**The Structure and Function of the Starch Utilization System in *Bacteroides thetaiotaomicron***

by

**Elizabeth Ann Cameron**

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in the University of Michigan
2014

Doctoral Committee:

Assistant Professor Eric C. Martens, Chair
Assistant Professor Nicole M. Koropatkin
Associate Professor Maria B. Sandkvist
Associate Professor Mark A. Saper
Professor Michele S. Swanson

**Acknowledgements**

I would first like to thank my thesis mentor Dr. Eric Martens for all of his support and guidance over the past five years. Thank you for pushing me to test my hypotheses from every angle and always listening to and valuing my ideas. You have always been my biggest advocate, which has given me the confidence to become a more independent scientist. I have been very lucky to have another fantastic mentor in Dr. Nicole Koropatkin. Nicole you have consistently challenged me to think about things in new and different ways and have taught me to always consider the bigger picture. I also thank you for your constant support in lab and in life.

I would like to thank my committee members: Dr. Michele Swanson, Dr. Maria Sandkvist and Dr. Mark Saper. Thank you for your feedback and guidance. I would like to thank Michele Swanson in particular for her advice and support during my search for a post-doctoral position.

I would like to thank all current and past members of the Martens and Koropatkin labs for your helpful advice and input over the years. I would particularly like to thank Nick Pudlo for helping me in countless ways and making the lab a fun place to work.

I am grateful to the entire University of Michigan Microbiology and Immunology Department for the opportunity to complete my doctoral work at such a fantastic training institution. I would like to thank all of the faculty and staff for useful discussions and advice throughout my dissertation work.

The work described here would not have been possible without funding from the University of Michigan Genetics Training Program, the Rackham Predoctoral Fellowship and the NIH.

Finally, I would like to thank my family and friends for their love and support. In particular I am grateful to my husband, Jamie Cameron for his constant encouragement, patience and love.

## Table of Contents

## List of Figures

**Figure**

## List of Tables

**Table**

## Abstract

Degradation of polysaccharides is an important function performed by the human gut microbiota. Bacterial carbohydrate metabolism in the gut not only provides the host with a significant portion of their daily nutrients, but is also a major factor shaping the composition of the microbial community. The Bacteroidetes, one of the two dominant bacterial taxa in the human gut, degrade a large number of carbohydrates via expression of unique multi-protein complexes, each targeting a different glycan. The first such system described was the starch utilization system (Sus) in *Bacteroides thetaiotaomicron* (*Bt*), an eight protein system required for the bacterium to metabolize starch. Homologous "Sus-like" systems are found in the majority of gut Bacteroidetes with some species devoting up to 20% of their genome to encoding them. The *Bt* Sus is a model for glycan acquisition by the Bacteroidetes, and the work presented here addresses several important questions regarding the structure and function of individual Sus proteins as well as how these components function together to efficiently acquire and degrade the abundant dietary glycan starch.

The crystal structures of two Sus outer-membrane proteins (OMPs), SusE and SusF, were solved revealing that they both contain multiple starch binding sites. In total the Sus OMPs (SusD,E,F and G) contain eight non-enzymatic starch binding sites that we demonstrate serve unique functions in starch catabolism. The SusD binding site is uniquely involved in initial sensing of available starch, leading to upregulation of the *sus* locus. Conversely, the SusE,F and G binding sites are important during starch catalysis, enhancing starch growth rate in a manner dependent on expression of the *Bt* polysaccharide capsule. We hypothesize these binding sites help overcome the barrier created by the bacterial capsule, which may obstruct access to starch. *In vivo* studies show that the Sus binding sites confer a fitness advantage to *Bt* on a starch-rich diet. Finally, we present the first single-molecule imaging studies performed with live *Bt* cells that provides evidence for a highly dynamic starch-induced Sus complex.

These studies provide important insight into the mechanisms of carbohydrate metabolism by gut symbionts, a process that significantly affects human health.

# Chapter I
# Introduction

## The human gut microbiota

The human gastrointestinal tract is home to trillions of bacterial cells that are collectively known as the gut microbiota. This community is established shortly after birth and has a profound effect on health and physiology, providing benefits such as modulation of immune development[1-4], digestion of recalcitrant dietary nutrients[5], and inhibition of pathogen colonization[6]. However, abnormalities in microbiota composition (dysbiosis) have been implicated in several disease states, including inflammatory bowel disease (IBD)[7-9], colon cancer[10-12], antibiotic-associated colitis[13] and obesity[14,15]. Dysbiosis is postulated to result when a typically healthy microbial community becomes unbalanced, due to either increased abundance of potentially harmful microorganisms[9,16,17] or increased flux through harmful metabolic pathways. The normal composition of the gut microbiota, both at single time points and over longer periods of human life, has only been deeply probed within the last several years[18-23]. Therefore, the definition of normal in the human gut microbiota as well as which changes, if any, are causal to the diseases noted above are still active areas of investigation. Current investigations seek to define the dominant forces shaping the microbiota to better understand the causes of dysbiosis and develop strategies to restore a healthy community.

The gut microbiota of a healthy human is home to hundreds of bacterial species, with the majority belonging to just two phyla: the Gram-positive Firmicutes and the Gram-negative Bacteroidetes[18]. Proteobacteria, Actinobacteria, Fusobacteria and Verrucomicrobia are also

common constituents of the gut microbiota although these are found at lower abundances. There is not a single "healthy" gut community structure, and in fact the species present and their abundances differs considerably between individuals[18,21,23]. It has been suggested that inter-individual variation is not continuous and individuals fall into 'enterotypes' or community structures with similar ratios of certain common taxonomic groups[21]. However, the discrete nature of these groups has been challenged, with some studies finding that enterotypes are more fluid than originally described. It was found that the enterotype of an individual can change over time[24] and is influenced by factors like host diet[25]. Additionally, different sequencing and analytical methods used by researchers can affect how individuals are categorized into these groups[26]. Therefore, more research is needed to determine whether gut microbial communities indeed fall into discrete enterotypes and, if so, whether certain enterotypes are linked to particular states of health or disease.

**Bacterial glycan utilization shapes the gut microbiota**

One major factor shaping the composition and physiology of the microbiota is the influx of glycans into the intestine, mostly from diet and host mucosal secretions. Humans consume dozens of different plant and animal-derived dietary glycans, most of which cannot be degraded by enzymes encoded in the human genome. In fact, the human genome contains only 17 potential digestive enzymes (of which only half are confirmed digestive enzymes) that confer the ability to degrade only three dietary glycans: starch, lactose and sucrose[27,28]. In contrast, certain members of the microbiota encode significantly higher numbers of glycosidic enzymes and have much more diversity in the substrates they can degrade. For example in the genome of the human fecal isolate *Bacteroides cellulocyliticus* strain WH2 there are 424 carbohydrate active

enzymes (CAZymes) encoded, belonging to 76 different CAZyme families, 56 of which are not found in the human genome[29]. Computational analysis of a simplified but representative model microbiome found that the Bacteroidetes on average contain more CAZymes per genome than any of the other common phyla in the human microbiome[28]. Because of the more expansive saccharolytic capacity of the microbiota compared to the host, the majority of carbohydrate breakdown occurs by bacteria in the lower intestinal tract. Microbial fermentation transforms these indigestible glycans into short chain fatty acids (SCFA), which serve as nutrients for colonocytes and other gut epithelial cells. Gut microorganisms therefore play a pivotal symbiotic role in helping humans access calories from otherwise indigestible nutrients[30]. Individual microorganisms prefer different glycans. Thus, selective consumption of these nutrients can influence which microbial groups proliferate and persist in the gastrointestinal tract, pointing to dietary glycans as a non-invasive strategy with which humans can directly influence the balance of species in the gut.

