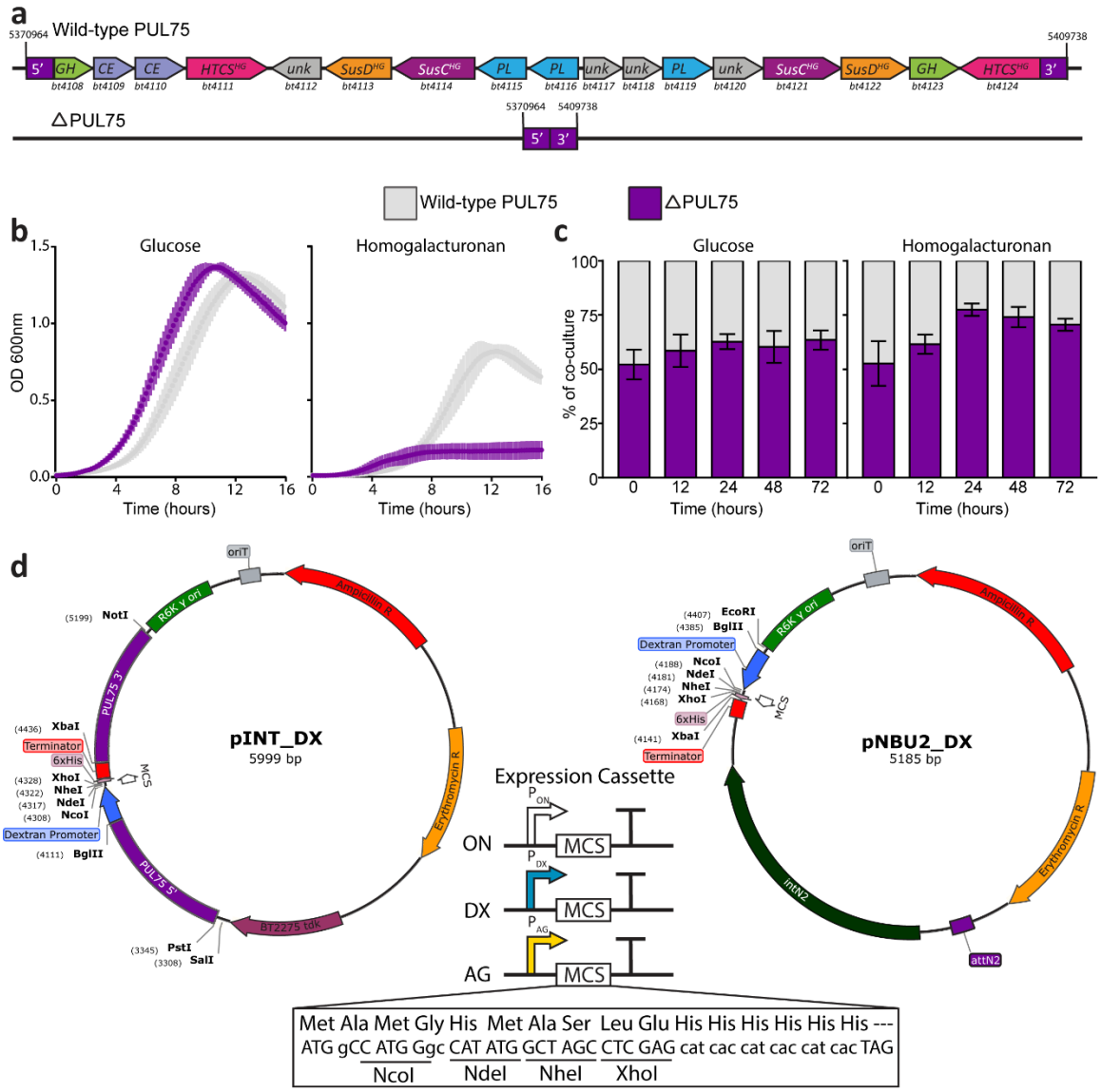


Figure 2.1: Target genomic locus and integrative expression vectors. (a) Schematic of *B.* theta genomic locus PUL75 containing genes required for homogalacturonan (HG) utilization (top). These genes are deleted in the *B. theta* ΔPUL75 strain (bottom). **(b)** Growth kinetics of wild-type (WT) *B. theta* (grey) and the ΔPUL75 strain (purple) in MM containing either glucose (GLC) or (HG) as sole carbon source. Growth was measured by OD_{600} readings over the course of 16 hours. **(c)** WT and ΔPUL75 were grown in co-culture over 72 hours on MM containing either GLC or HG as sole carbon source. The proportion of each strain was determined using strain specific primers and quantitative polymerase chain reaction **(d)** Plasmid maps of the pINT and pNBU2 vectors and reading frame of the multiple cloning site used in both vectors. The orientation of the DX, AG, and ON, expression cassettes are indicated. Values in (b) and (c) represent the average of three biological replicates and standard deviation.

(GalA) [145]. To access the impact of PUL75 deletion (*B. theta*-ΔPUL75) on bacterial growth, *B. theta* was cultured in minimal medium (MM) containing HG as a sole carbon source (Fig. 2.1b). Compared to the wild-type strain (WT), *B. theta*-ΔPUL75 was severely restricted in its ability to grow on HG. Surprisingly, when *B. theta*-ΔPUL75 was co-cultured with *B. theta* wild-type, the mutant strain grows to higher relative density than the wild-type (Fig. 2.1c), indicating that the *B. theta*-ΔPUL75 can utilize HG products, such as GalA, released into the medium by the WT strain. Sharing products within a community is referred to as "distributive metabolism" [143] and occurs with other glycans between unique strains and species. The apparent selective advantage for *B. theta*-ΔPUL75 in competition with *B. theta* wild-type suggests that there is a metabolic cost to maintaining the pathway and the deletion mutant is provided with unfettered access to fermentation products released from HG. This relationship may enable the selective proliferation of *B. theta*-ΔPUL75 in more complex communities and *in vivo* when supplemented with HG.

The vacated genome space in *B. theta*-ΔPUL75 (5371714…5408988) was targeted for chromosomal integration using a complementary PUL75-5' flank and PUL75-3' flank sequence inserted into the pExchange backbone (Fig. 2.1d). In addition, an MCS that is under the control of three unique promoters and flanked by an upstream RBS and downstream terminator was placed between the PUL75-5' and PUL75-3' sequences. This vector series, referred to as pINTEGRATE (pINT), enables precise and unmarked insertion of expression cassettes into the chromosome of *B. theta*. To complement this vector series, Darryl Jones developed a pNBU2 plasmid series containing similar inducible expression cassettes [110]. The pNBU2 vectors undergo single-cross over integration of the entire plasmid into the *B. theta* genome at one of two $tRNA^{ser}$ genes. This system enables faster

evaluation of chromosomal transgene function from a second genome locus, but the modified strain is left with contaminating plasmid DNA and selectable markers in its genome.

2.2.2 Validation of *B. theta* promoters for regulation of transgenes.

Three different promoters were selected for insertion into pINT and pNBU2 vectors based upon previous studies. These include a strong constitutive promoter P_{ON} (*bt1311* promoter and the rpiL* RBS [112]), and two inducible promoters: the DX responsive "P_{DX}", which drives expression of the PUL48 *susC*-like gene *bt3090* [131] (Fig. 2.2a); and the AG responsive "P_{AG}", which drives expression of the PUL5 *susC*-like gene *bt0268* [131] (Fig. 2.2b). DX and AG were selected because they induce some of the strongest glycan responsive expression in *B. theta*, they have distinct chemical structures (Fig. 2.2c), they are from diverse biological sources (DX is from a bacterial glycan; AG is a plant cell wall glycan), and they are induced by distinct regulatory mechanisms; PUL48 uses a SusR-like system (Fig. 2.2a) and PUL5 uses a HTCS system (Fig. 2.2b). The ability of each promoter to drive transgene expression was determined using NanoLuc, a luciferase reporter gene [112]. To evaluate proximity effects and differential expression platforms, NanoLuc-pINT and NanoLuc-pNBU2 vectors were conjugated into *B. theta*. NanoLuc assays determined that genome location did not result in a significant difference in luciferase activity for constitutive and glycan response promoters (Fig.S1). In contrast, glycan-specific patterns of induction were observed. When grown on glucose (GLC), DX, or AG, the constructs with the constitutive P_{ON} promoter consistently produced luminescence three orders of magnitude above the baseline produced by strains without a NanoLuc gene (Fig. 2.2d), and constructs with the inducible P_{DX} and P_{AG} responded to DX

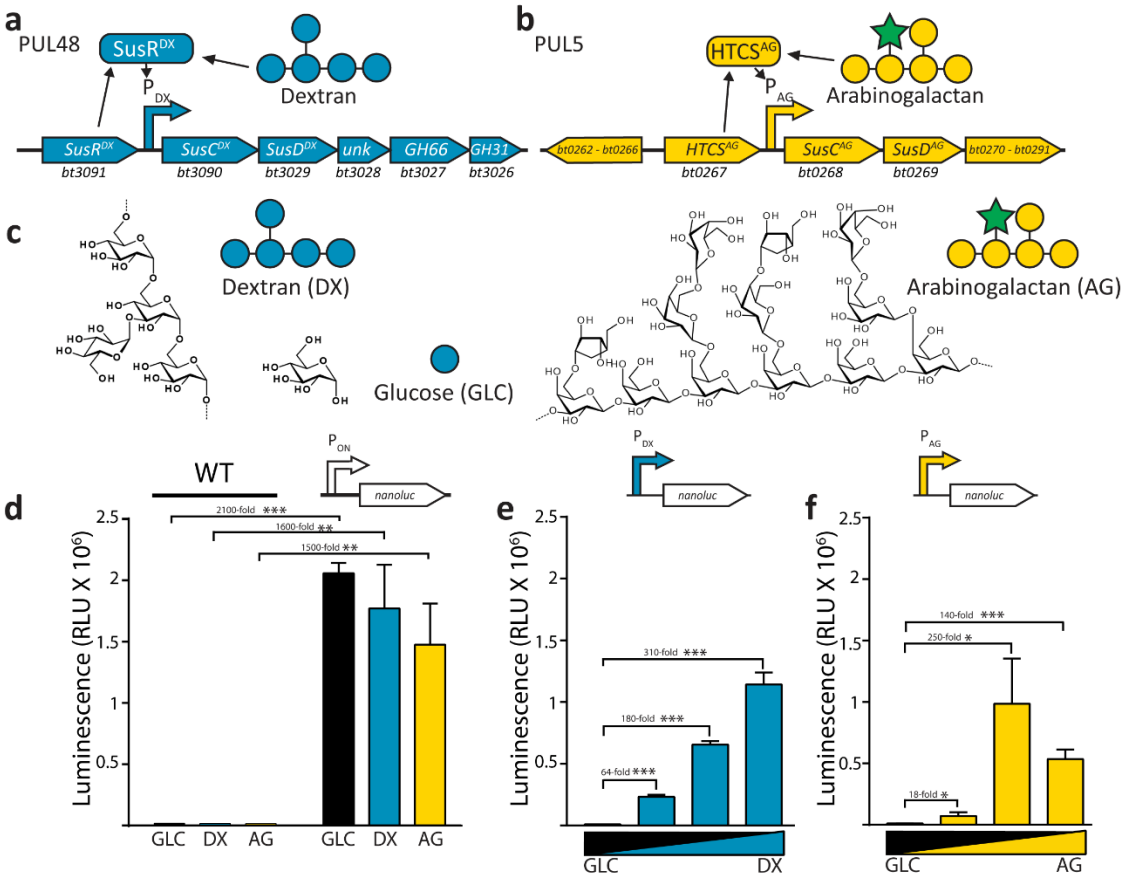


Figure 2.2: Inducible expression cassettes generate luminescence in a dose dependent manner under wild-type regulation in *Bacteroides thetaiotaomicron*. (a) Schematic of wild-type regulation by PUL48., P_{DX} is activated by the SusR-like protein $SusR^{DX}$ in the presence of DX. **(b)** Schematic of wild-type AG regulation., P_{AG} from PUL5 is activated by the HTCS protein $HTCS^{AG}$ in response to the presence of AG. **(c)** Schematic and chair representations of monomeric glucose (GLC; blue circle); the polymer dextran (DX) composed of a chain of GLC; and arabinogalactan (AG), compose dof a galactose (yellow circle), backbone and arabinose in side chains, (green star). **(d)** NanoLuc activity under constitutive expression by the P_{ON} promoter is measured post induction with GLC (black bar), DX (blue bar), and AG(yellow bar). Luminescence signal for NanoLuc production under the (e) DX or (f) AG inducible promoters in the presence of increasing concentrations of DX or AG respectively. Values in (d), (e), and (f) represent the average and standard error of the mean for three biological replicates. Asterisks represent levels of significance (*= p< 0.05, **=p<0.005, ***=p<0.0005).

(Fig. 2.2e) or AG (Fig. 2.2f) in a dose responsive fashion, respectively. Intriguingly, P_{AG} displayed higher activity in the presence of GLC and AG than with AG alone (Fig. 2.2f), suggesting that production of heterologous proteins can be tuned by adjusting glycan concentrations.

2.2.3 Engineering tunable expression of glycan-responsive PUL regulator proteins in *B. theta*

Next, I attempted to increase the heterologous production of NanoLuc in *B. theta*, by modifying the expression of the PUL48 and PUL5 regulatory proteins within the cell. Typically, regulatory proteins, such as HTCS and SusR-like homologs, are expressed at basal levels and are not induced along with CAZymes and transport proteins during glycan metabolism [131]. I reasoned that placing highly constitutive or inducible promoters upstream of the regulator gene in PUL5 and PUL48 may create feedback loops that result in higher levels of expression. Therefore, *B. theta* strains with a series of different promoter structures were generated (Table S1). These included: (1) a constitutive promoter, (2) a positive feedback loop (DX→DX; AG→AG), and (3) alternate forms of hybrid-induction (i.e. DX→AG and AG→DX). (1) The DX constitutive strain was created by placing the *bt1311* promoter (P_{ON}) [112] upstream of *bt3091*, the PUL48 SusR-like gene ($SusR^{DX}$) to generate the strain *B. theta*-$P_{ON}SusR^{DX}$. The AG constitutive strain was created by placing P_{ON}, upstream of *bt0267*, the PUL5 HTCS gene ($HTCS^{AG}$) to generate *B. theta*-$P_{ON}HTCS^{AG}$. (2) The positive feedback systems were created by placing P_{DX} upstream of $SusR^{DX}$ to create *B.theta*-$P_{DX}SusR^{DX}$; and P_{AG} [112] upstream of $HTCS^{AG}$ to create *B.theta*-$P_{AG}HTCS^{AG}$. (3) The hybrid-induction systems were created by inserting P_{AG} in front of $SusR^{DX}$ and P_{DX} in front of $HTCS^{AG}$ to create *B. theta*-$P_{AG}SusR^{DX}$ and *B. theta*-

P_{DX}HTCS^{AG}, respectively. In addition, a third hybrid strain was created, *B. theta*-P_{AG}SusR^{DX}+P_{DX}HTCS^{AG}, which contained both hybrid regulatory modifications.

To determine the effect of modifications to the promoter regions of the regulatory genes in PUL5 and PUL48, the growth profiles of each *B. theta* strain were determined (Fig. 2.3a). Strains with constitutive regulation (*B. theta*-P_{ON}SusR^{DX} and *B. theta*-P_{ON}HTCS^{AG}) and positive feedback regulation both performed similar to the wild-type. Each hybrid strain displayed a growth defect when cultured solely on the glycan sensed by the regulatory protein (DX for P_{AG}SusR^{DX}, or AG for P_{DX}HTCS^{AG}, Fig. 2.3a). This defect was mitigated when the strains were cultured on a mixture of DX and AG. This phenotype may be a result of the hybrid strains having a lower basal level of expression when regulated by a glycan-responsive promoter. On the mixture, the promoters are fully activated and lead to sufficient regulatory protein production to effectively activate their PULs preventing a growth defect.

To examine the effects of the PUL5 and PUL48 modifications to the expression of a transgene, NanoLuc activity was measured in the engineered *B. theta* strains. In the DX systems, constitutive and positive feedback regulation increased expression three-fold over wild-type in both the pINT and pNBU2 platforms when cultured on DX. The hybrid strain, *B.theta*-P_{AG}SusR^{DX}, displayed the lowest NanoLuc signal on DX but similar levels on the DX-AG mixture (Fig. 2.3b), which is consistent with what was observed in the growth profiles (Fig. 2.3a). The AG-responsive systems displayed similar patterns, with a lower increase in NanoLuc activity in the P_{ON} and P_{AG} systems (Fig. 2.3c).

