Discovery and Characterization of an Antimicrobial Toxin of the Gut Symbiont Bacteroides fragilis

Discovery and Characterization of an Antimicrobial Toxin of the Gut Symbiont *Bacteroides fragilis*

A dissertation presented

by

Andrew McKinley Shumaker

to

The Committee on Higher Degrees in Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University

Cambridge, Massachusetts

July 2018

Dissertation Advisor: Professor Pamela Silver
Co-Advisor: Professor Laurie E. Comstock

Andrew McKinley Shumaker

Discovery and Characterization of an Antimicrobial Toxin

of the Gut Symbiont *Bacteroides fragilis*

Abstract

The gut microbiota represents one of most densely populated ecosystems on earth. The trillions of bacteria that populate the gastrointestinal tract carry out essential tasks for the host, such as harvesting energy from food, synthesizing micronutrients, and participating in the development of the immune system. Gut bacteria are also implicated in disease, with certain bacteria linked to inflammatory diseases, obesity and depression. Given the significance of the gut microbiota, understanding the factors that shape this complex ecosystem is of great scientific importance.

One of the underappreciated forces that shapes the gut microbiota is antagonism: the process by which species kill or otherwise interfere with other species. Previous work in the Comstock group showed that the gut symbiotic bacterium *Bacteroides fragilis* 638R produces two diffusible toxins that each kill a subset of other strains of *B. fragilis*. I have shown that *B. fragilis* 638R produces a third toxin, BSAP-4, that kills a different subset of strains of *B. fragilis*. I characterized the activity of this toxin by demonstrating its activity spectrum and quantifying its killing behavior. I identified the toxin's receptors in target strains and bioinformatically analyzed the distribution of the toxin and its receptors in natural microbiotas. Finally, I used gnotobiotic mouse models of gut colonization to observe the fitness benefits conferred by the toxin and receptor genes.

To my father, who has always worked hard, encouraged me to do the same, and been a role model for how to be a good man.

To my mother, who has always been the most supportive, loving, patient, and enduring person I know.

To my brothers, with whom I have shared the most memories, the most hardships, and the most laughter.

Table of Contents

Acknowledgements

First, I would like to thank my advisor, Pam Silver, and her husband, Jeff Way, for supporting me throughout this work.

I would also like to thank my collaborator and co-advisor Laurie Comstock for her unwavering dedication and enthusiasm.

Thanks to Ahmad Khalil and Suzanne Walker, of my advisory committee, for their patience and vision.

Thanks to Wade Hicks, Jordan Kerns and David Riglar for their guidance as I embarked on my graduate work.

Thanks to Mike, Maria, Valentina, Leigh, and Leonor, my excellent collaborators who helped me survive the turbulent and wildly entertaining final year of my research.

Thanks to Jason Millberg of the Chemical Biology PhD program for making sure we were always in the right place, at the right time, eating the right food.

Thanks to Bernardo da Costa and Kevin Holden, and all my other colleagues at LS9, for teaching me and helping me apply to graduate school.

Thanks to the adventure crew: Ari, Allison, and Ben for hiking in the mountains and running around on the soccer pitch with me.

Thanks to the Silverinos: Isaac, Tyler, Steph, Sasha, Alina, Bryan, Shannon, Finn, Elena, Aashna, Yasmin, Johannes, and Georg. My time in and out of the lab was always engaging and fun thanks to you.

Special thanks to Marika: you have been a patient and thoughtful companion for the entire ride, and you know that laughter and chocolate make the best rewards and remedies.

Thanks to my peers in the Chemical Biology PhD program, an intimate group of great scientists and excellent people: Dave, Zeb, Matt, Eileen, Javier, Nienke, Liz, Yolanda, Blake, Emma, Daniel, Zhi, Marina, Alix, Sami, Tim, Leigh, Liv, and Angela.

Thanks to GROSS: Dan, Brendan, Karan, Chris, Ian, and Casey. Life is good with a glass of scotch, a rainbow road and some Hyrule.

Thanks to all of the fluffy bunnies and Danasaurs, especially Dee Dee, who was there the day I arrived in Cambridge. I am proud of the community we built and the barbecues we had.

Thanks to Allyson, Richard, Gaelen and Madeleine for your adventurous spirits, your willingness to go further and climb higher, and your bottomless appetites.

Thanks to my family for the love, support, and security you provide, and for equipping me with the humor required to live well. Thank you, Grandpa: I carry your voice with me everywhere. Thank you, Gammy: see you at Rocky.

Thanks to all of my friends from Lafayette, especially Ian, Luke, and David: you always remind me where I come from and what I should aspire to be.

Thanks to Yuan: you are one of the most trustworthy, caring, and smartest people I know, and I am truly grateful to call you my friend.

Thanks to Chris, John, Jon, Max, and Raiden: thank you for bringing out the best in me. Except when you bring out the worst in me.

Chapter 1: The role of antimicrobial proteins in the gut microbiota

Andrew M. Shumaker and Laurie E. Comstock

1.1 Introduction: The composition and structure of the gut microbiota

The human gut harbors an ecosystem of microbes that is among the most complex and densely populated on earth. The gastrointestinal tract contains an estimated 10^{14} bacteria, archaea and eukarya that are collectively referred to as the "gut microbiota." The human individual is composed of roughly the same number of bacterial cells as human cells[1], and comprises 100x the number of bacterial genes (the gut "microbiome") as human genes[2].

The gut microbiota plays an instrumental role in the health of the host. Over the past several decades, researchers have elucidated several key functions of the gut microbiota, including strengthening the gastrointestinal epithelial layer[3], harvesting energy from food[4], and regulating host immunity[5]. The importance of a functioning gut microbiota is highlighted by the reduced density, diversity, and representation of keystone microbial species in the diseased gut[6]. Hallmark gut microbial compositions have been correlated with inflammatory bowel diseases, Crohn's disease, ulcerative colitis[7], cancer[8], depression[9] and obesity[10]. Thus, the factors that organize, maintain and disrupt the gut microbiota are of great scientific and clinical importance.

Researchers have taken advantage of recent developments in nucleic acid sequencing technologies to gain insights into the spatial and temporal mutability and stability of gut microbiotas. Nucleic acids are routinely isolated from fecal samples and analyzed by amplification of 16S rRNA; shotgun sequencing of metagenomes; and transcriptomics from reverse-transcribed mRNAs. These methods reveal the relative abundances of various species and operational taxonomic units (OTUs); the functional capacities of whole microbiotas; and changes in gene expression, respectively. The knowledge we have accumulated through these

techniques provides essential context for ongoing studies, and serves to orient future inquiries about the structure and function of the gut microbiota.

Host physiology and behavior are the primary determinants of the microbial biogeography of the gut[11]. The gastrointestinal tract represents a diversity of habitats that are alternately hospitable and harsh toward microbial life. Between the oral cavity and the colon, the GI tract is characterized by constantly shifting and widely variable conditions, including pH; oxygen content; viscosity; host secretions; bile salts; microbial diversity and density; and availability of every conceivable nutrient. These multidimensional environmental gradients lead to complex niches. Deep sequencing of 16S rRNAs isolated from the mouth, stomach/duodenum, colon, and stool of four human subjects revealed 32-171 OTUs shared in all four sites; significant overlap between colon and stool; and hundreds of OTUs unique to each site[11]. Thus, the gut is populated both by flexible and specialized microbes that occupy either long-ranging or restricted physical sites.

The distribution of gut microbes reveals that their functional capacities correlate with certain environmental gradients—especially oxygen content and nutrient availability—leading to succession-like patterns along the length of the gut. For instance, oxygen availability is positively correlated with representation of facultative aerobes such as the Lactobacillaceae, which are enriched in the stomach and small intestine; strict anaerobes such as the Clostridiaceae are confined to the distal colon, where oxygen is extremely scarce[12]. The distribution of gut bacteria also reflects their digestive preferences. *Streptococcus* spp. and *Escherichia coli* are most abundant in the small intestine, where they can feed on simple carbohydrates; meanwhile, complex plant polysaccharides and dietary fibers remain undigested until they reach the distal colon, where *Bacteroides* spp. are able to metabolize them[13]. In addition to this longitudinal variability, microbial composition varies radially from the interior lumen outward toward the mucosal barrier coating the epithelial cells that line the gut. Studies of folivorous flying squirrels found different species representations within the food bolus and mucus layer of the proximal

small intestine, cecum, and distal large intestine[14]. Disentangling the correlative and causal relationships between microbial localization and effects on host physiology and gut ecology will remain central questions in the study of the gut microbiota.

1.2. Robustness and diversity of the gut microbiota

Studying the spatial and temporal stability of the gut microbiota has yielded insights into how bacteria are able to adapt to changing environments. Dietary shifts can cause rapid changes in the relative abundance of microbiota species[15]. Human subjects have significantly altered microbiotas as quickly as two days after shifting between animal- and plant-based diets. Animal-based diets increase the abundance of bile-tolerant micro-organisms (*Alistipes*, *Bilophila*, and *Bacteroides*) and decrease the abundance of bacteria that specialize in the consumption of plant-based polysaccharides (*Roseburia*, *Eubacterium rectale*, *Ruminococcus bromii*). Unnatural interventions, such as antibiotics treatment, have also been shown to affect microbiota composition. Studies tracking the microbiotas of subjects receiving two courses of ciprofloxacin over 10 months revealed profound community changes only 3-4 days after drug initiation, with partial but incomplete community restoration after drug cessation[16]. The consequences of significantly altered microbiota steady states in patients receiving antibiotics treatment are not well understood in adults; however, antibiotics courses in children are associated with increased risk of inflammatory bowel disease (IBD), with greater hazards in younger children and with multiple rounds of antibiotics[17]. Illustrations of how the environment affects microbiotas, and how microbiotas in turn exert influences on the host organism, will continue to be crucial areas of inquiry in this field.

Despite the remarkable and reproducible short-term mutability of gut microbiotas, they are also characterized by long-term stability, robustness and functional redundancy. Infants acquire their first microbiotas at birth, and the composition of the gut microbiota undergoes several marked changes over the first several years of life. The microbiota reaches relative

stability by age 3[18], typically represented by ~100 species and ~200 strains in an individual[19], but with only 30-40 species representing 99% of all bacterial biomass[20]. These strains are remarkably stable over time—healthy adults sampled 5 years apart retained 60% of the same strains, suggesting that some strains remain resident for decades[19].

While the species contained in individual gut microbiotas can differ greatly from person to person, their metagenomes are characterized by functional redundancy[21]. Because multiple bacterial species are specialized to occupy the various niches of the mammalian gastrointestinal tract, individual microbiotas tend to contain a "core microbiome" at the level of metabolic functions, irrespective of taxonomy. This feature was initially illustrated by studies of obese and lean individuals, which show phylum-level differences in gut microbiota composition contrasting with a high degree of functional microbiome redundancy. The microbiotas of obese individuals tended to be enriched with members of the phyla Firmicutes and Actinobacteria and were characterized by lower overall levels of microbial diversity, while the microbiotas of lean individuals tended to be enriched with members of the phylum Bacteroidetes and had greater levels of microbial diversity. Despite these gross phylogenetic differences, analyzing gene contents according to database annotations of broad functional categories (COG) and metabolic pathways (KEGG) revealed a "core microbiome" of shared functions and pathways. In 18 microbiomes analyzed, 26-53% of enzyme-level functional groups were found in all 18 microbiomes, while 8-22% were unique to a single microbiome. The core groups were also highly abundant, representing 93-98% of sequences collected. Indeed, this extensive overlap facilitates statistical identification of factors that correlate with the phenotype of the superorganism. Analysis of the differences of non-core functional groups between obese and lean individuals revealed 383 genes that were significantly enriched in obese or lean microbiomes. 75% of obesity-enriched genes belong to Actinobacteria, 42% of lean-enriched genes belong to Bacteroidetes, and their functional annotations suggest the genes' involvement in carbohydrate, lipid and amino-acid metabolism. The redundancy of microbiomes across

individuals highlights the outsize influence of the non-redundant sequences in metagenomes, which in turn underscores the imperative of examining the contents and functions of individual strains at the sub-species level.

1.3. The dynamic gut microbiota: sensing, responding, and competing

Despite the abundance of data harvested from these studies, there remain significant gaps in our understanding of how gut bacteria interact and behave in real time. Most of the aforementioned studies have characterized gut microbiotas and microbiomes by observing the relative abundances of species and genes. Differences in these abundances, or changes in these abundances over time, may be inferred to result from increased fitness of enriched species in a given environment. These correlative studies fail to capture microbial behavior and interactions in steady states or in response to perturbation, and are thus incapable of showing causal impacts of genes. Only by measuring specific behaviors and interactions can we appreciate the richness of gut microbial ecosystems[1]. Studying changing microbiotas at the level of OTUs and microbiomes allows researchers to generate hypotheses as to the fitness contributions of various genes. Transcriptomic analysis can lend support to these hypotheses by showing how microbes react to various stimuli. Finally, targeted studies of the metabolic and biochemical capacities of specific bacteria can isolate the *in vitro* and *in vivo* effects of genes and adaptations.

Mechanistic studies have characterized specific adaptations undergirding the abilities of gut bacteria to respond to changing conditions. In times of nutrient scarcity, some bacteria migrate from the interior of the colonic lumen toward the host epithelial layer, where they are able to forage on host glycans and mucins[22]. Other bacteria are capable of forming spores in times of stress, which allows them to persist in a dormant state until conditions improve, or to

[1] By analogy, a drought on the African savannah may lead to the depletion of antelopes and enrichment of lions, but these observations fail to capture the mechanism for the change. Only by identifying the mechanisms by which the perturbation (drought) affects the environment (less plant biomass, fewer watering holes) and its organisms (antelopes become vulnerable, lions benefit from easy predation) does a complete and nuanced understanding of the ecosystem emerge.

disperse to new habitats[23]. Gut bacteria also engage in cooperative and altruistic behavior. In lieu of maintaining over-large genomes encoding the machinery to metabolize any conceivable source of energy, some bacteria specialize and rely on other members of their community for cross-feeding. For instance, *Bacteroides ovatus* secretes enzymes that digest inulin extracellularly, which increases its fitness through reciprocal benefits *via* cross-feeding of *Bacteroides vulgatus*[24]. These mechanistic studies have expanded our conception of the gut microbiome as a truly dynamic and intricate ecosystem.

In addition to the suite of adaptations that allow them to thrive across shifting nutritional landscapes, bacteria seek ecological advantages by interfering with their neighbors to gain exclusive access to niches. Bacteria engage in interference competition—killing or inhibiting the growth of other bacteria—by producing small molecules, peptides or proteins that have either indirectly or directly antagonistic effects. Indirect antagonism comprises inhibitory influences with non-specific mechanisms of action, or that are exerted *via* intermediate processes. Direct antagonism involves molecules with defined targets or receptors in sensitive cells. Given the complexity and density of the gut microbial ecosystem, it is likely that our understanding of the prevalence of antagonistic processes in the gut will continue to grow with further study.

One ubiquitous manifestation of indirect antagonism involves bacterial production of small molecules that render an environment non-specifically inhospitable to other microbes. This relationship is typified by *Clostridium scindens*, which converts primary bile acids to secondary bile acids, which in turn have been shown to inhibit the germination of spores of *Clostridium difficile*[25]. Similarly, some bacterial species produce high concentrations of acetic acid and butyric acid, which interfere with the pathogenicity in *Salmonella*, *E. coli* and *C. difficile*[26]. Another form of indirect antagonism consists of microbial stimulation of a host process that in turn negatively impacts the growth or activity of a subset of gut microbes. *Bacteroides fragilis* PSA-mediated stimulation of regulatory T cells and production of anti-

inflammatory IL-10 by CD4+ T cells strengthens gut barrier function[5] and inhibits the ability of pathogenic bacteria to intoxicate and invade epithelial cells[27]. Conversely, many facultative aerobic pathogenic bacteria indirectly antagonize commensal anaerobic bacteria by stimulating the host inflammatory response[28], which leads to high local concentrations of reactive oxygen species[29]. Many other types of indirect antagonism are obscured by the sheer complexity of the gut microbiota: researchers may infer positive and negative interactions within many-membered consortia of bacteria, but the "active ingredients"—i.e. the specific biomolecules that mediate these interactions—are not readily educed.

Direct antagonism among gut bacteria is mediated by molecules, peptides and proteins that have specific targets in other gut bacteria. In general, antagonistic molecules are extremely diverse in origin, structure and mechanism of action; however, relatively few have been proven to have activity in the gut[2].

1.4. Antimicrobial peptides, proteins, and protein complexes

The majority of antagonistic molecules identified in gut bacteria are ribosomally produced peptides, proteins or protein complexes. Antimicrobial peptides, proteins and protein complexes range in size from class I bacteriocins (<5kDa)[30] to R-type bacteriocins[31] and type VI secretion systems (T6SS)[32], which are phage tail-like multi-protein complexes that can be >100nm in length. Gut-active antimicrobial proteins can be sub-classified into diffusible or contact-dependent categories, which is determined by their mode of production and dispersal. Diffusible toxins are secreted or actively transported by live cells (e.g. microcins with dedicated ABC transporters)[31], or may be released upon cell lysis (e.g. colicins of *E. coli*)[33]. Contact-dependent systems, including the T6SS of *Bacteroides* spp., require the predator cell to

[2] Bacteriophages and their mechanisms, genetics, distributions, and effects on gut evolution and ecology represent a fascinating and important area of study. However, given the complexity of this subject, discussion of phages is beyond the scope of this dissertation.

physically contact the target cells in order to transmit toxic effectors[34]. The activity spectra of these antimicrobial systems are controlled by the receptors in sensitive strains, as well as the cellular processes targeted by toxic effectors. Small secreted bacteriocins typically target cell-surface receptors for attachment and entry into the periplasm or cytoplasm, where they can induce pore formation; degrade nucleic acids or peptidoglycan; or inhibit gyrase activity, cell wall synthesis or protein synthesis[33]. Cells bearing T6SS fire under variable circumstances that are incompletely understood— triggers may include territorial invasion, community defense, and counter-attack[35]—and inject toxic effectors that can degrade nucleic acids, peptidoglycan, or membrane[36].

The receptors and targeted cellular processes of toxins and toxic effectors are striking for their high degrees of conservation, contributions to cellular fitness, or essentiality. For instance, a large fraction of bacteriocins bind to receptors for scarce or essential nutrients or co-factors. Colicins bind to receptors for nucleosides, siderophores, and vitamins[33]; nine different colicins bind to the vitamin-B12 receptor BtuB[37]. Similarly, the cellular processes targeted by toxins, though mediated by different proteins in different species, are essential and universally conserved. Common targets of enzymatic toxins include nucleic acids, the ribosome and the cell wall. Due to the universal requirement for membrane integrity, the mechanisms of pore-forming toxins are also highly general.

The generalities observed in toxin targeting and activity pose challenges for toxin-producing cells, which must retain auto-immunity in order to secure ecological advantages. Cells accomplish this either by synthesis of immunity proteins that block the activity of toxins[34], or by eliminating toxin receptors or targets from their genome. Immunity proteins are co-expressed with toxic effectors, inhibiting their activity until they reach target cells; they may also be constitutively expressed as a defense mechanism. Antimicrobials that do not have immunity proteins include some pore-forming toxins, which are produced by cells containing non-targeted orthologs of their receptors[38].

The aforementioned characteristics of antagonistic molecules have important consequences for their effects on gut ecology. The exposure of any given toxin to the various physiological and geographical features of the gut likely depends on toxin size, structure, mode of production, and characteristics of the producing organism. For instance, small diffusible toxins are likely to access a much greater fraction of the volume of the GI tract than large, contact-dependent secretion systems. Likewise, since species colonize at varying densities in different parts of the gut, the antimicrobial molecules of the most abundant species in a given area are likely to have a greater effect on gut microbial homeostasis than those of less abundant or transient species. The full extent to which antagonistic behaviors contribute to the maintenance of steady states of microbial composition in the gut—or allow species to invade or colonize new niches—is a relatively new area of study, but begs further exploration.

Some of the most abundant species of gut bacteria are the *Bacteroides*, which have recently been shown to host a diversity of antagonistic processes with incredibly prolific and important effects on gut ecology. *Bacteroides* species have been demonstrated to contain three genetic architectures of T6SS[32]. In gnotobiotic mouse models, a *B. fragilis* strain equipped with a T6SS transmitted toxic effectors into target cells of *B. thetaiotaomicron* at a rate exceeding 1 billion events per minute per gram of colonic contents[39]. Other types of toxins have been identified in various *Bacteroides* strains, usually with intra-species spectra of activity[38,40]. Characterization of these antagonistic interactions has increased our understanding of the gut microbiota and accelerated our ability to identify novel toxins and ecological interactions.

1.5. Discovery of secreted toxins of *Bacteroides* species

In 2014, researchers reported non-contact dependent inhibition among strains of *Bacteroides fragilis*[41]. Transposon mutagenesis was used to identify BF638R_1646 as the gene encoding the inhibitory molecule, a predicted MACPF (membrane attack complex/perforin)-domain containing protein. Further investigation confirmed that the MACPF gene encoded a

secreted antimicrobial toxin, which was named BSAP-1 ("*Bacteroides* secreted antimicrobial protein 1"); this marked the first demonstration of an active, bacterially-produced MACPF toxin targeting other bacteria.

MACPFs are pore-forming toxins that can be found in all 3 kingdoms of life. MACPF proteins were initially discovered in components of the immune system. One of the proteins of the complement system, C9, and the pore-forming protein perforin from cytotoxic T lymphocytes were shown to polymerize into pores with a diameter of ~100 Angstroms[42]. Subsequent studies connected the mechanisms to common structural features, anchored by the universally conserved MACPF domain[43]. Other examples of eukaryotic MACPFs have been identified and characterized: sea anemone venom contains hemolytic MACPF proteins[44], and the *Plasmodium* sporozoite microneme protein is a MACPF protein essential for traversing the sinusoidal cell layer during mosquito hindgut infection[45].

Pore forming toxins have also been identified in pathogenic bacteria. The cholesterol dependent cytolysins (CDCs), which belong to the MACPF/CDC superfamily, are a class of PFTs found in diverse Gram positive bacteria including members of the genera *Bacillus, Clostridium, Listeria,* and *Streptococcus*[43]. Several MACPF proteins have been found in pathogenic bacteria, including Plu-MACPF of the insect pathogen *Photorhabdus luminescens*[43].