***Changes in the human infant microbiota***

The human gut microbiota is established in the first few days of life and is initially seeded from microorganisms encountered during passage through the birth canal and incidental environmental exposures. The bacteria found in the human intestine are typically quite distinct from those found in non-gut environments[31,32] and thus are likely passed from human to human via the fecal oral route. After initial colonization, the gut experiences a series of progressive changes (ecological successions) in the richness and diversity of its inhabitants[33,34]. Early in life, fewer types of glycans transit the gut, as diet is restricted to mother's milk or formula. Many more glycans become available to the microbiota post-weaning as a diet rich in plant and animal

matter is introduced. Despite potential fluctuations in dietary carbohydrates, endogenous host glycans represent a stable source of nutrients for the microbiota over our lifespan. However, at both points (pre- vs. post-weaning), the carbohydrate composition of the gut is one important factor that guides the establishment of the microbial community.

Immediately after birth, infants consume a steady diet of breast milk or infant formula. Several hundred different glycan structures have been identified in human breast milk[35-37] with the primary components being lactose, glucose, galactose, *N*-acetylglucosamine, fucose, sialic acid, and a mixture of complex human milk oligosaccharides (HMOs). This latter class of highly diverse glycans, which seems to be uniquely abundant in human breast milk but not that of other mammals[38], is composed of repeating and variably branched lactose or *N*-acetyllactosamine units that are often decorated with sialic acid and fucose monosaccharides[35,36]. HMOs share structural similarities with human blood group antigens and the *O*-linked structures present in mucus. In contrast to the simpler lactose, most HMOs are not digested by human enzymes, suggesting that they have evolved as natural prebiotics to guide the development of the infant gut microbiota by selectively feeding certain species[39-41].

Studies using culture independent techniques to sample the infant microbiota report substantial temporal and inter-individual variation compared to adults,[33,34,42] and the nascent microbiota typically exhibits an abundance of bacteria from four phyla: Bacteroidetes, Proteobacteria, Firmicutes and Actinobacteria[33]. Although members of these same groups also dominate the adult gut, their proportions are different and more variable in the infant intestine, often varying in genus or species level taxa[33]. At the genus level, higher proportions of *Lactobacillus* and *Bifidobacterium* are observed in infants that are exclusively fed breast milk, suggesting that they may have co-evolved to occupy this niche, outcompeting other colonizers

for available HMOs. Consistent with this, some species of *Bifidobacterium* (*B. infantis* and *B. bifidum*) directly metabolize HMOs[43-48]. Conversely, Lactobacilli seem to prefer the monosaccharide components of HMOs, which may suggest a synergistic effect of Lactobacilli and Bifidobacteria digestion of human milk in the infant gut[49-51].

In contrast to breast-fed infants, formula-fed infants display lower abundance of Lactobacilli and Bifidobacteria and show increased abundance of *Clostridium*, *Bacteroides*, and members of the Enterobacteriaceae[42,52-57]. These observations reveal that cow's milk-based formula, which lacks the amount and diversity of oligosaccharides present in human milk[38], selects for different microorganisms during infancy. The potential long-term effects of these differences remain to be fully evaluated. However, one study found that adult mice exhibited variable susceptibility to chemically induced colitis based on the structure of milk sialyl-lactose oligosaccharides that they consumed during infancy[38]. This study, which found measurable differences in the bacteria that were present in adult mice several weeks after weaning, suggests that some members of the mammalian microbiota with the capacity to impact host health can be selected based on early nutritional conditions and persist in the gut after these conditions have been removed.

The ability of some members of the microbiota to access glycans attached to mucus may also have a role in early colonization by providing some bacteria with a source of endogenous nutrients during a period when dietary glycans are still absent. Due to their chemical similarity bacterial strategies for degrading HMOs and *O*-linked mucin glycans are likely to overlap. Outer membrane enzyme systems in the Gram-negative human gut symbiont *Bacteroides thetaiotaomicron*, which are used to degrade host mucus *O*-glycans[58-60], are also deployed during metabolism of HMOs[48]. Germfree mice colonized with a *B. thetaiotaomicron* mutant lacking

expression of 5 different gene clusters implicated in host glycan foraging in the adult gut was outcompeted by >200-fold relative to wild-type bacteria in a model of natural inter-generational transmission[60]. In this study, pre-weaned pups were exposed to similar amounts of these bacteria from their mother's fecal microbiota, but selectively retained the mucin-degrading strain. Thus, the ability to forage host glycans in the neonatal gut prior to introduction of a more complex diet may be one key parameter that helps species establish colonization.

***Changes post-weaning and adulthood***

The carbohydrate composition of the human diet undergoes a somewhat abrupt change at ~6 months, when complex foods such as cereals, fruits and vegetables are introduced. When such complex plant glycans enter the gut, the composition of the microbiota shifts and microorganisms that prefer these glycans, such as the Gram-negative Bacteroidetes and new species of Firmicutes, become more prevalent[61,33,34]. Recent culture independent metagenomic studies characterizing the functionality of microbial genes present at various times points in the developing human microbiota have noted the presence of genes for plant carbohydrate degradation prior to the introduction of solid food[34,56]. These genes may be harbored in the genomes of glycan generalists like *Bacteroides thetaiotaomicron*, which degrade milk oligosaccharides or host mucin glycans prior to weaning and shift their metabolism to dietary glycans as they are introduced. The presence of glycan-adaptable species pre-weaning suggests that the gut microbiota is primed for the post-weaning dietary change perhaps because the cyclical, fecal-oral transmission of microorganisms from parent to child selects for species that can target glycans present in both the infant and adult gut.

As a fully omnivorous diet is achieved post-weaning, the composition of the microbiota, as measured by abundance of broad taxonomic groups, stabilizes and experiences fewer temporal changes[62-64]. Studies using culture-independent techniques to enumerate the human gut microbiota have found that two bacterial phyla, the Firmicutes and Bacteroidetes, are numerically dominant in the adult microbiota[18,20,65]. However, it has been documented that a third phylum (Actinobacteria) is frequently underestimated using molecular approaches and "universal" primers for the 16S rRNA gene and is therefore also likely to be more abundant than reported[66]. The abundance of Firmicutes is usually greatest. However, the ratios of Bacteroidetes and Firmicutes can change over time and be influenced by different diets, especially those that promote changes in host adiposity[14,65], although a mechanistic explanation for these changes remains to elucidated. Another human study examined the differences between the gut microbiota of African children consuming a predominantly vegetarian, fiber-rich diet and European children consuming a lower-fiber diet that is more typical of Western societies. This study found a higher prevalence of Bacteroidetes/Actinobacteria compared to Firmicutes/Proteobacteria in the African children, and the opposite trend in European children, suggesting that the higher-fiber African diet was conducive to growth of specific fiber-degrading species[67]. Interestingly, the Bacteroidetes genera were quite different between groups: the microbiota of African children contained members of *Prevotella* and *Xylanibacter*, the latter being a genus that is very rarely, if ever, detected in Western samples[14,18,68]. By contrast, European children harbored *Bacteroides* and *Alistipes* as the dominant Bacteroidetes genera. In light of these clear genus differences, an interesting question for future work will be to measure the glycan degrading abilities of these different Bacteroidetes to determine if they have evolved to specialize on the different glycans contained in each diet.