In contrast to the strains with single modifications, the strain with both hybrid regulator systems displayed noticeable increases in NanoLuc activity (Fig. 2.3d). The strain

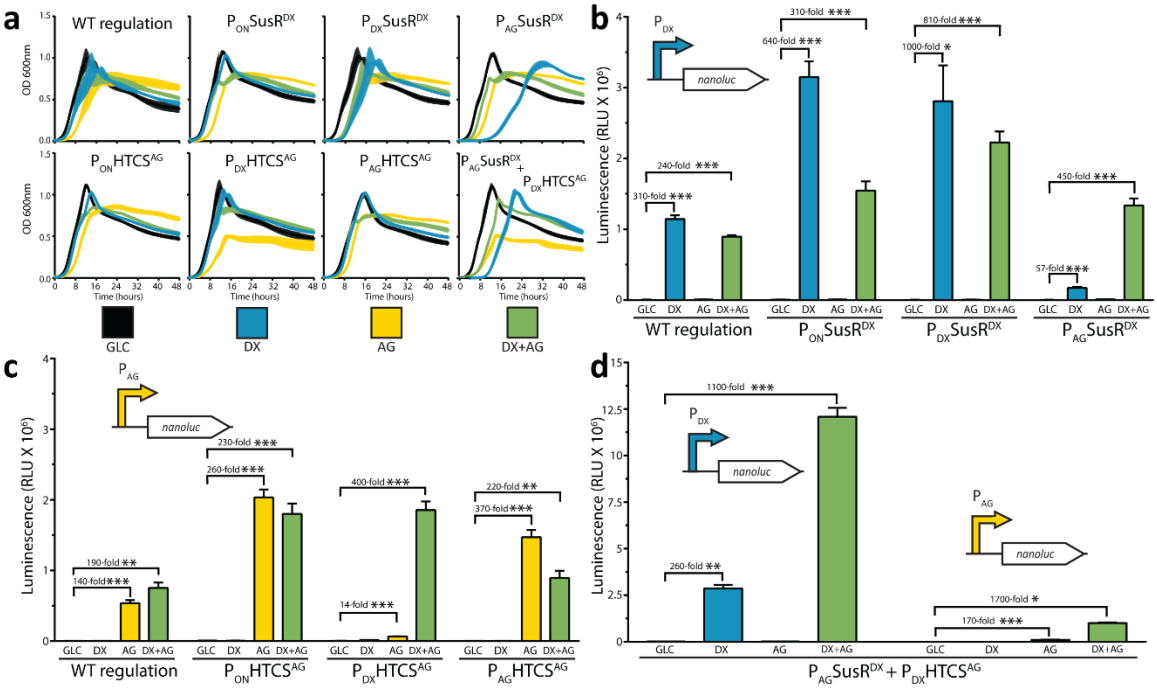


Figure 2.3: Regulatory engineering changes growth kinetics, levels of gene activity, and gene induction requirements. (a) Growth kinetics of WT *B. theta* and engineered regulatory strains on GLC (black), DX (blue), AG (yellow), and a mixture of DX and AG (green). **(b)** Luminescence detected from DX-responsive NanoLuc in strains with engineered *SusR*^{DX} when grown on GLC (black), DX (blue), AG (yellow), and the mixture of DX and AG (green). **(c)** Luminescence detected from AG-responsive NanoLuc in strains with engineered *HTCS*^{AG} when grown on GLC (black), DX (blue), AG (yellow), and the mixture of DX and AG (green). **(d)** Luminescence detected from both DX- and AG-responsive NanoLuc in a strain with engineered *SusR*^{DX} and *HTCS*^{AG} when grown on GLC (black), DX (blue), AG (yellow), and the mixture of DX and AG (green). Error bars in (a) represent standard deviation and error bars in (b), (c) and (d) represent standard error of the mean for three biological replicates. Asterisks represent levels of significance (*= p< 0.05, **=p<0.005, ***=p<0.0005).

examined produced the highest luminescence values for chromosomal NanoLuc when the transgene was placed behind P_{DX} and the strain was cultured on mixed sugars. Though it did not reach the same absolute value for reporter activity, NanoLuc behind P_{AG} had the greatest relative difference compared to baseline luminescence on GLC, owing to the high baseline of P_{DX}NanoLuc.

2.2.4 Augmenting transgene expression using a *B. theta* episomal expression vector

Plasmid-based expression systems hold several advantages over chromosomal expression systems, including efficient transformation, increased gene copy numbers, and shortened timelines due to lack of requirement for double cross-over. Several expression vectors compatible with *Bacteroides* spp. have been reported [109, 113]. To compare the potency of the DX and AG regulated expression cassettes developed here with previously described vector expression systems, I took advantage of the the pEP series of plasmids for *B. theta* (created by Darryl Jones) (Fig 2.4a). The pEP series contains *mobA* and *repA* genes derived from the *Bacteroides* plasmid pBI143 DNA [147] to enable the vector to exist as a plasmid in *B. theta*. Vectors are maintained by erythromycin selection and are equipped with an MCS for gene insertion and swappable promoter for changing glycan responsiveness. As a proof of concept, the DX promotor was ligated into the pEP vector upstream of NanoLuc and expression was induced by culturing the cells with GLC, DX, AG or a DX/AG mix. Combining plasmid-based expression with reciprocal modifications to hybrid promoter regulation amplified effects for selective expression when compared with chromosomal NanoLuc expression (Fig 2.4c). The *B.theta*-P_{DX}SusR^{DX} and *B.theta*-P_{ON}SusR^{DX} strains transformed with a P_{DX}-pEP vector had similar levels of luminescence (60-75 x 10^6 RLU) on DX and DX-AG mix,

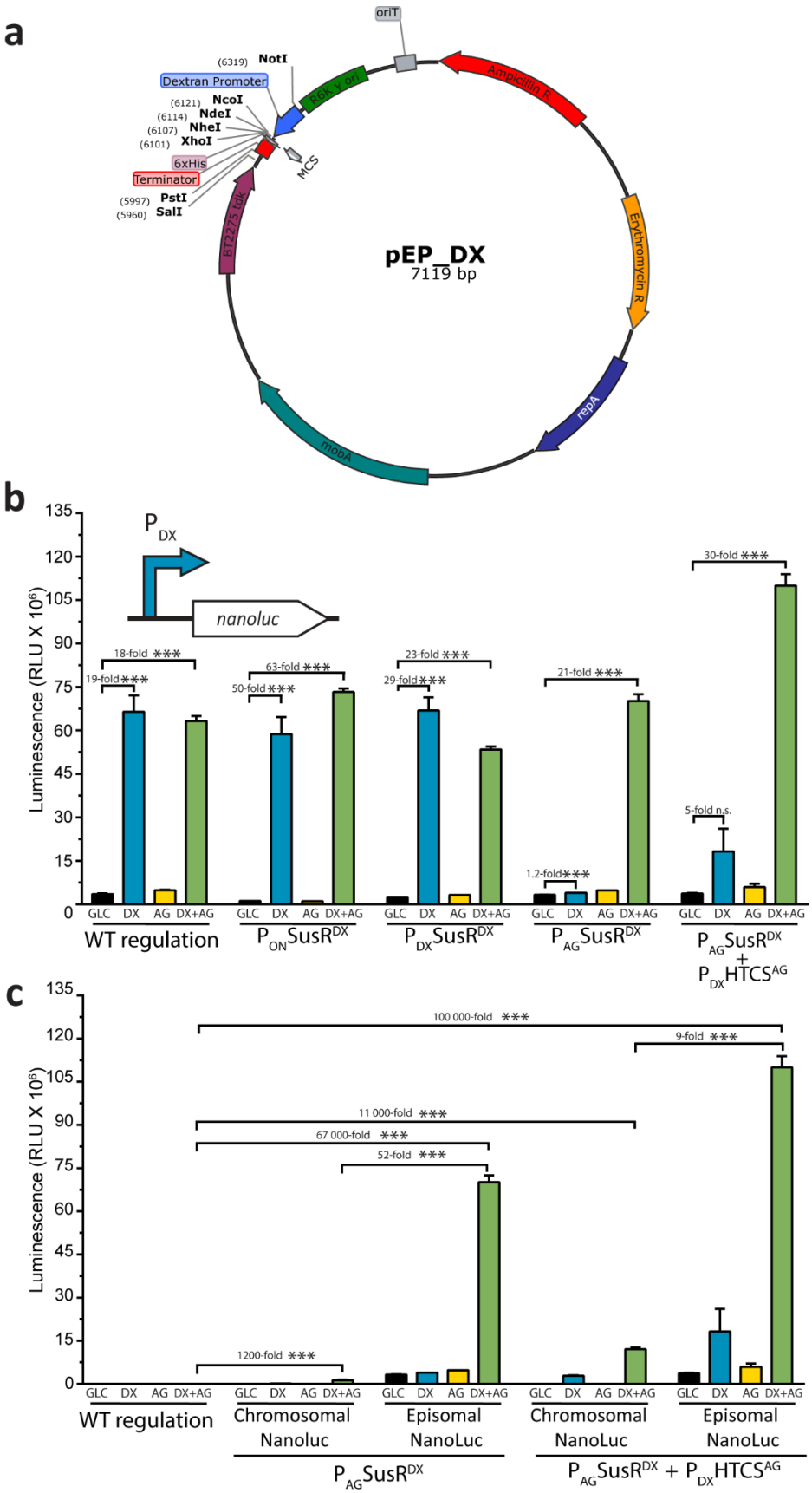

a
oriT
(6319) NotI
Dextran Promoter
(6121) NcoI
(6114) NdeI
(6107) NheI
(6101) XhoI
6xHis
Terminator
(5997) PstI
(5960) SalI
RSF 1 ori
Ampicillin R
MCS
BT2275 tdk
Erythromycin R
pEP_DX
7119 bp
repA
mobA
b
P DX
nanoluc
Luminescence (RLU X 10⁶)
18-fold ***
19-fold***
63-fold ***
50-fold***
23-fold ***
29-fold***
21-fold ***
1.2-fold***
30-fold ***
5-fold n.s.
GLC DX AG DX+AG
WT regulation
P ONSusR DX
P DXSusR DX
P AGSusR DX
P AGSusR DX + P DXHTCS AG
c
100 000-fold ***
9-fold ***
11 000-fold ***
67 000-fold ***
52-fold ***
1200-fold ***
Chromosomal NanoLuc
Episomal NanoLuc
P AGSusR DX
P AGSusR DX + P DXHTCS AG

Figure 2.4: Episomal transgene expression increases gene expression compared to chromosomal equivalent. (a) Vector map of the extra-chromosomal expression pEP_DX vector. **(b)** Luminescence signal detected from wild-type (WT), $P_{ON}SusR^{DX}$, $P_{DX}SusR^{DX}$, $P_{AG}SusR^{DX}$, and $P_{DX}HTCS^{AG}$+ $P_{AG}SusR^{DX}$ strains containing the NanoLuc under the control of the episomal pEP_DX promotor, in the presence of GLC (black bars), DX (blue bars), AG (yellow bars) or the DX and AG mixture (green bars). **(c)** Comparison of luminescence from engineered regulatory strains when NanoLuc is encoded in a plasmid or in the chromosome. Error bars represent standard error of the mean for three biological replicates. Asterisks represent levels of significance (*= $p< 0.05$, **=$p<0.005$, ***=$p<0.0005$).

suggesting DX levels were sufficient to maximize expression in both systems. In comparison, the *B.theta*-$P_{AG}SusR^{DX}$ strain only displayed activation to similar levels when grown on a mixture of DX and AG glycans. This result suggests that basal levels of $SusR^{DX}$ expression when grown on pure DX was not enough to drive plasmid expression but was complemented by increasing $HTCS^{AG}$ activity by adding AG to the medium (Fig. 2.4b). A ~2-fold amplification of NanoLuc activity over the other strains was observed in the *B. theta*-$P_{AG}SusR^{DX}$+$P_{DX}HTCS^{AG}$ strain when grown on the AG-DX mix. This suggests that increasing expression of both regulators results in higher levels of expression heterologous genes using the pEP-plasmid system. *B. theta*-$P_{AG}SusR^{DX}$+$P_{DX}HTCS^{AG}$ transformed with P_{DX}-pEP had the highest levels of reporter activity in this study. Use of the plasmid with this strain culminated in a 9-fold increase over its chromosomal activity and a 100,000-fold increase over the background luminescence of *B. theta* WT (Fig. 2.4c).

2.2.5 Heterologous production of an agarase in modified *B. theta* strains

To determine whether or not the dual-glycan expression systems can be used to produce heterologous enzymes, I used the pINT vectors to integrate a family 16 glycoside hydrolase from *B. uniformis* NP1 (*Bu*GH16) into the *B.theta*-$SusR^{DX}$ strain series. *Bu*GH16 is an endo-β-agarase that cleaves agarose to produce neoagarotetraose as a terminal product [148]. Wild-type *B. theta* does not possess this catalytic activity. *Bu*GH16, possessing an N-terminal outer membrane anchoring tag, was cloned into the pINT- P_{DX} vector and integrated into the *B. theta*-$SusR^{DX}$ strain. The outer membrane tag enabled whole-cell agarolysis assays to be performed. For the *B. theta*-$P_{ON}SusR^{DX}$ and *B. theta*-$P_{DX}SusR^{DX}$ strains, transgenic *Bu*GH16 protein was detected (Fig. 2.5a) and agarase activity was

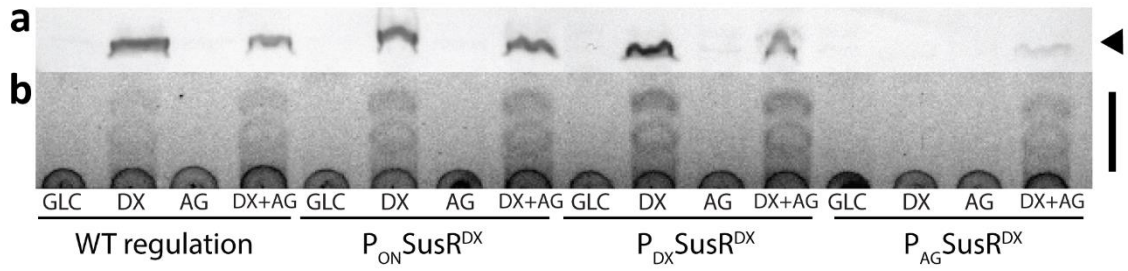


Figure 2.5: The heterologous BuGH16 agarase can be produced in engineered regulatory strains. (a) pINT integrated hexa-histidine tagged *Bu*GH16 (black arrow) from induced engineered *SusR*^{DX} strains detected by Western blot using anti-6-his antibodies conjugated to horse radish peroxidase. **(b)** Thin-layer chromatography of culture supernatant from induced engineered strains incubated with agarose shows the presence of neoagarooligosaccharides (black bar) generated from agarose by active *Bu*GH16.

observed in all conditions containing DX (Fig. 2.5b). Production of *Bu*GH16 in the $P_{AG}SusR^{DX}$ dual glycan induction strain only displayed *Bu*GH16 activity when the cultures were treated with both DX and AG. This pattern is similar to NanoLuc induction from the dual glycan pINT system (Fig.2.3b) and pEP vector system (Fig.2.4b). Taken together, these results suggest that engineering HTCS and SusR-like regulator proteins to respond to discrete glycan inducers in chromosomal and plasmid-based systems can improve the regulatory fidelity of heterologous enzyme production in *B. theta*.

2.3 Discussion

Engineering probiotic bacteria to express heterologous proteins is a promising approach for improving host health and animal production. Previous studies have developed expression cassettes that are constitutive [112], or respond to common glycans [112, 113] or synthetic chemicals [109]. This palette of genetic parts has been expanded with pINT and pNBU2-based gene expression systems. pINT constructs specifically integrate heterologous gene expression cassettes into the vacated genome space of *B. theta*-ΔPUL75, a deletion that severely restricts growth on HG but does not appear to compromise fitness when co-cultured with the *B. theta* wild-type strain (Fig. 2.1a-c). Sharing polysaccharide breakdown products within a community is referred to as "distributive metabolism" [143] and occurs with other glycans between unique strains and species such as *B. adolescentis* and *B. ovatus*. *B. adolescentis* cannot grow on xylans by itself, but co-culture with the xylan degrading *B. ovatus* can enable *B. adolescentis* to grow on some xylans [149]. Research suggests that simple, easy to digest glycans are shared among community members, but that structurally complex glycans are metabolised "selfishly" by *Bacteroides* that possess the enzymes required to dismantle them [150].

Since HG is a relatively simple polysaccharide, "public goods" [150], i.e. accessible monosaccharides and simple oligosaccharides, may be available for other bacteria to metabolize. Taken together with the results presented here involving the growth of the ΔPUL75 in co-culture, the concept of a distributive metabolism suggests that deletion of PUL75 and use of the pINT integration vector may not compromise the fitness of *B. theta in vivo*.

The integration vectors developed here have been designed to enable rapid subcloning of target genes, and for tailoring promotor selectivity, and altering the targeted integration site within the genome (Fig. 2.1d). Three different validated promoters were inserted into these expression cassettes, including P_{ON} (*bt1311*), a strong constitutive promoter; and two selective glycan responsive elements, P_{DX} (*bt3090*) and P_{AG} (*bt0267*), that respond to structurally distinct glycans (Fig. 2.2d-f). Constitutive expression of NanoLuc driven by P_{ON} was 1,500 to 2,000-fold greater than baseline auto-luminescence, on three different carbon sources (Fig. 2.2d). These results are consistent with previous findings [112], and underpin that this promoter operates independent of glycan-specific regulatory networks. Somewhat surprisingly, DX and AG induction of NanoLuc resulted in different expression profiles. DX displays a conventional dose-dependent relationship when mixed with glucose (Fig. 2.2e); whereas, the AG-responsive promoter resulted in optimal NanoLuc activity at 50 : 50 (GLC : AG), and there was an inhibitory effect observed when treated solely with AG (Fig. 2.2f). The AG-responsive promoter used in this study is from PUL5 and was chosen due to high levels of expression [103] and proven ability to drive transgene expression [112], but PUL5 is one of two AG responsive PULs. *B. theta* PUL65 has also been shown to be involved in AG utilization [103]. When the

PUL65 HTCS is deleted and the PUL5 HTCS is left intact, the mutant strain grows to a higher density (150%) on AG than the WT does *in vitro* [103], indicating that PUL65 partially represses PUL5. Additionally, the presence of GLC is known to downregulate expression of genes from PUL65 [131]. The results shown here would be explained if the repressive effect of PUL65 on PUL5 is alleviated when PUL65 is in turn repressed by the presence of GLC. Nevertheless, these effects can be used to tailor the dose-dependent expression profiles of heterologous proteins.