Genes encoding predicted MACPF proteins are readily identifiable in published genomes, which can be searched for the presence of proteins belonging to the MACPF protein family (PF01823.14). Assignment of uncharacterized proteins to this protein family is based on the presence of a conserved MACPF domain, which consists of several invariant amino acids, 6 random amino acids, and a di-glycine hinge[43]. Phyla with a large fraction of genomes containing at least one MACPF protein include Bacteroidetes (106 genomes containing a MACPF protein out of 466 genomes total) and Chlamydiae (120/137); MACPFs are rare in the rest of the phyla[41]. The diversity of MACPF protein genomes of the Bacteroidetes, which are Gram-negative bacilli with aerobic, facultative anaerobic, and anaerobic representatives, suggests they may play an

important role in gut ecology. 163 protein sequences from the Bacteroidetes have a significant match to the MACPF protein family (PF01823.14), representing 71 groups of 99% similar sequences. Many MACPF proteins are found in sequenced genomes of *B. fragilis*. Despite the bioinformatic ubiquity of MACPF proteins in the Bacteroidetes, BSAP-1 was the first MACPF protein found in gut bacteria with demonstrated killing activity against other gut bacteria, positing the existence of a novel class of interactions among members of the gut microbiota[41].

Additional phenotypic screens for antagonistic interactions between pairs of gut bacteria have expanded the list of putative secreted antimicrobial molecules. In the course of studying BSAP-1, researchers found that *B. fragilis* 638R ΔBSAP-1 retained the ability to antagonize some strains, suggesting that this 638R encoded another antimicrobial toxin in addition to BSAP-1. Transposon mutagenesis identified a ubiquitin-like protein that functioned as an antimicrobial toxin[40]. Named "ubb," this toxin had a distinct spectrum of activity to BSAP-1, augmenting the number of bacteria targeted by wild-type *B. fragilis* 638R.

B. fragilis 638R ΔBSAP-1 Δubb double-knockouts lost activity against some strains, but retained activity against another subset of strains, suggesting the presence of yet another secreted toxin. In this dissertation, I will describe the discovery and characterization of this, the third diffusible toxin of *B. fragilis* 638R. I will highlight the significance of this toxin and its receptor by describing its distribution in natural microbiotas. Finally, I will discuss the implications of this work as it relates to microbial evolution, gut microbial ecology, and applications in biotechnology and pharmaceuticals.

1.6 References

1. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biol.* **14,** e1002533 (2016).
2. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312,** 1355–9 (2006).

3. Berg, R. D. The indigenous gastrointestinal microflora. *Trends Microbiol.* **4,** 430–435 (1996).
4. Topping, D. L. & Clifton, P. M. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiol. Rev.* **81,** 1031–1064 (2001).
5. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122,** 107–18 (2005).
6. Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M. & Owen, L. J. Dysbiosis of the gut microbiota in disease. *Microb. Ecol. Health Dis.* **26,** (2015).
7. Borody, T. J., Warren, E. F., Leis, S., Surace, R. & Ashman, O. Treatment of ulcerative colitis using fecal bacteriotherapy. *J. Clin. Gastroenterol.* **37,** 42–47 (2003).
8. McIntyre, A., Gibson, P. R. & Young, G. P. Butyrate production from dietary fibre and protection against large bowel cancer in a rat model. *Gut* **34,** 386–391 (1993).
9. Fond, G. *et al.* Anxiety and depression comorbidities in irritable bowel syndrome (IBS): a systematic review and meta-analysis. *Eur. Arch. Psychiatry Clin. Neurosci.* **264,** 651–660 (2014).
10. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–3 (2006).
11. Stearns, J. C. *et al.* Bacterial biogeography of the human digestive tract. *Sci. Rep.* **1,** (2011).
12. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489,** 220–30 (2012).
13. Zoetendal, E. G. *et al.* The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *ISME J.* **6,** 1415–1426 (2012).
14. Lu, H.-P. *et al.* Spatial heterogeneity of gut microbiota reveals multiple bacterial communities with distinct characteristics. *Sci. Rep.* **4,** (2015).

15. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505,** 559–563 (2014).
16. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6,** e280 (2008).
17. Kronman, M. P., Zaoutis, T. E., Haynes, K., Feng, R. & Coffin, S. E. Antibiotic exposure and IBD development among children: a population-based cohort study. *Pediatrics* **130,** e794-803 (2012).
18. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–227 (2012).
19. Faith, J. J. *et al.* The long-term stability of the human gut microbiota. *Science* **341,** 1237439 (2013).
20. Savage, D. C. Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.* **31,** 107–133 (1977).
21. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–4 (2009).
22. Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307,** 1955–9 (2005).
23. Swick, M. C., Koehler, T. M. & Driks, A. Surviving Between Hosts: Sporulation and Transmission. *Microbiol. Spectr.* **4,** (2016).
24. Rakoff-Nahoum, S., Foster, K. R. & Comstock, L. E. The evolution of cooperation within the gut microbiota. *Nature* **533,** 255–9 (2016).
25. Sorg, J. A. & Sonenshein, A. L. Inhibiting the Initiation of Clostridium difficile Spore Germination using Analogs of Chenodeoxycholic Acid, a Bile Acid. *J. Bacteriol.* **192,** 4983–4990 (2010).

26. Sun, Y. & O'Riordan, M. X. D. Regulation of Bacterial Pathogenesis by Intestinal Short-Chain Fatty Acids. *Adv. Appl. Microbiol.* **85,** 93–118 (2013).
27. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–5 (2008).
28. Winter, S. E. *et al.* Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature* **467,** 426–429 (2010).
29. Rivera-Chavez, F. *et al.* Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella. *Cell Host Microbe* **19,** 443–54 (2016).
30. Delves-Broughton, J., Blackburn, P., Evans, R. J. & Hugenholtz, J. Applications of the bacteriocin, nisin. *Antonie Van Leeuwenhoek* **69,** 193–202 (1996).
31. Gebhart, D. *et al.* Novel High-Molecular-Weight, R-Type Bacteriocins of Clostridium difficile. *J. Bacteriol.* **194,** 6240–6247 (2012).
32. Coyne, M., Roelofs, KG, Comstock, LE. Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements. *BMC Genomics* **17,** (2016).
33. Cascales, E. *et al.* Colicin biology. *Microbiol Mol Biol Rev* **71,** 158–229 (2007).
34. Chatzidaki-Livanis, M., Geva-Zatorsky, N. & Comstock, L. E. Bacteroides fragilis type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species. *Proc Natl Acad Sci U A* **113,** 3627–32 (2016).
35. Gallique, M., Bouteiller, M. & Merieau, A. The Type VI Secretion System: A Dynamic System for Bacterial Communication? *Front. Microbiol.* **8,** (2017).
36. Russell, A. B., Peterson, S. B. & Mougous, J. D. Type VI secretion system effectors: poisons with a purpose. *Nat Rev Microbiol* **12,** 137–48 (2014).
37. Sabet, S. F. & Schnaitman, C. A. Purification and properties of the colicin E3 receptor of Escherichia coli. *J. Biol. Chem.* **248,** 1797–1806 (1973).

38. Roelofs, K. G., Coyne, M. J., Gentyala, R. R., Chatzidaki-Livanis, M. & Comstock, L. E. Bacteroidales Secreted Antimicrobial Proteins Target Surface Molecules Necessary for Gut Colonization and Mediate Competition In Vivo. *mBio* **7,** e01055-16 (2016).
39. Wexler, A. G. *et al.* Human symbionts inject and neutralize antibacterial toxins to persist in the gut. *Proc Natl Acad Sci U A* **113,** 3639–44 (2016).
40. Chatzidaki-Livanis, M. *et al.* Gut Symbiont Bacteroides fragilis Secretes a Eukaryotic-Like Ubiquitin Protein That Mediates Intraspecies Antagonism. *MBio* **8,** (2017).
41. Chatzidaki-Livanis, M., Coyne, M. J. & Comstock, L. E. An antimicrobial protein of the gut symbiont Bacteroides fragilis with a MACPF domain of host immune proteins. *Mol Microbiol* **94,** 1361–1374 (2014).
42. Young, J. D., Cohn, Z. A. & Podack, E. R. The ninth component of complement and the pore-forming protein (perforin 1) from cytotoxic T cells: structural, immunological, and functional similarities. *Science* **233,** 184–190 (1986).
43. Rosado, C. J. *et al.* The MACPF/CDC family of pore-forming toxins. *Cell. Microbiol.* **10,** 1765–1774 (2008).
44. A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone Actineria villosa. - PubMed - NCBI. Available at: https://www.ncbi.nlm.nih.gov/pubmed/15019483. (Accessed: 28th June 2018)
45. Kaiser, K. *et al.* A member of a conserved Plasmodium protein family with membrane-attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. *Mol. Biochem. Parasitol.* **133,** 15–26 (2004).

Chapter 2: Identification of a MACPF toxin in *B. fragilis*

Andrew M. Shumaker, Maria Chatzidaki-Livanis, Michael J. Coyne and Laurie E. Comstock

2.1 Antimicrobial interactions in the gut

The interference competition mediated by contact-dependent and secreted antimicrobial toxins is likely to play a significant role in the makeup of the gut microbiota. Antimicrobial proteins and peptides have been extensively studied in *Pseudomonas*[1], *E. coli*[2], Gram positive lactic acid bacteria[3], and streptococci[3]. The toxins of these organisms have been pursued primarily due to their well understood physiology, facile genetic tractability, and/or status as opportunistic pathogens. However, these organisms do not colonize the gut in high concentrations. Until recently, the antagonistic behaviors of the most abundant gut bacteria—the Firmicutes and the Bacteroidetes—have remained unappreciated.

Over the past decade, researchers have leveraged sequencing technologies to scour the gut microbiome for genes that encode antimicrobial peptides, proteins, and protein complexes. *Bacteroides* spp. have been revealed to contain three different genetic architectures of type-VI secretion systems (T6SS)[4]; numerous biosynthetic gene clusters with predicted NRPS/PKS/bacteriocin products have been identified in the Firmicutes, but few have been characterized[5]; and hundreds of MACPF-domain containing proteins have been bioinformatically identified but remain to be functionally characterized[6].

In parallel with the explosion of genetic data emanating from gut microbiome research, an increased interest and capability to culture and genetically manipulate gut bacteria has shed light on their behavior *in vitro* and *in vivo*. As the molecular biology and genetic toolkit available to microbiologists expands, our ability to rapidly conduct sophisticated experiments has accelerated our understanding of how gut bacteria collaborate and compete. The findings of

these experiments will expand our conception of how these behaviors affect gut ecology and microbial evolution.

Antimicrobial interactions can be functionally identified *in vitro* through a variety of assays that track bacterial growth in various conditions. Strains producing diffusible antimicrobial toxins can be identified using the spot agar test[6], in which a bacterial colony of killer cells is grown for 24hr on solid media. Bacterial cells are removed by swabbing, and any remaining cells killed by chloroform vapor, leaving any secreted molecules embedded in the agar matrix. Sensitive cells are grown to mid-log phase, suspended in molten "top agar" media containing 0.75% agar at 50degC, and poured over the petri dishes. Any secreted toxins in the petri dish will kill cells in the overlaid top agar, resulting in a clear spot of no growth in the lawn where the toxin-producing strain was spotted. Thus, the spot agar test delivers a qualitative, binary result of whether or not the growth of the sensitive strain is inhibited by toxin-producing strain.

Toxins that are secreted into the extracellular milieu can also be studied quantitatively by measuring the dose response of sensitive cells in the presence of varying concentrations of toxin. This assay, known as the liquid survival titration assay[1], involves incubation of a known concentration of target cells with a known concentration of toxin. After a fixed period of time, sensitive cells are titered and colony forming units measured to infer the fraction of cells killed by the toxin during co-incubation. In some instances, the liquid survival titration assay is an effective method of comparing the potencies of different toxins.

Toxin activity may also be evaluated in ecological and evolutionary terms. The competitive advantage provided by toxin expression may be evaluated in animal models, wherein toxin-producing and toxin-sensitive cells are inoculated together or in isolation. These experiments offer simplified measures of how toxin expression limits the ability of a sensitive strain to colonize the gut.

Close study of antagonistic interactions among specific gut bacteria will be crucial to understanding how these interactions affect the composition of the gut microbiota and its stability. In this chapter, I will describe transposon mutagenesis and its application in the identification of novel bacterial toxins. Next, I will describe a transposon mutant that led to the identification of a MACPF-domain containing toxin in *B. fragilis* 638R. Finally, I will describe the genetic and biochemical methods used to identify the receptor of the toxin and its killing spectrum.

2.2 Transposon mutagenesis identifies genes involved in antimicrobial production

Identification of the genetic determinants of empirically-observed antimicrobial interactions is straightforwardly accomplished by mutagenesis. Transposon mutagenesis constitutes a method in which target cells receive by conjugation a vector that cannot replicate. The vector is equipped with a transposon containing an antibiotic resistance gene that integrates randomly into the target cell genome. Cells with insertions are selected on the corresponding antibiotic, yielding a library of unique insertions in the host genome. This library may then be screened for a desired phenotype: in the case of antimicrobial production, loss of the ability to kill target cells.

Transposon mutagenesis can be a powerful technique, but it relies on several prerequisites. First, the cells to be mutagenized must be genetically tractable. Second, the cells must be able to express both the transposase and the antibiotic resistance marker—and third, they must be sensitive to the corresponding antibiotic.

The most efficient means of introducing DNA into *Bacteroides* species is by conjugation. The efficiency of conjugal mating into *Bacteroides* species is highly variable for reasons that are not fully understood. Parameters that may affect conjugation efficiency include envelope physiology and defense mechanisms such as restriction-modification systems[7].

Random transposition of conjugally transferred DNAs is efficiently catalyzed by mariner transposases. Mariner transposases are able to catalyze insertion in target genomes without accessory host factors, and have been shown to function in all three domains of life[8]. The himar1C9 transposase, a hyperactive mutant of the Himar1 mariner transposase which excises and inserts DNA fragments flanked by 29-base pair inverted repeat (IR) sequences[9], was initially shown to function in *Bacteroides thetaiotaomicron*[10]. The plasmid described in Goodman et al., pSAM_Bt, contains a *B. thetaiotaomicron* promoter (BT_1311) driving the himar1C9 transposase. In order to study the prolific foraging strain *Bacteroides cellulosilyticus* WH2, McNulty et al. modified pSAM_Bt by introducing the *BWH2_3183* promoter—which drives expression of EF-Tu in *B. cellulosilyticus* WH2—in place of the native promoter driving expression of the ermG antibiotic resistance gene[11]. The addition of this promoter dramatically increased the number of antibiotic-resistant mutants recovered in the transposon screen. The efficiency increases effected by the introduction of strong, constitutive promoters upstream of transposase and antibiotic-resistance genes highlights the importance of native transcriptional machinery when designing transposon mutagenesis experiments.

Finally, transposon mutagenesis is enabled by selection of transposon mutants on antibiotics. The study of natural bacterial isolates is often hampered at this step because bacteria possess manifold resistances to common selection markers[12]. Genes encoding antibiotic resistance are prevalent in the human gut microbiota, and are readily transferred among species by horizontal gene transfer of plasmids, integrative conjugative elements (ICE), and other mobile elements[13]. The widespread resistance of bacteria to many classes of antibiotics is abetted by the selection pressure applied by rampant deployment of antibiotics in the medical, food and agriculture industries[14]. Exposure of organisms to antibiotics decreases the fraction of bacteria lacking antibiotic resistance genes and selects for resistant members of microbiotas[15]; furthermore, antibiotic exposure activates excision and transfer of ICE[16], which increases HGT and the distribution of antibiotic resistance genes in microbiomes.

The use of antibiotics in genetic studies of *Bacteroides* is particularly challenging. The primary type strains of *Bacteroides*, which were isolated earlier in the 20th century prior to the antibiotics explosion (one type strain, *B. fragilis* NCTC 9343, was isolated in 1955 from an abdominal abscess in London), are naturally resistant to the most frequently utilized antibiotics in the lab, such as aminoglycosides (gentamicin, kanamycin, streptomycin and spectinomycin), chloramphenicol, and beta-lactams (ampicillin, carbenicillin, cefoxitin)[12]. Furthermore, contemporary *Bacteroides* species isolated from humans are likely to contain genes that confer resistance to clindamycin (ermG, e.g.), macrolides such as erythromycin (ermG, e.g.), tetracyclines (tetQ, e.g.), and fluoroquinolones (ciprofloxacin, nalidixic acid)[17,18]. Transposon mutagenesis is thus limited to those strains that either a) are sensitive to antibiotics for which a functioning resistance cassette exists, or b) have been otherwise manipulated to become sensitive to antibiotics while retaining viability and functions of interest. One such manipulation involves chemical mutagenesis by UV irradiation or ethyl methanesulfonate (EMS) mutagenesis, which may find applications in rendering wild bacterial strains genetically tractable (Appendix A).

Three erythromycin-sensitive strains of *Bacteroides fragilis* were shown to express antimicrobial activity in spot agar tests. The first strain we selected for mutagenesis was *B.f.* 638R ΔBSAP-1 Δubb, the genesis of which is described in the introductory chapter of this dissertation[6]. This strain had antimicrobial activity against *B. fragilis* strains 3_1_12 and J38-1 (fig. 2.1a), which we suspected would simplify receptor identification in subsequent experiments. We also selected *B.f.* NCTC9343 and *B.f.* CM13 for mutagenesis because had robust antimicrobial activity against *B.f.* 3_1_12 in spot agar tests. I subjected these putative toxin-producing strains to transposon mutagenesis by mating *E. coli* S17 lambda pir cells harboring the plasmid pWH2 (pSAM_Bt with the WH2 promoter driving ermG, as described above[11]) with selection on erythromycin. Colonies of the resulting libraries were picked into 8x12 arrays on rectangular petri dishes, grown overnight and then replica-plated. The original

arrayed plates were then subjected to spot agar tests by removing cells with Whatman paper, killing remaining cells with chloroform vapor, and overlaying with the sensitive strain *B.f.* 3_1_12. Overlay plates were incubated overnight and then analyzed for abrogated antimicrobial activity. In total, 768 mutants were screened in the 638R library; 1,344 were screened in the 9343 library; and 1,152 were screened in the CM13 library. While most transposon mutants retained antimicrobial activity, some appeared to lose antimicrobial activity as observed by microbial growth above specific sectors of the spot agar dish. Candidate transposon mutants were selected from the replicated plates and the spot agar test was repeated. One mutant had confirmed abrogration of antimicrobial activity in the 638R library (fig. 2.1b); three in the 9343 library (data not shown); and two in the CM13 library (data not shown). Mutants with confirmed loss of antimicrobial activity were saved for further study.

B. fragilis host strain	Gene disrupted	Nucleotide start / gene length (bp)	Protein annotation
638R	BF638R_2714	240/1521	MACPF-domain containing protein
9343	9343_2668	73/1296	putative two component system histidine kinase
9343	9343_2073	105/2766	Conserved hypothetical exported protein
9343	9343_2077	134/2355	Putative TonB-related exported protein
CM13	CM_01202	2/1386	Putative phosphoglucomutase/ phosphomannomutase family protein
CM13	CM_01343	248/1224	Putative transmembrane protein

Table 2.1. Genetic loci disrupted in transposon mutants of *B. fragilis* with abrogated antimicrobial activity against *B. f.* 3_1_12. Nucleotide start indicates precise junction of transposon insertion within the disrupted gene. Protein annotation represents the top-scoring homologies resulting from HHpred search of the PDB, pfam, and TIGRFAMS databases[19].

The genetic loci disrupted by transposon insertion were identified by arbitrary PCR and Sanger sequencing. Arbitrary PCR—also known as nested semi-random PCR—involves a semi-random primer paired with a primer homologous to a sequence within the transposon, which together hybridize in the region of the transposon and generate a small amount of short amplicons with chimeric genomic-transposon sequences[20]. Subsequent PCR cycles with nested

primers enrich these amplicons, allowing them to serve as templates for Sanger sequencing initiated by the nested PCR primers. The resulting sequence aligns to the end of the transposon cassette before reading into the genome, revealing the precise location of transposon insertion. Table 2.1 details the loci disrupted by transposon insertions in *B.f.* mutants with disrupted antimicrobial activity against *B.f.* 3_1_12.

2.3 BF638R_2714 is a MACPF protein that kills several strains of *B. fragilis*

I selected the transposon mutant in *B.f.* 638R for further study because it localized to a gene encoding a protein annotated as a known class of toxins: the MACPF (membrane attack complex/perforin) domain-containing proteins (fig. 2.1c). To further show that BF638R_2714 indeed encodes toxin activity, the gene was cloned into the expression vector pMCL140, which drives transcription from a strong constitutive promoter, and mated into *Bacteroides thetaiotaomicron* VPI-5482. The resulting strain showed strong inhibition of *B.f.* 3_1_12 and several other strains in the agar spot assay, while *B. thetaiotaomicron* with the empty vector did not (fig. 2.1d).