Beyond the influence of certain types of diets in shaping the composition of the microbiota, supplementing the diet with particular glycans can impact species abundance. Not all species that possess the potential to degrade a given glycan will do so successfully *in vivo*. For example, inulin and smaller fructo-oligosaccharides (FOS) selectively increase the abundance of Bifidobacteria[69], although many *Bacteroides* species are also able to use these glycans[70]. More recently attention has focused on the ability of starch that is not readily digested by mammalian enzymes, known as resistant starch (RS), to direct changes in the composition of the microbiota. Based on human and animal feeding studies, some microbial species may be more adept than others at degrading various forms of RS and are responsive to diets augmented with this nutrient[71,72].

***Responses to rapid diet changes***

In contrast to long-term changes between infancy and adulthood, our diets can also elicit rapid changes in microbiota composition as dietary glycans and other nutrients fluctuate from meal-to-meal[70,73-75]. In germfree mice colonized with a transplanted human microbiota, a rapid shift from a high fat diet to a high carbohydrate diet resulted in community changes that were observable after just one day, but took several days to stabilize[73]. In addition, 10 human subjects who were fed either high-fat/low-fiber or low-fat/high-fiber diets in a controlled setting exhibited detectable changes in the microbiota within 24 hours of a dietary shift[25]. Observations like these underscore the relationship between the microbiota and diet, suggesting that some proportion of our gut microorganisms is constantly fluctuating in abundance as a result of meal-to-meal variations. In contrast to protein and fat, which are more readily targeted by human absorptive systems, the low digestibility of non-starch dietary glycans suggests that changes in their

abundance may exert a major impact on the microbiota. With this in mind, the effects on the microbiota of the high-fat diets mentioned above could serve to enrich for species that are capable of digesting host mucosal glycans by reducing dietary fiber. Indeed, in germfree mice colonized with a simplified microbiota, consisting of just *Eubacterium rectale* and *Bacteroides thetaiotaomicron*, the latter microorganism increased expression of host glycan-degrading genes when colonized animals were switched to a high-fat/low-fiber diet[76]. Much work is still needed to determine the precise relationships governing these diet-microbiota interactions, the locations along the length and width of the gut that are influenced by different dietary glycans, and the microbial populations that should be targeted for enrichment or depletion during certain states of dysbiosis.

**Glycan diversity in the human gut**

The biochemistry of the various host and dietary glycans that enter the gut is exceptionally diverse (**Figure 1.1)**. Many different glycosidic linkages may be incorporated into a single polymer, which correspondingly require several linkage-specific degradative enzymes. The human genome is capable of fully degrading a very small subset of glycans, namely starch, lactose and sucrose, each of which contains only one or two different linkages. In contrast, some microorganisms in the intestinal tract target dozens of glycans and possess the corresponding enzymatic tools for depolymerizing each of these molecules into their component sugars. Gut microorganisms vary widely in the number of different glycans that they are capable of degrading. For example, the human gut symbionts *Bacteroides thetaiotaomicron*, *Bacteroides ovatus* and *Bacteroides cellulosilyticus* can all degrade over a dozen different types of glycans[29,48,61,77], while some species are restricted to one or a few[61,78] (Urs, Pudlo and Martens

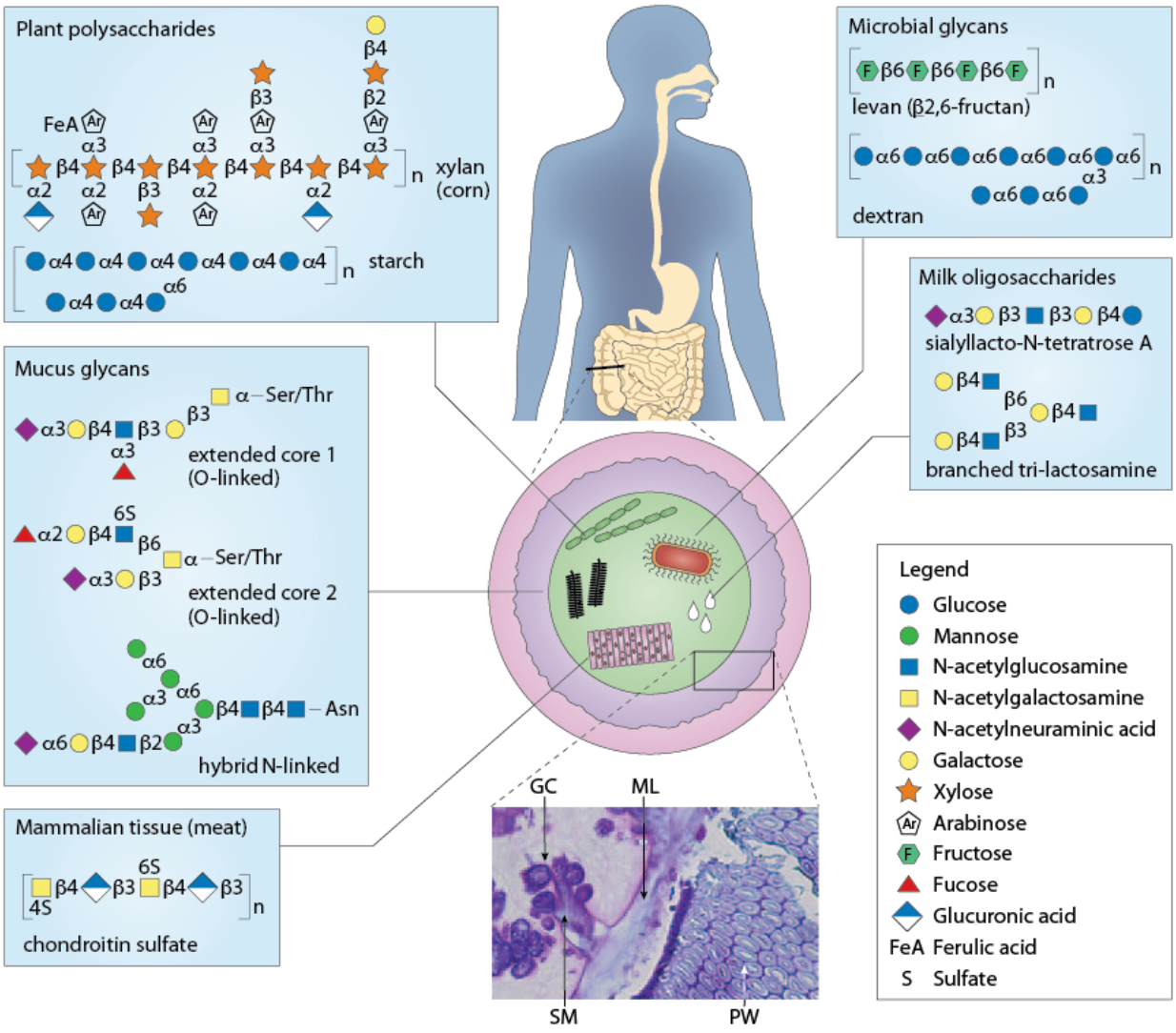



**FIGURE 1.1. Sources and chemical variation of glycans in the gut.**

The center illustration shows a cross-sectional view of the intestine depicting five different sources of glycans: dietary plants, dietary animal tissue, endogenous microorganisms (*e.g.*, capsules), mucus and breast milk. Some representative glycan structures are shown for each source. However, the complexity of all possible glycans in each category is much more expansive than shown. Monosaccharides are schematized according to the legend and interconnecting linkages are also indicated[79]. Brackets at the end of horizontal glycan chains indicate that they may extend further with a similar linkage pattern. The inset at the bottom shows a section of germfree mouse colon stained with periodic acid-Schiff base and Alcian blue stains for various carbohydrates. The section is oriented similarly as the corresponding box in the gut illustration in the center and highlights the locations of host mucus-secreting goblet cells (GC), secreted mucus (SM), the mucus layer (ML) and a fragment of plant cell wall (PW) located immediately adjacent to the mucus layer.