To explore the potential of augmenting glycan responsive expression levels and providing tighter control, the promoter elements of $SusR^{DX}$ and $HTCS^{AG}$ were engineered to be constitutively expressed under control of P_{ON}, and in response to DX and AG (Fig. 2.3). For both regulators, constitutive expression and positive feedback had the largest effect and boosted NanoLuc activity three-fold and two-fold above WT for DX (Fig. 2.3b) and AG (Fig. 2.3c), respectively. This may be due to increases in the pool of regulatory protein available to sense sugar and upregulate target genes. Supporting this, it has previously been observed that increasing the copy number of the canonical *sus* regulatory gene by incorporating it into a plasmid leads to increases in reporter gene activity [127]. Interestingly, hybrid-regulation ($P_{AG}SusR^{DX}$ and $P_{DX}HTCS^{AG}$) required the presence of both DX and AG for maximal expression. These results suggest that the regulators are not induced on the single sugars, and when both DX and AG are supplied the promoters become activated, leading to higher levels of downstream gene expression, which could be explained if promoter activation results in higher levels of regulatory proteins in the cell. This effect was greatly increased when hybrid regulation of both $SusR^{DX}$ and $HTCS^{AG}$ were introduced into the same strain (*B. theta*-$P_{AG}SusR^{DX}$+$P_{DX}HTCS^{AG}$) and NanoLuc was

induced by DX (Fig. 2.3d). However, similar expression levels were not observed in this strain when NanoLuc was under AG regulation. This suggests a possible repression of AG associated genes by metabolism of DX.

The observation that *B. theta* grows to a higher density on DX than it does on AG (Fig. 2.3a), suggests that DX metabolism is preferential to AG. Previous literature examining growth kinetics on different pairs of sugars has also identified DX being preferred over AG [151]. The observation that this hybrid-system requires two chemically distinct glycans for maximum expression represents a novel approach for regulating transgene expression in *B. theta*. Other expression systems requiring two unique molecules for induction have been described, including the single chain tetracycline repressor in *E. coli* that requires both tetracycline and 4-dedimethylamino-anhydrotetracycline [152], or the classic *lac* operon that requires both lactose and cyclic adenosine monophosphate bound catabolite activator protein for full expression [153, 154]. The use of a dual glycan regulatory mechanism is an important development that does not require the addition of broad spectrum antibiotics, which can disrupt the delicate balance of the DGM.

In order to determine if the dual glycan expression effect is conserved with episomal expression, I transformed the pEP vector (Fig. 2.4a) containing NanoLuc behind P_{DX} into the $SusR^{DX}$ strain. The hybrid regulation strain *B. theta*-$P_{AG}SusR^{DX}$ produced a ~20-fold increase on the mixed glycans compared to expression on pure glycans (Fig. 2.4b). In the *B. theta*-$P_{AG}SusR^{DX}$+$P_{DX}HTCS^{AG}$ strain this effect was further amplified, which is similar to what was observed with the pINT expression platforms (Fig. 2.3d). The higher luminescence activity observed using pEP-expression is attributed to the potential for higher copy numbers of the expression cassettes within the cell. These vectors replicate

within the cell independently of the genome, and the pBI143 plasmid, from which the mobilization and replication genes in pEP originate, has been estimated to be maintained at a copy number of 20 per cell [147]. SusR-like regulatory proteins are integral membrane proteins, which likely has implications for expression of genes located within the chromosome. It is currently unknown how membrane bound regulator proteins, such as SusRs and HTCSs, come into contact with their DNA targets, but it is likely that the chromosome must diffuse towards the regulator protein in the membrane. Episomal expression vectors would then be able to overcome this deficiency by more readily diffusing to the cytoplasmic DNA-binding domain of BT3091.

To demonstrate that the dual-glycan expression system is effective for other transgenes, I replaced NanoLuc with *Bu*GH16. This enzyme cleaves the D-galactose-(β1,4)-3,6-anhydro-L-galactose bond in agarose, which is a common structural polysaccharide in red algae and used as a thickening agent in the food industry. I compared *Bu*GH16 activity with the P_{DX} promoter in *B. theta* strains with modified $SusR^{DX}$. *Bu*GH16 production (Fig. 2.5a) and agarase activity (Fig. 2.5b) was observed as expected on DX and the mixture in the control and $P_{ON}SusR^{DX}$ and $P_{DX}SusR^{DX}$ strains; however, dual-glycan regulation was only observed when $SusR^{DX}$ was controlled by the P_{AG} promoter. This result underpins that hybrid engineering of glycan-responsive regulatory proteins can provide unprecedented control of transgene expression in *B. theta*.

2.4 Conclusion

Novel gene expression systems with unique induction requirements have been created in *B. theta* by modifying the promoter elements of regulatory genes from two distinct PULs. These systems have been shown to increase expression of transgenes. Most

significantly, engineering promoters that are activated by chemically distant glycans (i.e. DX and AG), resulted in dual glycan expression systems that require both glycans to maximize gene expression. These trends are reproducible across different vectors systems and reporter genes. The systems detailed in this study produce increases in reporter activity in line with other reports in the literature which achieve reporter gene activity across dynamic ranges from 2,000-fold [113] to 100,000-fold [109]. These systems will help pave the way for more sophisticated glycan-directed regulatory engineering and rationally designed bacterial tools that respond to dietary glycans.

2.5 Methods

2.5.1 Vector construction

To generate the pINTEGRATE vectors (pINT_DX) a 750 bp region upstream of PUL75 (PUL75 5') and the promoter and initiator codon of *bt3090* were amplified from *B. theta* genomic DNA (primers in Table 2.1) and assembled by overlap PCR to contain a BglII cut site between the two fragments (to facilitate future promoter variants of pINT) with a PstI cut site on the 5' end of the insert and an NcoI site on the 3' end. Concurrently, the terminator from pNBU2 [110] was amplified with a 5' forward primer containing an MCS with 6xHis tag and a 3' reverse XbaI primer. After digest with PstI/NcoI and NcoI/XbaI respectively, these products were simultaneously ligated into pExchange-tdk (pEX-tdk) [110] digested with PstI and XbaI. The resultant construct was then digested with XbaI and ligated to a 750 bp PCR fragment corresponding to the downstream region of PUL75 (PUL75 3') that was amplified with an XbaI cut site on the 5' end and a NheI (XbaI isocaudomer) cut site on the 3' end. The orientation of the insert was confirmed by restriction digest and the entire cloned region was confirmed by sequencing. pINT_AG and

Table 2.1: Primers used in this study.

Vector creation primers	**Sequence (5' - 3')**
pET28_MCS_In_For	TCGCGGATCCGAATTCGAGC
pET28_MCS_In_Rev	GCTCGAATTCGGATCCGCGA
Pul75_5'_For_XhoI	TTCTCGAGGAATTCAATGTAGCAAATGATAG
Pul75_5'_Rev	GGTTTATTGTTTTAAGGATTATACATGCTA
PUL75 3' XbaI For	TATATATCTAGAAATATCGGTGTTTTTTGTTAGAAC
PUL75 3' NheI Rev	TATATAGCTAGCTTTCAGTCCTTCCAGTTCTG
pBI143_KpnI_For	TTTGGTACCTGATTTTGTGCTGTTTTGGGGG
pBI143_KpnI_Rev	AAAGGTACCCACAACCGTTTTCCGAACCC
NanoLuc_Fwd	TTAAGCTAGCATGGTTTTTACTCTGGAAGATTTTG
NanoLuc_rev_XhoI_HIS	TTAACTCGAGTGCCAGGATGCGCTCGC
Bt_4116_For	CGTTGGCTGTTTCCCATTCG
GH16_XbaI_Rev	TTTTTCTAGAGAAATCATTTCTCCAAATAGTTCC
AG+4116-GH16_stitch_FWD	CTTTATAAATCTAATAGTATGAATAAAACTTTTTTAG
AG+4116-GH16_stitch_REV	CTAAAAAAGTTTTATTCATACTATTAGATTTATAAAG
Promoter engineering primers	**Sequence (5' - 3')**
BT267_Pro_insert_5'_BamHi_For	AATTAAGGATCCGAATTTGGAGTGAATGCTTTG
BT267_Pro_insert_5'_BglII_Rev	TATATAAGATCTAAATCACTTAAAATATATTAATTTAAAACTACC
BT267_Pro_insert_3'_XbaI_Rev	TATATATCTAGACGCTTTCCCATGTTAAAATC
Ter_PRO_0268_BglII_For	TATATAAGATCTTGAACTGCACTTGCTTTGATAATTAATGATAAACAATCTAAAAGCACTCTAATCGTTATCGGAGTGCTTTTAGATTACTAATAAAATACACAAGTACGCGTCTTAATG
Ter_PRO_1311_BglII_For	TATATAAGATCTTGAACTGCACTTGCTTTGATAATTAATGATAAACAATCTAAAAGCACTCTAATCGTTATCGGAGTGCTTTTAGATTACTAATGATCTGGAAGAAGCAATGAAAG
Ter_PRO_3090_BglII_For	TATATAAGATCTTGAACTGCACTTGCTTTGATAATTAATGATAAACAATCTAAAAGCACTCTAATCGTTATCGGAGTGCTTTTAGATTACTAAGAAATCCACTACTTTTTTAGCACG
PRO0268-BT0267 stitch For	CTAATACTTATTTGGTCTAAATAACTTTATAAATCTAATAGTATGAATAGTGATAGCTATCTTTATTTTAATATTG
BT0267-PRO0268 stitch Rev	CAATATTAAAATAAAGATAGCTATCACTATTCATACTATTAGATTTATAAAGTTATTTAGACCAAATAAGTATTAG
PRO1311-BT0267 stitch For	CCCGCATTTTAAAATAAAATAAATTATTTATTTAATTAAACGAATATGAATAGTGATAGCTATCTTTATTTTAATATTG
BT0267-PRO1311 stitch Rev	CAATATTAAAATAAAGATAGCTATCACTATTCATATTCGTTTAATTAAATAAATAATTTATTTTATTTTAAAATGCGGG
PRO3090-BT0267 stitch For	GAAAACAAAAAGTAATCCTAATATTAACTTTAAATTGATGTACATGAATAGTGATAGCTATCTTTATTTTAATATTG
BT0267-PRO3090 stitch Rev	CAATATTAAAATAAAGATAGCTATCACTATTCATGTACATCAATTTAAAGTTAATATTAGGATTACTTTTTGTTTTC
BamHI-3092Flank-For	ATTATTGGATCCGGCCATATGTATGATGTACACAATTATCC
3091Flank-XbaI-Rev	ATATATTCTAGAGAAAGTGTTTCTGCCCTTTCTTGTC
Sequencing primers	**Sequence (5' - 3')**
PUL75-Ext-Alt_F	CCAATCACGATACCCAGGAATAA
PUL75-Ext-Alt_R	CCACGAATCAACTGTAACTGATG
pEx-MobA_Seq_For	AAAGCTTTAAAACATGCGAGAA
pEx-MobA_Seq_Rev	GGAAGTTGAATGTCCGGTGA
BT0267_Pro_Insert_Int_SeqF	GTTGTGCTAATTGTCAGCAATGTC
BT0267_Pro_Insert_Int_SeqR	CAATTATTTCCGTGTGCCTTTGCG
3091ProInsert_Seq_For	CAACGTGGGTAATCTTCAATGAAGGATG
3091ProInsert_Seq_Rev	CATTTCAGATAGTTGTATGCACGGTC
3091ProInsert_internal_Seq_For	GGGAAACTAAAAAGATGAGCTTTGATTTCG
3091ProInsert_internal_Seq_Rev	CATTTCTTGCGCAAAGCAAACAGGG
qPCR primers	**Sequence (5' - 3')**
Q-deltaPUL75_F4	CAAAATTAGTCGTTTCCGTGC
Q-deltaPUL75_R4	ATCCGTTCAATGTAGCGTCC
Q-intactPUL75_F4	ATGTGGGGAATCACGTCTAC
Q-intactPUL75_R4	CACGAAGCAGGATATTGCAG
qBT2161_F	AAAGAAGACATCGTAAACTTGGG
qBT2161_R	GAAACGCCTTCCAACTTAGC

pINT_ON were created by replacing the BglII to NcoI fragment in pINT_DX with the promoter and initiator codon from *bt0269* genomic DNA and *bt1311* and the rpiL* RBS from pAT593 [112] respectively.

The single-step integrative vector pNBU2_DX was generated by amplifying the Promoter-MCS fragment of pINT_DX and ligating it into a variant of pNBU2 amplified to create BglII and XbaI cut sites between the R6K origin of replication and the terminator of pNBU2 [112]. The promoter and MCS from pINT_AG and pINT_ON were amplified with BglII and XbaI respectively and cloned into the reciprocal sites of pNBU2_DX to create pNBU2_AG and pNBU2_ON.

To create the episomal expression constructs (Table 2.2) pEP_DX, pEP_AG, and pEP_ON, the promoter to terminator fragments from pINT_DX, pINT_AG, and pINT_ON were amplified with NotI and PstI restriction enzyme sites respectively and cloned into a variant of pEx-tdk containing the *repA* and *mobA* genes (Genbank AAB39963.1 and AAB39964.1 respectively) derived from the *Bacteroides* plasmid pBI143 DNA [147] (generous gift of Dr. Jeffery Smith) , which allow the plasmid to exist in *B.theta* extra-chromosomally.

The reporter genes NanoLuc [155] and *Bu*GH16 (*np1_8* [148]) were fused to respective promoters by overlapping PCR and cloned into the BglII/XhoI or NotI/XhoI sites of the respective vectors to create reporter constructs. In order to target *Bu*GH16 to the cell surface and to avoid interference with the translocation of endogenous proteins, the native N-terminal signal peptide of *Bu*GH16 (nucleotides 1-67) was replaced with that of the putative polysaccharide lyase family 6 (PL6) (*bt4116*; nucleotides 1-63) from PUL75, predicted to be expressed on the surface of *B. theta* [156, 157].

Table 2.2: Plasmids and Strains used in this study

Plasmid	Notes	Source
pExchange_tdk	parent vector to pINT and pEP.	PMID_18611383
pINT_ON	Base vector with no reporter gene.	This study
pINT_DX	Base vector with no reporter gene.	This study
pINT_AG	Base vector with no reporter gene.	This study
pNBU2_ON	Base vector with no reporter gene.	This study
pNBU2_DX	Base vector with no reporter gene.	This study
pNBU2_AG	Base vector with no reporter gene.	This study
pEP_DX	Base vector with no reporter gene.	This study
pINT_ON_nanoluc	Contains NanoLuc under PON.	This study
pINT_DX_nanoluc	Contains NanoLuc under PDX.	This study
pINT_AG_nanoluc	Contains NanoLuc under PAG.	This study
pNBU2_ON_nanoluc	Contains NanoLuc under PON.	This study
pNBU2_DX_nanoluc	Contains NanoLuc under PDX.	This study
pNBU2_AG_nanoluc	Contains NanoLuc under PAG.	This study
pEP_DX_nanoluc	Contains NanoLuc under PDX.	This study
pINT_DX_BuGH16	Contains BuGH16 under PDX.	This study
pExchange_tdk_ONsusR	Used to change promoter of b*t3091* to PON.	This study
pExchange_tdk_DXsusR	Used to change promoter of b*t3091* to PDX.	This study
pExchange_tdk_AGsusR	Used to change promoter of b*t3091* to PAG.	This study
pExchange_tdk_ONHTCS	Used to change promoter of bt0267 to PON	This study
pExchange_tdk_DXHTCS	Used to change promoter of bt0267 to PDX	This study
pExchange_tdk_AGHTCS	Used to change promoter of bt0267 to PAG	This study
Strain	**Notes**	**Source**
E. coli S17-1 lambda pir	*recA pro hsdR RP4-2-Tc::Mu-Km::Tn7 λ-pir*	PMID_6340113
B. thetaiotaomicron VPI-5482	Wild-type.	ATCC 29148
B. thetaiotaomicron VPI-5482 Δtdk	Deletion of bt2275 thymidine kinase to allow for counter selection by FUDR.	PMID_18611383
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75	Deletion of bt4108-bt4124.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pON-nanoluc	NanoLuc inserted into empty PUL75 locus.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pDX-nanoluc	NanoLuc inserted into empty PUL75 locus.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pAG-nanoluc	NanoLuc inserted into empty PUL75 locus.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pON-nanoluc	NanoLuc integration with pNBU2 vector.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pDX-nanoluc	NanoLuc integration with pNBU2 vector.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pAG-nanoluc	NanoLuc integration with pNBU2 vector.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pbt3091::pON	Dextran regulator under constitutive promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pbt3091::pDX	Dextran regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pbt3091::pAG	Dextran regulator under arabinogalactan promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pbt0267::pON	Arabinogalactan regulator under constitutive promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pbt0267::pDX	Arabinogalactan regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pbt0267::pAG	Arabinogalactan regulator under arabinogalactan promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pbt3091::pAG, pbt0267::pDX	Dextran regulator under arabinogalactan promoter, and arabinogalactan regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pDX-nanoluc, pbt3091::pON	NanoLuc inserted into empty PUL75 locus. Dextran regulator under constitutive promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pDX-nanoluc, pbt3091::pDX	NanoLuc inserted into empty PUL75 locus. Dextran regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pDX-nanoluc, pbt3091::pAG	NanoLuc inserted into empty PUL75 locus. Dextran regulator under arabinogalactan promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pAG-nanoluc, pbt0267::pON	NanoLuc inserted into empty PUL75 locus. Arabinogalactan regulator under constitutive promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pAG-nanoluc, pbt0267::pDX	NanoLuc inserted into empty PUL75 locus. Arabinogalactan regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pAG-nanoluc, pbt0267::pAG	NanoLuc inserted into empty PUL75 locus. Arabinogalactan regulator under arabinogalactan promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pDX-nanoluc, pbt3091::pAG, pbt0267::pDX	NanoLuc inserted into empty PUL75 locus. Dextran regulator under arabinogalactan promoter, and arabinogalactan regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pAG-nanoluc, pbt3091::pAG, pbt0267::pDX	NanoLuc inserted into empty PUL75 locus. Dextran regulator under arabinogalactan promoter, and arabinogalactan regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75. pNBU2-pDX-nanoluc, pbt3091::pON	NanoLuc integration with pNBU2 vector. Dextran regulator under constitutive promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pDX-nanoluc, pbt3091::pDX	NanoLuc integration with pNBU2 vector. Dextran regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pDX-nanoluc, pbt3091::pAG	NanoLuc integration with pNBU2 vector. Dextran regulator under arabinogalactan promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pAG-nanoluc, pbt0267::pON	NanoLuc integration with pNBU2 vector. Arabinogalactan regulator under constitutive promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pAG-nanoluc, pbt0267::pDX	NanoLuc integration with pNBU2 vector. Arabinogalactan regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pAG-nanoluc, pbt0267:pAG	NanoLuc integration with pNBU2 vector. Arabinogalactan regulator under arabinogalactan promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pDX-nanoluc, pbt3091::pAG, pbt0267::pDX	NanoLuc integration with pNBU2 vector. Dextran regulator under arabinogalactan promoter, and arabinogalactan regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75, pNBU2-pAG-nanoluc, pbt3091::pAG, pbt0267::pDX	NanoLuc integration with pNBU2 vector. Dextran regulator under arabinogalactan promoter, and arabinogalactan regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pDX-buGH16	BuGH16 inserted into empty PUL75 locus.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pDX-buGH16, pbt3091::pON	BuGH16 inserted into empty PUL75 locus. Dextran regulator under constitutive promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pDX-buGH16, pbt3091::pDX	BuGH16 inserted into empty PUL75 locus. Dextran regulator under dextran promoter.	This study
B. thetaiotaomicron VPI-5482 Δtdk, ΔPUL75::pDX-buGH16 pbt3091::pAG	BuGH16 inserted into empty PUL75 locus. Dextran regulator under arabinogalactan promoter.	This study

2.5.2 Regulatory gene promoter engineering

pEx-tdk was used to insert one of three promoters in front of the start codon of *bt3091* (*SusR*^{DX}) or *bt0267* (*HTCS*^{AG}). The three promoters were P_{bt1311}(promoter from pAT593 [112], P_{bt3090} (defined as the sequence between the stop codon of *bt3091* and the start codon *bt3090*), or P_{bt0268} (promoter from pMM660 [112]) and were denoted as P_{ON}, P_{DX}, and P_{AG} respectively. A terminator sequence was appended to the 5' end of each promoter to reduce context dependence. The insertions were targeted by cloning the 750bp flanks on either side of the desired insertion site into the pEx-tdk vector. Sequences required to modify *SusR*^{DX} were created by gene synthesis (BioBasic). Sequences required to modify *HTCS*^{AG} were created by overlap PCR.