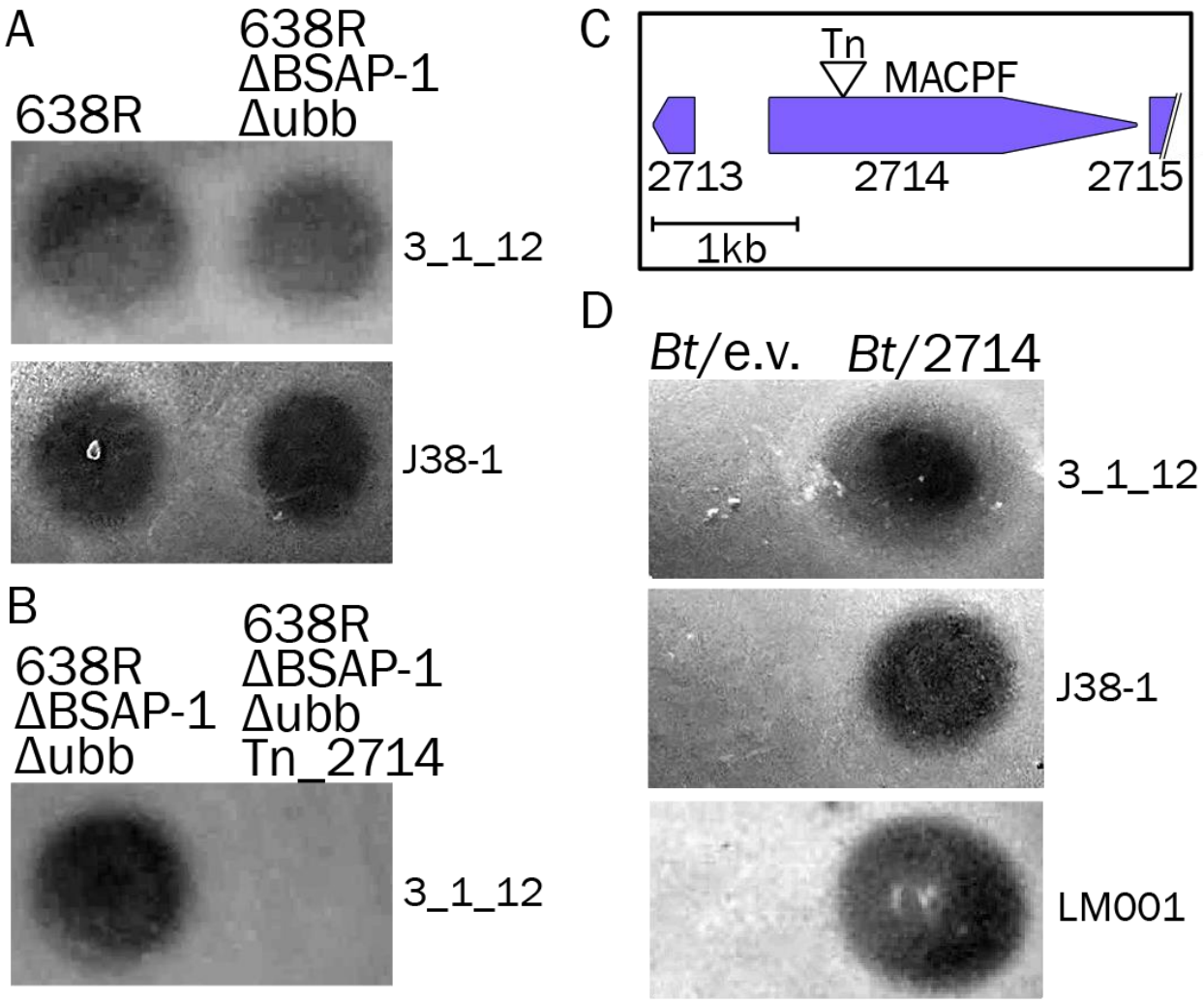


Figure 2.1. Transposon mutagenesis of *B. fragilis* 638R identifies BF638R_2714 as gene implicated in production of a secreted toxin. a) 638R with deleted BSAP-1 and ubiquitin-like toxin retains the ability to kill *B.f.* 3_1_12 and J38-1. b) A transposon mutant displayed abrogated antimicrobial activity against 3_1_12. c) Sanger sequencing showed that the transposon cassette integrated in BF638R_2714, a gene encoding a predicted MACPF-domain containing protein. d) Heterologous expression of BF638R_2714 in *B. thetaiotaomicron* confers the ability to kill several strains of *B.f.*, while the empty vector does not.

To exclude the possibility that polar effects were responsible for the loss of antimicrobial activity in the transposon mutant, a clean non-polar deletion of BF638R_2714 was made in *B.f.* 638R ΔBSAP-1 Δubb to generate *B.f.* 638R ΔBSAP-1 Δubb Δ2714. This was accomplished by amplifying 2,500bp fragments upstream and downstream of BF638R_2714, stitching them together to retain the first 25 amino acids and last 9 amino acids of the gene, and cloning into a prepared pKNOCK vector backbone. After mating into *B.f.* 638R ΔBSAP-1 Δubb, erythromycin-

resistant co-integrates were allowed to recombine in the absence of antibiotic selection and plated to BHIS plates. These colonies were replica-plated to BHIS plates supplemented with erythromycin and erythromycin-sensitive clones screened for recombination events leading to the scarless, in-frame, non-polar deletion. The resulting triple-deletion strain lost the ability to inhibit the growth of *B.f.* 3_1_12 in the spot agar test. Complementation of the triple-deletion strain with a plasmid expressing BF638R_2714 restored growth inhibition against *B.f.* 3_1_12 (fig. 2.2).

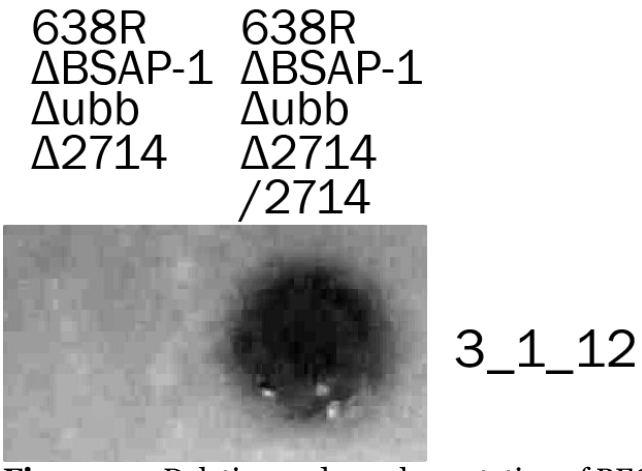


Figure 2.2 Deletion and supplementation of BF638R_2714 abolishes and restores killing of *B. fragilis* 3_1_12.

I next sought to show that the purified protein product of BF638R_2714 is sufficient to inhibit the growth of *B.f.* 3_1_12. The predicted product of BF638R_2714 sports a 29-aa N-terminal signal peptidase II (SpII) signal sequence, suggesting it is exported as a 477-amino acid, 53kDa protein. First, the gene was PCR amplified to omit the native SpII signal sequence and cloned into the *E. coli* expression vector pET16b to introduce an N-terminal poly-histidine tag. The resulting construct was transformed into *E. coli* BL21(DE3), induced with IPTG and protein purified over nickel-NTA resin. The purified, His-tagged protein inhibited the growth of *B.f.* 3_1_12 and in the agar spot assay in a dose-dependent fashion (fig. 2.3). This result supported the genetic results from transposon mutagenesis, heterologous expression, deletion

and complementation: that BF638R_2714 encodes a secreted antimicrobial protein that acts against *B.f.* 3_1_12.

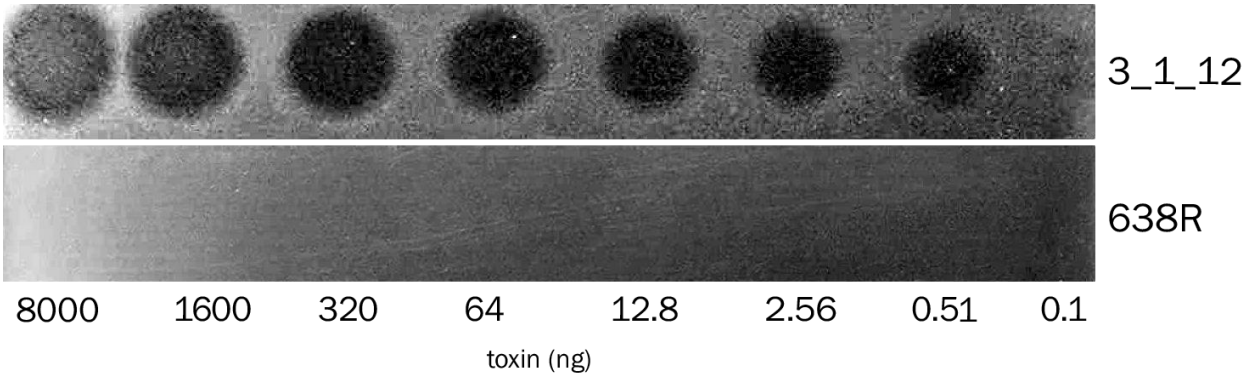


Figure 2.3. Purified His-tagged BF638R_2714 inhibits growth of *B.f.* 3_1_12 (top) in dose-dependent fashion. Toxin was added to petri dish by dispensing 5µl of His-tagged toxin serially diluted in PBS. *B.f.*638R (bottom) is not targeted by the toxin.

A previous study did not detect toxin activity by BF638R_2714 because the in panel of six *B.f.* strains against which it was tested for sensitivity, none were sensitive[6]. My results confirm that BF638R_2714 is a toxin. Given BF638R_2714's robust inhibitory activity against *B.f.* 3_1_12, I sought to measure its spectrum of activity against a battery *B.f.* species and strains. 6 out of 13 *B.f.* strains tested were sensitive. Spot agar tests against *B.f.* 3_1_12, S36L11, J38-1, CL03T12C007, LM46, and LM001 showed varying levels of sensitivity as observed by zones of clearing of varying sizes and characters (fig. 2.4). Notably, the set of strains inhibited by BF638R_2714 partially but incompletely overlaps with the set of strains inhibited by the other two toxins in the *B.f.* 638R genome (fig. 2.5)

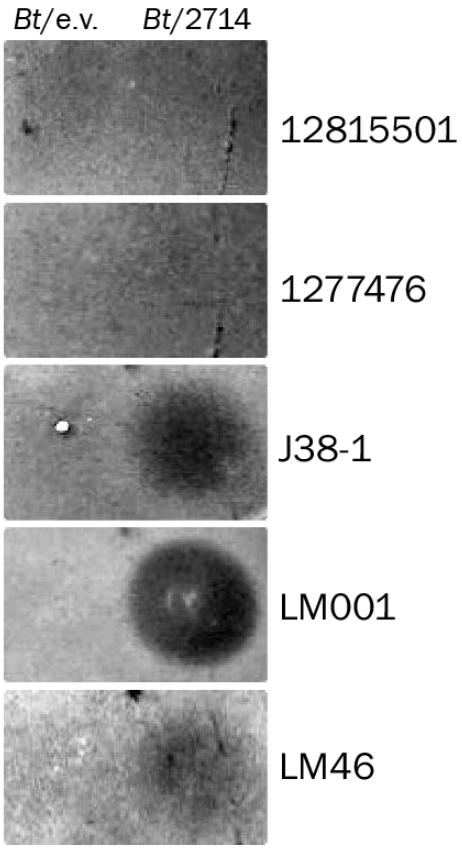


Figure 2.4 *B. thetaiotaomicron* expressing BF638R_2714 kills several wild-type strains of *B. fragilis*. Strains J38-1, LM001, and LM46 were killed by *B.t.* expressing toxin, but not by *B.t.* with the empty vector; 12815501 and 1277476 were not targeted.

2.4 Distribution of MACPF toxins in nature and in Bacteroidetes

MACPF-domain containing proteins belong to a family of pore-forming toxins that can be found in every kingdom of life. Examples of MACPF proteins include the C6-C9 complement proteins of the human immune system[21]; vertebrate perforin, which is delivered by natural killer cells and cytotoxic T cells[22]; sea anemone toxins[23]; the *Plasmodium sporozoite* microneme protein essential for traversing the sinusoidal cell layer[24]; and the Clostridial cholesterol-dependent cytolysins (CDCs), which target mammalian cells[25]. MACPF proteins intoxicate target cells by assembling in 15-18mers before undergoing a conformational change involving an unfolding of beta sheets that insert into the membrane of the target cell, forming a pore ~10nm in diameter[26]. While MACPF domains are highly divergent in sequence, they tend to share a thin

L-shaped structure of four antiparallel beta sheets, and contain a highly conserved motif Y/W-G-T/S-H-F/Y-X6-GG[27]. Upon polymerization, the pre-pore undergoes a conformational change wherein the central beta sheets unravel and insert into the membrane as amphipathic beta hairpins (fig. 2.5)[28]. Within the phylum Bacteroidetes, MACPF proteins are widespread, with hundreds of representatives in sequenced genomes[6]. Some strains encode up to 7 MACPF proteins—*B.f.* 638R encodes two: BSAP-1 and and BF638R_2714. While BF638R_2714 had been previously identified bioinformatically, researchers failed to demonstrate its antimicrobial activity. The transposon mutants I generated were the first indication that BF638R_2714 encodes a functional antimicrobial protein.

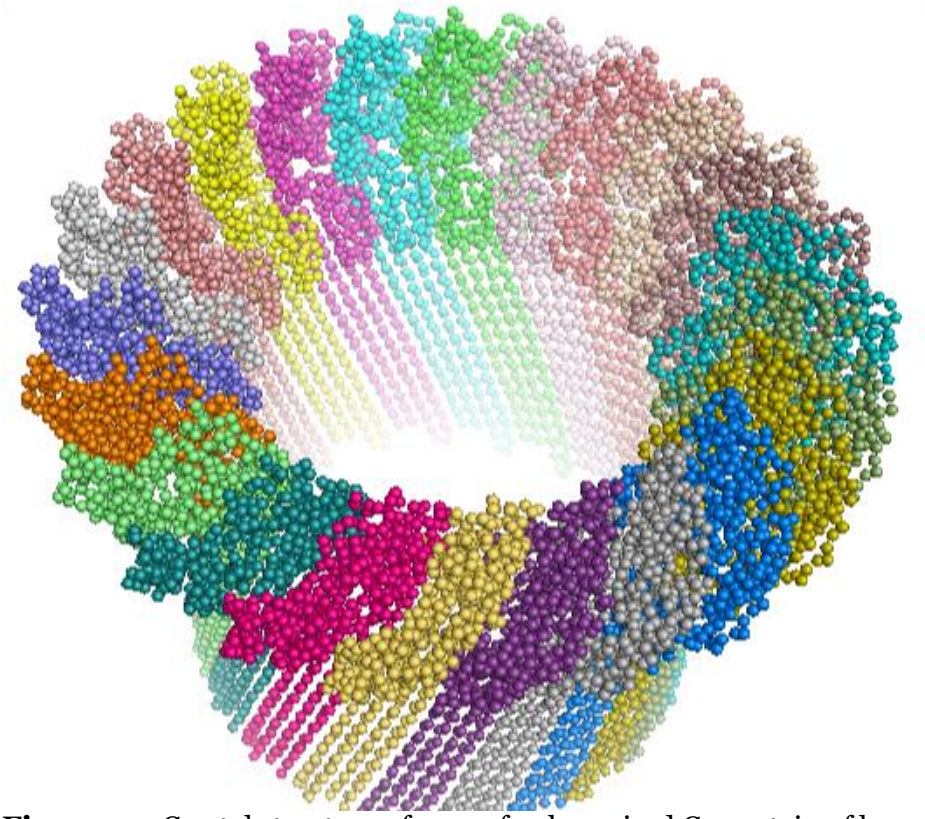


Figure 2.5 Crystal structure of pore of polymerized C9 protein of human complement system. MACPF-domain containing proteins of *Bacteroides* spp. are predicted to disrupt target cell membranes by forming polymeric ring structures similar to that of the poly-C9 ring pictured above. Each monomer C9 is shown in a different color. (PDB: 5FMW)

The status of this protein as a toxin was genetically and biochemically confirmed; thus, I designate this toxin BSAP-4. BSAP-4 is the second MACPF toxin found in the genome of *B.f.*

638R. The first is BSAP-1, which was shown to target a porin-like outer membrane protein of other strains of *B. fragilis*. BSAP-1 is only 41% similar to BSAP-4, but is identifiable by its MACPF motif. In addition to BSAP-1 and BSAP-4, two other MACPF toxins have been characterized and named (Table 2.2). BSAP-1 and BSAP-4 genes are present in many species of *B.f.*: BSAP-1 is more abundant, with orthologs with 99% identity in 29 sequenced *B.f.* genomes, while homologs of BSAP-4 in were found in 21 sequence *B.f.* genomes[6].

Name	Producing strain	Gene name	Length (aa)	Sensitive strains
BSAP-1	*B. fragilis* 638R	BF638R_1646	372	Any species-matched strain without BSAP-1
BSAP-2	*B. uniformis* CL03T00C23	HMPREF1072_01167	508	Any species-matched strain without BSAP-2
BSAP-3	*B. vulgatus* CL09T03C04	HMPREF1058_01765	491	Any species-matched strain without BSAP-3
BSAP-4	*B.f.* 638R	BF638R_2714	506	Species-matched strains without BSAP-4 **and** a targeted receptor ortholog

Table 2.2 Active MACPF toxins of *Bacteroides*. All BSAPs ("*Bacteroides* secreted antimicrobial proteins") target sensitive strains within the same species that lack the corresponding BSAP. Each BSAP is encoded adjacent to a non-targeted ortholog of the receptor. BSAP-4 is the only BSAP with many characterized receptor variants (see chapter 3).

The fact that the *B.f.* 638R genome encodes two functional MACPF toxins with different activity spectra, as well as a ubiquitin-like toxin with yet another distinct spectrum of activity[29], suggests that antagonism is a successful evolutionary strategy (fig. 2.6). With a comparatively small fraction of its genome dedicated to polysaccharide utilization (~50 polysaccharide utilization loci, or "PUL"), the catabolic capabilities of *B.f.* are relatively limited compared with other members of the genus. Most strains of *B.f.* digest only the simplest polysaccharides, preferring monosaccharides, amylose, and amylopectin[30] in addition to host mucin glycans[31]. By comparison, the *B. thetaiotaomicron* genome contains hundreds of glycosylhydrolases and SusC/D homologs in over 100 polysaccharide utilization loci (PUL), allowing it to digest a wide variety of plant polysaccharides[30,32] in addition to host glycans[33]. Given its relatively limited nutritional niche, *B. fragilis* may be physically restricted to the mucosal layer, and may rely on

antimicrobial toxins to defend its territory from other *Bacteroides* species that are better able to utilize numerous polysaccharides.

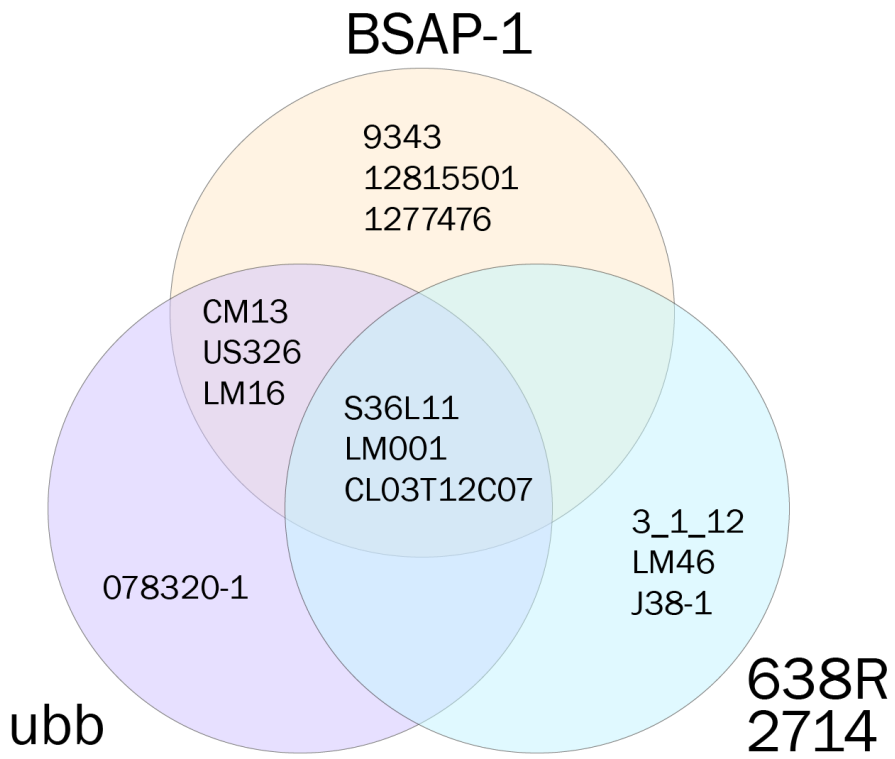


Figure 2.6 Activity spectrum of the three secreted antimicrobial toxins of *B. fragilis* 638R against selected strains of *B. fragilis*. Sensitivities to BSAP-1, ubb, and the BF638R_2714 toxin (BSAP-4) were determined by spot agar tests of sensitive strains over heterologously expressed or purified His-tagged toxins.

The wide distribution of MACPF-domain containing proteins across the phylum Bacteroidetes, as well as the presence of three secreted and one contact-dependent antimicrobial systems in the genome of *B.f.* 638R alone, supports the idea that antagonism is an important evolutionary strategy. The variable distribution of toxins, MACPFs and other antimicrobials may depend on their relative fitnesses, which may in turn depend on the distribution of the receptors they target within the species.

2.5 Conclusion

In this chapter, I have discussed the identification and characterization of the antimicrobial toxin BSAP-4 (BF638R_2714). Subsequent research was focused on identification of the receptor of this toxin in sensitive strains; the distribution of the toxin and its receptor in nature; and implications of toxin-mediated antagonism in an experimental animal model. Future studies of gut bacterial toxins must continue to identify novel toxins and seek deep understanding of their mechanisms of activity. Transposon libraries rapidly yielded a panel of toxin-deficient mutants in a variety of backgrounds, indicating that this methodology is apt to spawn many avenues of research. Mutagenesis and downstream genetic manipulations require sensitivity to antibiotics or the use of alternative selective markers, meaning that undomesticated strains with broad antibiotic resistances will remain difficult to study. This may be addressed through chemical mutagenesis, which may sensitize resistant strains without affecting toxin activity (Appendix A). Alternatively, bioinformatic screens coupled with targeted evaluation *via* heterologous expression in related, genetically tractable species may be an effective way to study toxin candidates from wild strains. Characterization of specific toxins in great depth will facilitate the study of the prevalence and implications of antagonistic behaviors more broadly. Each novel class of confirmed toxin affords researchers the ability to search for related toxins among the plethora of proteins annotated in gut microbiomes with unknown or merely “putative” or “predicted” functions. The mountains of microbiome data that are continuously generated and deposited in our databases will remain mysterious until empirical studies shed light on their mechanisms, structures and functions.