*Submitted*). From an ecological perspective, species with broad glycan-degrading abilities may be thought of as "generalists" that shift their metabolism from meal-to-meal, while species with narrower glycan degrading potential may be considered "specialists" that focus on one or a few glycans. Specialists run the risk of becoming extinct in a host if their preferred nutrients wane for too long, thus, it is most likely that such microorganisms would only evolve to degrade ubiquitously abundant dietary glycans or host derived mucins. Evidence is emerging that even in glycan generalists there is a hierarchy of preferred substrates. *Bacteroides thetaiotaomicron* for example prioritizes utilization of plant glycans over host glycans when grown in a mixture of carbohydrate sources[80]. Other glycan generalists likely display similar preferences for particular substrates, and determining the molecular basis for this prioritization will be important for understanding the behavior of these organisms.

The task of degrading glycans in the gut is further complicated by the fact that many of these substrates are sequestered within larger structures like the plant cell wall, or regional microhabitats like the mucus layer, which may be difficult for some species to access. Plant cell wall glycans (cellulose, hemicellulose and pectin) are intertwined in a polysaccharide matrix in many foods. In addition, hemicelluloses and pectins vary substantially in their fine-level structure between plant sources[81,82]. Thus, the dietary glycans available in whole-wheat bran differ from those available in a potato skin or in an apple. Intracellular plant glycans such as starch may be contained in either insoluble granules or as chemical forms that are resistant to degradative enzymes. Cooking, milling and other food preparation processes can all influence the abundance of these "resistant starch" forms and the availability of other plant glycans to intestinal microbes. Finally, the chemical diversity of endogenous *O*- and *N*-linked glycans (hundreds of different structures may be attached to a single mucin glycoprotein[83]) requires that

mucosal bacteria produce many different degradative enzymes, a substantial metabolic investment, to effectively utilize these heterogeneous polymers. Indeed, one reason why such glycan diversity exists in secreted mucus could be to deter microbial species from evolving to be too efficient at harvesting these structures, thus protecting the integrity of this important barrier.

Regardless of the particular glycan substrate degraded by gut microorganisms, the colonic epithelium benefits from the end result of this microbial metabolism by absorbing short chain fatty acids, such as butyrate, propionate and acetate. Butyrate that is produced in the colon exerts local effects on the colonic epithelium because it is a preferred energy source of colonocytes and has also been associated with suppressed growth of colonic tumors[84]. Acetate and propionate are absorbed into the bloodstream and travel to the liver where they are incorporated into lipid and glucose metabolism, respectively[85]. In addition to being absorbed by the host, the presence of acetate is also manifest in the colonic environment, where it may augment butyrate production by some species[86,87] and prevent colonization of some enteric pathogens[88].

**Starch is an important dietary glycan**

Starch is the primary energy storage molecule synthesized by the majority of plants[89] and one of the most abundant carbohydrates in the human diet. Plant starch is composed of two glucose polymers, amylose and amylopectin, that can be covalently linked in the same macromolecule[89]. Although starch is composed of a single monosaccharide, it displays a high level of structural diversity (**Figure 1.2**). Starch is organized into higher order structures called granules in plants that contain roughly 25% amylose and 75% amylopectin. Amylose is composed almost exclusively of α-1,4 linked glucose chains that take on a helical shape and is

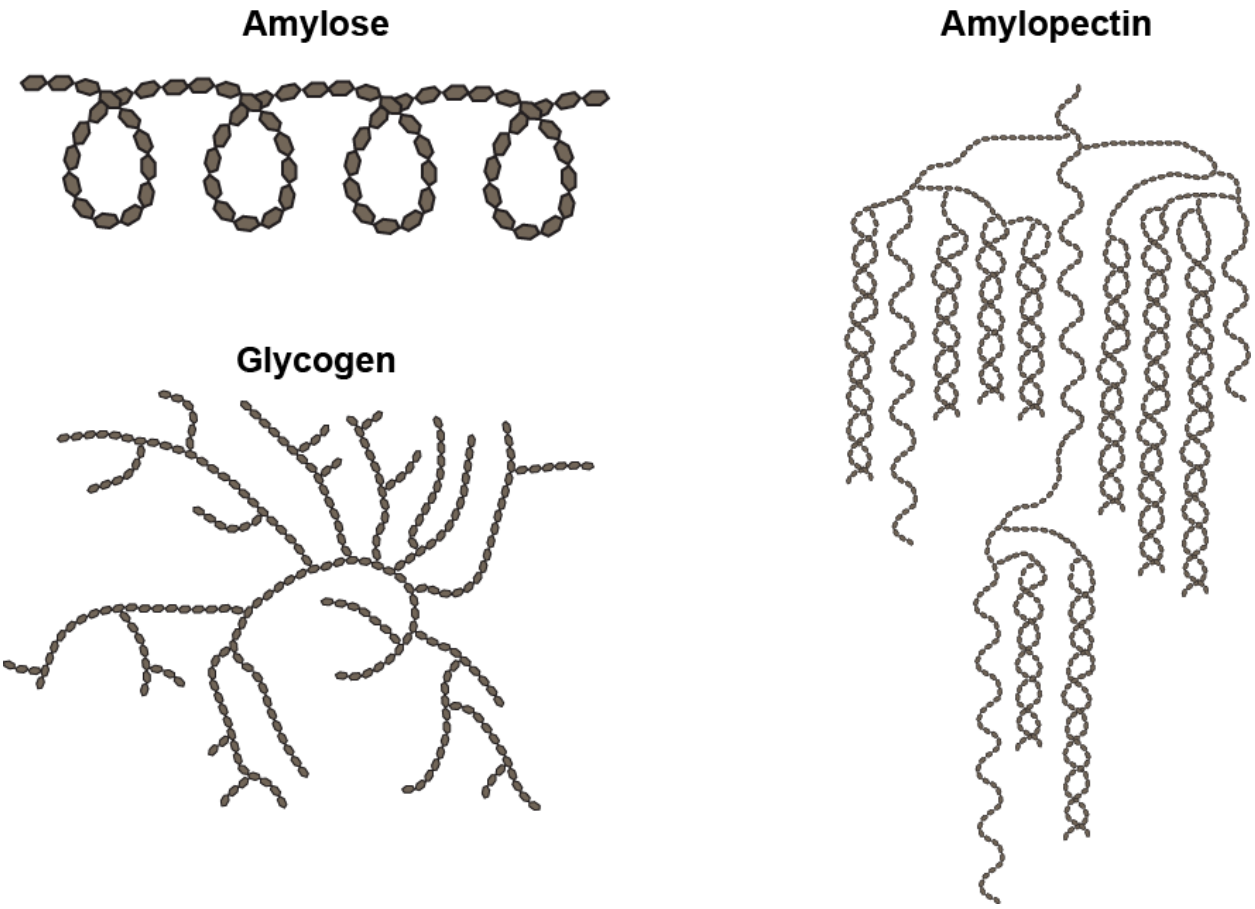



**FIGURE 1.2. Starch is a structurally diverse carbohydrate.**

Structures of three starch types are depicted (note glucose residues are not depicted at the same scale across the three different structures). Amylose is a plant starch composed almost exclusively of α-1,4 linked glucose which adopt a helical shape. Amylopectin is another starch produced in plants. Amylopectin contains α-1,6 branch points that connect α-1,4 linked helices. Amylopectin contains clusters of single and double helices connected by longer α-1,4 chains. Amylopectin is typically much larger than amylose, with a molecular weight on the order of $10^8$ Daltons where amylose is typically on the order of $10^5$ Daltons[90]. Glycogen, a starch analogue produced by animals, fungi and bacteria, is even more heavily branched than amylopectin[91]. In glycogen the α-1,6 branches are more randomly distributed than in amylopectin resulting in a web-like structure. The microbiota has the capacity to degrade all three of these starch macro-structures and in fact glycoside hydrolase family 13 (GH13), a family of α-glucosydic enzymes, is one of the most abundant CAZyme families in the human microbiome[28].