2.5.3 Conjugations

Donor cultures of *E. coli* strain S17-1λpir were grown in 5 mL lysogeny broth (LB) [158] with 100 μg/mL ampicillin. Strains of *B. theta* (Table 2) were routinely grown in 5 mL Tryptone Yeast Extract Glucose (TYG) [159] at 37°C in an anaerobic atmosphere (85% N_2, 10% CO_2, 5% H_2). Donor and recipient cultures were pelleted by centrifugation and resuspended together in 1 mL of TYG and plated on Supplemented Brain Heart Infusion (BHIS) [160] agar. To allow cell mating to occur, plates were grown agar side down at 37°C overnight under aerobic conditions. During cell mating, the vector is transferred from the donor *E. coli* and integrated into the chromosome of recipient *B. theta* through homologous recombination.

After 16 – 24 h, the resulting biomass was scraped from the plate, suspended in TYG broth and serially diluted. Cell suspensions were plated on BHI agar with 200 μg/mL gentamycin to select against *E. coli*, and 25 μg/mL erythromycin to select against

B. theta that had not received vector. Pure cultures were prepared from randomly selected resistant colonies using the streak plate method.

pNBU2 based conjugations were screened by colony PCR to confirm vector integration. For the double crossover conjugations using pEx-tdk and pINT, eight resistant colonies were arbitrarily chosen and grown in TYG broth. At this stage, a second recombination event may occur that ejects the vector backbone and unwanted DNA from the chromosome but retains the sequence of interest. Each cell culture was plated on BHI agar with 200 μg/mL floxuridine (FUDR) to select against cells that had not ejected the plasmid backbone. Pure cultures were obtained by streak plating and were screened by colony PCR. Positive clones underwent genome extraction for sequence confirmation.

2.5.4 Growth curves

Strains of *B. theta* were inoculated from a glycerol stock into 5 mL TYG medium and grown for 16 – 24h under anaerobic conditions (85% N_2, 10% CO_2, 5% H_2). TYG cultures were diluted 1 in 50 into prewarmed anaerobic two times *Bacteroides* Minimal Medium (MM)[146]. 100 μL of diluted cells were plated in a transparent 96 well plate containing 100 μL of prewarmed anaerobic 1% (w/v) carbohydrate solution for final concentrations of 1X MM and 0.5% (w/v) sugar. Carbohydrates were 0.5% (w/v) glucose (GLC), 0.5% homogalacturonan (HG), 0.5% (w/v) dextran (DX), 0.5% (w/v) arabinogalactan (AG), or a mixture of 0.25% (w/v) DX combined with 0.25% (w/v) AG (MIX). Plates were sealed with clear, gas permeable membranes and measured in a BioTek Synergy HT plate reader programmed to read absorbance at 600 nm every 10 minutes for 48 h. Three replicates, each being the average of three observations, were

tested for each condition. Negative controls consisted of wells containing 100 µL medium and 100 µL water. The average of the replicates minus the negative controls was plotted with the standard error of the mean.

2.5.5 *B. theta* wild-type and ΔPUL75 co-culture and qPCR

TYG pre-cultures were created as described above for growth curves, three for wild-type (WT) and three for ΔPUL75. These dense cultures were mixed together 1:1. 100 µL from each mixture was used to inoculate 10 mL of 1X MM with either 0.5% (w/v) GLC or HG. The MM cultures were grown at 37°C in an anaerobic environment. At 0, 12, 24, 48, and 72 h a 1 mL sample was taken from each MM culture for DNA extraction using a GeneJet genomic extraction kit (Thermo Fisher Scientific). 0.1 ng of DNA was assayed in triplicate in a QuantStudio 6 Flex Real-Time PCR system (Thermo Fisher Scientific) using PerfeCta SYBR green low-ROX master mix (Quanta Biosciences) and strain specific primers, and primers for a housekeeping gene present in both strains (Table 1) for 40 cycles of 95°C for 15s and 56°C for 45s. DNA standards (22.4 ng, 5.6 ng, 0.112 ng, 0.0224 ng, and 0.0028 ng) were run in triplicate. A standard curve was used to calculate the relative quantities of each strain present in each sample. Data were analyzed using qbase+ (Biogazelle).

2.5.6 NanoLuc assays

TYG pre-cultures were created as described above for growth curves. TYG cultures were diluted 1 in 100 into prewarmed MM with 0.5% carbohydrate solution in borosilicate glass tubes. After 24 h incubation at 37°C under anaerobic conditions, cells were pelleted by centrifugation and the supernatant was discarded. Cells were lysed by resuspending in 1/10th culture volume of BugBuster® (Millipore Sigma) and incubating

at room temperature for 10 min. Phosphate buffered saline (PBS) was used to dilute lysate to twenty times the original culture volume. Some samples produced too much luminescence for the plate reader to accurately measure and required dilution to two-hundred the original culture volume. 30 μL samples of dilute lysate were aliquoted in opaque 96 well microplates. 30 μL of NanoGlo reagent (Promega) freshly prepared according to manufacturer's specifications was added to each well and the resulting luminescence was measured using a Synergy HT Multi-detection plate reader. Technical replicates, each being the average of three observations, were tested for each condition. The luminescence of the negative controls (30 μL PBS + 30 μL NanoGlo) were subtracted from each replicate. The resulting value was then multiplied by the dilution factor and normalized to the density of the original culture as shown in equation (1).

Equation (1)

$$\frac{((\text{luminescent reading} - \text{negative control}) \text{ x dilution factor})}{\text{OD}_{600} \text{ of culture}}$$

The technical replicates were then averaged and plotted with the standard error of the mean.

2.5.7 Statistical analysis

To examine whether or not changes in NanoLuc activity between strains and conditions were significant, data were analyzed using GraphPad Prism 7.00. T-tests were used to compare conditions due to simplicity of interpretation and providing an appropriate level of granularity compared to an ANOVA. Each condition had three replicates and tests were performed without assuming a consistent standard deviation. Probability values of 0.05, 0.005, and 0.0005 were used to define levels of significance.

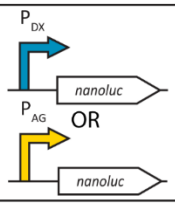
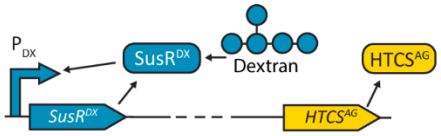
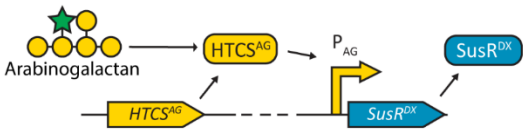
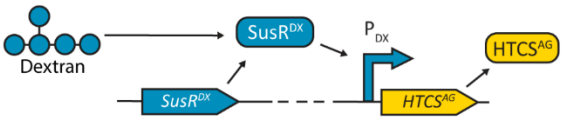
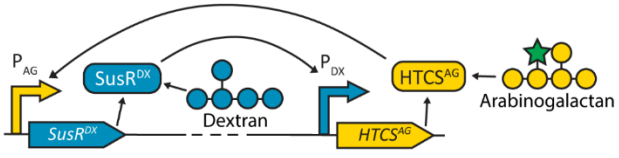
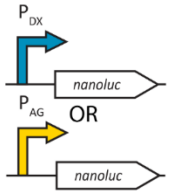
2.5.8 BuGH16 agarolysis assays

Cells were prepared as described above for NanoLuc assays but were not lysed. Cells were washed and resuspended in two times MM before being incubated with an equal volume of 0.8% (w/v) agarose under aerobic conditions at 37°C for 24 h. The liquid phase was clarified by centrifugation before heat killing at 100°C for 10 min. Nine microliters of the resulting solution was spotted onto a silica thin layer chromatography (TLC) plate. The TLC plate was run in a 2:1:1 (v/v) solution of butanol : acetic acid : water. After drying, the plate was stained with one part 0.2% (w/v ethanol) dihydroxynaphthalene to two parts 3.75:1 ethanol : sulfuric acid solution, and colourized by heating at 100°C for 5 min.

2.5.9 Western blotting

Cells were prepared as described above for NanoLuc assays, save that they were lysed in 1/45th culture volume BugBuster. Cell lysate was mixed 1:1 with SDS loading buffer (10 μL of 2 M glycerol, 1.28 M β-mercaptoethanol, 125 mM Tris pH 6.8, 140 mM SDS, 60 μM bromophenol blue) and denatured at 95°C for 5 min. Samples were run on 15% sodium dodecyl sulfate polyacrylamide gels at 200 V for 1 hour in a Novex Mini Cell (Invitrogen). Protein was transferred from gels to a Sequi-Blot polyvinylidene fluoride (PVDF) membrane (BioRad) in an X Cell II Blot Module (Invitrogen) at 30 V for 16 h. Membranes were incubated with blocking buffer (5% (w/v) skim milk powder in a solution of 50 mM Tris, 150 mM NaCl and 0.1% (v/v) Tween 20 (TBS-T)) for 2 h followed by incubation with a 1:2500 dilution of rabbit anti-6-his IgG conjugated to horse radish peroxidase (Bethyl) in blocking buffer for 6 h. Membranes were washed three times by rinsing in water and rocking in TBS-T for five min. An Opti-4CN (BioRad) substrate kit was used to colorize the membranes.

Table S1: Novel engineered regulatory systems.

Strain	Regulatory Schematic	Reporter Regulation
WT regulation		
$P_{ON}SusR^{DX}$ Constitutive boost.		
$P_{DX}SusR^{DX}$ Positive feedback.		
$P_{AG}SusR^{DX}$ Hybrid regulation.		
$P_{ON}HTCS^{AG}$ Constitutive boost.		
$P_{DX}HTCS^{AG}$ Hybrid regulation.		
$P_{AG}HTCS^{AG}$ Positive feedback.		
$P_{AG}SusR^{DX} + P_{DX}HTCS^{AG}$ Double hybrid regulation.		

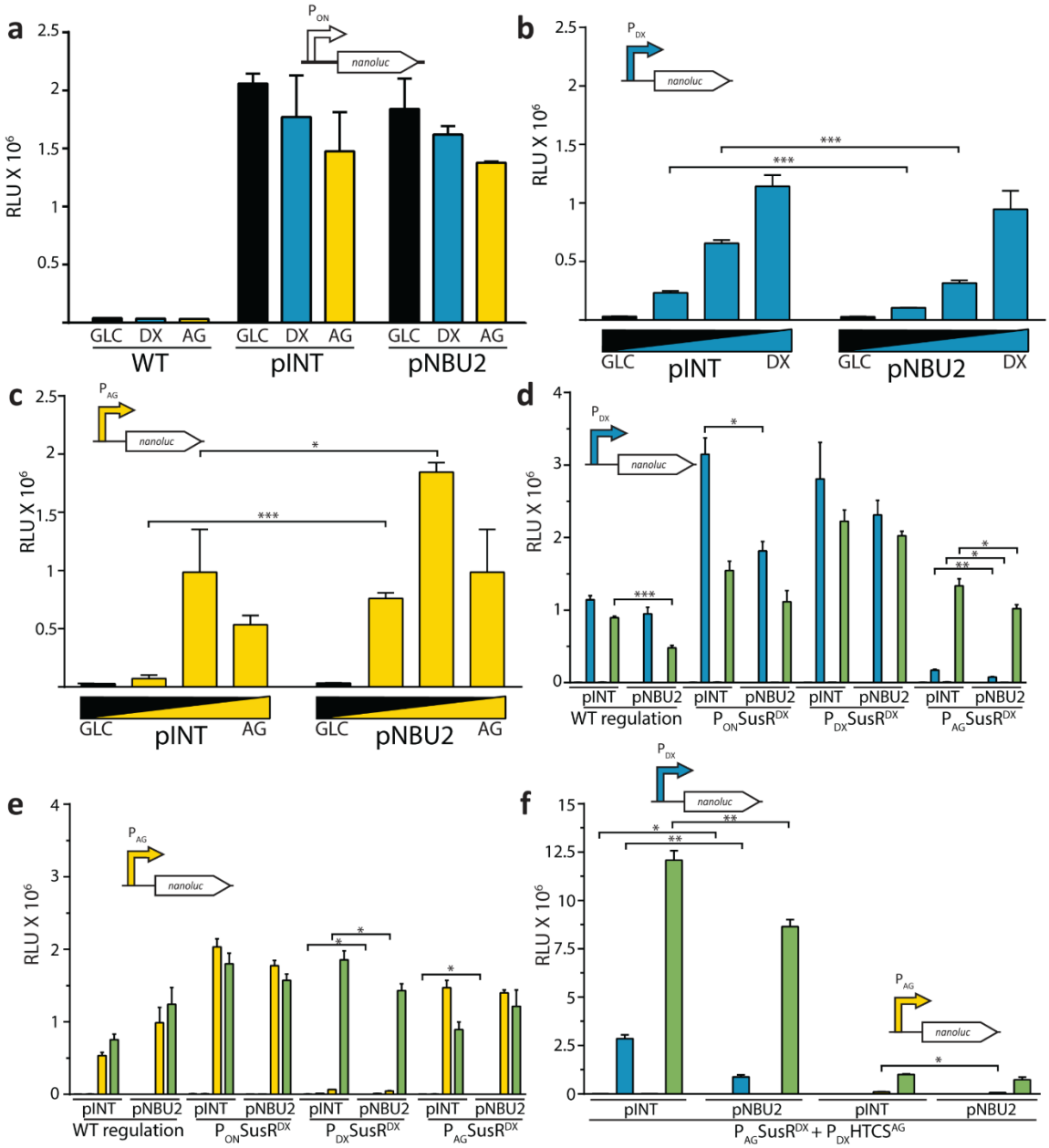


Figure S1: Comparison of NanoLuc activity from two different genomic loci. pINT integrates targets into the vacant PUL75 locus. pNBU2 integrates targets, and the entire pNBU2 plasmid, into a $tRNA^{ser}$ locus. Every pINT condition was compared to its pNBU2 counterpart. **(a)** NanoLuc activity under control of the P_{ON} promoter on GLC (black), DX (blue), or AG (yellow). **(b)** NanoLuc activity under control of the P_{DX} promoter on varying concentrations of GLC and DX. **(c)** NanoLuc activity under control of the P_{AG} promoter on varying concentrations of GLC and AG. **(d)** NanoLuc activity under control of the P_{DX} promoter in modified $SusR^{DX}$ strains grown on GLC, DX, AG or a mixture of DX and AG (green). **(e)** NanoLuc activity under control of the P_{AG} promoter in modified $HTCS^{AG}$. **(f)** NanoLuc activity under control of either P_{DX} or P_{AG} in strains with modified $SusR^{DX}$ and $HTCS^{AG}$. Significant differences between pINT and pNBU2 counterparts are indicated with brackets. 32/48 (67%) of comparisons between the two loci do not result in statistically significant difference. Asterisks represent level of significance (*= p< 0.05, **=p<0.005, ***=p<0.0005).

Chapter 3 Conclusions and Future Directions

In this thesis I have used genetic elements native to *B. theta* to generate several strains with altered regulatory circuits and novel growth and reporter gene activity phenotypes. To the best of my knowledge these are the first gene expression systems to require two structurally distinct polysaccharides. These are also the first expression systems to manipulate PUL regulators in order to increase downstream gene expression and change the induction requirements of gene expression. These regulatory systems are important as proof-of-concept for the idea of engineered microorganisms with carbohydrate tunable activities, which will aid in the development of engineered probiotics and investigations into polysaccharide relevant metabolic and signaling networks. My research also compares levels of reporter gene activity from integrations with the unpublished pINT vector system against integrations performed with the previously described pNBU2 vector system. These advancements represent relevant new tools for the expanding field of microbiome science.