While these methodologies are certain to identify more toxins in the as-yet-uncharacterized genetic material of gut bacteria, the specific activities and ecological effects of these toxins will be more difficult to decipher. What is the concentration of every toxin in every part of the gut? What is the effective radius of secreted toxins produced by a discrete population? Are toxins primarily used for ecosystem defense, or for ecosystem invasion? How

rapidly do toxin genes and their receptors evolve, and how often are they transferred horizontally? Understanding the kinetics, timing, biogeography, and evolutionary significance of antimicrobial toxin activity will be paramount to a comprehensive understanding of the gut microbiota.

2.6 Methods

Oligonucleotides

All oligonucleotides used in this study are listed in Appendix B.

Bacterial strains and growth conditions

Bacteroides strains used in this study were previously described and are listed in Appendix B. *Bacteroides* were grown anaerobically at 37 deg C in supplemented basal medium for liquid cultures, or on supplemented brain-heart infusion (BHIS) plates. Antibiotics (erythromycin 5ug/ml; gentamicin 200ug/ml; tetracycline 6ug/ml) were added when indicated. *E. coli* strains (DH5a, BL21(DE3), RK231, S17 lambda pir) were grown in L broth or on L plates supplemented with antibiotics (carbenicillin 100ug/ml; kanamycin 100ug/ml) when indicated.

Agar spot test for growth inhibition analysis

Antimicrobial interactions among bacteria were tested using the agar spot test. Bacterial cells were scraped from petri dishes into 500ul of phosphate-buffered saline (PBS) and re-suspended to an approximate density of $3x10^9$ cells/ml and 5ul spotted on warm, dry BHIS plates. These plates were incubated overnight to allow producing strains to secrete antimicrobial proteins into the agar medium. Cells were removed using a cotton swab, and any remaining cells killed by incubating for 15 minutes in chloroform vapor. Alternatively, purified antimicrobial molecules were dispensed directly onto overlay plates and allowed to dry. Strains to be tested for growth inhibition were grown to mid-log phase (OD600 = 0.5) and 50ul-100ul added to 4ml of molten BHIS top agar (0.75% Bacto agar) at 50 degrees C, briefly mixed, and poured over the prepared plates. These agar overlay plates were then incubated overnight and zones of clearing analyzed after ~20-24hr of growth.

Transposon mutagenesis

Bacteroides fragilis 638R Δ638R_1646 Δ638R_3923 strain was mutagenized using the mariner transposon plasmid pWH2 as previously described (ref?). Transposon mutants were screened for loss of inhibitory activity against *B.f.* 3_1_12 using the agar overlay assay. The transposon insertion site was identified by arbitrary PCR amplification across the transposon junction with the *B.f.* 638R genome. Amplicons were purified from agarose gels and analyzed by Sanger sequencing using an oligonucleotide directed outward from the transposon cassette, as previously described[20].

Cloning and heterologous expression of genes

Genes destined for complementation and heterologous expression studies were PCR amplified and cloned into prepared backbones of the expression plasmids pFD340 or pMCL140. Genes were also cloned into the integrative plasmid pNBU2 for heterologous expression from a single genomic copy. Sequence-confirmed clones recovered in *E. coli* DH5a were transferred into *Bacteroides* strains by conjugal mating using the helper plasmid RK231. Clones recovered in S17 lambda pir were mated directly into *Bacteroides* strains.

Creation of deletion mutants

Scarless, nonpolar genomic deletions were obtained using the pKNOCK vector[34]. Briefly, 2500bp flanking regions upstream and downstream of the gene to be deleted were PCR-amplified and cloned by Gibson assembly into pKNOCK vector backbone linearized with BamHI. The resulting plasmid was transferred by conjugal mating using the helper plasmid RK231 into the host strain of *Bacteroides fragilis* and co-integrates were selected for 72hr on BHIS plates supplemented with gentamicin and erythromycin (BHIS+GE). 10 co-integrates were picked off initial selection plates, struck for isolation on BHIS+GE plates, and incubated for 24hr. A single colony from each BHIS+GE isolation plate was picked into 1ml supplemented basal media without antibiotics and grown for 24hr to allow outgrowth and secondary recombination. These cultures were back-diluted 1:100 in supplemented basal media, grown to

OD=0.5 (corresponding to $3x10^8$ cfu/ml), diluted and plated on BHIS plates, and incubated for 24hr. These BHIS plates were replica-plated to BHIS plates supplemented with erythromycin and incubated for 24hr. BHIS+erythromycin plates were analyzed for "ghosts" and corresponding colonies on original BHIS plates picked as putative double cross-outs. Erythromycin-sensitive clones were confirmed to have the desired genotype by colony PCR and Sanger sequencing.

Cloning and Purification of His-tagged 638R_2714

638R_2714 was PCR-amplified using primers designed to omit the first 84nt of the ORF to eliminate its 28-aa N-terminal signal sequence. The amplified fragment was digested with NdeI and BamHI and ligated into a prepared pET16b vector to introduce an N-terminal poly-His tag. Transformants of *E. coli* BL21(DE3) were confirmed by Sanger sequencing. Expression of the recombinant protein was induced by addition of IPTG, and the recombinant protein was purified using the Probond Nickel-NTA purification system according to the manufacturer's instructions for native protein purification. Eluted fractions were analyzed by SDS PAGE and dialyzed against PBS before analysis of activity.

2.7 References

1. Scholl, D. & Martin, D. W. Antibacterial Efficacy of R-Type Pyocins towards Pseudomonas aeruginosa in a Murine Peritonitis Model. *Antimicrob. Agents Chemother.* **52,** 1647–1652 (2008).
2. Cascales, E. *et al.* Colicin biology. *Microbiol. Mol. Biol. Rev. MMBR* **71,** 158–229 (2007).
3. Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocins - a viable alternative to antibiotics? *Nat Rev Microbiol* **11,** 95–105 (2013).
4. Coyne, M., Roelofs, KG, Comstock, LE. Type VI secretion systems of human gut Bacteroidales segregate into three genetic architectures, two of which are contained on mobile genetic elements. *BMC Genomics* **17,** (2016).

5. Jack, R. W., Tagg, J. R. & Ray, B. Bacteriocins of gram-positive bacteria. *Microbiol. Rev.* **59,** 171–200 (1995).
6. Chatzidaki-Livanis, M., Coyne, M. J. & Comstock, L. E. An antimicrobial protein of the gut symbiont Bacteroides fragilis with a MACPF domain of host immune proteins. *Mol Microbiol* **94,** 1361–1374 (2014).
7. Tock, M. R. & Dryden, D. T. F. The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8,** 466–472 (2005).
8. Lampe, D. J., Churchill, M. E. & Robertson, H. M. A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J.* **15,** 5470–5479 (1996).
9. Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J. & Robertson, H. M. Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 11428–11433 (1999).
10. Goodman, A. L. *et al.* Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6,** 279–289 (2009).
11. McNulty, N. P. *et al.* Effects of diet on resource utilization by a model human gut microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome. *PLoS Biol.* **11,** e1001637 (2013).
12. Wexler, H. M. Bacteroides: the Good, the Bad, and the Nitty-Gritty. *Clin. Microbiol. Rev.* **20,** 593–621 (2007).
13. Huddleston, J. R. Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infect. Drug Resist.* **7,** 167–176 (2014).
14. Llor, C. & Bjerrum, L. Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. *Ther. Adv. Drug Saf.* **5,** 229–241 (2014).
15. Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M. & ESAC Project Group. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet Lond. Engl.* **365,** 579–587 (2005).

16. Johnson, C. M. & Grossman, A. D. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu. Rev. Genet.* **49,** 577–601 (2015).
17. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6,** e280 (2008).
18. Hecht, D. W., Jagielo, T. J. & Malamy, M. H. Conjugal transfer of antibiotic resistance factors in Bacteroides fragilis: the btgA and btgB genes of plasmid pBFTM10 are required for its transfer from Bacteroides fragilis and for its mobilization by IncP beta plasmid R751 in Escherichia coli. *J Bacteriol* **173,** 7471–80 (1991).
19. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Comput. Resour. Mol. Biol.* **430,** 2237–2243 (2018).
20. Tang, Y. P. & Malamy, M. H. Isolation of Bacteroides fragilis mutants with in vivo growth defects by using Tn4400', a modified Tn4400 transposition system, and a new screening method. *Infect Immun* **68,** 415–9 (2000).
21. Shinkai, Y., Takio, K. & Okumura, K. Homology of perforin to the ninth component of complement (C9). *Nature* **334,** 525–527 (1988).
22. Rosado, C. J. *et al.* A common fold mediates vertebrate defense and bacterial attack. *Science* **317,** 1548–51 (2007).
23. Oshiro, N. *et al.* A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone Actineria villosa. *Toxicon* **43,** 225–228 (2004).
24. Kaiser, K. *et al.* A member of a conserved Plasmodium protein family with membrane-attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. *Mol. Biochem. Parasitol.* **133,** 15–26 (2004).
25. Rosado, C. J. *et al.* The MACPF/CDC family of pore-forming toxins. *Cell. Microbiol.* **10,** 1765–1774 (2008).

26. McCormack, R., de Armas, L., Shiratsuchi, M. & Podack, E. R. Killing machines: three pore-forming proteins of the immune system. *Immunol. Res.* **57,** 268–278 (2013).
27. Ponting, C. P. Chlamydial homologues of the MACPF (MAC/perforin) domain. *Curr Biol* **9,** R911-3 (1999).
28. Dudkina, N. V. *et al.* Structure of the poly-C9 component of the complement membrane attack complex. *Nat. Commun.* **7,** 10588 (2016).
29. Chatzidaki-Livanis, M. *et al.* Gut Symbiont Bacteroides fragilis Secretes a Eukaryotic-Like Ubiquitin Protein That Mediates Intraspecies Antagonism. *MBio* **8,** (2017).
30. Salyers, A. A., Vercellotti, J. R., West, S. E. & Wilkins, T. D. Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. *Appl Env. Microbiol* **33,** 319–22 (1977).
31. MacFarlane, G. T. & Gibson, G. R. Co-utilization of polymerized carbon sources by Bacteroides ovatus grown in a two-stage continuous culture system. *Appl Env. Microbiol* **57,** 1–6 (1991).
32. Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299,** 2074–6 (2003).
33. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4,** 447–57 (2008).
34. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16,** 1105–15 (2008).

Chapter 3: Characterization of BSAP-4 receptors in sensitive strains of *B. fragilis*

Andrew M. Shumaker, Maria Chatzidaki-Livanis, Michael J. Coyne and Laurie E. Comstock

3.1 Introduction to antimicrobial toxin receptors

Despite their structural and mechanistic diversity, all secreted antimicrobial toxins initiate intoxication by attaching to a target cell. The initial binding partner of a toxin on the target cell is referred to as the receptor, and is typically displayed on the outer surface of target cells. Most toxin receptors are proteins, but some are lipopolysaccharides or other extracellular structure[1]. Intoxication can proceed through inhibition of this primary receptor, or through subsequent steps, which may include toxin polymerization and pore formation[2], and translocation into the periplasm or cytosol[3]. Intoxication culminates with the interaction of the toxin with its ultimate target, resulting in the inhibition of a cellular process or gross disruption of cellular physiology.

Toxins are more easily identified than toxin receptors and targets for a variety of genetic reasons. Toxins are not essential, do not contribute strongly to fitness in axenic cultures in vitro, and are typically produced via gene products of a single operon. In contrast, toxin receptors and/or targets tend to be either essential or strong contributors to target cell fitness (a consequence of the strong selective pressure against toxin sensitivity), and the phenomenon of toxin sensitivity may depend on several genes that are not necessarily genetically linked. This means that while transposon mutagenesis is an effective tool for the identification of toxins, it can fail to identify toxin receptors because toxin-resistant mutants are i) less readily generated and ii) unlikely to be mutants of the true receptor but rather of regulatory genes.

In circumstances under which transposon mutagenesis is not suitable, toxin-resistant mutants are best isolated by taking advantage of natural mutagenesis or chemical mutagenesis.

Single amino acid substitutions that are neutral with respect to receptor function may disrupt toxin binding. The receptor may then be identified by deep sequencing to identify single nucleotide polymorphisms that are enriched in a population selected in the presence of toxin[4]. While this method can robustly generate toxin-resistant mutants, it can be time- and cost-intensive. In addition, receptor mutants are not necessarily selected because the fitness landscapes for toxin resistance and receptor function may not intersect—i.e. abrogation of toxin binding may inevitably interfere with receptor function—while the fitness landscapes of other genes involved in the intoxication process may be more readily traversed.

Gene capture screening approaches have been used to identify the factors contributing to toxin sensitivity. This method involves shearing, cloning, and heterologous expression of fragments of a sensitive genome in a resistant host. The resulting library is evaluated for toxin sensitivity, and sensitive clones subjected to further study via process of elimination and more targeted libraries. This strategy requires the expression host to be genetically tractable, yet physiologically similar to the sensitive strain. This strategy has been successfully used to determine the target of the *Lactococcus lactis* bacteriocin LsbB: a cosmid library of the sensitive strain was transformed into resistant mutants of the sensitive strain, and the receptor identified as a Zn-dependent membrane bound metalloprotease *yvjB*[5].

There are also numerous biochemical methods that can capture the interactions of a toxin with its receptor, including affinity purification; small molecule and isotopic labeling; and quantitative proteomics[6]. While these techniques are valuable in specific cases, their applications have thus far seen limited use in the field of microbial toxins, so a more thorough discourse is beyond the scope of this dissertation.

In the absence of robust biochemical or forward genetic screening approaches, genomic and bioinformatic analyses can generate candidate receptors for toxins of interest. Even in the absence of structural information, analysis of a novel toxin's primary sequence or operon can

provide clues as to its target. Identification of previously characterized domains within novel toxins may constrain the set of candidate receptors. Candidates may also be revealed through analysis of sensitive strains. Genomic alignments of closely related strains that differ in resistance or sensitivity to a toxin can highlight genes that are absent in either strain, or that display greater-than-average sequence diversity in either strain.

In recent studies of toxins that target closely-related species of *Bacteroides*, researchers have observed a compelling trend wherein the genomic loci of toxin genes tend to be adjacent to a toxin-resistant ortholog of the toxin-sensitive receptor[1,7]. In sensitive strains, this region is marked by the absence of the toxin and the presence of the toxin-sensitive receptor. This pattern is readily observable through genomic alignments, and proved applicable in my search for the receptor of BSAP-4.

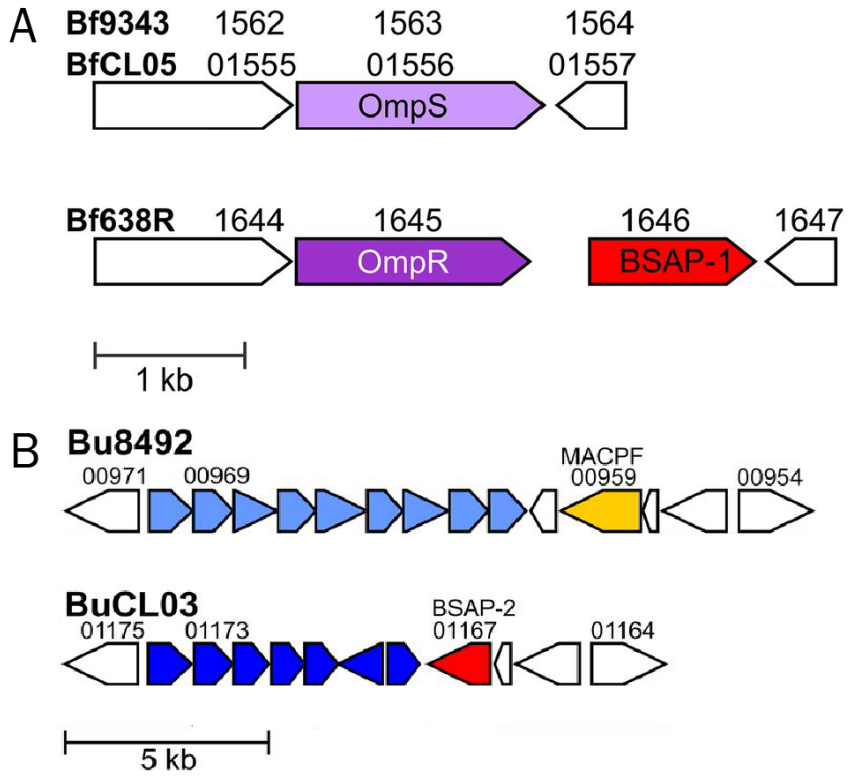


Figure 3.1 BSAP-1 and BSAP-2 are linked to non-targeted orthologs of toxin receptor in sensitive strains. A) In *B.f.*638R, the BSAP-1 locus is adjacent to the locus of BF638R_1645, a non-targeted ortholog of the sensitive receptor in *B.f.*9343 and *B.f.*CL05.

Figure 3.1 (continued) In *B. uniformis* CL03, BSAP-2 is encoded next to an O-antigen biosynthesis operon that is highly divergent from an O-antigen biosynthesis operon in sensitive strains. Note that *B.u.*8492, which is targeted by BSAP-2, encodes a MACPF without proven antimicrobial activity (yellow). Figure adapted from [1].

3.2 The receptor of BSAP-4 in *B. fragilis* 3_1_12 is BFAG02253

Previous studies of *Bacteroides* antimicrobial proteins have shown that antimicrobial proteins may be acquired in tandem with a resistant version of the receptor in sensitive strains. BSAP-1, the first MACPF toxin of *B. fragilis* to be identified, is encoded directly upstream of a porin-like outer membrane protein (BF638R_1645)[1] (fig. 3.1A). Comparing genomic alignments 638R with the sensitive strain NCTC9343 at this locus, the latter lacked the toxin and possessed an orthologous porin-like outer membrane protein. This ortholog was confirmed to be necessary for toxin sensitivity, suggesting that BF638R_1645 is a non-targeted ortholog of the targeted receptor[1]. BSAP-2, another MACPF toxin identified in *B. uniformis* CL03T00C23, is encoded adjacent to an O-antigen biosynthesis operon that was highly divergent when compared with an O-antigen biosynthesis operon at the same locus in sensitive strains[1] (fig. 3.1B). Subsequent analyses confirmed that this toxin targeted O-antigens in sensitive strains, and that the O-antigen biosynthesis operon encoded adjacent to the toxin is not targeted. This bioinformatic approach to identifying candidate receptors is facilitated by the fact that the antagonism in question involves strains of the species. Uncovering the mechanisms underlying an antimicrobial phenotype is simplified by the significant overlap between the two genomes.

Having confirmed the identity of BSAP-4 as an antimicrobial toxin, we turned to genomic alignments with sensitive strains to identify candidate receptors. As expected, sensitive strains did not contain a BSAP-4 ortholog, but did contain an ortholog of the gene adjacent to the toxin gene, BF638R_2715 (fig. 3.2). This gene and its orthologs in sensitive strains encode calycin-like beta barrel outer membrane proteins (protein family PF13944). Reasoning that BF638R_2715 may encode a resistant version of the sensitive receptors in 3_1_12 and J38-1, we selected BFAG02253 and M068_02717 as candidates in our search for the receptor of BSAP-4.

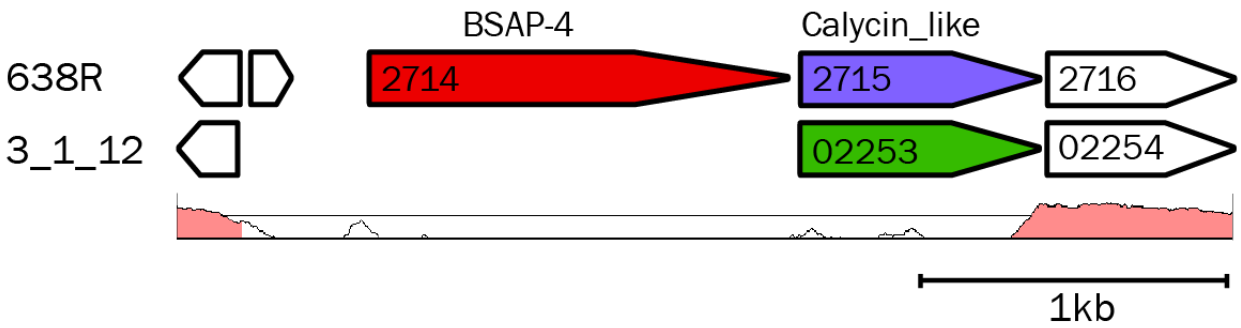


Figure 3.2 Genome alignment reveals receptor candidate at BSAP-4 locus. The sensitive strain *B. fragilis* 3_1_12 did not contain an ortholog of BSAP-4. BF638R_2715 (purple), a predicted outer membrane protein belonging to the calycin_like protein family, had an ortholog in 3_1_12 (BFAG02253, green). Percent identity over 100-bp spans is graphed below the alignment, with >70% identity colored pink. We hypothesized that BFAG02253 is the receptor of BSAP-4.

To interrogate the hypothesis that BFAG02253 is the toxin receptor in *B.f.* 3_1_12, this gene was cloned into the expression vector pFD340 and mated into *B. thetaiotaomicron*. Heterologous expression of BFAG02253 in *B. thetaiotaomicron* rendered that strain sensitive to BSAP-4 as shown in spot agar tests against 638RΔBSAP-1Δubb, *B. theta* expressing BSAP-4 (fig. 3.3), and His-tagged protein. This experiment was repeated with the receptor from J38-1 (M068_02717) and several other orthologous receptors, all of which rendered *B. thetaiotaomicron* sensitive to BSAP-4, although to different extents (data not shown). These data strongly suggested that BFAG02253 and its orthologs function as the receptor for BSAP-4.

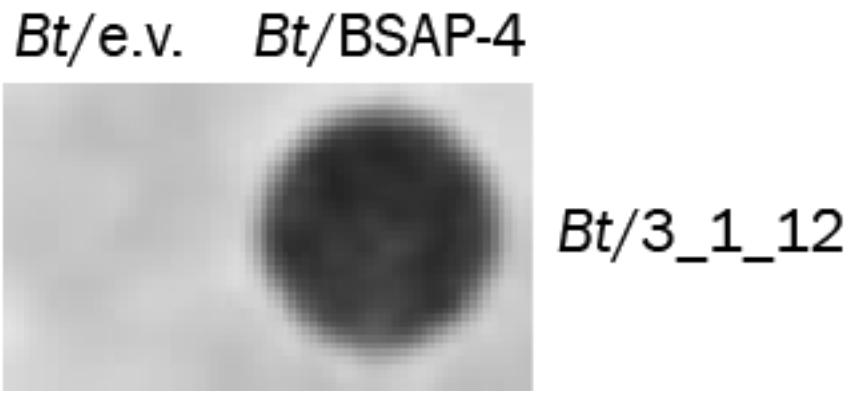


Figure 3.3 Heterologous expression of BFAG02253 in *B. thetaiotaomicron* confers sensitivity to BSAP-4. Spotted *B. theta* containing the empty vector (*Bt*/e.v.) did not inhibit overlaid *B. theta* heterologously expressing the sensitive receptor from 3_1_12 BFAG02253 (*Bt*/3_1_12), while spotted *B. theta* expressing BSAP-4 (*Bt*/BSAP-4) did inhibit *Bt*/3_1_12. Similar plates overlaid with *B. theta* containing the empty vector were not sensitive to *Bt*/BSAP-4 (not shown).