largely insoluble in water. In contrast, amylopectin contains α-1,6 branch points which disrupt the helices and increase the solubility of this molecule. Amylopectin branch points tend to occur in close proximity to each other, resulting in clusters of α-1,6 bonds connected by helices[89,92]. Glycogen is a branched starch-like molecule made by animals, fungi and bacteria[89,93]. In contrast to amylopectin, the α-1,6 linkages in glycogen are more randomly distributed, resulting in a web-like structure[89]. Glycogen is synthesized for energy storage in animal muscle cells and may represent a significant component of the omnivorous human diet. Additionally, bacterial glycogen released by lysed cells may represent another 'endogenous' gut carbohydrate; however, the role of glycogen as a carbon source for the microbiota has been largely unexplored.

In addition to the structural diversity observed between classes of starch molecules, structural aspects like degree of branching, chain lengths connecting branch points, and degree of polymerization can differ significantly between plant/animal sources[89].

Starch is the only complex carbohydrate that human enzymes are capable of targeting, and thus the majority of its accessible, soluble form is degraded and absorbed in the upper digestive tract. However, starch that is not readily digested by human enzymes, known as resistant starch (RS), can reach the colon[94,95] and be fermented by the colonic microbiota. Therefore, recent studies have investigated the effect of RS on the composition of the microbiota and the metabolites produced to evaluate the utility of RS as a prebiotic[96].

RS is categorized into four types (RS1 through RS4) based on structure and degree of resistance to enzymatic degradation. Consumption of some RS forms by humans preferentially results in increases in the short chain fatty acid butyrate. Butyrate has been reported to exert anti-inflammatory[84,97-100] and anti-tumorigenic effects[84,101-104], and has been suggested as a possible therapeutic for inflammatory bowel disease[105]. Butyrate is produced by members of the

Firmicutes but is rarely associated with the SCFA profiles of Bacteroidetes[106,107]. Human volunteers consuming RS2 (starch in its natural granular form) experienced increases in the Firmicutes *Ruminococcus* spp. and *Eubacterium rectale*[71]; likewise, overweight individuals consuming a diet high in RS3 (retrograded starch) exhibited increases in *Eubacterium rectale*, *Roseburia* spp. and *Ruminococcus bromii*[108]. These findings are consistent with *in vitro* observations that these species bind directly to insoluble starch particles and may be primary components of bacterial food-chains that target starch[109,110]. RS4 (chemically altered RS) was found to increase levels of Actinobacteria and Bacteroidetes while decreasing Firmicutes[71]. This demonstrates how structural heterogeneity within one polysaccharide can have significant effects on the composition and metabolism of the microbiota; even within a seemingly simple molecule like starch, composed of a single monosaccharide.

**The *Bacteroides thetaiotaomicron* Starch Utilization System**

Among the many bacteria in the human microbiota the Gram-negative Bacteroidetes are distinguished by an expanded capacity for carbohydrate utilization[61,77]. For example, the genome of *Bacteroides thetaiotaomicron* (*B. theta*) contains 261 predicted glycoside hydrolases and polysaccharide lyases[111,112], an enormous number for an organism. Its ability to degrade a large number of glycans and its genetic manipulability has made *B. theta* a useful model for studying bacterial carbohydrate utilization.

As starch is one of the most abundant carbohydrates in the human diet, one would expect that this would be a common nutrient source targeted by human gut bacteria. Indeed, a study measuring the carbohydrate utilization profiles of 354 Bacteroidetes strains revealed that of 30 plant and host derived complex glycans starch was catabolized by the most strains (Urs, Pudlo