3.1 Future directions

More testing will be required before these dual-glycan regulation systems can be implemented in more complex systems. The changes in growth properties and heterologous gene expression observed here are likely due to alterations in the amount of regulatory protein being produced, but transcriptomic or proteomic studies will be required to confirm that this is occurring. It will be important to learn to what extent the quantity of regulatory protein is correlated with reporter gene activity level because that knowledge will be used to inform future circuit development. For example, if there is a point at which increasing regulatory protein concentration does not lead to further

increases in reporter activity, then it should be possible to design a more efficient system that achieves the optimum production of regulatory protein. However, the optimum concentration of regulatory protein may be different under different conditions. Growth and reporter gene activity in this study was only tested in minimal medium and limited combinations of inducing carbohydrates. The colon is a much more complex microbial ecosystem and it will be important to test production levels when cells are exposed to other microorganisms and stimuli. Host mucin derived polysaccharides will always be available in the colon and the availability of other polysaccharides will change with each meal. Metabolism of other available polysaccharides could affect the metabolism of DX or AG and this may alter induction of regulatory proteins and reporter genes. Techniques such as simulating the DGM using different ratios of inducing and non-inducing sugars will be important first steps to determine ranges of activity that can be expected from these regulatory systems *in vivo*. Furthermore, though most of the strains developed here did not display noticeable growth defects, the *in vitro* conditions used for these experiments may not accurately reflect their fitness in competition with other microorganisms. Colonization studies will be necessary to determine if the engineered strains can persist or proliferate with a community and become established in a gut. The heterologous protein production of these strains must also be assessed *in vivo.* The amount of production that is required will vary for different applications. Additionally, it will be important to assess how producing heterologous proteins affects *B. theta*'s persistence and stability in microbial communities.

DX and AG responsive systems will likely be appropriate for applications in which bacterial activity is desired most of the time. Because AG is found in plant cell

walls, it would be difficult to find a livestock diet that does not contain AG, and diets considered healthy for humans often contain a large proportion of plant-based foods [179]. DX is a bacterial glycan made by fermentation and is less ubiquitous than AG, but it can be provided to livestock in the form of silage [180]. For humans, a variety of fermented foods are available. nuAttempting to create systems that take advantage of other glycan-responsive elements may lead to the development of systems with wider production ranges or more desirable specificities. There is room to elaborate on the systems created in this study. Currently, some of my engineered systems require the presence of two unique glycans for full activity, but it may be possible to create systems that require three or more glycans, or to create multiple dual-glycan systems in the same strain.

3.2 Public opinion on genetically modified organisms

A challenge for the field of engineered probiotics, and biotechnology in general, is public opinion on genetic engineering and genetic modification (GM). A 2013 survey found that 53% of Americans believe that GM foods are unsafe [161], and the preliminary results of a 2018 survey found that 34% of Canadians share that sentiment [162]. The 2013 survey of Americans only allowed participants to accept or reject the statement that GM food is unsafe, whereas the Canadian survey allowed a neutral option, which may account for the apparent disparity in opinion between the two nations. Nevertheless, widespread belief that genetically modified organisms (GMOs) are harmful would likely translate to a rejection of engineered probiotics and meat raised with the aid of GM bacteria.

Anti-GMO beliefs may stem from misunderstanding the nature of genetic technologies which could be alleviated with more education and science outreach [163]. Indeed, general attitudes towards GMO's seem to have improved since the early 2000's [164]. Anecdotally, in an interview with Nature News, Neal Carter, the innovator behind the non-browning GM 'Arctic apple', said "I rarely get e-mails saying we are Satan any more, now we have people asking where they can buy the apples" [165].

The regulatory engineering in this thesis was done with genetic parts native to *B. theta* which may assuage some public concern, as it is known that people are generally more accepting of cisgenic modifications than they are of transgenic modifications [166]. But regardless of public opinion, genetically engineered plants, animals, and microorganisms intended for human consumption are likely here to stay. GMOs have become a mainstay in modern agriculture, with an estimated 95% of canola, 80% of corn, 60% of soybeans, and nearly all sugar beets grown in Canada being GM [167]. The benefits of genetic engineering are simply too powerful to ignore. For example, GM yeast can produce a beer that tastes 'hoppier' than beer made with actual hops [168], reducing the need for the hops plant which is expensive and water intensive to produce. For meat production, the AquAdvantage salmon, which made a 2017 market debut in Canada, grows twice as fast as non-GMO Atlantic salmon [169]. It is likely that GMOs will become more pervasive in everyday life as genetic engineering technologies become more sophisticated.

3.3 Biocontainment of genetically engineered microbes

Because engineered bacteria can be used to colonize a host [71], an important topic for future research is controlling the spread of genetically engineered microorganisms. This

is referred to as biocontainment and can be divided into passive and active strategies. Passive strategies delete essential functions and create a dependency on supplied molecules so that if the microorganism exits the containment area where the essential molecule is supplied, or the molecule is withheld, the microorganism will no longer be able to perform necessary functions and will die. Active biocontainment strategies involve the construction of a kill switch in the engineered microorganism so that when a specific signal is received the microorganism will self-destruct.

A popular passive strategy for biocontainment is to knockout thymidylate synthase (*thyA*), a gene required for thymidine synthesis. Without *thyA,* cells must be supplemented with thymine or thymidine in order to replicate their DNA and survive. This was the strategy in use for the first human clinical trial of *L. lactis* [59] where it was observed that engineered cells recovered from stool were still thymine auxotrophs, indicating that the containment method was functioning as intended. However, *thyA* deletion was tested more recently with *B. ovatus* [170]. In combination with the anaerobic requirements of *B. ovatus*, *thyA* deletion was hypothesized to be an acceptable containment method. Thymine auxotrophic *B. ovatus* was fed to mice and it was demonstrated that horizontal gene transfer took place *in vivo* and allowed *B. ovatus* to overcome its gene deletion. *ThyA* horizontal gene transfer has also been observed in *E. coli* [171] and is a possibility in lactococci as well due to the lactococcal sex factor [172]. Because of this, thymine auxotrophy by itself is not an acceptable biocontainment method for future engineered probiotics.

Biocontainment systems apply selective pressure on the organism to subvert the containment mechanisms. More sophisticated passive containment strategies are being developed that attempt to eliminate the possibility of evolving out of containment. For

example, knocking out every phosphate transporter and introducing a phopshite transporter and a phosphite dehydrogenase to make cells dependent on phosphite, which is rarely detected in nature [173]. The authors claim the lowest reported escape mutation rate but admit that they did not test their system in situations where horizontal gene transfer would be a concern. Strategies that are not easily circumvented by random mutation or horizontal gene transfer involve making modifications to numerous essential proteins to make escape mutants less likely. Examples of this include making essential proteins dependent on a non-standard amino acid [174], or a special ligand [175]. These make escape mutants less likely but require a substantial amount of labor.

Active containment systems are frequently based on the natural toxin/antitoxin (TA) systems found on plasmids. TA systems consist of an unstable antitoxin and a stable toxin and help to ensure plasmid maintenance. If a daughter cell does not receive a copy of the plasmid, then the remaining antitoxin in its cytoplasm will continue to degrade, and the stable toxin will result in cell death. TA systems are sometimes chromosomally located where they are thought to provide a defense against phages. When cellular resources are consumed by phage replication, the antitoxin is no longer produced in quantities large enough to counteract the toxin, which results in cell death and reduces the number of phages that can attack neighboring cells. The first model containment system was made using part of the *hok*/*sok* TA system from *E. coli* [176], which functions via the *sok* antisense RNA preventing the translation of the toxic Hok protein [177]. Systems that rely on toxin production for cell death tend to generate toxin-resistant mutants, even when multiple toxins with unique targets are simultaneously used [178].

The style of regulatory systems developed here lend themselves to the creation of containment systems. Active systems are simple to envision but may not be effective. Simply placing a toxin gene under a dual glycan system would provide fine control over when it is induced, but the low levels of expression from uninduced PUL promoters may lead to enough toxin production to quickly generate selective pressure for toxin resistance. A more comprehensive biocontainment solution would be to place antitoxins under a dual glycan system and constitutively produce toxin proteins fused to other proteins that convey a selective advantage. For example, fusing toxins to agarolytic enzymes and supplementing the host diet with agarose would provide a niche for the engineered bacteria to fill and would create selection pressure against frameshift and nonsense mutations that disable the enzyme-toxin fusion protein. Cell death would be achieved by withholding either or both of the glycans necessary to induce antitoxin production. This type of system would be simpler to implement than passive strategies that rely on extensive modifications and may result in lower escape frequency than typical TA systems.

References

1. Sender, R., S. Fuchs, and R. Milo, Revised estimates for the number of human and bacteria cells in the body. PLoS Biology, 2016. **14**(8): p. e1002533.

2. Hooper, L.V. and J.I. Gordon, Commensal host-bacterial relationships in the gut. Science, 2001. **292**(5519): p. 1115-8.

3. Hooper, L.V., T. Midtvedt, and J.I. Gordon, How host-microbial interactions shape the nutrient environment of the mammalian intestine. Annual Review of Nutrition, 2002. **22**: p. 283-307.

4. Cummings, J.H. and G.T. Macfarlane, The control and consequences of bacterial fermentation in the human colon. Journal of Applied Bacteriology, 1991. **70**(6): p. 443-59.

5. den Besten, G., et al., The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. Journal of Lipid Research, 2013. **54**(9): p. 2325-2340.

6. Bergman, E.N., Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. Physiological Reviews, 1990. **70**(2): p. 567-90.

7. Stappenbeck, T.S., L.V. Hooper, and J.I. Gordon, Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(24): p. 15451-15455.

8. Macpherson, A.J. and E. Slack, The functional interactions of commensal bacteria with intestinal secretory IgA. Current Opinion in Gastroenterology, 2007. **23**(6): p. 673-678.

9. Round, J.L. and S.K. Mazmanian, The gut microbiota shapes intestinal immune responses during health and disease. Nature Reviews Immunology, 2009. **9**(5): p. 313-323.

10. Peloquin, J.M. and D.D. Nguyen, The microbiota and inflammatory bowel disease: Insights from animal models. Anaerobe, 2013. **24**: p. 102-106.

11. Callaway, T.R., et al., Probiotics, prebiotics and competitive exclusion for prophylaxis against bacterial disease. Animal Health Research Reviews, 2008. **9**(2): p. 217-25.

12. Frank, D.N., et al., Investigating the biological and clinical significance of human dysbioses. Trends in Microbiology, 2011. **19**(9): p. 427-34.

13. Brunkwall, L. and M. Orho-Melander, The gut microbiome as a target for prevention and treatment of hyperglycaemia in type 2 diabetes: from current human evidence to future possibilities. Diabetologia, 2017. **60**(6): p. 943-951.

14. Davis, C.D., The gut microbiome and its role in obesity. Nutrition Today, 2016. **51**(4): p. 167-174.

15. Qin, J., et al., A human gut microbial gene catalogue established by metagenomic sequencing. Nature, 2010. **464**: p. 59.

16. Ju Young, C., et al., Decreased diversity of the fecal microbiome in recurrent *Clostridium difficile*—associated diarrhea. Journal of Infectious Diseases, 2008. **197**(3): p. 435-438.

17. Schenck, L.P., P.L. Beck, and J.A. MacDonald, Gastrointestinal dysbiosis and the use of fecal microbial transplantation in *Clostridium difficile* infection. World Journal of Gastrointestinal Pathophysiology, 2015. **6**(4): p. 169-180.

18. Carabotti, M., et al., The gut-brain axis: Interactions between enteric microbiota, central and enteric nervous systems. Annals of Gastroenterology : Quarterly Publication of the Hellenic Society of Gastroenterology, 2015. **28**(2): p. 203-209.

19. Park, A.J., et al., Altered colonic function and microbiota profile in a mouse model of chronic depression. Neurogastroenterology & Motility, 2013. **25**(9): p. 733-e575.

20. Zheng, P., et al., Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. Molecular Psychiatry, 2016. **21**(6): p. 786-96.

21. Neufeld, K.M., et al., Reduced anxiety-like behavior and central neurochemical change in germ-free mice. Neurogastroenterology & Motility, 2011. **23**(3): p. 255-64, e119.

22. Strati, F., et al., New evidences on the altered gut microbiota in autism spectrum disorders. Microbiome, 2017. **5**: p. 24.

23. Fecteau, M.E., et al., Dysbiosis of the fecal microbiota in cattle infected with *Mycobacterium avium* subsp. paratuberculosis. PLoS One, 2016. **11**(8): p. e0160353.

24. Zeineldin, M., B. Aldridge, and J. Lowe, Dysbiosis of the fecal microbiota in feedlot cattle with hemorrhagic diarrhea. Microbial Pathogenesis, 2018. **115**: p. 123-130.

25. Gresse, R., et al., Gut microbiota dysbiosis in postweaning piglets: Understanding the keys to health. Trends in Microbiology, 2017. **25**(10): p. 851-873.

26. Teirlynck, E., et al., Morphometric evaluation of "dysbacteriosis" in broilers. Avian Pathology, 2011. **40**(2): p. 139-44.

27. Di Criscio, T., et al., Production of functional probiotic, prebiotic, and synbiotic ice creams. Journal of Dairy Science, 2010. **93**(10): p. 4555-4564.

28. Hill, C., et al., The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nature Reviews Gastroenterology & Hepatology, 2014. **11**: p. 506.

29. O'Mahony, L., et al., *Lactobacillus* and *Bifidobacterium* in irritable bowel syndrome: symptom responses and relationship to cytokine profiles. Gastroenterology, 2005. **128**(3): p. 541-551.

30. Bravo, J.A., et al., Ingestion of *Lactobacillus* strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(38): p. 16050-16055.

31. Nemani, K., et al., Schizophrenia and the gut-brain axis. Progress in Neuro-Psychopharmacology & Biological Psychiatry, 2015. **56**: p. 155-160.

32. Motevaseli, E., A. Dianatpour, and S. Ghafouri-Fard, The role of probiotics in cancer treatment: emphasis on their *in vivo* and *in vitro* anti-metastatic effects. International Journal of Molecular and Cellular Medicine, 2017. **6**(2): p. 66-76.

33. Awad, W.A., et al., Effects of dietary inclusion of probiotic and synbiotic on growth performance, organ weights, and intestinal histomorphology of broiler chickens. Poultry Science, 2009. **88**(1): p. 49-56.

34. Timmerman, H.M., et al., Health and growth of veal calves fed milk replacers with or without probiotics. Journal of Dairy Science, 2005. **88**(6): p. 2154-2165.

35. Huyghebaert, G., R. Ducatelle, and F. Van Immerseel, An update on alternatives to antimicrobial growth promoters for broilers. Journal of Veterinary Science, 2011. **187**(2): p. 182-188.

36. Sherman, P.M., J.C. Ossa, and K. Johnson-Henry, Unraveling mechanisms of action of probiotics. Nutrition in Clinical Practice, 2009. **24**(1): p. 10-14.

37. Hutkins, R.W., et al., Prebiotics: why definitions matter. Current Opinion in Biotechnology, 2016. **37**: p. 1-7.

38. Yazawa, K., K. Imai, and Z. Tamura, Oligosaccharides and polysaccharides specifically utilizable by bifidobacteria. Chemical and Pharmaceutical Bulletin, 1978. **26**(11): p. 3306-11.

39. Lindsay, J.O., et al., Clinical, microbiological, and immunological effects of fructo-oligosaccharide in patients with Crohn's disease. Gut, 2006. **55**(3): p. 348-55.

40. Jacobsen, H., et al., Carbohydrate digestibility predicts colon carcinogenesis in azoxymethane-treated rats. Nutrition and Cancer, 2006. **55**(2): p. 163-70.

41. Clark, M.J., K. Robien, and J.L. Slavin, Effect of prebiotics on biomarkers of colorectal cancer in humans: a systematic review. Nutrition Reviews, 2012. **70**(8): p. 436-43.

42. Coudray, C., et al., Dietary inulin intake and age can significantly affect intestinal absorption of calcium and magnesium in rats: a stable isotope approach. Nutrition Journal, 2005. **4**: p. 29-29.

43. Kelly, G., Inulin-type prebiotics: a review. (Part 2). Alternative Medicine Reviews, 2009. **14**(1): p. 36-55.

44. Whisner, C.M. and L.F. Castillo, Prebiotics, bone and mineral metabolism. Calcified Tissue International, 2018. **102**(4): p. 443-479.

45. Cani, P.D., et al., Gut microbiota fermentation of prebiotics increases satietogenic and incretin gut peptide production with consequences for appetite sensation and glucose response after a meal. American Journal of Clinical Nutrition, 2009. **90**(5): p. 1236-43.

46. Parnell, J.A. and R.A. Reimer, Weight loss during oligofructose supplementation is associated with decreased ghrelin and increased peptide YY in overweight and obese adults. American Journal of Clinical Nutrition, 2009. **89**(6): p. 1751-9.

47. Slavin, J., Fiber and prebiotics: mechanisms and health benefits. Nutrients, 2013. **5**(4): p. 1417-1435.

48. Femia, A.P., et al., Antitumorigenic activity of the prebiotic inulin enriched with oligofructose in combination with the probiotics *Lactobacillus rhamnosus* and *Bifidobacterium lactis* on azoxymethane-induced colon carcinogenesis in rats. Carcinogenesis, 2002. **23**(11): p. 1953-60.

49. Dinleyici, E.C., et al., The effect of a multispecies synbiotic mixture on the duration of diarrhea and length of hospital stay in children with acute diarrhea in Turkey: Single blinded randomized study. European Journal of Pediatrics, 2013. **172**(4): p. 459-464.

50. Dewi, M.R., Y. Soenarto, and I.P.G. Karyana, Efficacy of synbiotic treatment in children with acute rotavirus diarrhea. . Paediatrica Indonesiana, 2015. **55**(2): p. 74-78.