Further genetic confirmation of the identity of the receptor was desired through a knockout of the sensitive receptor. Genetic manipulation of *B.f.* 3_1_12 is hindered by its natural resistance to erythromycin and tetracycline. After confirming that *B.f.* J38-1 is sensitive to erythromycin and therefore amenable to genetic manipulation, a knockout plasmid targeting the putative receptor M068_02717 was constructed. The resulting knockout strain, J38-1 ΔM068_02717, was not inhibited by His-tagged protein in spot agar tests, confirming it is necessary for toxin sensitivity (fig. 3.4).

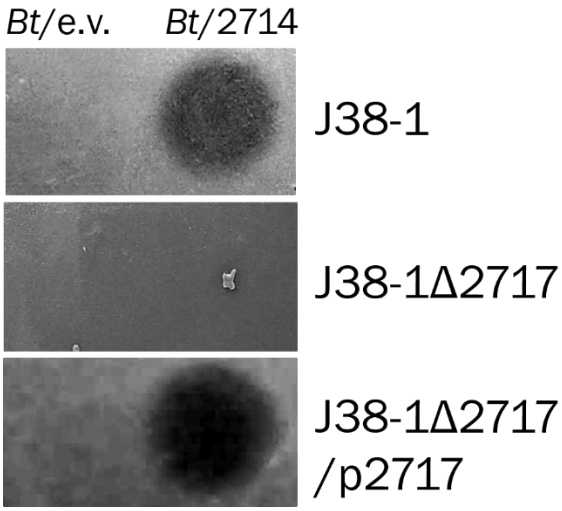


Figure 3.4 Deletion and complementation of BSAP-4 receptor abolishes and restores sensitivity to BSAP-4. The receptor ortholog, M068_02717, was deleted in the sensitive strain J38-1. Unlike wild-type J38-1, this strain (J38-1Δ2717) was not sensitive to BSAP-4 produced by *B. thetaiotaomicron* heterologously expressing BSAP-4 (*Bt*/2714). Heterologous expression of the receptor ortholog *in trans* in the deletion mutant (J38-1Δ2717/p2717) restored BSAP-4 sensitivity.

3.3 Orthologs of BFAG02253 display variable toxin sensitivities

BFAG02253 belongs to a poorly characterized protein family found in all sequenced genomes of *B. fragilis*, albeit with extensive sequence diversity. BFAG02253 and its orthologs are the first gene in a predicted operon containing 8 genes including a cobalamin biosynthesis related gene. BFAG02253 is a predicted calycin-like beta barrel outer membrane protein, suggesting it belongs to a family of proteins that transport small hydrophobic molecules. Sensitive strains contained diverse orthologs of the sensitive receptor in 3_1_12 (table 3.1) A

protein BLAST search for similar proteins showed that BFAG02253 is relatively rare in sequenced genomes: only five sequenced strains contain a perfect match. By contrast, there are numerous *B.f.* genomes containing genes encoding identical proteins to M068_02717; 17 sequenced genomes contained a gene with 100% identity to M068_02717; 33 others contained a gene with one of two specific amino acid changes. Indeed, two of the *B.f.* strains tested for sensitivity to BSAP-4 contained one of these gene variants: S36L11 and CL03T12C07.

	BF638R 2715	BFAG 02253	M068 2717	ADC61 RS00515	CQW36 03505	M136 2706	BSAP-4 Sensitivity	Strain
BF638R_2715	100	33.6	30.9	30.9	31.3	31.3	resistant	638R
BFAG_02253	33.6	100	77.6	77.6	77.9	77.9	sensitive	3_1_12
M068_2717	30.9	77.6	100	99.6	99.3	99.6	sensitive	J38-1
ADC61_RS00515	30.9	77.6	99.6	100	99.6	99.3	sensitive	894 BFRA
CQW36_03505	31.3	77.9	99.3	99.6	100	99.6	sensitive	US326
M136_2706	31.3	77.9	99.6	99.3	99.6	100	sensitive	S36L11

Table 3.1 Amino acid identity heat map of BSAP-4 receptor orthologs. Orthologs of the four most similar proteins (>99% identity) differed by only 1 or 2 amino acid substitutions.

I speculated that the receptor sequence diversity might affect toxin binding affinity or overall toxin sensitivity. Spot agar tests suggested that wild-type strains expressing different receptor orthologs had slight differences in sensitivity to natively-secreted as well as purified, His-tagged toxin: *B.f.* 3_1_12 displayed large and clear zones of clearing, while J38-1 displayed smaller and fainter zones of clearing (data not shown). *B.f.* S36L11 and CL03T12C07, meanwhile, typically displayed moderately clear zones of clearing (data not shown).

To reduce the impact of strain-specific membrane differences or adjuncts, orthologous receptors were cloned into expression vectors for heterologous expression in *B. thetaiotaomicron*. The resulting strains demonstrated dose-dependent sensitivity to BSAP-4 in spot agar tests (fig. 3.5).

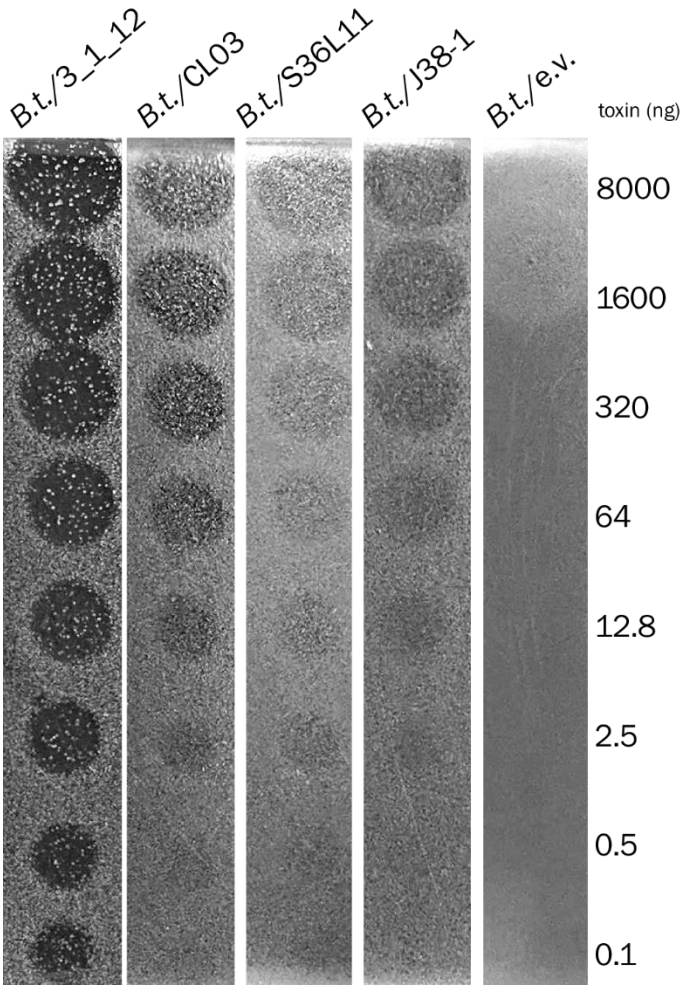


Figure 3.5 Heterologous expression of BSAP-4 receptors in *B. thetaiotaomicron* incurs dose-dependent inhibition by His-tagged BSAP-4 in spot agar tests. *B. theta* heterologously expressing BFAG02253, the receptor from *B.f.*3_1_12, was qualitatively most sensitive to BSAP-4 ("*B.t.*/3_1_12," left-most). Other variants were qualitatively less sensitive to BSAP-4. *B. theta* with the empty vector was not sensitive to BSAP-4.

Because of the qualitative nature of the spot agar test, I sought to develop a quantitative assay that could better quantify sensitivity to the toxin. I adapted the liquid survival titration assay (LSTA) to quantify how varying toxin concentrations affect the sensitivities of cells expressing variants of the toxin receptor[8]. Briefly, a known concentration of sensitive cells is incubated in the presence of purified toxin or liquid culture supernatants conditioned by toxin-secreting cells. The mixture is plated to petri dishes and colony-forming units are used to infer the fraction of cells that survived the toxin treatment, generating a dose-response curve.

Differing responses of identically treated strains thus reveal quantitative differences in toxin sensitivity at the cellular level.

Toxin for the LSTA was generated by overnight growth of a 50ml basal medium culture of *B. thetaiotaomicron*/pMCL177, a strain that strongly expresses and secretes BSAP-4. Cells were pelleted by centrifugation and the supernatant was filter sterilized. As a negative control, the supernatant of a culture of *B. thetaiotaomicron* containing the empty vector control was prepared in parallel. Supernatants were analyzed for protein content by SDS PAGE: the supernatant of *B. thetaiotaomicron*/pMCL177 displayed a band at 53kDa, the predicted size of BSAP-4, while the supernatant of the empty vector control did not. The concentration of BSAP-4 in supernatant was quantified by image analysis of the 53kDA band by interpolation of a standard curve of bovine serum albumin using ImageJ. Sensitive cells were grown to early log phase, diluted to roughly 10^5 cells/ml, mixed 1:1 with *B. thetaiotaomicron* supernatants in 96-well microtiter plates, and incubated anaerobically for 30min at 37degC. Colony forming units were determined by plating 50µl of the mixture on BHIS plates using the 50µl 6-log mode on the EddyJet 2 spiral plating device (Neutec), and colonies counted using the Flash&Go colony counter (Neutec). Samples were normalized to untreated controls.

The liquid survival titration assay identified differences in toxin sensitivity among wild-type *B.f.* strains. As expected due to the large zones of clearing in the spot agar test, *B.f.* 3_1_12 was the most sensitive in the LSTA, as evidenced by the low number of colonies recovered in the most dilute toxin treatment. Surprisingly, despite the moderate response of *B.f.* J38-1 in the

spot agar test, it did not respond strongly in the LSTA (data not shown).

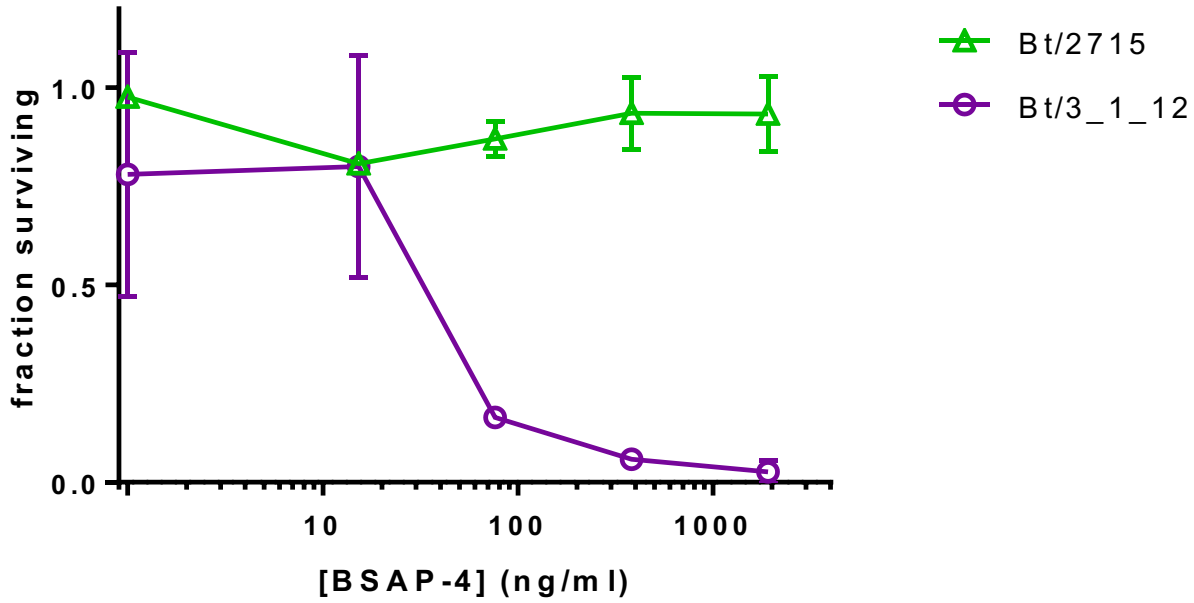


Figure 3.6 Liquid survival titration assay of receptor orthologs heterologously expressed in *B. thetaiotaomicron* demonstrate quantitative differences in dose-response to BSAP-4. *B. theta* expressing the non-targeted receptor from 638R ("Bt/2715") did not show a dose-response. *B. theta* expressing the orthologous receptors from 3_1_12 ("Bt/3_1_12") showed the strongest dose response. Error bars represent standard deviation (n=3) in experiments performed on separate days. Statistical significance was determined by two-tailed paired t-test (P = 0.0338).

Toxin receptors heterologously expressed in *B. thetaiotaomicron* also revealed differences in toxin dose response (fig. 3.6). *B.t.*/3_1_12 was the most sensitive to the toxin. Heterologous expression of the resistant receptor, BF638R_2715, did not confer toxin sensitivity. Given the roughly ~53kDa size of BSAP-4, 100ng/ml corresponds to ~2nM, indicating that the EC50 in these experiments falls in the low- to mid-nanomolar range.

3.4 BSAP-4 receptor sequences suggest diversifying selection

The tremendous diversity of receptors of BSAP-4 stands in stark contrast to the receptors of other MACPF toxins characterized in the *Bacteroides*. Like BSAP-4, the MACPF toxins BSAP-1 and BSAP-2 are encoded alongside the non-targeted orthologs of their receptors in sensitive strains[1]. However, while BSAP-4 has numerous confirmed receptors with manifold

homologs and orthologs in numerous strains of *B. fragilis*, the genes encoding the receptors of BSAP-1 and BSAP-2 are highly conserved. All *B.f.* strains have either BSAP-1 alongside a toxin-resistant receptor ortholog, or the sensitive locus; likewise, all *B. uniformis* strains have either BSAP-2 alongside an operon encoding the non-targeted O-antigen, or the sensitive locus[1]. Given the variable sensitivities of different strains to BSAP-4, and the variable sequences of their receptors, I speculated that the selective pressure exerted by the toxin caused the sensitive receptor to undergo diversifying selection.

To interrogate this hypothesis, I analyzed the primary sequences and predicted structures of the receptor orthologs. Alignments of highly homologous receptors revealed two variable alleles within otherwise identical proteins (figure 3.7). Using the J38-1 orthologous receptor M068_02717 as a reference, variability was identified at the amino acid positions L137 and P254. Some strains had the substitution L137P, while others had the substitution P254T. All four permutations of these two alleles (L137, P254; P137, P254; L137, T254; and P137, T254) exist in published genomes of *B. fragilis* strains. Indeed, the orthologous receptors I tested had either L137, T254 (J38-1) or P137, T254 (S36L11, CL03T12C07). The sequences of the orthologous receptors were confirmed by Sanger sequencing, and did not necessarily reflect the sequences in published genome sequences of the corresponding strains as shown in figure 3.7.

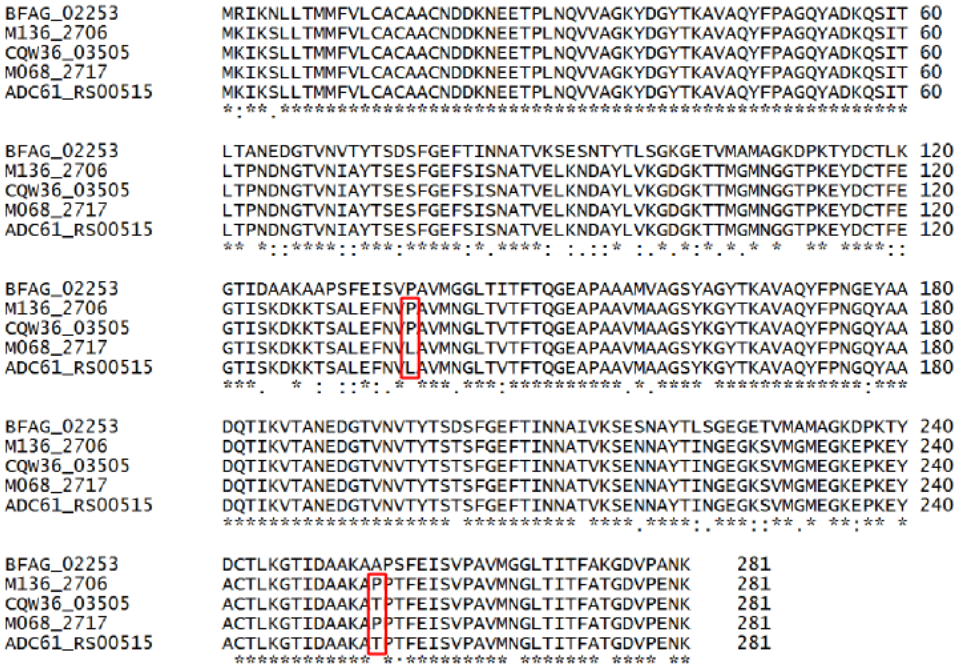


Figure 3.7 Alignment of BFAG02253 and four other receptor orthologs in sensitive strains. While BFAG02253 is only 77% identical to the other aligned orthologs, the rest are >99% identical, with two variable positions (red boxes). Table 3.1 reflects the strains in which each orthologous receptor protein are found. Alignment performed using Clustal Omega.

The natures of these alleles were intriguing because proline substitutions tend to induce relatively large changes in protein structure. I hypothesized that these substitutions entail structural changes that affect the ability of BSAP-4 to interact with the receptor. To interrogate this hypothesis, I generated structural homology models using I-TASSER[9] and Phyre2[10]. Models were based on the 2.4A crystal structure of BF2706, a calycin_like beta barrel outer membrane protein from *B.f.* NCTC9343 (PDB: 3RWX). The homology models scored very well and had similar structures with the exception of two loops, one of which has predicted extracellular topologies (Appendix C). Further biochemical, genetic, and structural experiments will be required to determine how sequence diversity in the orthologous receptors of BSAP-4 affects their sensitivities to BSAP-4 and their native functions. While it is tempting to assume that amino acid substitutions directly affect toxin binding, there are many other explanations for the differences in toxin sensitivity. For instance, mutations may affect receptor expression levels,

stability, efficiency of membrane insertion, diffusion rate within membrane, polymerization kinetics, interaction with other membrane structures and proteins, and many other parameters.

To investigate the toxin sensitivities of natural receptor orthologs with P137, P254, I conducted site-directed mutagenesis on the orthologous receptor from *B. fragilis* CL03T12C07 (M136_2706). In the published genome sequence, this receptor has a proline at positions 137 and 254 (figure 3.7); however, Sanger sequencing revealed that the receptor in my stock of this strain had a single SNP introducing a threonine substitution at position 254. I used PCR to introduce the T254P mutation and transferred the resulting plasmid into *B. thetaiotaomicron*. I performed the liquid survival titration assay (LSTA) against empty vector and toxin-expressing supernatants as described above. The results of the LSTA showed that the receptor bearing the T254P mutation rendered the expressing strain significantly more sensitive to BSAP-4 compared with the strain expressing the wild-type receptor (figure 3.8).

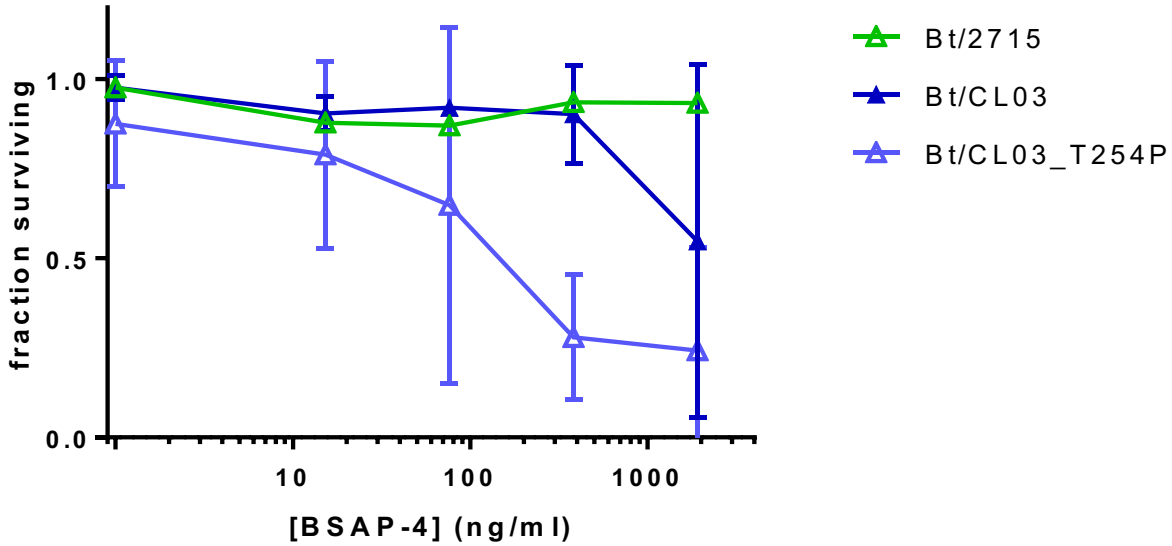


Figure 3.8 Liquid survival titration assay of heterologously expressed, site-directed mutant of orthologous receptor of BSAP-4. Site-directed mutagenesis to introduce the mutation T254P in the background of M136_2706 (the receptor gene of *B.f.*CL03T12C07) sensitized the expressing strain to BSAP-4. *B. theta* expressing the site-directed mutant ("Bt/CL03_T254P") was significantly more sensitive to BSAP-4 relative to *B. theta* expressing the wild-type receptor ("Bt/CL03").