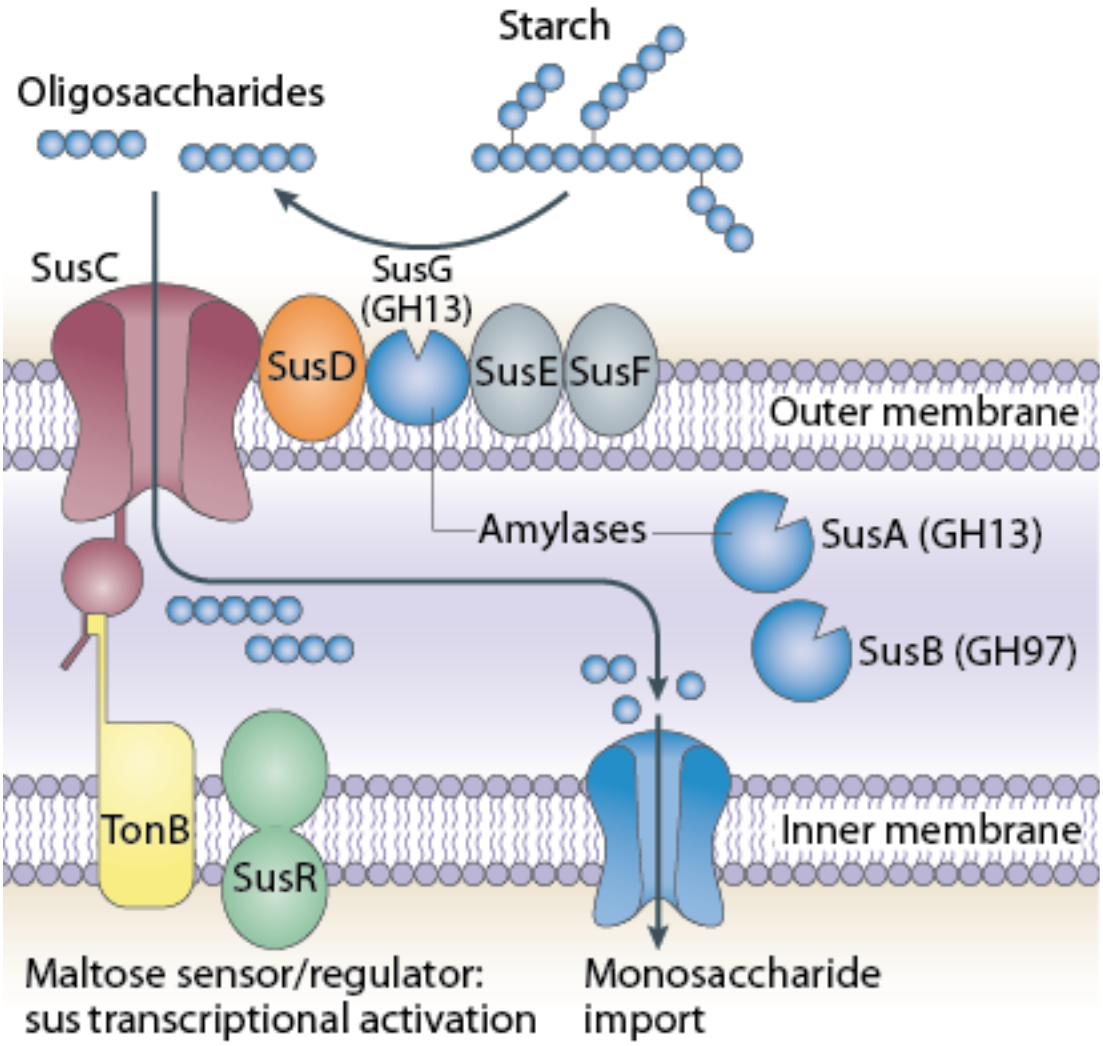



**FIGURE 1.3. A model of the *Bacteroides thetaiotaomicron* Starch Utilization System.**

The TonB-dependent transporter SusC works in concert with the starch-binding lipoproteins SusD, SusE, SusF and SusG, which is a glycoside hydrolase family 13 (GH13) α-amylase. Starch binding is initiated by SusD, E and F, followed by initial degradation by SusG. Oligosaccharides are transported into the periplasm via SusC. In the periplasm, maltooligosaccharides are further degraded by another GH13 enzyme (SusA, neopullulanse) and a GH97 enzyme (SusB, α-glucosidase). Homologs of the proteins SusC (TonB-dependent transporter) and SusD (carbohydrate-binding protein) are a hallmark of every Sus-like system, but carbohydrate-binding proteins akin to SusE and SusF as well as glycoside hydrolases, vary substantially between Sus-like loci.

and Martens, *Submitted*). In the 1990's and early 2000's Abigail Salyers and her lab described an eight-gene locus in the *B. theta* genome that is essential for starch utilization; this locus was deemed the starch utilization system (Sus) (**Figure 1.3**). The Sus contains the sensor/regulator protein SusR, the three starch degrading enzymes SusA, B and G, the three non-enzymatic starch-binding proteins SusD, E and F and the outer membrane spanning transporter, SusC. These proteins are the sole mechanism that *B. theta* employs to sense, bind, degrade and import starch into the cell[113-117].

SusD is a predicted lipoprotein that contains a signal peptidase II recognition motif at its N- terminus. An N-acyl-S-diacylglycerol moiety is covalently linked to an N-terminal cysteine before the signal peptide is cleaved and then the cysteine residue is acetylated[118]. The protein is anchored in the outer membrane, exposed to the cell's external environment where it can interact with starch molecules. SusD is essential for growth of *B. theta* on starch substrates of more than six glucose residues, and its loss (by in frame deletion) results in a growth defect on maltopentaose and maltotetraose[119]. The crystal structure of SusD showed that the protein has a single starch-binding site which induces a curvature in bound linear maltoligosaccharides[119]. Additionally, isothermal titration calorimetry revealed that SusD bound cyclic substrates α and β-cyclodextrin with ~20-fold higher affinity than their linear counterparts[119]. This suggests that SusD does not recognize the individual glucose residues, but binds by recognizing the structure of the starch helix itself, which may increase the promiscuity of the protein and allow it to bind a more diverse array of starch molecules.

SusE and SusF, which share ~38% amino acid homology at their C-termini[120], are the two remaining non-enzymatic starch binding proteins in the Sus complex. They are lipidated and tethered to the outer membrane at their N-terminal cysteine residues via the same mechanism

as SusD. Unlike SusD, SusE and SusF are not essential for growth on starch; however they do contribute to the overall ability of *B. theta* cells to bind this substrate[117]. A metagenomic analysis comparing human gut bacteria to related environmental species revealed that SusEF-like proteins are members of a large protein family that is significantly enriched in the human microbiome[121], suggesting they play an important role specifically in the gut.

Initial enzymatic degradation of starch by the Sus is performed by SusG, an endo-acting α-amylase belonging to the glycoside hydrolase family 13 (GH13), one of the largest carbohydrate degrading enzyme families[112]. The crystal structure of SusG revealed that, in addition to a single active site, the protein has an independently folding carbohydrate-binding module (CBM), known as CBM58, and a surface-binding site. *In vitro* assays demonstrated that CBM58 actually decreases enzymatic activity on soluble starch, but significantly enhances SusG's ability to degrade insoluble starch, suggesting it functions to increase affinity for less accessible substrates and allow more efficient degradation[122]. The surface-binding site is oriented such that the pitch of the starch helix is directed into the protein, which likely limits the size of substrates the surface site can accommodate[122]. Perhaps this surface-binding site binds intermediate products that are created during degradation, which can then be imported via SusC.

SusC is a TonB-dependent β-barrel transporter that spans the *B. theta* outer-membrane and serves to import starch breakdown products into the periplasm of the bacterium. Unlike other TonB-dependent transporters, SusC cannot bind and transport its substrate alone, but requires SusD[117,119]. When the transported maltooligosaccharides enter the periplasm they are further degraded by two additional enzymes, the neopullulanase SusA and α-glucosidase SusB[114,123], before being imported into the bacterial cytoplasm. Biochemical studies of SusB revealed that the enzyme prefers shorter substrates such as maltotriose, providing evidence for a

model in which SusA first degrades the chains imported by SusC, which are subsequently degraded by SusB[123].

SusR, the inner-membrane spanning sensor/regulator protein, controls transcription of *susA-G* in response to maltooligosaccharides present in the periplasm[116]. Deletion of *susR* abolishes all *sus* transcription, while increasing *susR* copies increases *sus* transcription levels, confirming that this regulator acts as an activator of the *sus* locus[124]. The C-terminal, cytoplasmic domain of SusR contains a helix-turn-helix motif, which is presumed to bind DNA and initiate transcription of the *sus* locus, although this has never been directly shown. The N-terminal domain extends into the periplasm and is presumed to sense starch breakdown products after import by SusC. An interesting feature of this system is that *sus* transcription is activated only in response to maltose (two α-1,4 linked glucose units) or longer maltooligosaccharides, not glucose. The lack of glucose recognition demonstrates that SusR senses not only the sugar monomers of starch but the connecting α-1,4 linkage as well[116], allowing for a more fine tuned, specific response to the constantly changing carbohydrate environment in the gut.

**Expansion of Sus-like systems in the Bacteroidetes**

The *B. theta* Sus was the first glycan acquisition system described in detail in the Bacteroidetes; however, when the complete genome of *B. theta* was sequenced (the first among the Bacteroidetes[111]) it was revealed that this organism encodes a high number of similarly patterned systems. These partially homologous loci were termed Polysaccharide Utilization Loci (PULs), and the protein systems they encode "Sus-like systems"[59]. PULs are defined by homologs of the *susC* and *susD* genes but differ in the number and type of enzymes they have as well as the number of SusEF-like carbohydrate-binding proteins they encode. Most PULs do not

have regulators related to SusR but are controlled by hybrid two component regulators or ECF-σ/anti-σ factors, or other regulators which work through various mechanisms[60]. In the *B. theta* genome alone there are 88 PULs; that comprise 866 genes, 18% of the *B. theta* genome[60].

Transcriptional profiling of *B. theta* grown on purified carbohydrate sources, or isolated from the ceca of mice, demonstrates that each PUL targets a unique carbohydrate substrate. *B. theta* encodes systems to target a number of plant polysaccharides as well as host-derived mucosal glycans[58-60,77].