51. Steed, H., et al., Clinical trial: the microbiological and immunological effects of synbiotic consumption - a randomized double-blind placebo-controlled study in active Crohn's disease. Alimentary Pharmacology & Therapeutics, 2010. **32**(7): p. 872-883.

52. Yasuda, K., et al., A new synbiotic consisting of *Lactobacillus casei* subsp casei and dextran improves milk production in Holstein dairy cows. Journal of Veterinary Medical Science, 2007. **69**(2): p. 205-208.

53. Krumbeck, J.A., J. Walter, and R.W. Hutkins, Synbiotics for improved human health: recent developments, challenges, and opportunities. Annual Review of Food Science and Technology, 2018. **9**(1): p. 451-479.

54. Fujimori, S., et al., A randomized controlled trial on the efficacy of synbiotic versus probiotic or prebiotic treatment to improve the quality of life in patients with ulcerative colitis. Nutrition, 2009. **25**(5): p. 520-525.

55. Šmid, A., et al., Randomized clinical trial: the effect of fermented milk with the probiotic cultures *Lactobacillus acidophilus* La-5® and *Bifidobacterium* BB-12® and Beneo dietary fibres on health-related quality of life and the symptoms of irritable bowel syndrome in adults. Journal of Functional Foods, 2016. **24**: p. 549-557.

56. Worthley, D.L., et al., A human, double-blind, placebo-controlled, crossover trial of prebiotic, probiotic, and synbiotic supplementation: effects on luminal, inflammatory, epigenetic, and epithelial biomarkers of colorectal cancer. American Journal of Clinical Nutrition, 2009. **90**(3): p. 578-586.

57. Le, D.T., et al., A live-attenuated *Listeria* vaccine (ANZ-100) and a live-attenuated *Listeria* vaccine expressing mesothelin (CRS-207) for advanced cancers: phase I studies of safety and immune induction. Clinical Cancer Research, 2012. **18**(3): p. 858-868.

58. Lagenaur, L.A., et al., Prevention of vaginal SHIV transmission in macaques by a live recombinant *Lactobacillus.* Mucosal Immunology, 2011. **4**(6): p. 648-57.

59. Braat, H., et al., A phase I trial with transgenic bacteria expressing interleukin-10 in Crohn's disease. Clinical Gastroenterology and Hepatology, 2006. **4**(6): p. 754-9.

60. Takiishi, T., et al., Reversal of diabetes in NOD mice by clinical-grade proinsulin and IL-10–secreting *Lactococcus lactis* in combination with low-dose anti-CD3 depends on the induction of foxp3-positive T cells. Diabetes, 2017. **66**(2): p. 448-459.

61. Gregg, K., et al., Genetically modified ruminal bacteria protect sheep from fluoroacetate poisoning. Applied and Environmental Microbiology, 1998. **64**(9): p. 3496-3498.

62. Tayyab, U., et al., Ruminal and intestinal protein degradability of various seaweed species measured in situ in dairy cows. Animal Feed Science and Technology, 2016. **213**: p. 44-54.

63. Bach, S.J., Y. Wang, and T.A. McAllister, Effect of feeding sun-dried seaweed (Ascophyllum nodosum) on fecal shedding of Escherichia coli O157 : H7 by feedlot cattle and on growth performance of lambs. Animal Feed Science and Technology, 2008. **142**(1-2): p. 17-32.

64. Choct, M. and G. Annison, Antinutritive Effect of Wheat Pentosans in Broiler Chickens - Roles of Viscodity and Gut Microflora. British Poultry Science, 1992. **33**(4): p. 821-834.

65. van Bruchem, J. et al., Intake, rumination, reticulo-rumen fluid and particle kinetics, and faecal particle size in heifers and cattle fed on grass hay and wilted grass silage. Livestock Produciton Science, 1991. 27: p. 297-308.

66. Monod, J. and F. Jacob, Teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harbor Symposia on Quantitative Biology, 1961. **26**: p. 389-401.

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

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

69. Tabor, S. and C.C. Richardson, A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proceedings of the National Academy of Sciences of the United States of America, 1985. **82**(4): p. 1074-1078.

70. Wassenaar, T.M., Insights from 100 Years of research with probiotic *E. Coli.* European Journal of Microbiology & Immunology, 2016. **6**(3): p. 147-161.

71. Riglar, D.T., et al., Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. Nature Biotechnology, 2017. **35**(7): p. 653-658.

72. Hwang, I.Y., et al., Reprogramming microbes to be pathogen-seeking killers. ACS Synthetic Biology, 2014. **3**(4): p. 228-237.

73. Human Microbiome Project, C., et al., Structure, function and diversity of the healthy human microbiome. Nature, 2012. **486**(7402): p. 207-214.

74. Rosano, G.L. and E.A. Ceccarelli, Recombinant protein expression in *Escherichia coli*: advances and challenges. Frontiers in Microbiology, 2014. **5**: p. 172-172.

75. Li, H., et al., The outer mucus layer hosts a distinct intestinal microbial niche. Nature Communications, 2015. **6**: p. 13.

76. Bourdichon, F., et al., Food fermentations: Microorganisms with technological beneficial use. International Journal of Food Microbiology, 2012. **154**(3): p. 87-97.

77. Tarr, P.I., C.A. Gordon, and W.L. Chandler, Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. Lancet, 2005. **365**(9464): p. 1073-1086.

78. Bureau of Chemical Safety, F.D., Health Products and Food Branch, Health Canada's proposal to enable the use of *Lactococcus lactis* DSM 11037 as an oxygen scavenger in certain modified atmosphere-packaged (MAP) and vacuum-packaged preserved meat products. 2016.

79. Wessels, S., et al., The lactic acid bacteria, the food chain, and their regulation. Trends in Food Science & Technology, 2004. **15**(10): p. 498-505.

80. Ricci, A., et al., Update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 7: suitability of taxonomic units notified to EFSA until September 2017. EFSA Journal, 2018. **16**(1): p. e05131.

81. Steidler, L., et al., Treatment of murine colitis by *Lactococcus lactis* secreting interleukin-10. Science, 2000. **289**: p. 1352.

82. Steidler, L., et al., Mucosal delivery of murine interleukin-2 (IL-2) and IL-6 by recombinant strains of Lactococcus lactis coexpressing antigen and cytokine. Infection and Immunity, 1998. **66**(7): p. 3183-9.

83. Klijn, N., A.H. Weerkamp, and W.M. Devos, Genetic marking of *Lactococcus lactis* shows its survival in the human gastrointestinal tract. Applied and Environmental Microbiology, 1995. **61**(7): p. 2771-2774.

84. Oakley, B.B., et al., The chicken gastrointestinal microbiome. FEMS Microbiology Letters, 2014. **360**(2): p. 100-112.

85. Goldstein, E.J.C., Anaerobic bacteremia. Clinical Infectious Diseases, 1996. **23**: p. S97-S101.

86. Wexler, H.M., *Bacteroides*: the good, the bad, and the nitty-gritty. Clinical Microbiology Reviews, 2007. **20**(4): p. 593-621.

87. Hehemann, J.-H., A.B. Boraston, and M. Czjzek, A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. Current Opinion in Structural Biology, 2014. **28**: p. 77-86.

88. Salazar, N., et al., Exopolysaccharides produced by intestinal *Bifidobacterium* strains act as fermentable substrates for human intestinal bacteria. Applied and Environmental Microbiology, 2008. **74**(15): p. 4737-45.

89. Abbott, D.W., et al., Coevolution of yeast mannan digestion: convergence of the civilized human diet, distal gut microbiome, and host immunity. Gut Microbes, 2015. **6**(5): p. 334-9.

90. Bjursell, M.K., E.C. Martens, and J.I. Gordon, 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, 2006. **281**(47): p. 36269-79.

91. Tancula, E., et al., Location and characterization of genes involved in binding of starch to the surface of *Bacteroides thetaiotaomicron.* Journal of Bacteriology, 1992. **174**(17): p. 5609-5616.

92. Martens, E.C., et al., Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. Journal of Biological Chemistry, 2009. **284**(37): p. 24673-7.

93. Cameron, E.A., et al., Multifunctional nutrient-binding proteins adapt human symbiotic bacteria for glycan competition in the gut by separately promoting enhanced sensing and catalysis. mBio, 2014. **5**(5): p. 12.

94. Glenwright, A.J., et al., Structural basis for nutrient acquisition by dominant members of the human gut microbiota. Nature, 2017. **541**(7637): p. 407-+.

95. Hamady, Z.Z., et al., Identification and use of the putative *Bacteroides ovatus* xylanase promoter for the inducible production of recombinant human proteins. Microbiology, 2008. **154**(Pt 10): p. 3165-74.

96. Farrar, M.D., et al., 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, 2005. **98**(5): p. 1191-7.

97. Hamady, Z.Z., et al., Xylan-regulated delivery of human keratinocyte growth factor-2 to the inflamed colon by the human anaerobic commensal bacterium *Bacteroides ovatus.* Gut, 2010. **59**(4): p. 461-9.

98. Hamady, Z.Z., et al., Treatment of colitis with a commensal gut bacterium engineered to secrete human TGF-beta1 under the control of dietary xylan 1. Inflammatory Bowel Diseases, 2011. **17**(9): p. 1925-35.

99. Distaso, A., Contribution à l'étude sur l'intoxication intestinale. Centralb f Bakteriol, 1912. **62**: p. 433-468.

100. Moore, W.E.C. and L.V. Holdeman, New names and combinations in the genera *Bacteroides* Castellani and Chalmers, *Fusobacterium* Knorr, *Eubacterium* Prevot, *Propionibacterium* Delwiche, and *Lactobacillus* Orla-Jensen. International Journal of Systematic Bacteriology, 1973. **23**(1): p. 69-74.

101. Cato, E.P. and J.L. Johnson, Reinstatement of species rank for *Bacteroides fragilis, B. ovatus, B. distasonis, B. thetaiotaomicron*, and *B. vulgatus*: Designation of neotype strains for *Bacteroides fragilis* (Veillon and Zuber) Castellani and Chalmers and *Bacteroides thetaiotaomicron* (Distaso) Castellani

and Chalmers. International Journal of Systematic Bacteriology, 1976. **26**(2): p. 230-237.

102. Kelly, D., et al., Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA. Nature Immunology, 2004. **5**(1): p. 104-12.

103. Martens, E.C., et al., Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biology, 2011. **9**(12): p. e1001221.

104. Bayley, D.P., E.R. Rocha, and C.J. Smith, Analysis of cepA and other *Bacteroides fragilis* genes reveals a unique promoter structure. FEMS Microbiology Letters, 2000. **193**(1): p. 149-154.

105. Vingadassalom, D., et al., An unusual primary sigma factor in the Bacteroidetes phylum. Molecular Microbiology, 2005. **56**(4): p. 888-902.

106. Mastropaolo, M.D., M.L. Thorson, and A.M. Stevens, Comparison of *Bacteroides thetaiotaomicron* and *Escherichia coli* 16S rRNA gene expression signals. Microbiology, 2009. **155**: p. 2683-2693.

107. Wegmann, U., N. Horn, and S.R. Carding, Defining the *Bacteroides* ribosomal binding site. Applied and Environmental Microbiology, 2013. **79**(6): p. 1980-1989.

108. Smith, C.J., M.B. Rogers, and M.L. McKee, Heterologous gene expression in *Bacteroides fragilis*. Plasmid, 1992. **27**(2): p. 141-54.

109. Lim, B., et al., Engineered regulatory systems modulate gene expression of human commensals in the gut. Cell, 2017. **169**(3): p. 547-558.e15.

110. Koropatkin, N.M., et al., Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure, 2008. **16**(7): p. 1105-1115.

111. Wang, J., et al., Characterization of a *Bacteroides* mobilizable transposon, NBU2, which carries a functional lincomycin resistance gene. Journal of Bacteriology, 2000. **182**(12): p. 3559-3571.

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

113. Horn, N., et al., A novel tightly regulated gene expression system for the human intestinal symbiont *Bacteroides thetaiotaomicron.* Frontiers in Microbiology, 2016. **7**: p. 1080.

114. Ravcheev, D.A., et al., Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: comparative genomics reconstruction of metabolic and regulatory networks. BMC Genomics, 2013. **14**: p. 873.

115. Paget, M.S.B. and J.D. Helmann, The σ(70) family of sigma factors. Genome Biology, 2003. **4**(1): p. 203-203.

116. Helmann, J.D., The extracytoplasmic function (ECF) sigma factors. Advances in Microbial Physiology, 2002. **46**: p. 47-110.

117. Capra, E.J. and M.T. Laub, The evolution of two-component signal transduction systems. Annual review of microbiology, 2012. **66**: p. 325-347.

118. Sanders, D.A., et al., Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY. Journal of Biological Chemistry, 1989. **264**(36): p. 21770-8.

119. Xu, J., et al., A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science, 2003. **299**(5615): p. 2074-6.

120. Townsend, G.E., 2nd, et al., Intramolecular arrangement of sensor and regulator overcomes relaxed specificity in hybrid two-component systems. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(2): p. E161-9.

121. Xu, J., et al., Message from a human gut symbiont: sensitivity is a prerequisite for sharing. Trends in Microbiology, 2004. **12**(1): p. 21-8.

122. Sonnenburg, E.D., et al., 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 of the United States of America, 2006. **103**(23): p. 8834-8839.

123. Bolam, D.N. and N.M. Koropatkin, Glycan recognition by the Bacteroidetes Sus-like systems. Current Opinion in Structural Biology, 2012. **22**(5): p. 563-9.

124. Lowe, E.C., et al., A scissor blade-like closing mechanism implicated in transmembrane signaling in a *Bacteroides* hybrid two-component system. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(19): p. 7298-7303.

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

126. Lynch, J.B. and J.L. Sonnenburg, Prioritization of a plant polysaccharide over a mucus carbohydrate is enforced by a *Bacteroides* hybrid two-component system. Molecular Microbiology, 2012. **85**(3): p. 478-91.

127. Della, J.N. and A.A. Salyers, Effect of regulatory protein levels on utilization of starch by *Bacteroides thetaiotaomicron.* Journal of Bacteriology, 1996. **178**(24): p. 7180-7186.

128. Naessens, M., et al., *Leuconostoc* dextransucrase and dextran: production, properties and applications. Journal of Chemical Technology and Biotechnology, 2005. **80**(8): p. 845-860.

129. Guggenheim, B. and H.E. Schroeder, Biochemical and morphological aspects of extracellular polysaccharides produced by cariogenic streptococci. Helvetica Odontologica Acta, 1967. **11**(2): p. 131-+.

130. AN, D.B., Medical applications of dextran and its derivatives, in Polysaccharides in Medicinal Applications, D.S.M. Dekker, Editor. 1996: New York. p. 505 - 523.

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

132. Fincher, G.B., B.A. Stone, and A.E. Clarke, Arabinogalactan-proteins: Structure, biosynthesis, and function. Annual Review of Plant Physiology and Plant Molecular Biology, 1983. **34**: p. 47-70.

133. Faith, J.J., et al., The long-term stability of the human gut microbiota. Science (New York, N.Y.), 2013. **341**(6141): p. 1237439-1237439.

134. Larsen, O.F.A. and E. Claassen, The mechanistic link between health and gut microbiota diversity. Scientific Reports, 2018. **8**(1): p. 2183.

135. Xu, J. and J.I. Gordon, Honor thy symbionts. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(18): p. 10452-10459.

136. Sonnenburg, E.D., et al., Diet-induced extinctions in the gut microbiota compound over generations. Nature, 2016. **529**(7585): p. 212-5.

137. Bartlett, J.G., Antibiotic-associated diarrhea. New England Journal of Medicine, 2002. **346**(5): p. 334-339.

138. Perez-Cobas, A.E., et al., Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. Gut, 2013. **62**(11): p. 1591-1601.

139. Marques, T.M., et al., Programming infant gut microbiota: influence of dietary and environmental factors. Current Opinion in Biotechnology, 2010. **21**(2): p. 149-156.

140. Looft, T., et al., In-feed antibiotic effects on the swine intestinal microbiome. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(5): p. 1691-1696.

141. Wallace, R.J., Ruminal microbiology, biotechnology, and ruminant nutrition: progress and problems. Journal of Animal Science, 1994. **72**(11): p. 2992-3003.

142. Shepherd, E.S., et al., An exclusive metabolic niche enables strain engraftment in the gut microbiota. Nature, 2018. **557**(7705): p. 434-438.

143. Grondin, J.M., et al., Polysaccharide utilization loci: Fueling microbial communities. Journal of Bacteriology, 2017. **199**(15).

144. Wexler, A.G. and A.L. Goodman, An insider's perspective: *Bacteroides* as a window into the microbiome. Nature Microbiology, 2017. **2**: p. 17026.

145. Luis, A.S., et al., Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides.* Nature Microbiology, 2018. **3**(2): p. 210-219.

146. Martens, E.C., H.C. Chiang, and J.I. Gordon, Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host & Microbe, 2008. **4**(5): p. 447-457.

147. Smith, C.J., L.A. Rollins, and A.C. Parker, Nucleotide sequence determination and genetic analysis of the *Bacteroides* plasmid, pBI143. Plasmid, 1995. **34**(3): p. 211-222.

148. Pluvinage, B., et al., Molecular basis of an agarose metabolic pathway acquired by a human intestinal symbiont. Nature Communcations, 2018. **9**(1): p. 1043.

149. Rogowski, A., et al., Glycan complexity dictates microbial resource allocation in the large intestine. Nature Communications, 2015. **6**: p. 7481.

150. Rakoff-Nahoum, S., M.J. Coyne, and L.E. Comstock, An ecological network of polysaccharide utilization among human intestinal symbionts. Current Biology, 2014. **24**(1): p. 40-49.

151. Schwalm, N.D., III, G.E. Townsend, II, and E.A. Groisman, Prioritization of polysaccharide utilization and control of regulator activation in *Bacteroides thetaiotaomicron.* Molecular Microbiology, 2017. **104**(1): p. 32-45.