Figure 3.8 (continued) *B. theta* expressing the resistant receptor is included as a negative control ("Bt/2715"). Error bars represent standard deviation (n=3) in experiments performed on separate days. Statistical significance was determined by two-tailed paired t-test (P = 0.0397).

The fact that most toxin receptors make strong contributions to microbial fitness led us to speculate about the function of receptors of BSAP-4. Orthologs of the sensitive receptor BFAG02253 are annotated as calycin-like outer membrane proteins. The calycins are an extremely diverse superfamily of beta-barrel transporters including lipocalins, fatty acid binding proteins (FABPs), metalloprotease inhibitors (MPIs) and avidins. Lipocalins, the largest of these families, bind small hydrophobic molecules, including for the transport of steroids, bilins, retinoids and lipids.

BFAG02253 and its orthologs are the first gene in an operon containing 13 genes (14 in the case of *B.f.*638R, with BSAP-4 as the first gene). The largest gene in this operon is annotated as CobN-like cobalamin biosynthesis related gene. Cobalamin—also known as vitamin B12—belongs to a class of small molecules called corrinoids, which are characterized by tetrapyrrole corrin rings that coordinate cobalt ions. Apart from the conserved metal coordination moiety, corrinoids vary in the identity of a "lower ligand" variable group, which can be one of several biosynthetically installed bases including benzimidazoles, purines or phenols[11]. Several novel classes of cobalamin biosynthesis and modification genes have recently been characterized, and different *Bacteroides* species have been shown to preferentially utilize different corrinoids in their cellular metabolism[12]. Indeed, corrinoids have been shown to be significant modulators the compositions of gut microbiotas, suggesting that corrinoids represent a source of nutrient competition[13].

In addition to the CobN-like protein, other genes in this operon are related to metabolism of heme or cobalamin. These include a HmuY-like protein, which is similar to a heme-binding protein found in *Porphyromonas gingivalis*[14]; a CirA-like TonB-dependent transporter[15]; and the ribonucleoside diphosphate reductase components NrdA and NrdB, the

expression of which are known to be regulated by cobalamins. These annotations suggest that this operon is involved in the acquisition or export of nutrients, cofactors or co-enzymes.

3.5 Conclusion

In this chapter, I identified the receptor of BSAP-4 in sensitive strains *B.f.*3_1_12, J38-1, and others. These receptors were easily identifiable due to the pattern observed in other BSAPs: the toxin is linked to the non-targeted ortholog of the receptor in sensitive strains. This pattern was observed for BSAP-4, with one crucial difference: while all the other BSAPs have either the resistant or the sensitive receptor, BSAP-4 has many receptor orthologs. This diversity adds complexity to the study of the fitness contributions made by BSAP-4 and each receptor ortholog. In subsequent experiments, I proceeded to evaluate these fitness contributions in an animal model of gut colonization. I also searched natural genomes and metagenomes for the presence and co-occurrence of BSAP-4 and its various receptors.

3.6 Methods

Oligonucleotides

All oligonucleotides used in this study are listed in Appendix B.

Bacterial strains and growth conditions

Bacteroides strains used in this study were previously described and are listed in Appendix B. *Bacteroides* were grown anaerobically at 37 deg C in supplemented basal medium for liquid cultures, or on supplemented brain-heart infusion (BHIS) plates. Antibiotics (erythromycin 5ug/ml; gentamicin 200ug/ml; tetracycline 6ug/ml) were added when indicated. *E. coli* strains (DH5a, BL21(DE3), RK231, S17 lambda pir) were grown in L broth or on L plates supplemented with antibiotics (carbenicillin 100ug/ml; kanamycin 100ug/ml) when indicated.

Agar spot test for growth inhibition analysis

Antimicrobial interactions among bacteria were tested using the agar spot test. Bacterial cells were scraped from petri dishes into 500ul of phosphate-buffered saline (PBS) and re-suspended

to an approximate density of $3x10^9$ cells/ml and 5ul spotted on warm, dry BHIS plates. These plates were incubated overnight to allow producing strains to secrete antimicrobial proteins into the agar medium. Cells were removed using a cotton swab, and any remaining cells killed by incubating for 15 minutes in chloroform vapor. Alternatively, purified antimicrobial molecules were dispensed directly onto overlay plates and allowed to dry. Strains to be tested for growth inhibition were grown to mid-log phase (OD600 = 0.5) and 50ul-100ul added to 4ml of molten BHIS top agar (0.75% Bacto agar) at 50 degrees C, briefly mixed, and poured over the prepared plates. These agar overlay plates were then incubated overnight and zones of clearing analyzed after ~20-24hr of growth.

Cloning and heterologous expression of genes

Genes destined for complementation and heterologous expression studies were PCR amplified and cloned into prepared backbones of the expression plasmids pFD340 or pMCL140. Genes were also cloned into the integrative plasmid pNBU2 for heterologous expression from a single genomic copy. Sequence-confirmed clones recovered in *E. coli* DH5a were transferred into *Bacteroides* strains by conjugal mating using the helper plasmid RK231. Clones recovered in S17 lambda pir were mated directly into *Bacteroides* strains. Site-directed mutagenesis of previously cloned orthologous receptor genes was performed by whole-plasmid PCR amplification using partially-overlapping oligonucleotides introducing the desired mutations. Linear amplification products were circularized by isothermal assembly and cloned as described above.

Creation of deletion mutants

Scarless, nonpolar genomic deletions were obtained using the pKNOCK vector[16]. Briefly, 2500bp flanking regions upstream and downstream of the gene to be deleted were PCR-amplified and cloned by Gibson assembly into pKNOCK vector backbone linearized with BamHI. The resulting plasmid was transferred by conjugal mating using the helper plasmid RK231 into the host strain of *Bacteroides fragilis* and co-integrates were selected for 72hr on BHIS plates supplemented with gentamicin and erythromycin (BHIS+GE). 10 co-integrates were picked off initial selection

plates, struck for isolation on BHIS+GE plates, and incubated for 24hr. A single colony from each BHIS+GE isolation plate was picked into 1ml supplemented basal media without antibiotics and grown for 24hr to allow outgrowth and secondary recombination. These cultures were back-diluted 1:100 in supplemented basal media, grown to OD=0.5 (corresponding to $3x10^8$ cfu/ml), diluted and plated on BHIS plates, and incubated for 24hr. These BHIS plates were replica-plated to BHIS plates supplemented with erythromycin and incubated for 24hr. BHIS+erythromycin plates were analyzed for "ghosts" and corresponding colonies on original BHIS plates picked as putative double cross-outs. Erythromycin-sensitive clones were confirmed to have the desired genotype by colony PCR and Sanger sequencing.

Liquid survival titration assay

Toxin for the LSTA was generated by overnight growth of a 50ml basal medium culture of *B. thetaiotaomicron*/pMCL177. Cells were pelleted by centrifugation and the supernatant was filter sterilized. The supernatant of a culture of *B. thetaiotaomicron* containing the empty vector control was prepared in parallel as a negative control. Supernatants were analyzed for protein content by SDS page. The concentration of BSAP-4 in supernatant was quantified by image analysis of the 53kDA band by interpolation of a standard curve of bovine serum albumin using ImageJ. Sensitive cells were grown to early log phase, diluted to roughly 10^5 cells/ml, mixed 1:1 with *B. thetaiotaomicron* supernatants in 96-well microtiter plates, and incubated anaerobically for 30min at 37degC. Colony forming units (cfu) were determined by plating 50µl of the mixture on BHIS plates using the 50µl 6-log mode on the EddyJet 2 spiral plating device (Neutec), and colonies counted using the Flash&Go colony counter (Neutec). Experimentally determined cfu/ml of samples were presented as the fraction of cells surviving by normalization against cfu/ml of untreated controls.

3.7 References

1. Roelofs, K. G., Coyne, M. J., Gentyala, R. R., Chatzidaki-Livanis, M. & Comstock, L. E. Bacteroidales Secreted Antimicrobial Proteins Target Surface Molecules Necessary for Gut Colonization and Mediate Competition In Vivo. *mBio* **7,** e01055-16 (2016).
2. Bischofberger, M., Gonzalez, M. R. & van der Goot, F. G. Membrane injury by pore-forming proteins. *Curr. Opin. Cell Biol.* **21,** 589–595 (2009).
3. Cascales, E. *et al.* Colicin biology. *Microbiol. Mol. Biol. Rev. MMBR* **71,** 158–229 (2007).
4. Ekblad, B., Nissen-Meyer, J. & Kristensen, T. Whole-genome sequencing of mutants with increased resistance against the two-peptide bacteriocin plantaricin JK reveals a putative receptor and potential docking site. *PLOS ONE* **12,** e0185279 (2017).
5. Uzelac, G. *et al.* A Zn-dependent metallopeptidase is responsible for sensitivity to LsbB, a class II leaderless bacteriocin of Lactococcus lactis subsp. lactis BGMN1-5. *J. Bacteriol.* **195,** 5614–5621 (2013).
6. Lemichez, E. & Barbieri, J. T. General Aspects and Recent Advances on Bacterial Protein Toxins. *Cold Spring Harb. Perspect. Med.* **3,** (2013).
7. Chatzidaki-Livanis, M., Coyne, M. J. & Comstock, L. E. An antimicrobial protein of the gut symbiont Bacteroides fragilis with a MACPF domain of host immune proteins. *Mol Microbiol* **94,** 1361–1374 (2014).
8. Scholl, D. & Martin, D. W. Antibacterial Efficacy of R-Type Pyocins towards Pseudomonas aeruginosa in a Murine Peritonitis Model. *Antimicrob. Agents Chemother.* **52,** 1647–1652 (2008).
9. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12,** 7–8 (2015).
10. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).

11. Allen, R. H. & Stabler, S. P. Identification and quantitation of cobalamin and cobalamin analogues in human feces. *Am. J. Clin. Nutr.* **87,** 1324–1335 (2008).
12. Degnan, P. H., Barry, N. A., Mok, K. C., Taga, M. E. & Goodman, A. L. Human gut microbes use multiple transporters to distinguish vitamin B12 analogs and compete in the gut. *Cell Host Microbe* **15,** 47–57 (2014).
13. Degnan, P. H., Taga, M. E. & Goodman, A. L. Vitamin B12 as a modulator of gut microbial ecology. *Cell Metab.* **20,** 769–778 (2014).
14. Wójtowicz, H. *et al.* Unique Structure and Stability of HmuY, a Novel Heme-Binding Protein of Porphyromonas gingivalis. *PLOS Pathog.* **5,** e1000419 (2009).
15. Buchanan, S. K. *et al.* Structure of colicin I receptor bound to the R-domain of colicin Ia: implications for protein import. *EMBO J.* **26,** 2594–2604 (2007).
16. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16,** 1105–15 (2008).

Chapter 4: Ecological significance of BSAP-4 and its receptors in the mammalian gut

Andrew M. Shumaker, Michael J. Coyne and Laurie E. Comstock

4.1 Gnotobiotic mouse models of gut colonization

We turned to gnotobiotic mouse models of gut colonization to study how BSAP-4 and its receptors contribute to microbial fitness *in vivo*. Gnotobiotic (from the Greek "gnotos" (known) and "biotic" (life)) mouse models provide a more realistic environment for the evaluation of gut symbiotic bacteria than the test tube, and do not suffer from the confounding factors imposed by native microbiotas in conventionally raised mice[1]. In an effort to isolate the fitness contributions of BSAP-4 and its receptors, we performed a series of competition experiments by co-administration of two isogenic strains at a time to germ-free mice, followed by quantification of their relative abundances in feces. The results of such an experiment reflect the differential abilities of isogenic strains to colonize, proliferate and survive in the GI tract of each animal. Strains to be tested were grown to mid-log phase, adjusted to identical optical densities, and gavaged into germ-free Swiss-Webster mice at least 5 weeks of age. Each experiment was performed in 3 male and 3 female mice. After one week, bacteria were isolated from mouse feces and plated for single colonies, which were analyzed by colony PCR.

Our experiments were driven by two primary hypotheses: first, that BSAP-4 provides a fitness advantage by out-competing sensitive, isogenic strains, and second, that the sensitive receptor provides a fitness advantage in the absence of BSAP-4-producing strains. To confirm the fitness advantage conferred by the sensitive receptor, we first conducted a competition experiment between wild-type *B.f.* J38-1 and J38-1ΔM068_2717. Our results showed that the wild-type strain possessing the intact receptor modestly out-competed strains with the receptor deleted in 4 of the 6 mice, although a one-sample *t* test of arcsine transformed values indicated that this difference was not statistically significant (P value of 0.1066 for wild-type and 0.2641

for receptor deletion) (fig. 4.1). Thus, this experiment was not conclusive with respect to the fitness advantage conferred by the M068_2717 receptor in its native context.

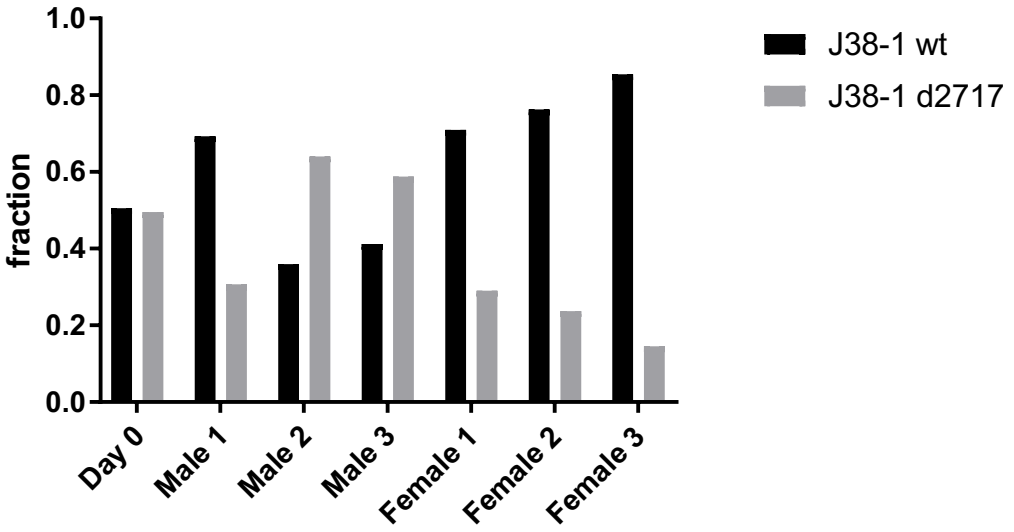


Figure 4.1 Mouse gut competition between *B. fragilis* J38-1 wild-type ("wt") vs. deleted receptor ("d2717"). The wild-type strain out-competed the strain with the deleted receptor in 4 out of 6 mice, but this difference was not statistically significant. Mouse fecal samples were collected on day 7 post-gavage.

Next, we conducted a competition experiment between a strain expressing BSAP-4 (*B. thetaiotaomicron*/pMCL140_BSAP-4) and an isogenic strain expressing the sensitive receptor (*B. thetaiotaomicron*/pFD340_BFAG02253). These specific recombinant strains were selected for the competition experiment due to the robust antimicrobial phenotype observed in the spot agar test and liquid survival titration assays. We hypothesized that the activity observed *in vitro* would be recapitulated *in vivo*, allowing the BSAP-4-expressing strain to out-compete the receptor-expressing strain. To our surprise, we found that the receptor-expressor out-competed the BSAP-4-expressor, with 87% of all colonies arising after the one week co-colonization identified as receptor-expressors ($P = 0.0018$ for BFAG02253; $P = 0.0011$ for BSAP-4 by one-sample *t* test) (fig. 4.2). Skeptical of these results, I selected two fecal sample isolates of each isogenic strain for further characterization. Sequencing confirmed that the isolates retained the BSAP-4 or the receptor plasmid. Additionally, I subjected these colonies to the spot agar test by

spotting the BSAP-4-expressors and overlaying recovered receptor-expressors. The isolates displayed characteristic zones of clearing, confirming that their antagonistic behavior was retained (data not shown). We concluded that either i) the receptor conferred a sufficiently large fitness advantage *in vivo* to the heterologous species *B.thetaiotaomicorn* to overcome any fitness deficit exerted by BSAP-4 secretion of co-resident bacteria, or ii) the toxin places a fitness disadvantage on cells that express it.

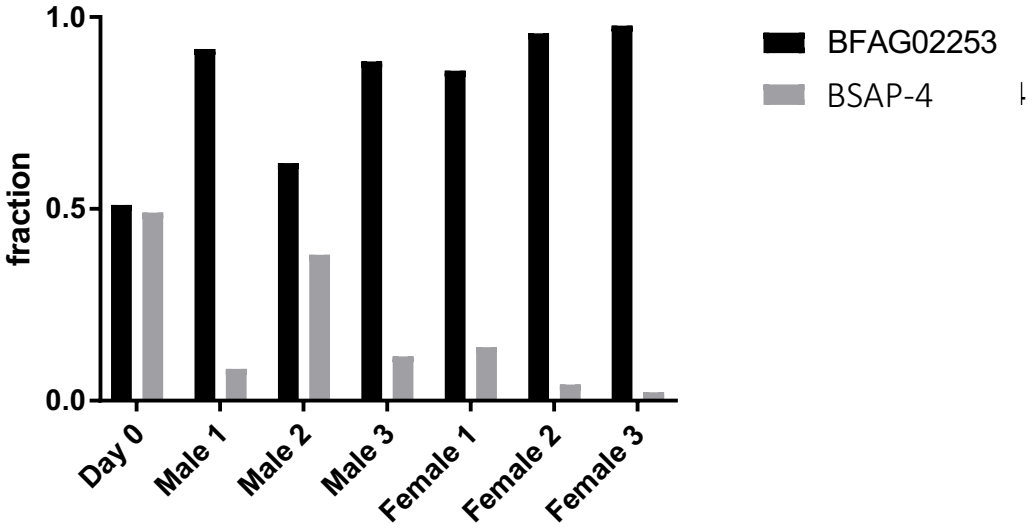


Figure 4.2 Mouse gut competition between isogenic strains of *B. thetaiotaomicron* expressing BSAP-4 ("BSAP-4") and BSAP-4 sensitive receptor ("BFAG02253"). Fecal samples were collected on day 7 post-gavage.

We further studied the fitness contribution of the receptor by competing the *B.thetaiotaomicron* strain expressing either the resistant receptor BF638R_2715 or the sensitive receptor BFAG02253. We hypothesized that the sensitive receptor would have equal or greater fitness compared with the resistant receptor. We were again surprised to find that BF638R_2715 significantly out-competed BFAG02253 (fig. 4.3). *B. thetaiotaomicron* expressing BF638R_2715 was present in 85% of analyzed colonies (P = 0.0002), while *B. thetaiotaomicron* expressing BFAG02253 was present in 15% of analyzed colonies (P = 0.0001).

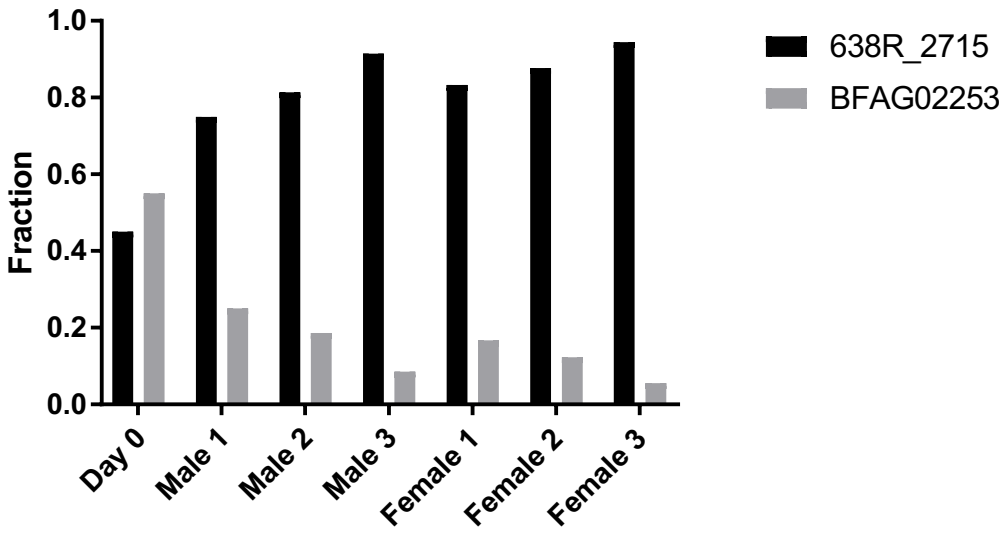


Figure 4.3 Mouse gut competition between isogenic strains of *B. thetaiotaomicron* heterologously expressing the sensitive receptor from 3_1_12 ("BFAG02253") or the non-targeted ortholog from 638R ("638R_2715"). Fecal samples were collected on day 7 post-gavage.

In addition, we sought to confirm that BSAP-4-expressors would fail to out-compete strains expressing the resistant receptor. We competed isogenic strains expressing either BSAP-4 or the resistant receptor BF638R_2715, hypothesizing that BSAP-4 would fail to antagonize the strain expressing the resistant receptor and that the latter would have a fitness advantage over the BSAP-4-expressor. As expected, the resistant receptor was found in 86% of colonies analyzed, indicating significant out-competition by the receptor-expressor ($P < 0.0001$); BSAP-4 was found in 14% of colonies analyzed ($P = 0.0128$) (fig. 4.4).