Analysis of all available Bacteroidetes genomes revealed that the Sus-like systems are prevalent throughout the phylum, with similarly patterned PULs appearing in the majority of Bacteroidetes lineages[60]. There is substantial overlap in the substrates that the gut Bacteroidetes can utilize; for example the ability to utilize starch is widespread among gut *Bacteroides*[59,77] (Urs, Pudlo and Martens *Submitted*). However it appears each species retains a certain level of specificity in the substrates it is able to, and prefers, to utilize. *B. ovatus*, which on the 16S rRNA level is 96.5% identical to *B. theta*, is an efficient degrader of hemicelluloses, whereas *B. theta* grows very poorly, if at all, on these substrates[77]. *B. theta*, on the other hand, has evolved several systems to degrade host *N*- and *O*-linked glycans, chondroitin sulfate and heparin. In comparison *B. ovatus* is a poor utilizer of host glycans, especially *O*-linked structures attached to secreted mucus[77]. This highlights how organisms that are phylogenetically closely related can display significant phenotypic differences and therefore behave very differently in the human gut. Acquisition of unique glycan degrading systems may allow the bacteria to inhabit an unused nutrient niche, helping them thrive in the highly competitive environment of the human gut. An intriguing example of this is found in the microbiota of Japanese individuals, where *Bacteroides plebius* has acquired a PUL targeting porphyran, a carbohydrate abundant in nori, edible seaweed

that is common in the Japanese diet[125,126]. Porphyran degradation genes are found far less frequently in the gut microbiome of westerners, underscoring how diet can shape evolution of gut species. Analysis of the genomic region carrying the genes required for porphyran degradation revealed a conjugative transposon (cTn) was responsible for the transfer of this system to *B. plebius*, most likely from marine bacteria[125]. A similar cTn is likely responsible for acquisition of a PUL targeting α–mannan in *B. theta*[60], a carbohydrate found in yeast cell walls. However, not all PULs seem to be acquired via mobile elements, and many appear to be the result of duplications or recombination events[127]. Additional studies will be critical to understand the evolutionary forces that shape the saccharolytic capacity of this important group of gut bacteria.

The diversification of glycan utilization ability in the gut Bacteroidetes underscores that, in addition to phylogenetic studies of microbiota composition, phenotypic studies are needed to accurately assess how particular species are affected by various factors in the gut and how in turn they can influence host physiology.

**Summary and Chapter Outline**

The carbohydrates available in the intestinal tract and the ability of certain species to degrade them is a major factor that shapes the composition of the gut microbiota, which is known to have significant effects on human health and disease. Through the expansion of Sus-like systems the Bacteroidetes have evolved the ability to utilize a wide variety of glycans, making them key players in this process. Although a significant amount of work has been devoted to identifying Sus-like systems and their cognate substrates, there is still a crucial gap in our knowledge of how these systems function on the molecular level. Using the *B. theta* Sus as a

model we can decipher molecular mechanisms of glycan acquisition that can be extended to the huge number of Sus-like systems encoded in the human microbiome. The focus of this dissertation is to elucidate the molecular mechanisms by which the eight Sus proteins work together to bind, degrade and import starch.

In Chapter II I describe the x-ray crystallographic structures and biochemical properties for two of the *B. theta* Sus outer-membrane lipoproteins, SusE and SusF, demonstrating that they are multi-domain starch binding proteins that are structural homologs of one another. SusE and SusF both have multiple starch binding sites that display differences in binding affinity and substrate preference, suggesting they may aid in accommodating the diverse starch structures *B. theta* likely encounters in the intestinal tract.

In Chapter III I explore the roles of the eight non-catalytic binding sites present across the four Sus outer-membrane lipoproteins (SusD, E, F, G) that, despite having similar biochemical functions, play unique roles within the context of the multi-protein complex. The SusD binding site is important for sensing available starch, leading to increased expression of the *sus* locus. In contrast, the SusE, SusF and SusG binding sites are not as critical for the transcriptional response but offset a loss in starch affinity created by the *B. theta* polysaccharide capsule.

Finally, described in Appendix I are the first ever studies of Sus protein imaging in live *B. theta* cells to investigate the nature of Sus protein interactions and complex formation. Tracking the movement of fluorophore-labeled SusG revealed that this protein is highly motile on the cell surface and that its movement is influenced by the presence of starch or other Sus proteins. These studies suggest that there is indeed a starch-induced Sus complex but that it is highly dynamic.

Understanding the molecular mechanisms of starch acquisition by the Sus will inform our model of glycan acquisition by the Bacteroidetes, which comprise a large portion of the human gut microbiota. Further understanding of this process will allow us to more fully appreciate how bacterial carbohydrate metabolism shapes the gut environment and microbiota composition. This knowledge may lead to novel strategies for manipulating the gut microbiota via non-invasive routes like diet to maximize the benefits they provide to the host and maintain a healthy community structure.

**Notes**

Portions of this chapter were reprinted and modified with permission from Koropatkin, N.M., Cameron, E.A., Martens, E.C. How Glycan Metabolism Shapes the Human Gut Microbiota. *Nature Reviews Microbiology*. **10**, 323-35 (2012).

**References**


1 Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313-323 (2009).

2 Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* **341**, 569-573 (2013).

3 Hooper, L. V. & Macpherson, A. J. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat Rev Immunol* **10**, 159-169 (2010).

4 Ivanov, II, Zhou, L. & Littman, D. R. Transcriptional regulation of Th17 cell differentiation. *Semin Immunol* **19**, 409-417 (2007).

5 Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* **6**, 121-131 (2008).

6 Wardwell, L. H., Huttenhower, C. & Garrett, W. S. Current concepts of the intestinal microbiota and the pathogenesis of infection. *Curr Infect Dis Rep* **13**, 28-34 (2011).

7 Packey, C. D. & Sartor, R. B. Commensal bacteria, traditional and opportunistic pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases. *Curr Opin Infect Dis* **22**, 292-301 (2009).

8 Rooks, M. G. *et al.* Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. *Isme J* (2014).

9 Garrett, W. S. *et al.* Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* **8**, 292-300 (2010).

10 O'Keefe, S. J. *et al.* Products of the colonic microbiota mediate the effects of diet on colon cancer risk. *J Nutr* **139**, 2044-2048 (2009).

11 Kostic, A. D. *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207-215 (2013).

12 Warren, R. L. *et al.* Co-occurrence of anaerobic bacteria in colorectal carcinomas. *Microbiome* **1**, 16 (2013).

13 Khoruts, A., Dicksved, J., Jansson, J. K. & Sadowsky, M. J. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *J Clin Gastroenterol* **44**, 354-360 (2010).

14 Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022-1023, (2006).

15 Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214 (2013).

16 Sun, L., Nava, G. M. & Stappenbeck, T. S. Host genetic susceptibility, dysbiosis, and viral triggers in inflammatory bowel disease. *Curr Opin Gastroenterol* **27**, 321-327 (2011).

17 Bloom, S. M. *et al.* Commensal bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* **9**, 390-403 (2011).

18 Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638 (2005).

19 Turnbaugh, P. J. *et al.* Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc Natl Acad Sci U S A* **107**, 7503-7508 (2010).

20 Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65 (2010).

21 Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* (2011).

22 Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biol* **12**, R50 (2011).

23 Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214 (2012)

24 Rajilic-Stojanovic, M., Heilig, H. G., Tims, S., Zoetendal, E. G. & de Vos, W. M. Long-term monitoring of the human intestinal microbiota composition. *Environ Microbiol* (2012).

25 Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105-108 (2011).

26 Koren, O. *et al.* A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. *PLoS Comput Biol* **9**, e1002863 (2013).

27 Cantarel, B. L., Lombard, V. & Henrissat, B. Complex carbohydrate utilization by the healthy human microbiome. *PLoS One* **7**, e28742 (2012).

28 El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* **11**, 497-504 (2013).

29 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).

30 McNeil, N. I. The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr* **39**, 338-342 (1984).

31 Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920 (2005).