152. Kamionka, A., et al., Induction of single chain tetracycline repressor requires the binding of two inducers. Nucleic Acids Research, 2006. **34**(14): p. 3834-3841.

153. Jacob, F. and J. Monod, Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology, 1961. **3**(3): p. 318-356.

154. Malan, T.P., et al., Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter. Journal of Molecular Biology, 1984. **180**(4): p. 881-909.

155. Hall, M.P., et al., Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. ACS Chemical Biology, 2012. **7**(11): p. 1848-1857.

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

157. Nielsen, H., Predicting secretory proteins with SignalP. Methods in Molecular Biology, 2017. **1611**: p. 59-73.

158. Bertani, G., Studies on lysogenesis I. : The mode of phage liberation by lysogenic *Escherichia coli.* Journal of Bacteriology, 1951. **62**(3): p. 293-300.

159. Yokoyama, M.T., J.R. Carlson, and L.V. Holdeman, Isolation and characteristics of a skatole-producing *Lactobacillus* sp. from the bovine rumen. Applied and Environmental Microbiology, 1977. **34**(6): p. 837-842.

160. Bacic, M.K. and C.J. Smith, Laboratory maintenance and cultivation of *Bacteroides* species. Current Protocols in Microbiology, 2008. **0 13**: p. 2-24.

161. Hallman, W.K., C.L. Cuite, and X. Morin, Public perceptions of labeling genetically modified foods. 2013, Rutgers University Community Repository.

162. Charlebois, S., et al., Biotechnology in food: Canadian attitudes towards genetic engineering in both plant- and animal-based foods 2018, Dalhousie University.

163. Blancke, S., et al., Fatal attraction: the intuitive appeal of GMO opposition. Trends Plant Sci, 2015. **20**(7): p. 414-8.

164. Rzymski, P. and A. Królczyk, Attitudes toward genetically modified organisms in Poland: to GMO or not to GMO? Food Security, 2016. **8**(3): p. 689-697.

165. Maxmen, A., Genetically modified apple reaches US stores, but will consumers bite?, in Nature News. 2017, Nature Publishing Group.

166. Gaskell, G., et al., Europeans and biotechnology in 2010: Winds of change? - A report to the European Commission's Directorate-General for Research. 2013.

167. Network, C.B.A., Where in the world are GM crops and foods?, in GMO Inquiry. 2015.

168. Denby, C.M., et al., Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. Nature Communications, 2018. **9**(1): p. 965.

169. Clifford, H., AquAdvantage Salmon - a pioneering application of biotechnology in aquaculture. BMC Proceedings, 2014. **8**(Suppl 4): p. O31-O31.

170. Wegmann, U., et al., Use of genetically modified bacteria for drug delivery in humans: revisiting the safety aspect. Scientific Reports, 2017. **7**(1): p. 2294.

171. Ishibashi, M., Y. Sugino, and Y. Hirota, Chromosomal location of thymine and arginine genes in *Escherichia coli* and F' incorporating them. Journal of Bacteriology, 1964. **87**(3): p. 554-561.

172. Gasson, M.J., et al., Characterization and exploitation of conjugation in *Lactococcus lactis.* International Dairy Journal, 1995. **5**(8): p. 757-762.

173. Hirota, R., et al., A novel biocontainment strategy makes bacterial growth and survival dependent on phosphite. Scientific Reports, 2017. **7**: p. 44748.

174. Mandell, D.J., et al., Biocontainment of genetically modified organisms by synthetic protein design. Nature, 2015. **518**(7537): p. 55-+.

175. Lopez, G. and J.C. Anderson, Synthetic auxotrophs with ligand-dependent essential genes for a BL21(DE3) biosafety strain. Acs Synthetic Biology, 2015. **4**(12): p. 1279-1286.

176. Molin, S., et al., Conditional suicide system for containment of bacteria and plasmids. Nature Biotechnology, 1987. **5**(12): p. 1315-1318.

177. Gerdes, K., P.B. Rasmussen, and S. Molin, Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. Proceedings of the National Academy of Sciences of the United States of America, 1986. **83**(10): p. 3116-3120.

178. Torres, B., et al., A dual lethal system to enhance containment of recombinant micro-organisms. Microbiology (Reading), 2003. **149**(12): p. 3595-3601.

179. Canada, H., Eating well with Canada's Food Guide - A resource for educators and communicators, G.o. Canada, Editor. 2011.

180. Langston, C.W. and C. Bouma, A study of the microorganisms from grass silage. II. The *lactobacilli.* Applied Microbiology, 1960. **8**(4): p. 223-234.

Appendix: The AppA Phytase

Background

One of the protein targets I attempted to produce in *B. theta* was the AppA phytase from *E. coli*. The goal for this project was to mitigate the impact of phytate in the diets of monogastric animals. Phytate is a storage form of phosphorus in cereals and legumes [1]. It is manufactured in plant cells by myo-inositol synthases [2], and is degraded by a class of enzymes called phytases [3]. Phytases catalyze the hydrolysis of phytate into inositol and bioavailable inorganic phosphate. Monogastric animals, such as poultry and swine, are unable to efficiently metabolise dietary phytic acid due to a lack of phytase in their gastrointestinal tracts [4]. Poultry and pigs bred for agricultural purposes are commonly fed grain-based diets in which up to 80% of the phosphorus content is present in the form of phytic acid [5]. Grain based diets lack bioavailable phosphorus for monogastric livestock, so their diets must be fortified with mineral phosphorus as it is an essential nutrient for growth [6]. This extra step imparts an additional effort and cost. Phytic acid is also known to inhibit absorption of some other nutrients including amino acids and minerals [7], thereby reducing feed efficiency.

In addition to its detrimental effect in livestock feed, phytate is implicated in environmental problems as well. Excess nutrients from animal manure, particularly phosphorus, can run-off into ground water and cause over-enrichment, termed eutrophication, of water systems. Eutrophication can result in the explosive growth of algae and cyanobacteria known as an algal bloom. These organisms can deplete the dissolved oxygen content of water sources, and produce dangerous toxins [8], and ultimately reducing the economic and ecological value of water systems [9].

Historically, several solutions have been proposed in order to lessen the impacts of phytate in agriculture. Proposals to feed livestock with diets derived from meat and bone, which contain more bioavailable phosphorus, have been met with strong resistance due to concerns over disease. An alternative approach is to supplement feed with an exogenous phytase, such as one derived from fungi [10]. This strategy improves the digestibility of phytic acid by as much as 75% [11], but it is not optimal, as exogenous phytase production is expensive, and feed must be supplemented continuously. To address these concerns, a novel approach called EnviroPig [12] was developed in which a strain of genetically modified pigs was able to express phytase in its salivary glands. This approach was successful, with recorded fecal phosphorus reductions of up to 75% [12]. Unfortunately, public concerns about the safety of genetically modified animals for human consumption resulted in the cancellation of the program in 2012 [13].

The guts of ruminant animals, such as cattle, contain bacteria that produce phytases in sufficient quantity to allow these animals to indirectly metabolize phytic acid [14]. As the gut bacteria of monogastric animals do not produce phytase on a large enough scale to allow for host utilization of phytic acid, a synbiotic approach could be taken to emulate the activity of rumen bacteria. Commensal monogastric gut bacteria can be genetically modified to express phytase, and the feed of the animal can be supplemented with a prebiotic designed to induce the production of this recombinant phytase. The subsequent production of phytase by the monogastric gut bacteria would allow these animals to metabolize phytic acid, thus reducing the amount of phosphorus present in the animal's manure and reducing the detrimental nutrient absorption effects of phytic acid.

This solution would provide a number of practical and economic benefits to the agriculture and livestock industry. Health concerns over disease associated with animal-based feed, production costs associated with feed supplementation, and public hesitancy toward genetically modified animals would be greatly reduced if not eliminated with a synbiotic approach because bacteria which undergo normal proliferation and retention in the gut would be directly responsible for phytic acid metabolism.

Results and Discussion

The AppA phytase was originally produced in *E. coli* with six different signal peptides. Four of the signal peptides were predicted to target the periplasm (Table A.1). 1161 and 1789 are predicted to use the Sec pathway, and TAT1 and TAT2 are predicted to use the twin arginine translocase pathway. The two other peptides; Anch and Sec are predicted to anchor proteins to the outside of the outer membrane and secrete them into the extracellular environment respectively. Phytase with no signal peptide (NS) was also tested.

Five different fractions were taken from cell cultures, the media supernatant (M), periplasms (P), cytoplasms, whole cell soluble protein (S), and whole cell insoluble protein (I). Phytase activity was readily detectable by colorimetric phosphate release assay (Fig A1a), and protein was detected by Western blotting (Fig. A.1b). Dot blots and proteinase K digestion were used to demonstrate localization to the outer membrane (Fig. A.2).

Ultimately only Anch and Sec were chosen for use in *B. theta* due to the reasoning that AppA would be more effective in an animal host if it were outside of the bacterial cell than it would be in the periplasm. Despite preliminary experiments that indicated recombinant phytase activity from *B. theta*, this result could not be reproduced. Different

Table A.1: Signal peptides used to direct localization of the AppA phytase.

Name	Peptide Sequence	Gene	Predicted Localization
1161	MNKRILSVFVCCALYYSAQA	*bt1161* aminopeptidase	Periplasmic
1789	MKKSILIAALGLFSLSAMAQ	*bt1789* aminopeptidase	Periplasmic
TAT1	MDRRNFLKTGGIALLGSLA	*bt3648* putative modulator of DNA gyrase	Periplasmic
TAT2	MDRRDFLKTVAITGAAMTIQ	*bt0776* hypothetical protein	Periplasmic
Anch.	MNKHLHFLSLLWLSMLMAFMTACSDDKNI TDPAPEPEPPVEG	*bt3698 susG*	Outer Membrane
Sec.	MRKVLGLLLLLSVVSAAWAQ	*B. vulgatus bvmpk0229* pectinesterase	Secreted

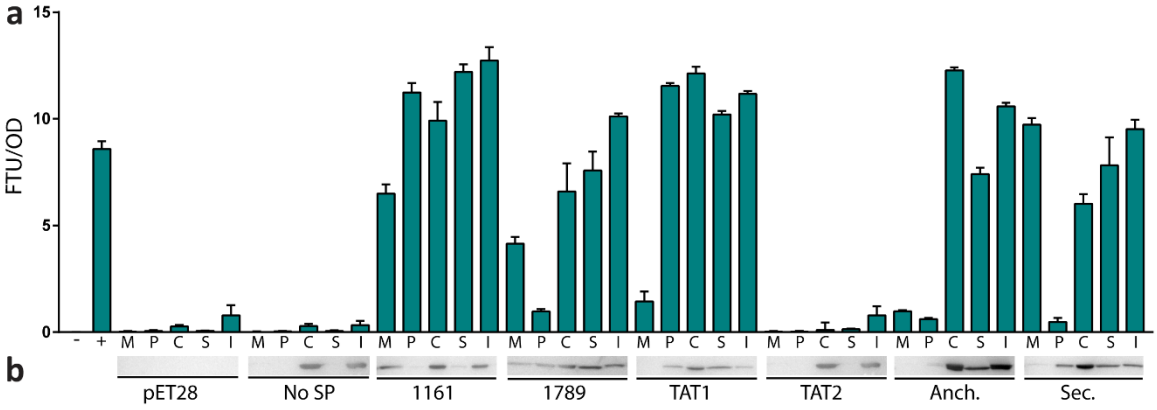


Figure A.1: AppA phytase activity and localization patterns when produced with different signal peptides in *E. coli*. (a) Phosphate release assays were performed with various cell and culture fractions of *E. coli* transformed with plasmids carrying signal peptide tagged variants of the *appA* gene. Activity is expressed in phytase units normalized to the optical density of the cell culture (FTU/OD). One FTU equals on μmol of phosphorus released per minute. Fractions are: concentrated media supernatant (M), periplasms (P), cytoplasms (C), whole cell soluble protein (S), and whole cell insoluble protein (I). **(b)** AppA protein was detected from fractions using Western blotting.

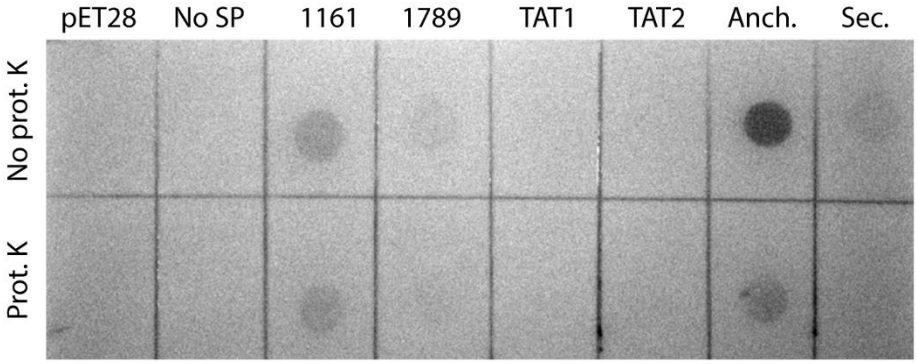


Figure A.2: AppA protein localized to the *E. coli* outer membrane by signal peptides. Whole cells with plasmids carrying signal peptide tagged variants of the *appA* gene were spotted onto a polyvinylidine fluoride membrane and detected by immunohistochemistry. Incubation with proteinase K degrades outer membrane protein and diminishes signal.

combinations of induction time, assay incubation temperature, assay incubation time, and volume of cell sample used were tested, but no activity above baseline could be demonstrated. These tests culminated in a time course experiment that also failed to demonstrate phytase activity in any cell fraction (Fig. A.3). AppA protein was never detected from *B. theta* by Western blot.

In *E. coli* AppA is known to require a chaperone protein in order to fold properly [15], and a BLAST search did not find any *B. theta* protein with homology to this chaperone. However, if lack of a proper chaperone were the barrier preventing AppA from being produced in *B. theta,* one would still expect to see protein by Western blot. qPCR experiments were carried out to gain insight on why AppA could not be produced. These experiments discovered that AppA mRNA was present in the cells where expected (Fig. A.4) which suggests that the barrier to AppA production could be translational.

Following this train of thought, investigations into codon usage differences between *E. coli* and *B. theta* were made. Data from the Codon Usage Table Database [16] was used to create a table comparing the most preferred codon for each amino acid in both species (Table A.2). In some cases, codon usage is very similar between the two species, such as with aspartic acid, but in other cases, codon usage preferences are reversed, such as with alanine. The AppA protein is, assuming that codon frequency equates to ease of translation, less favourable to produce in *B. theta* than it is in *E. coli* (Fig. A.5c). This is the expected result as AppA is a native *E. coli* protein. To examine whether *appA* could be improved, the preferred amino acid chart was then used to construct an optimized version of AppA for *B. theta* (Fig. A.5d). Replacing every codon in *appA* with the most common codon for that amino acid in *B. theta* did result in a sequence that is theoretically easier for

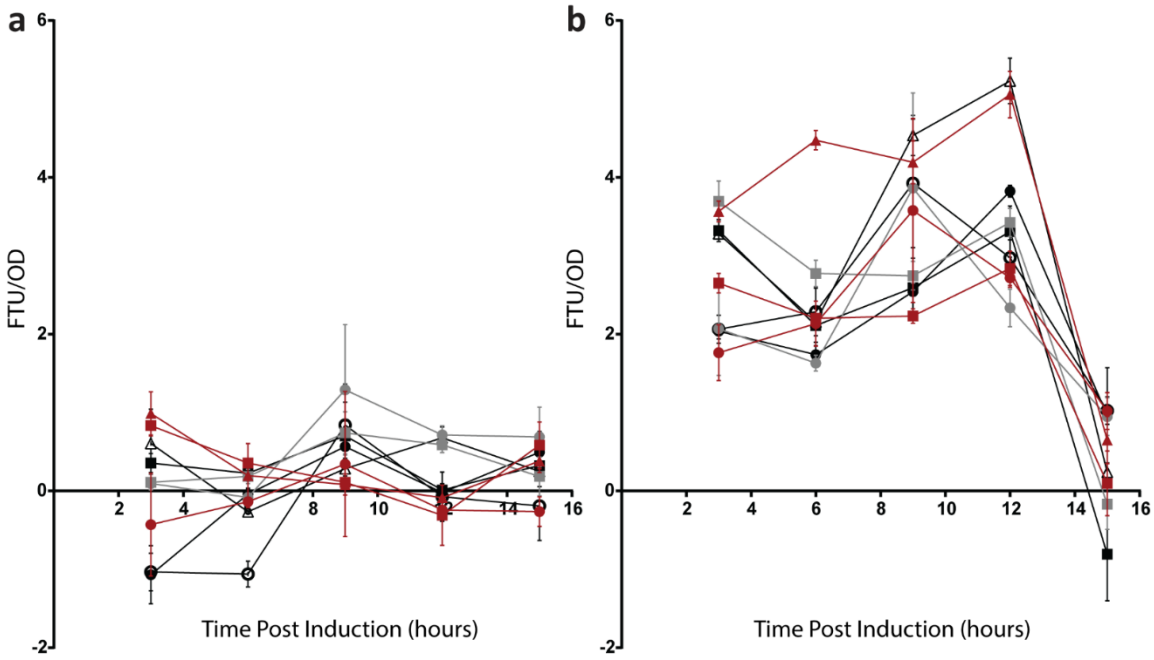


Figure A.3: Lack of meaningful change in phytase activity detected from transgenic *B. theta* harbouring the *appA* phytase under control of different promoters.
Phosphate release assays were performed on the parental *B. theta* strain (red), and strains with *appA* under P_{ON} (grey), P_{DX} (solid black), or P_{AG} (white with black outline) at 3, 6, 9, 12, and 15 hours post induction. Cells were either **(a)** whole, or **(b)** lysed, and were grown on GLC (circles), DX (squares), or AG (triangles).