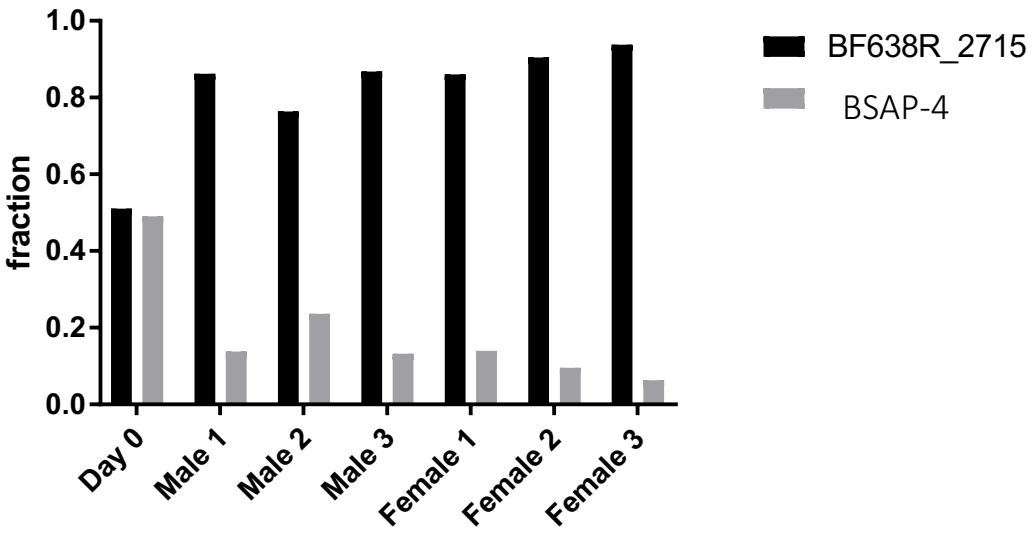


Figure 4.4 Mouse gut competition between isogenic strains of *B. thetaiotaomicron* heterologously expressing BSAP-4 ("BSAP-4") or the non-targeted receptor ortholog from *B.f.*638R ("BF638R_2715"). Fecal samples were collected on day 7 post-gavage.

The mouse experiments were surprising because they showed the opposite of the effect I expected. The results suggest that the receptor confers a fitness advantage to *B. thetaiotaomicron* that overcomes the fitness disadvantage incurred by toxin sensitivity. Given the identity of BFAG02253 as an outer membrane transporter, it is likely that the advantage is nutritional. *B. thetaiotaomicron* VPI-5482, the strain used as the heterologous expression strain in these studies, contains four members of the Calycin_like protein family (PF13944), one of which (BT_0498) is 34% identical with 83% coverage to BFAG02253. Like BFAG02253, BT_0498 is the first gene in a 13-gene operon containing a putative lipoprotein, a TonB-dependent receptor, a cobalamin biosynthesis gene, and a protoporphyrin IX magnesium chelatase (fig. 4.5). This means that in the background of *B. thetaiotaomicron*, BFAG02253 is either redundant; its overexpression is additive to the pre-existing Calycin_like proteins and their pathways; or that BFAG02253 transports different molecules than BT_0498. Whether BFAG02253 transports vitamin B12, heme, or other small molecule, my results are a reminder that toxin receptors do not exist solely to confer sensitivity to toxins.

Given the unexpected fitness benefit conferred to B. thetaiotaomicron expressing BFAG02253, a different set of experiments might capture the interactions of BF638R_2714 and its receptors with greater verisimilitude. Cloning the toxin and the sensitive receptor, from BF638R and therefor giving the toxin producing strain the equal fitness advantage would be predicted to allow for the toxin effect to be seen. However, when these two orthologous receptors were placed in *B. thetaiotaomicron*, the 638R receptor conferred a better fitness advantage than the receptor of 3_1_12.

The significant fitness benefits conferred by BFAG02253 and BF638R_2715 when heterologously expressed in *B. thetaiotaomicron* were not observed in the wild-type *B. fragilis* J38-1 in competition with its receptor knockout. The *B.f.* J38-1 genome does not contain additional calycin_like proteins, ruling out the possibility that a functional ortholog complements the deleted receptor. One explanation is that the basal level of receptor expression in *B. fragilis* J38-1 is lower than heterologous expression of receptors from a plasmid in *B. thetaiotaomicron*. This would be a provocative finding because cells are expected to exert precise control of expression of nutrient acquisition and other metabolic proteins. However, conditions under which the native metabolic controls of *B. fragilis* strains evolved are likely strongly divergent from the gnotobiotic mouse gut. Expression levels and expression control could explain the discrepancy between the fitness benefits of receptor expression in *B. fragilis* J38-1 and *B. thetaiotaomicron*.

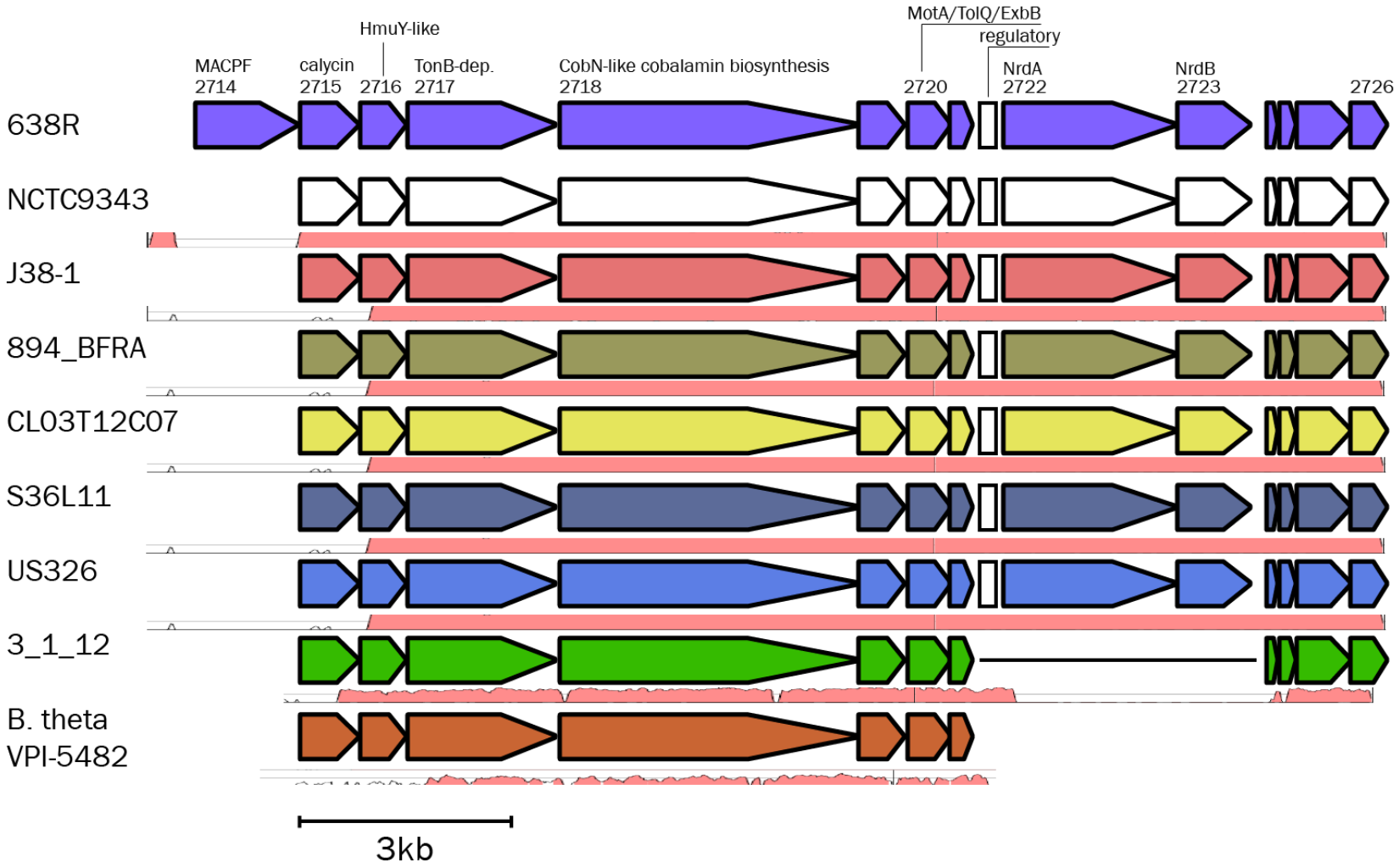


Figure 4.5 Alignment of biosynthesis operon containing BSAP-4 and calycin_like orthologs of the BSAP-4 receptor. Graph beneath each alignment represents % identity of the corresponding sequence compared with 638R, with pink indicating >70% identity. Note that NCTC9343 does not contain BSAP-4, but does contain the non-targeted receptor ortholog. A putative regulatory region is also annotated. 3_1_12 and *B. theta* VPI5482 have large gaps in this operon.

4.2 Distribution of BSAP-4 and receptors in nature

Our discovery and characterization of BSAP-4 and its receptors prompted us to interrogate the distribution of these genes in the human gut. We used BLASTP to search for BSAP-4, the resistant receptor, and the sensitive receptors BFAG02253 and M068_2717 in a database of full genomes and shotgun assemblies from the order Bacteroidales[2,3]. BSAP-4 and its receptors (both resistant and sensitive) had >85% orthologs only in strains of *B. fragilis;* the most similar receptor ortholog found in other species is BT_0498, which is 48% similar to the receptor from J38-1. All of the proteins had orthologs in numerous *B.f.* strains, with the exception of BFAG02253, which was only found in three other strains of *B.f.* The BSAP-4 protein sequence and the resistant receptor were found in 30 strains of *B.f.;* BFAG02253 was found in 5 strains; and orthologs of M068_2717 were found in ~50 strains. BSAP-4 was never present without a resistant receptor. No genomes contained both BSAP-4 and a sensitive receptor or the resistant receptor and a sensitive receptor. Most strains that contained the resistant receptor also contained BSAP-4, but some strains contained only the resistant receptor.

We were intrigued by genomes containing the resistant receptor without BSAP-4. One of these strains is the type strain *B.f.* NCTC9343 (fig. 4.5). Alignments of this region in 9343 and 638R revealed a high degree of homology at the nucleotide level, including 100% correspondence between 638R_2715 and 9343_2621. Indeed, the only difference between the two strains in this region appears to be an insertion of 10 nucleotides and the BSAP-4 ORF in BF638R, with the alignment resuming directly after the stop codon of 638R_2714.

These analyses revealed the distribution of these genes within single *B. fragilis* genomes, but did not capture the degree to which these genes co-exist in a single ecosystem. We next sought to characterize the prevalence of BSAP-4 and its receptors at the ecosystem level by analyzing metagenomes of gut microbiotas. We searched for BSAP-4, the resistant receptor,

sensitive receptor orthologs in the "3CGC" (three cohorts gene catalog) compendium of metagenomes[4]. This catalog contains 1,267 metagenomes originating from 1,070 individuals (including American, Chinese, Danish and Spanish), comprising 159,325,886 predicted proteins. We queried the catalog using BSAP-4, the resistant receptor BF638R_2715 and the sensitive receptors BFAG02253, and M068_2717. We used BLASTP to identify proteins with >95% identity and greater than 20% query coverage, and unique participant IDs (since some individuals contributed multiple metagenome samples). This relatively permissive cut-off yielded 387 returns, including 103 BSAP-4 orthologs, 171 resistant receptor orthologs, and 113 sensitive receptor orthologs. This dataset allowed us to examine the co-occurrence of all queries in a single metagenome. The resistant receptor was found without BSAP-4 in 68 metagenomes, suggesting that BSAP-4 and resistant receptor are not always found together as is the pattern with BSAP-1. Only 1 metagenome contained both a resistant and a sensitive receptor, showing the co-occurrence of two *B.f.* strains. Finally, none of the 103 metagenomes that contained BSAP-4 also contained any of the sensitive receptors. These data suggest that BSAP-4 producing strains and BSAP-4 sensitive strains do not co-occur in the human gut.

Figure 4.6 Heatmap of co-occurrence of BSAP-4, sensitive receptors, and non-targeted receptor orthologs in 3CGC dataset. 1,267 individual metagenomes are juxtaposed from left to right, with blue indicating present and red indicating absent. BSAP-4 and sensitive receptors never co-occur. When one sensitive receptor is present, all four sensitive receptors are shown as present due to their mutual >95% identity.

We next attempted to identify the number of unique proteins that are slight variants of each receptor type. In total, we identified 5 unique orthologs of BSAP-4; 14 orthologs of the resistant receptor ortholog BF638R_2715; 3 orthologs of BFAG02253, and 8 orthologs of M068_2717. All orthologs of BSAP-4 were greater than 99.6% identical to one another, with the exception of an ortholog with only 95.85% identity found in the metagenome of a Chinese individual. All orthologs of BFAG02253 were ~99% identical, while there appeared to be greater diversity in the other sensitive receptors.

Suspecting that the protein blast algorithm failed to return all orthologs of BSAP-4 receptors, we used HMMER to generate a profile hidden Markov model (HMM) of a Clustal Omega alignment of the sensitive receptor proteins[5,6]. This model was used to scan the 3CGC subset. The hmmscan returned 3,680 results. The results were filtered by cutoffs for i) full sequence e-values $>10^{-3}$; ii) best domain e-values $> 10^{-3}$; iii) proteins annotated as "complete" in

the 3CGC set; iv) protein lengths between 260-300 amino acids. The resulting 829 putative orthologs were further reduced by eliminating duplicate returns from the same participant ID, and made non-redundant by eliminating 100% identical proteins, leaving 122 unique sequences. To identify which of these sequences are potential functional orthologs of the receptors, we first identifed the best pfam hits for the original queries[7]. "Calycin_like" and "Peptidase_AF" were the best-scoring pfams for all the sensitive receptors. The HMM models of these pfams (PF13944.5 and PF01828.16, respectively) were extracted from Pfam 31.0, and hmmsearch was used to run the 122 non-redundant queries (and the original sensitive receptor protein queries) against these profile HMMs. This search returned 122 hits to Calycin_like and 10 hits to Peptidase_A4; 9 queries were hits to both (see neighbor-joining tree and protein identity matrix) (fig. 4.7). These results suggest that the metagenomic dataset contains many functional orthologs of the newly characterized sensitive receptors that were not identified in our BLAST search (Appendix D contains protein identity matrix of these orthologs). The broad distribution of receptor orthologs supports their status as important contributors to fitness in the mammalian gut, and may extend the landscape of toxin-mediated antagonism.

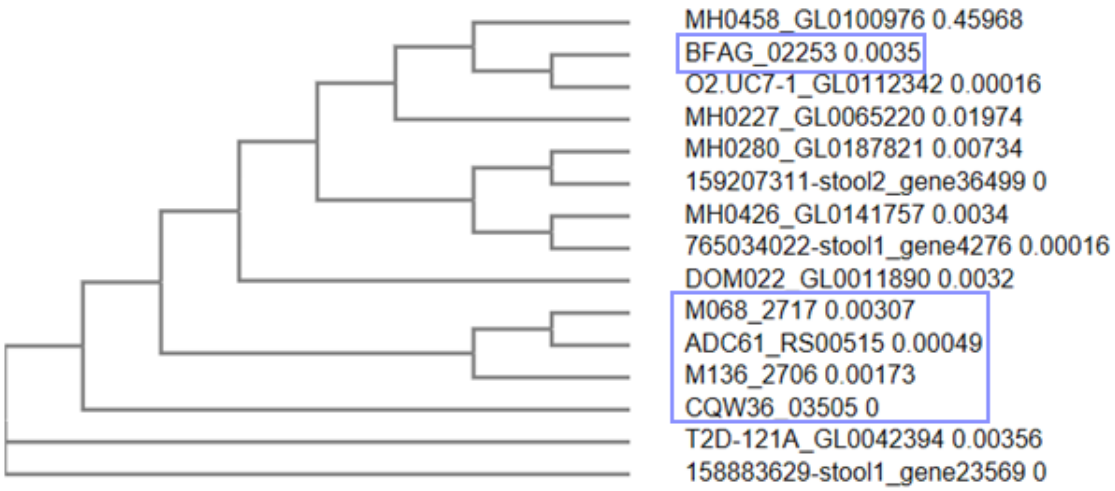


Figure 4.7 Neighbor-joining tree of 14 predicted orthologs of BSAP-4 receptors in gut metagenomes. The 14 proteins, including the 5 characterized in this study (blue boxes), were identified by HMM models against the 5 sensitive receptors. Tree generated by Clustal Omega alignment without distance corrections.

4.3 Conclusion

In this chapter, I describe gnotobiotic mouse experiments that suggest either unexpected fitness benefits conferred to *B. thetaiotaomicron* by predicted calycin_like outer membrane transporters, or fitness disadvantages conferred by toxin expression. Separating these fitness effects may be accomplished by competition of i) a strain expressing the toxin vs. a strain containing the empty vector, and ii) a strain expressing a receptor vs. a strain containing the empty vector. Future studies should seek to demonstrate the function of the operon to which the toxin and its receptors belong, which is likely to be consequential to microbial fitness in the mammalian gut.

Despite the inability of strains expressing BSAP-4 to out-compete sensitive strains in the mouse gut, the absolute mutual exclusivity of BSAP-4 and sensitive receptors in human gut metagenomes suggests that BSAP-4 is not inert in terms of ecological effects. The simplified character of the gnotobiotic mouse gastrointestinal tract may obscure the true nature of this newly identified antimicrobial interaction. The widespread distribution of these orthologs and BSAP-4 in natural metagenomes further supports their ecological importance. Given the importance of non-redundant functional capacities of microbes, it may be fruitful to look further into mutual exclusivity to assess the extent to which gross changes can result from specific antimicrobial interactions. Determining how many species are excluded—and the cascading effects that this can have on other more distantly related species of gut microbes—will help to develop a fuller picture of how the network of antagonistic behaviors affects gut ecology and ultimately the health of the host.

4.4 Methods

Gnotobiotic mouse experiments

Mouse studies were approved by the Harvard Medical Area Standing Committee on Animals. Germ-free Swiss-Webster mice (at least 5 weeks of age) were housed in sterile OptiMice cages

(Animal Care Systems). Pairs of bacterial strains destined for mouse gut colonization were grown to mid-log phase, adjusted to equivalent optical densities, and gavaged into mice. Each colonization experiment was performed in 3 male and 3 female mice. Male mice and female mice were housed separately. The initial ratio of each strain was determined by plating and analysis of 96 colonies by colony PCR. At 1 week post-gavage, fecal samples were collected, diluted in PBS and bacteria plated for single colonies, and at least 90 colonies per mouse were analyzed by colony PCR. For experiments using strains in which genes were expressed from plasmids, drinking water was supplemented with 1mg/ml of clindamycin to maintain plasmids.

Bioinformatic analysis of BSAP-4 receptor orthologs

Databases were searched for orthologs using BLASTP[8]. Alignments of orthologs were generated using Clustal Omega[9]. The 3CGC (three cohorts gene catalog) was used as a database of human gut metagenomes[4]. Profile hidden Markov models were generated using HMMER[5], and databases scanned with hmmscan. Profile HMMs were also extracted from Pfam 31.0[7]. Proteins were classified according to their highest scoring Pfam.

4.5 References

1. Martín, R., Bermúdez-Humarán, L. G. & Langella, P. Gnotobiotic Rodents: An In Vivo Model for the Study of Microbe–Microbe Interactions. *Front. Microbiol.* **7,** (2016).
2. Coyne, M. J., Zitomersky, N. L., McGuire, A. M., Earl, A. M. & Comstock, L. E. Evidence of Extensive DNA Transfer between Bacteroidales Species within the Human Gut. *mBio* **5,** e01305-14-e01305-14 (2014).
3. Priyam, A. *et al.* Sequenceserver: a modern graphical user interface for custom BLAST databases. (2015). doi:10.1101/033142
4. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* **32,** 834–41 (2014).

5. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39,** W29-37 (2011).
6. Sievers, F. *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7,** 539 (2011).
7. Bateman, A. *et al*. The Pfam protein families database. *Nucleic Acids Res* **30,** 276–80 (2002).
8. Ye, J., McGinnis, S. & Madden, T. L. BLAST: improvements for better sequence analysis. *Nucleic Acids Res* **34,** W6-9 (2006).
9. Chenna, R. *et al*. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31,** 3497–500 (2003).

Appendix A: EMS Mutagenesis

Antibiotic sensitization of a natural isolate of *B. thetaiotaomicron* using ethyl methanesulfonate (EMS) mutagenesis

A panel of spot agar tests between natural isolates of *B. thetaiotaomicron* revealed an antimicrobial interaction between *B. thetaiotaomicron* CL12 and *B. thetaiotaomicron* CL01. This suggested the presence of an uncharacterized diffusible toxin. Genetic studies in these strains were precluded by natural resistance of both strains to erythromycin and tetracycline. I attempted to sensitize these strains to antibiotics by using chemical mutagenesis to mutate latent genes that encode antibiotic resistance mechanisms[1,2]. I used ethyl methanesulfonate, a chemical mutagen that alkylates guanine, leading to transition mutations (G>A) during DNA replication, to mutagenize the toxin-producing strain BtCL012. I screened the resulting library for erythromycin- and tetracycline-sensitive mutants. Finally, I confirmed that antibiotic-sensitive mutants retained antimicrobial activity against BtCL01.

BtCL012 were grown to mid-log phase and treated with 1.4% EMS for various lengths of time, up to 60min. Cells were washed to remove residual EMS, which was neutralized with sodium thiosulfate (see methods for safety considerations), and recovered for ~6hr in supplemented basal medium and frozen in 20% glycerol. A killing curve was generated by comparing viability of EMS-treated cells against an untreated control. Recovered, frozen cells treated in conditions resulting in 95-99% killing were rapidly thawed, diluted and plated for single colonies on supplemented BHI medium. These colonies were replica-plated onto BHIS medium supplemented with 10ug/ml erythromycin, 6ug/ml tetracycline, or 5ug/ml erythromycin and 3ug/ml tetracycline. The resulting plates were screened for ghosts, generating a set of putative antibiotic-sensitive mutants. The growth rates of these mutants in the presence of antibiotics were measured to confirm sensitivity. Six mutants with significantly impaired growth in the presence of antibiotics were subjected to spot agar tests with overlaid BtCL01. All

six produced zones of clearing in the BtCL01 lawn, suggesting that they retained antimicrobial activity.

Safety considerations

Ethyl methanesulfonate, or EMS (CAS # 62-50-0), is acutely toxic (category 4), mutagenic (category 1B), and carcinogenic (category 2). EMS was handled exclusively in a chemical fume hood with personal protective equipment including nitrile gloves, lab coat and safety eyewear. The half-life of EMS in 10% sodium thiosulfate and 1% NaOH is 17 minutes at 37degC and 1.4 hours at 20degC (ref 2). EMS waste was transferred to a beaker containing excess neutralizing solution and left at room temperature for 24hr before labeling for disposal. Solid waste was double-bagged and discarded in the appropriate waste container.

Cells to be mutagenized were grown to mid-log phase, washed twice with PBS, and resuspended in 1.5ml of ice-cold PBS, placed on ice and transferred to a chemical fume hood. A 250ul aliquot was reserved as an untreated control. EMS was added to the remaining cells for a final concentration of 1.4% v/v, and 250ul aliquoted to 5 2ml round-bottom microcentrifuge tubes, which were closed, sealed with parafilm and vortexed briefly to mix. Tubes were incubated at 37degC with mild agitation for 0min, 15min, 30min, 45min or 60min; the untreated control was processed along with the 0min timepoint. After the indicated length of time, cells were washed twice with PBS and waste discarded in the beaker containing neutralization buffer. During wash steps, cells were centrifuged at 21,000g for 2min to ensure complete sedimentation of cells and minimal disturbance during washing.

Washed cells were resuspended in 500ul of pre-reduced supplemented basal media, titered for viable cells, and then incubated at 37degC anaerobically for 6 hours, enough for two doublings. The recovered cells were suspended in 20% glycerol and placed at -80degC. After 48hr of incubation, the concentration of viable cells in each EMS treatment was determined by counting colony forming units, and the killing percentage calculated by comparison with the

untreated control. The treatments yielding greater than 95% killing—the 45min and 60min timepoint treatments—were selected for screening for loss of antibiotic resistance. The corresponding frozen stocks were thawed, approximate cells/ml estimated by measuring OD600, and approximate viable cells/ml calculated by multiplying by the killing percentage for the corresponding condition. This figure was used to dilute and plate the library on BHIS plates with a desired density of ~300 cfu per petri dish.

The colonies recovered on permissive BHIS plates were replica-plated to the restrictive condition of BHIS supplemented with either 10ug/ml erythromycin, 6ug/ml tetracycline, or 5ug/ml erythromycin + 3ug/ml tetracycline. Out of ~5,000 screened mutant colonies, 2 mutants had significantly impaired growth rates in the presence of erythromycin (data not shown). Crucially, these mutants retained antimicrobial activity against BtCL01 as shown by spot agar tests. The erythromycin sensitivity and robust antimicrobial activity make render these mutants amenable to the transposon mutagenesis protocol described in Chapter 1 of this thesis.

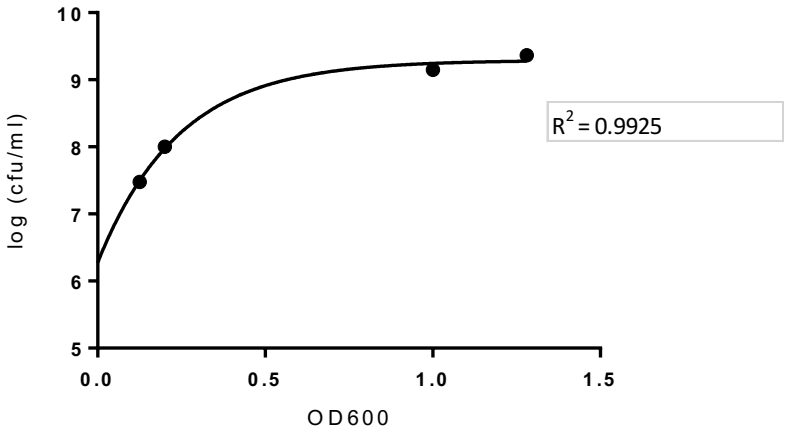


Figure A1.1 Colony forming units per milliliter (cfu/ml) of *Bacteroides thetaiotaomicron* as a function of optical density. Optical density of various cultures of mid- to late-log *B. thetaiotaomicron* CL012 (wild type, Brucella broth) were measured in 0.1cm cuvette in Nanodrop. Cultures were then titered on BHIS plates and colonies were counted. Curve was fit to one phase exponential decay in Graphpad yielding Y0 = 6.725, plateau = 9.292, K=4.142. (`Y=(Y0 - Plateau)*exp(-K*X) + Plateau`). These parameters were routinely used to estimate cfu/ml from OD600.

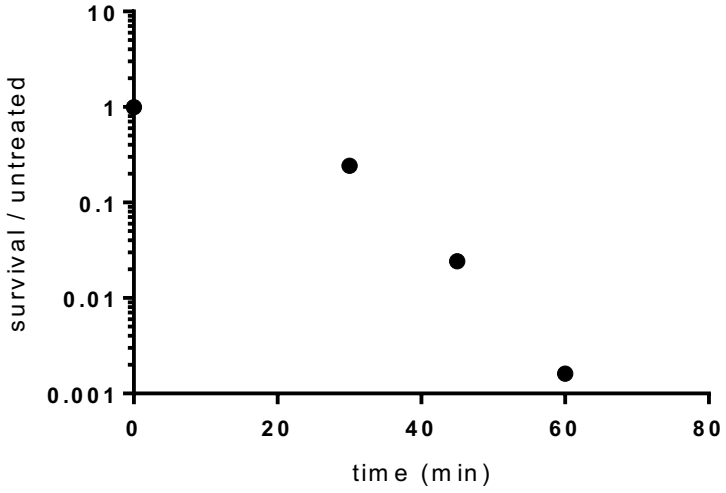


Figure A1.2 Killing curve of EMS mutagenesis of *B. theta* CL012. *B. theta* CL012 was subjected to EMS mutagenesis as described above. Survival ratio was calculated by dividing cfu/ml of a given sample by the cfu/ml of the untreated control.

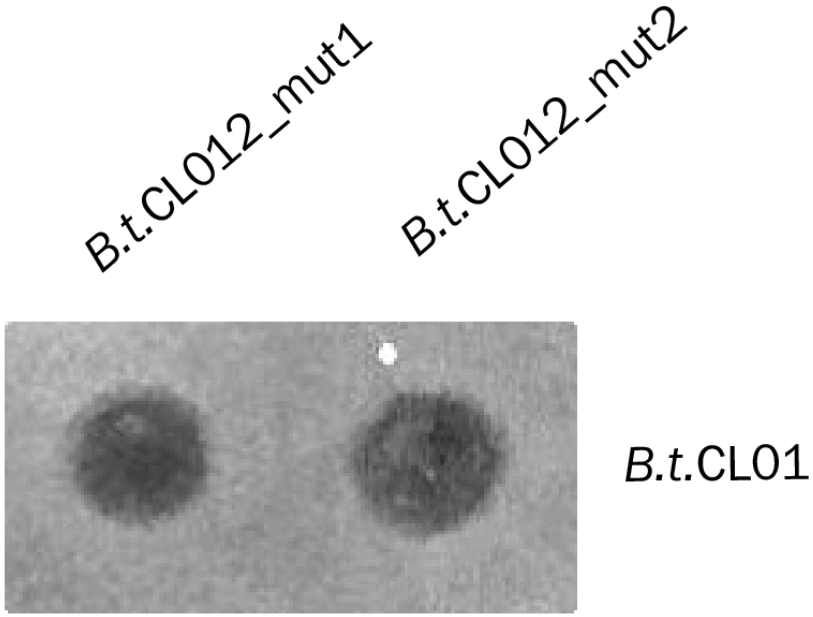


Figure A1.3 Erythromycin-sensitive mutants of *B. theta* CL012 retain killing activity against *B.t.*CL01

Appendix B: Primers, plasmids, and strains

Table B1: Primers described in this dissertation

Purpose	Primer Sequence[a]	Primer Orientation
Expression of BF638R_2714	tattggatccccacctacctttgcaaacagttat	Forward
	ctgcggatccgcccttcttattgtattgtatcgtc	Reverse
	cggccgctctagaactagtgatccgcccttcttattgtattg	Forward
	ttagtcgacgtcgacggtatcagcatcatgaagaagataaaaata g	Reverse
Expression of BF638R_2715	aatcagaattgactctagaggggatgacgatacaatacaataag	Forward
	attcgagctcggtacccgggataagcatcgttcccagc	Reverse
	cattgaacgaacgttggggatgacgatacaatacaataag	Forward
	attcgagctcggtacccgggataagcatcgttcccagc	Reverse
	ttagtcgacgtcgacggtatcagcatcatgaagaagataaaaata g	Forward
	cggccgctctagaactagtgataagcatcgttcccagcag	Reverse
Expression of BFAG02253 (3-1-12 receptor)	agctggatcctaggtattgtaagccctttttattg	Forward
	agctggatcctgagattgttctgttatcaggtc	Reverse
	cggccgctctagaactagtgtgctgcaaggcgattaatg	Forward
	ttagtcgacgtcgacggtatacaattacggctgacaatg	Reverse
	tggaatcgatcacaattacggctgacaatgga	Forward
	agctggatcctgagattgttctgttatcaggtc	Reverse
Expression of M068_2717 (J38-1 ortholog)	attgggatcctaggtattgtcaacccttttattg	Forward
	gcagggatccataagcatcgttcccagcag	Reverse
	cggccgctctagaactagtgcacaattacggctgacaatg	Forward
	ttagtcgacgtcgacggtatataagcatcgttcccagc	Reverse
Expression of M136_2706 (S36L11 ortholog)	attgggatcctaggtattgtcaacccttttattg	Forward
	gcagggatccataagcatcgttcccagcag	Reverse
	cggccgctctagaactagtgcacaattacggctgacaatg	Forward
	ttagtcgacgtcgacggtatataagcatcgttcccagc	Reverse
Expression of His-tagged BF638R_2714	cactcatatgtgcaataaaacagatctgatcac	Forward
	cacccatatgttattgtattgtatcgtcatcccac	Reverse

Table B1: Primers (continued)

Purpose	Primer Sequence[a]	Primer Orientation
Deletion of BF638R_2714	cggccgctctagaactagtggtataacctaaccggaaaaag	Upstream flank forward
	catcccacatctaattcccaaaagacttagtttg	Upstream flank reverse
	tgggaattagatgtgggatgacgatacaatac	Downstream flank forward
	cgaattcctgcagcccggggcctttcactatttccactcg	Downstream flank reverse
Deletion of M068_2717	cggccgctctagaactagtggtgcagtaagctccaggttg	Upstream flank forward
	gaaagtaatggtcagtccgtgcacaagcgcacagc	Upstream flank reverse
	gctgtgcgcttgtgcacggactgaccattactttc	Downstream flank forward
	cgaattcctgcagcccgggggatcgtccgaagctttg	Downstream flank reverse
Site-directed mutagenesis of CL03 receptor	aagcaccTcccactttcgaaatatcg	Forward P254 primer
	tgggAggtgctttggctgcatc	Reverse P254 primer

a. Restriction enzyme recognition sites are underlined

Table B2: Plasmids described in this dissertation

Plasmid	Description
pFD340	Mobilizable expression vector for heterologous expression of genes in *Bacteroides*
pMCL140	Mobilizable expression vector for heterologous expression of genes in *Bacteroides*
pMCL177	pMCL140 expressing BSAP-4
pKNOCK	For construction of scarless deletion mutants in *Bacteroides*
pET16b	Cloning and expression of His-tagged proteins
RK231	helper plasmid for tri-parental mating (KanR)
pWH2	Mariner transposon plasmid (maintained in s17 lambda pir)

Table B3: Bacterial strains described in this dissertation

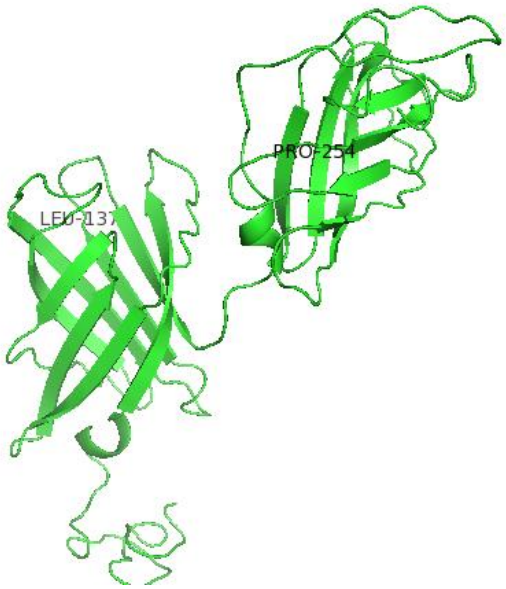
Species	**Strain**	**Description**
B. fragilis	638R	Produces BSAP-4, BSAP-1, ubb
B. fragilis	638RΔBSAP-1Δubb	Produces BSAP-4
B. fragilis	3_1_12	sensitive to BSAP-4
B. fragilis	J38-1	sensitive to BSAP-4
B. fragilis	S36L11	sensitive to BSAP-4
B. fragilis	CL03T12C07	sensitive to BSAP-4

Table B3: Bacterial strains (continued)

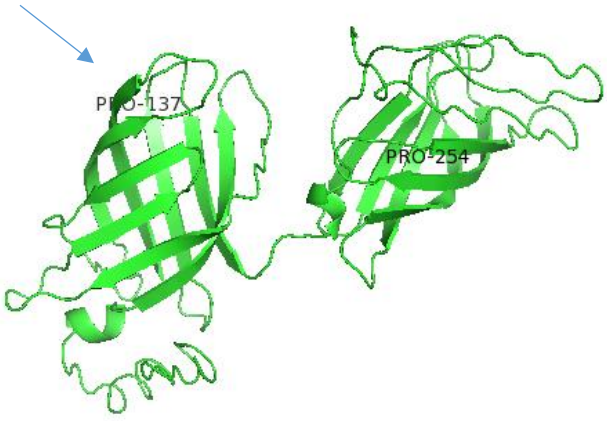
Species	**Strain**	**Description**
B. fragilis	LM46	resistant to BSAP-4
B. fragilis	LM001	resistant to BSAP-4
B. fragilis	NCTC9343	resistant to BSAP-4; contains 638R_2715 ortholog. Kills 3_1_12
B. fragilis	CM13	produces antimicrobial molecule vs. 3_1_12
B. fragilis	1277476	resistant to BSAP-4
B. fragilis	12815501	resistant to BSAP-4
B. thetaiotaomicron	CL01T08C14	Killed by CL012
B. thetaiotaomicron	CL12T00C04	Kills CL01
B. thetaiotaomicron	VPI-5482	Not sensitive to BSAP-4, does not kill *B. fragilis*
E. coli	s17-lambda pir	cloning and mating
E. coli	Dh5a	cloning strain
E. coli	BL21(DE3)	Protein production strain for His-tagged protein

Appendix C: Structural homology modeling

Structural homology models of BSAP-4 receptor orthologs were generated using I-TASSER[3] and Phyre2[4]. Models were based on the 2.4A crystal structure of BF2706, a calycin_like beta barrel outer membrane protein from *B.f.* NCTC9343 (PDB: 3RWX). I-tasser models are shown below. Phyre2 returned similar results.

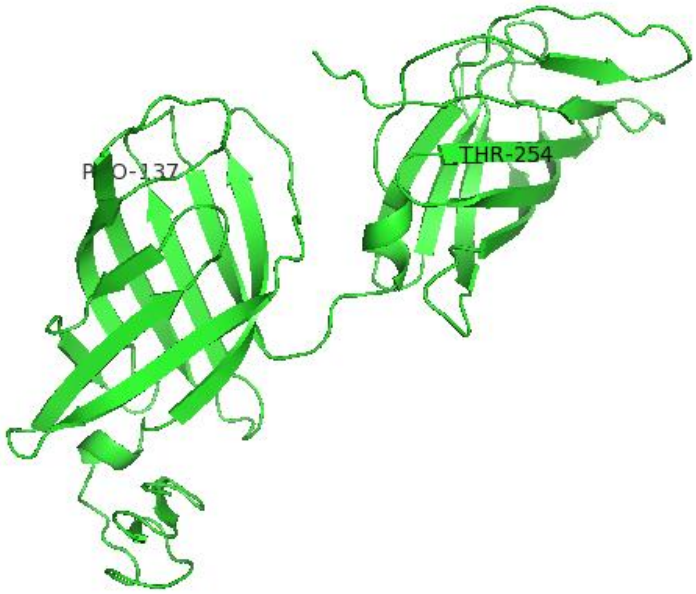


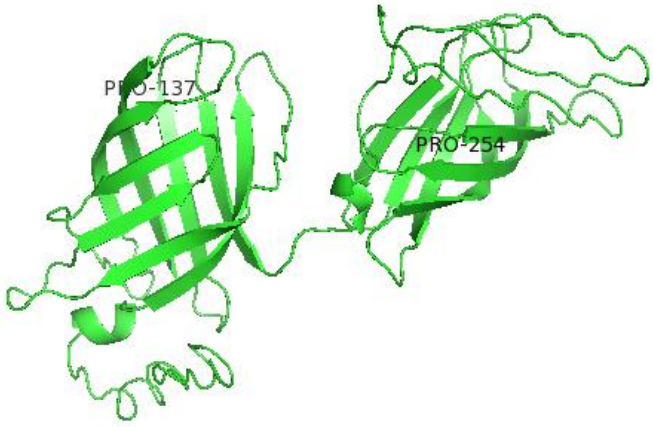
Above: *B. fragilis* J38-1. Leu137 and Pro254 both localize to the ends of beta sheets; prolines at these locations would likely affect loop topology or beta strand structure.



Above: Pro137 may cause different loop structure than Leu137.

Below: P254 and T254 may cause subtle beta sheet structural shifts.

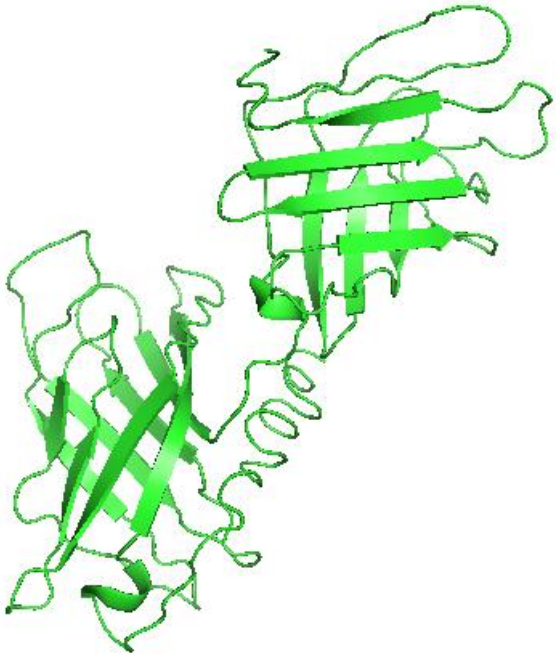




For reference, the homology models of BF638R_2715 (the non-targeted receptor ortholog) and BFAG02253 (the receptor from 3_1_12) are shown below.

Above: 638R_2715 calycin_like homology model

Below: BFAG02253 (from 3_1_12) calcyin_like homology model



Appendix D: Protein identity matrix of orthologous receptors

Protein identity matrix of orthologs identified by hidden Markov models generated against sensitive receptors. Non-redundant protein screening yielded a list of 15 unique proteins found in the 3CGC dataset. Highlighted proteins are the sensitive receptors characterized in this work.

	MH0458_GL0100976	BFAG_02253	O2.UC7-1_GL0112342	MH0227_GL0065220	MH0280_GL0187821	MH0426_GL0141757	DOM022_GL0011890	T2D-121A_GL0042394	765034022-stool1_gene427(	159207311-stool2_gene3649	M068_2717	ADC61_RS00515	CQW36_03505	158883629-stool1_gene2356	M136_2706
MH0458_GL0100976	100.0	43.3	43.8	41.8	42.0	42.6	42.2	42.2	42.6	42.6	42.2	42.2	42.6	42.6	42.6
BFAG_02253	43.3	100.0	99.6	76.1	77.3	77.9	78.3	77.6	77.9	77.6	77.6	77.6	77.9	77.9	77.9
O2.UC7-1_GL0112342	43.8	99.6	100.0	76.5	77.7	77.7	78.8	78.0	77.7	78.0	78.0	78.0	78.4	78.4	78.4
MH0227_GL0065220	41.8	76.1	76.5	100.0	96.7	96.8	97.1	97.1	97.1	97.5	96.8	97.1	97.5	97.5	97.1
MH0280_GL0187821	42.0	77.3	77.7	96.7	100.0	98.5	98.5	98.5	98.9	99.3	98.2	98.5	98.9	98.9	98.5
MH0426_GL0141757	42.6	77.9	77.7	96.8	98.5	100.0	98.6	98.6	99.6	99.3	98.2	98.6	98.9	98.9	98.6
DOM022_GL0011890	42.2	78.3	78.8	97.1	98.5	98.6	100.0	99.3	98.9	99.3	98.9	99.3	99.6	99.6	99.3
T2D-121A_GL0042394	42.2	77.6	78.0	97.1	98.5	98.6	99.3	100.0	98.9	99.3	98.9	99.3	99.6	99.6	99.3
765034022-stool1_gene4276	42.6	77.9	77.7	97.1	98.9	99.6	98.9	98.9	100.0	99.6	98.6	98.9	99.3	99.3	98.9
159207311-stool2_gene36499	42.6	77.6	78.0	97.5	99.3	99.3	99.3	99.3	99.6	100.0	98.9	99.3	99.6	99.6	99.3
M068_2717	42.2	77.6	78.0	96.8	98.2	98.2	98.9	98.9	98.6	98.9	100.0	99.6	99.3	99.3	99.6
ADC61_RS00515	42.2	77.6	78.0	97.1	98.5	98.6	99.3	99.3	98.9	99.3	99.6	100.0	99.6	99.6	99.3
CQW36_03505	42.6	77.9	78.4	97.5	98.9	98.9	99.6	99.6	99.3	99.6	99.3	99.6	100.0	100.0	99.6
158883629-stool1_gene23569	42.6	77.9	78.4	97.5	98.9	98.9	99.6	99.6	99.3	99.6	99.3	99.6	100.0	100.0	99.6
M136_2706	42.6	77.9	78.4	97.1	98.5	98.6	99.3	99.3	98.9	99.3	99.6	99.3	99.6	99.6	100.0

References

1. Cupples, C. G. & Miller, J. H. A set of lacZ mutations in Escherichia coli that allow rapid detection of each of the six base substitutions. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 5345–5349 (1989).
2. Foster, P. L. In vivo mutagenesis. *Methods Enzymol.* **204,** 114–125 (1991).
3. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12,** 7–8 (2015).
4. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).