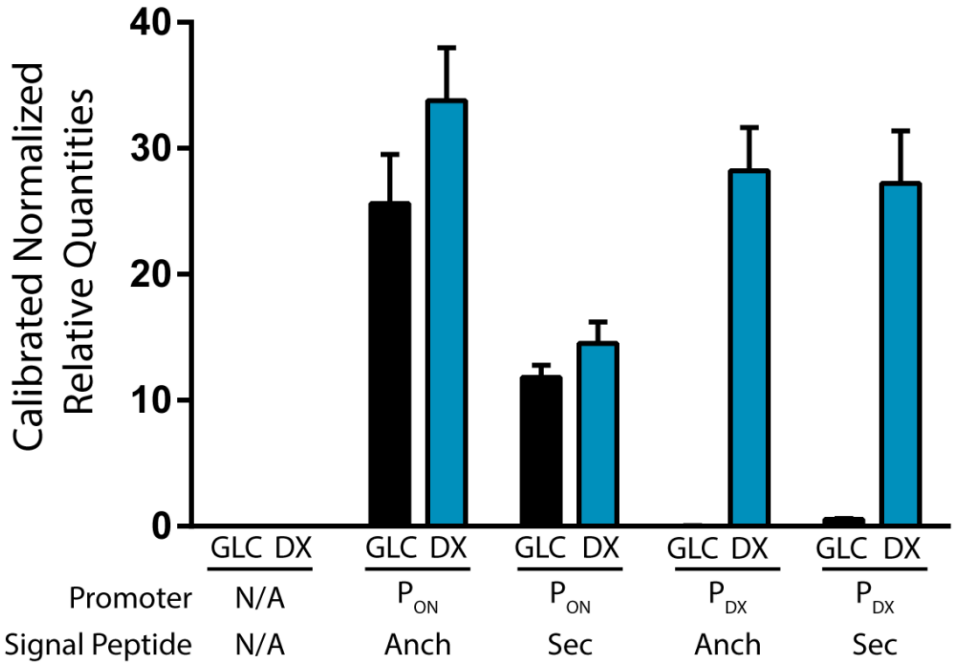


Figure A.4: Detection of *appA* transcript from transgenic *B. theta* by qPCR. Select transgenic *B. theta* strains were grown in minimal medium with either GLC or DX as sole carbon source and assessed to determine whether or not there was *appA* transcript present in the cells.

Table A.2: Codon usage preference for each amino acid compared between *E. coli* and *B. theta*.

Alanine				Arginine				Asparagine				Aspartic acid			
E. coli		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*	
GCG	35.6893	GCT	29.74137	CGC	40.1684	CGT	37.42709	AAC	55.05161	AAT	60.54832	GAT	62.74536	GAT	59.44508
GCC	26.8875	GCA	28.70373	CGT	38.2428	CGC	19.92058	AAT	44.94839	AAC	39.45168	GAC	37.25464	GAC	40.55492
GCA	21.33437	GCC	27.39068	CGG	9.46097	AGA	15.29623								
GCT	16.08883	GCG	14.16423	CGA	6.464214	CGG	13.75477								
				AGA	3.623438	CGA	7.350013								
				AGG	2.040168	AGG	6.251318								

Cysteine				Glutamine				Glutamic acid				Glycine			
E. coli		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*	
TGC	55.49633	TGT	55.55216	CAG	65.22658	CAG	57.06382	GAA	69.14466	GAA	71.99263	GGC	40.45916	GGA	39.69394
TGT	44.50367	TGC	44.44784	CAA	34.77342	CAA	42.93618	GAG	30.85534	GAG	28.00737	GGT	33.84018	GGT	30.73527
												GGG	14.96908	GGC	19.45789
												GGA	10.73158	GGG	10.1129

Histidine				Isoleucine				Leucine				Lysine			
E. coli		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*	
CAT	57.06474	CAT	66.43586	ATT	50.94473	ATC	38.45075	CTG	61.61385	CTG	31.49447	AAA	76.71138	AAA	61.61385
CAC	42.93526	CAC	33.56414	ATC	41.88063	ATT	37.9441	TTA	13.01639	TTG	23.31718	AAG	23.28862	AAG	38.38615
				ATA	7.174639	ATA	23.60515	TTG	12.89689	CTT	16.37113				
								CTC	10.37114	TTA	14.11515				
								CTT	10.36688	CTC	9.562352				
								CTA	3.634631	CTA	5.13972				

Methionine				Phenylalanine				Proline				Serine			
E. coli		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*	
ATG	100	ATG	100	TTT	57.4758	TTC	50.39677	CCG	52.85174	CCG	38.73399	AGC	27.6893	TCT	21.27923
				TTC	42.5242	TTT	49.60323	CCA	19.09257	CCT	30.72688	TCG	15.41953	AGT	19.03855
								CCT	15.78335	CCC	19.33171	AGT	15.01023	TCC	17.15002
								CCC	12.27233	CCA	11.20742	TCC	14.96825	AGC	16.38309
												TCT	14.57994	TCA	13.68785
												TCA	12.33274	TCG	12.46127

Threonine				Tryptophan				Tyrosine				Valine			
E. coli		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*		*E. coli*		*B. theta*	
ACC	43.49709	ACA	28.03709	TGG	100	TGG	100	TAT	57.00171	TAT	69.98407	GTG	37.23618	GTA	34.75624
ACG	26.86793	ACT	27.24653					TAC	42.99829	TAC	30.01593	GTT	25.87264	GTG	26.49868
ACT	16.67304	ACC	25.25097									GTC	21.38127	GTT	21.57518
ACA	12.96194	ACG	19.46542									GTA	15.50991	GTC	17.1699

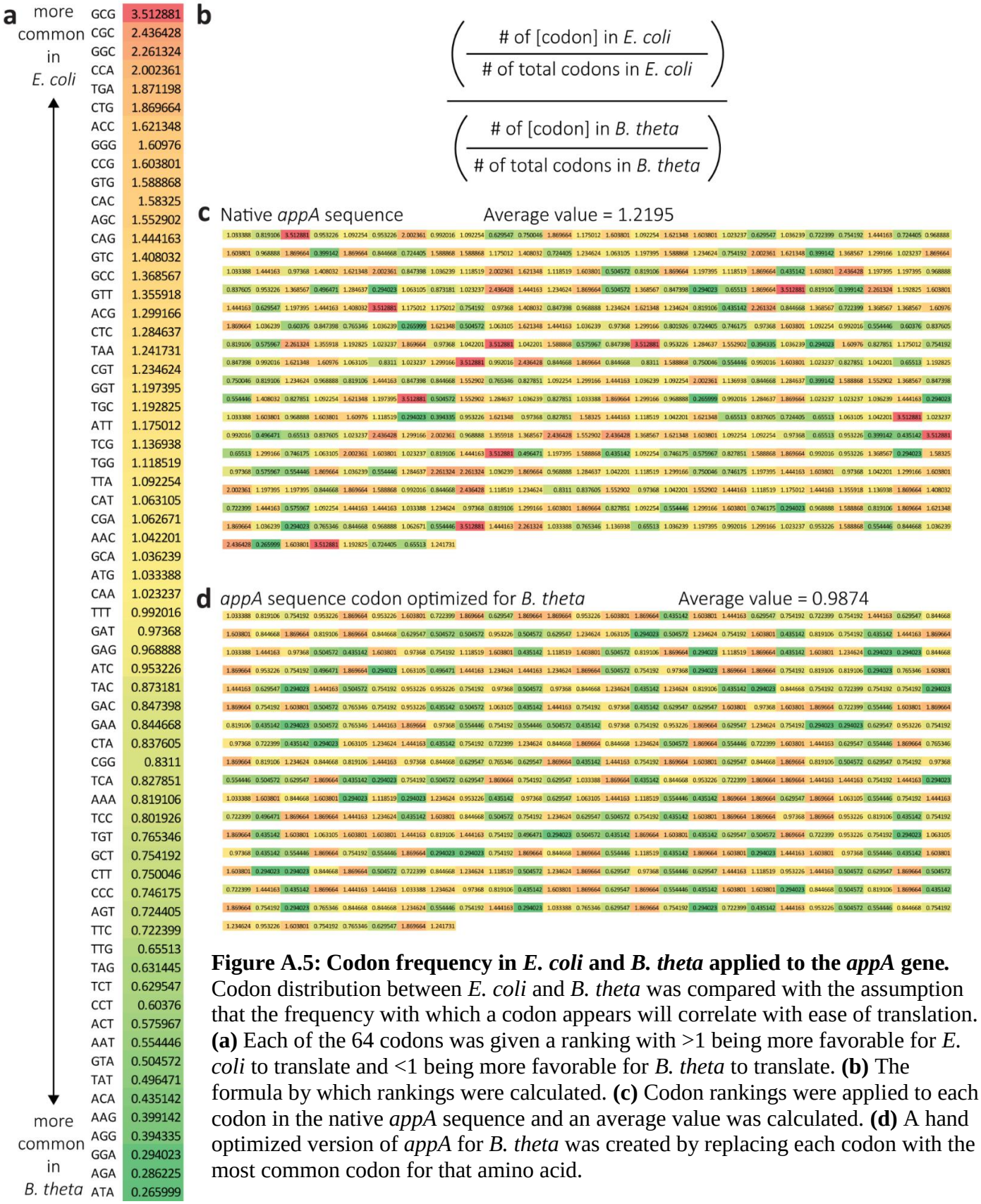


Figure A.5: Codon frequency in *E. coli* and *B. theta* applied to the *appA* gene. Codon distribution between *E. coli* and *B. theta* was compared with the assumption that the frequency with which a codon appears will correlate with ease of translation. **(a)** Each of the 64 codons was given a ranking with >1 being more favorable for *E. coli* to translate and <1 being more favorable for *B. theta* to translate. **(b)** The formula by which rankings were calculated. **(c)** Codon rankings were applied to each codon in the native *appA* sequence and an average value was calculated. **(d)** A hand optimized version of *appA* for *B. theta* was created by replacing each codon with the most common codon for that amino acid.

B. theta to translate. In the future, attempting to produce this version of *appA* in *B. theta* would help to determine if codon bias was preventing *B. theta* from producing the native version of *appA*, and could help to expand the range of heterologous genes that can be used in *B. theta.*

Methods

AppA activity and protein detection in *E. coli*

E. coli BL21 (DE3) cells were transformed with pET28 plasmids containing the *appA* gene and plated on LB agar with kanamycin. The following morning, transformants were used to inoculate LB broth with kanamycin and incubated at 37ºC with aeration at 300 rpm. When the OD600 nm of the broth cultures reach 0.8, 200 µL/L of 1 M isopropyl β-D-1-thiogalactopyranoside (IPTG) was used to induce AppA production. Samples were taken immediately post induction and 3 h post induction.

Periplasmic and cytoplasmic fractions were separated by osmotic shock as follows. A volume of cells were pelleted by centrifugation and resuspended in an ice-cold solution of 20% sucrose, 1 mM ethylenediaminetetraacetic acid, and 20 mM Tris at pH 8. After a 10 min incubation on ice, the cells were pelleted again and resuspended in $1/25^{th}$ volume of room temperature 5 mM $MgCl_2$ and incubated for 10 min followed by another centrifugation step to separate periplasms in the supernatant from intact cytoplasms in the pellet.

18 µL of cell sample was added to a 96 well plate containing 60 µL of substrate solution (freshly prepared 0.2% (w/v) dodecasodium salt phytate in 0.1M sodium acetate buffer, pH 5) and incubated at 37ºC for 1 h. 75 µL of 5% (w/v) trichloroacetic acid was added to each well to stop the reaction before adding 75 µL of freshly prepared colouring

reagent (4 parts of 1.5% (w/v) ammonium molybdate in 5.5% (v/v) H_2SO_4, and 1 part 2.7% (w/v) ferrous sulfate) and mixing by pipetting. Absorbance at 700 nm was promptly measured and, using a standard curve, was used to calculate phytase units (FTUs) with one FTU being equal to 1 µmol of phosphorus released per minute.

For Western blotting, samples were run on 15% sodium dodecyl sulfate gels, transferred to PVDF membranes, and detected with rabbit anti-6-his IgG as described in section 2.6.9 for BuGH16.

For dot blotting, live cells were incubated in Tris-buffered saline either with or without 100 µM proteinase K at 37ºC for 35 min with shaking at 100 rpm. Cells were then spotted onto a PVDF membrane and blocked and detected as described above for Western blotting.

Attempts to produce AppA in *B. theta*

The *appA* gene was integrated into the chromosome of *B. theta* using pINT vectors as described in section 2.6.3. Initially, phosphate release assays were carried out as described for *E. coli* with a few differences. *B. theta* were not freshly transformed with plasmid as the *appA* gene was integrated into their genome instead, *B. theta* could not be agitated because it is a containment level 2 organism and agitation would generate potentially hazardous aerosols, and transgene production was not induced by adding IPTG to the media. Instead, cells were pelleted and resuspended in fresh MM containing either the inducing sugar or a non-inducing control sugar. When the conditions that worked for *E. coli* did not produce results in *B. theta*, the following changes were tested:

- Varying induction time from 3 h to 24 h.
- Performing phosphate release assays at 55ºC instead of 37ºC

- Performing assays in a water bath instead of a dry oven.
- Scaling up assays to perform them in microcentrifuge tubes instead of microplates.
- Varying assay run-time from 30 min to 48 h.
- Varying the amount of cell material in the assay from one-tenth to double the amount used in *E. coli* assays.
- Using twenty times the cell material used in *E. coli* assays and centrifuging to remove cell debris before colourizing.

Unfortunately, reproducible phytase activity was not detectable in *B. theta*. Attempts to detect AppA protein by Western blot were likewise carried out, with variations in concentration of sample loaded, transfer voltage, transfer time, and antibody incubation time.

qPCR in B. theta

Dense TYG precultures were created as described for growth curves in section 2.6.4. 300 µL was used to inoculate 10 mL MM cultures and was incubated at 37°C in an anaerobic atmosphere for approximately 4 h. OD_{600} readings reached ~0.75 for DX cultures and ~1.0 for GLC cultures. 300 µL of the dense MM cultures was combined with 600 µL RNAprotect Cell Reagent (Qiagen) and incubated for 5 minutes before being pelleted by centrifugation and discarding the supernatant. RNAprotect treated cell pellets were stored at -20°C until ready to purify the RNA with a GeneJet RNA kit (Thermo Fisher). Purified RNA was run on an Agilent 2100 Bioanalyzer to ensure quality. The RNA integrity numbers of the samples ranged from 8.1 to 9.5 / 10. cDNA was created from RNA samples with a Quantitect Reverse Transcription Kit (Qiagen). Samples were

run on a QuantStudio 6 Flex Real-Time PCR system as described in 2.6.5. Data were analyzed using qbase+ (Biogazelle).

Codon usage analysis

The Codon Usage Table Database (CUTD) [16] draws information from the GenBank and RefSeq databases and uses that information to construct codon usage tables for each available species and strain. Tables were downloaded on September 4th 2018 from the July 24th 2018 update to the CUTD. Only information for *E. coli* BL21(DE3) and *B. theta* VPI-5482 was considered.

Appendix References

1. Reddy, N.R., S.K. Sathe, and D.K. Salunkhe, Phytates in legumes and cereals, in Advances in Food Research, E.M.M. C.O. Chichester and G.F. Stewart, Editors. 1982, Academic Press. p. 1-92.

2. Hitz, W.D., et al., Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiology, 2002. **128**(2): p. 650-660.

3. Mullaney, E.J. and A.H.J. Ullah, The term phytase comprises several different classes of enzymes. Biochemical and Biophysical Research Communications, 2003. **312**(1): p. 179-184.

4. Klopfenstein, T.J., et al., Animal diet modification to decrease the potential for nitrogen and phosphorus pollution. Council for Agricultural Science and Technology, 2002(21): p. 1-16.

5. Selle, P.H., A.R. Walker, and W.L. Bryden, Total and phytate-phosphorus contents and phytase activity of Australian-sourced feed ingredients for pigs and poultry. Australian Journal of Experimental Agriculture, 2003. **43**(5): p. 475-479.

6. NRC, Nutrient requirements of swine 10th revised rdition. 1998, Washington, D.C: The National Academies Press.

7. Gibson, R.S., et al., A review of phytate, iron, zinc, and calcium concentrations in plant-based complementary foods used in low-income countries and implications for bioavailability. Food Nutrition Bulletin, 2010. **31**(2 Suppl): p. S134-46.

8. Heisler, J., et al., Eutrophication and harmful algal blooms: A scientific consensus. Harmful Algae, 2008. **8**(1): p. 3-13.

9. Smith, V.H., G.D. Tilman, and J.C. Nekola, Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. Environmental Pollution, 1999. **100**(1–3): p. 179-196.

10. Yi, Z., et al., Effectiveness of Natuphos phytase in improving the bioavailabilities of phosphorus and other nutrients in soybean meal-based semipurified diets for young pigs. Journal of Animal Science, 1996. **74**(7): p. 1601-1611.

11. Dersjant-Li, Y., et al., Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. Journal of the Science of Food and Agriculture, 2015. **95**(5): p. 878-896.

12. Golovan, S.P., et al., Pigs expressing salivary phytase produce low-phosphorus manure. Nature Biotechnology, 2001. **19**(8): p. 741-5.

13. Schmidt, S. Genetically engineered pigs killed after funding ends. PostMedia News, 2012.

14. Raun, A., E. Cheng, and W. Burroughs, Ruminant nutrition, phytate phosphorus hydrolysis and availability to rumen microorganisms. Journal of Agricultural and Food Chemistry, 1956. **4**(10): p. 869-871.

15. Berkmen, M., D. Boyd, and J. Beckwith, The nonconsecutive disulfide bond of *Escherichia coli* phytase (AppA) renders it dependent on the protein-disulfide isomerase, DsbC. J Biol Chem, 2005. **280**(12): p. 11387-94.

16. Athey, J., et al., A new and updated resource for codon usage tables. BMC Bioinformatics, 2017. **18**(1): p. 391.