32 Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* **6**, 776-788 (2008).

33 Palmer, C., Bik, E. M., Digiulio, D. B., Relman, D. A. & Brown, P. O. Development of the Human Infant Intestinal Microbiota. *PLoS Biol* **5**, e177 (2007).

34 Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4578-4585 (2010).

35 Kunz, C., Rudloff, S., Baier, W., Klein, N. & Strobel, S. Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annu Rev Nutr* **20**, 699-722 (2000).

36 Ninonuevo, M. R. *et al.* A strategy for annotating the human milk glycome. *J Agric Food Chem* **54**, 7471-7480 (2006).

37 Zivkovic, A. M., German, J. B., Lebrilla, C. B. & Mills, D. A. Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4653-4658 (2011).

38 Fuhrer, A. *et al.* Milk sialyllactose influences colitis in mice through selective intestinal bacterial colonization. *J Exp Med* **207**, 2843-2854 (2010).

39 Chaturvedi, P., Warren, C. D., Buescher, C. R., Pickering, L. K. & Newburg, D. S. Survival of human milk oligosaccharides in the intestine of infants. *Adv Exp Med Biol* **501**, 315-323 (2001).

40 German, J. B., Freeman, S. L., Lebrilla, C. B. & Mills, D. A. Human milk oligosaccharides: evolution, structures and bioselectivity as substrates for intestinal bacteria. *Nestle Nutr Workshop Ser Pediatr Program* **62**, 205-218; discussion 218-222 (2008).

41 Gnoth, M. J., Kunz, C., Kinne-Saffran, E. & Rudloff, S. Human milk oligosaccharides are minimally digested in vitro. *J Nutr* **130**, 3014-3020 (2000).

42 Favier, C. F., Vaughan, E. E., De Vos, W. M. & Akkermans, A. D. Molecular monitoring of succession of bacterial communities in human neonates. *Appl Environ Microbiol* **68**, 219-226 (2002).

43 Sela, D. A. Bifidobacterial utilization of human milk oligosaccharides. *Int J Food Microbiol* (2011).

44 Miwa, M. *et al.* Cooperation of beta-galactosidase and beta-N-acetylhexosaminidase from bifidobacteria in assimilation of human milk oligosaccharides with type 2 structure. *Glycobiology* **20**, 1402-1409 (2010).

45 LoCascio, R. G., Desai, P., Sela, D. A., Weimer, B. & Mills, D. A. Broad conservation of milk utilization genes in Bifidobacterium longum subsp. infantis as revealed by comparative genomic hybridization. *Appl Environ Microbiol* **76**, 7373-7381 (2010).

46 Turroni, F. *et al.* Genome analysis of Bifidobacterium bifidum PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc Natl Acad Sci U S A* **107**, 19514-19519 (2010).

47 Garrido, D., Kim, J. H., German, J. B., Raybould, H. E. & Mills, D. A. Oligosaccharide binding proteins from Bifidobacterium longum subsp. infantis reveal a preference for host glycans. *PLoS One* **6**, e17315 (2011).

48 Marcobal, A. *et al.* Bacteroides in the Infant Gut Consume Milk Oligosaccharides via Mucus-Utilization Pathways. *Cell Host Microbe* (2011).

49 Martinez, R. C. *et al.* In vitro evaluation of gastrointestinal survival of Lactobacillus amylovorus DSM 16698 alone and combined with galactooligosaccharides, milk and/or Bifidobacterium animalis subsp. lactis Bb-12. *Int J Food Microbiol* **149**, 152-158 (2011).

50 Rodriguez-Diaz, J., Monedero, V. & Yebra, M. J. Utilization of natural fucosylated oligosaccharides by three novel alpha-L-fucosidases from a probiotic Lactobacillus casei strain. *Appl Environ Microbiol* **77**, 703-705 (2011).

51 Schwab, C. & Ganzle, M. Lactic acid bacteria fermentation of human milk oligosaccharide components, human milk oligosaccharides and galactooligosaccharides. *FEMS Microbiol Lett* **315**, 141-148 (2011).

52 Coppa, G. V. *et al.* Oligosaccharides in human milk during different phases of lactation. *Acta Paediatr Suppl* **88**, 89-94 (1999).

53 Favier, C. F., de Vos, W. M. & Akkermans, A. D. Development of bacterial and bifidobacterial communities in feces of newborn babies. *Anaerobe* **9**, 219-229 (2003).

54 Fallani, M. *et al.* Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology* **157**, 1385-1392.

55 Harmsen, H. J. *et al.* Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. *J Pediatr Gastroenterol Nutr* **30**, 61-67 (2000).

56 Kurokawa, K. *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**, 169-181 (2007).

57 Mariat, D. *et al.* The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol* **9**, 123 (2009).

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

59 Bjursell, M. K., Martens, E. C. & Gordon, J. I. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. *J Biol Chem* **281**, 36269-36279 (2006).

60 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**, 447-457 (2008).

61 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 Environ Microbiol* **33**, 319-322 (1977).

62 Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694-1697 (2009).

63 Franks, A. H. *et al.* Variations of bacterial populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes. *Appl Environ Microbiol* **64**, 3336-3345 (1998).

64 Zoetendal, E. G., Akkermans, A. D. & De Vos, W. M. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl Environ Microbiol* **64**, 3854-3859 (1998).

65 Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* **102**, 11070-11075 (2005).

66 Hill, J. E. *et al.* Improvement of the representation of bifidobacteria in fecal microbiota metagenomic libraries by application of the cpn60 universal primer cocktail. *Applied and environmental microbiology* **76**, 4550-4552 (2010).

67 De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A* **107**, 14691-14696 (2010).

68 Tap, J. *et al.* Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* **11**, 2574-2584 (2009).

69 Meyer, D. & Stasse-Wolthuis, M. The bifidogenic effect of inulin and oligofructose and its consequences for gut health. *Eur J Clin Nutr* **63**, 1277-1289 (2009).

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

71 Martinez, I., Kim, J., Duffy, P. R., Schlegel, V. L. & Walter, J. Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects. *PLoS One* **5**, e15046 (2010).

72 Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *Isme J* **6**, 1535-1543 (2012).

73 Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* **1**, 6ra14 (2009).

74 Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a Human Gut Microbiota's Response to Diet in Gnotobiotic Mice. *Science* (2011).

75 Goodman, A. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci U S A* (2011).

76 Mahowald, M. A. *et al.* Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc Natl Acad Sci U S A* **106**, 5859-5864 (2009).

77 Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).

78 Wegmann, U. *et al.* Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota ('Ruminococcus bicirculans') reveals two chromosomes and a selective capacity to utilize plant glucans. *Environ Microbiol* (2013).

79 Varki, A. *et al. Essentials of Glycobiology*. (Cold Spring Harbor Laboratory Press, 1999).

80 Rogers, T. E. *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol Microbiol* (2013).

81 Mohnen, D. Pectin structure and biosynthesis. *Curr Opin Plant Biol* **11**, 266-277 (2008).

82 Saulnier, L., Marot, C., Chanliaud, E. & Thibault, J.-F. Cell wall polysaccharide interactions in maize bran. *Carbohyd Polym* **26**, 279-287 (1995).

83 Larsson, J. M., Karlsson, H., Sjovall, H. & Hansson, G. C. A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19**, 756-766 (2009).

84 Hamer, H. M. *et al.* Review article: the role of butyrate on colonic function. *Aliment Pharmacol Ther* **27**, 104-119 (2008).

85 Rombeau, J. L. & Kripke, S. A. Metabolic and intestinal effects of short-chain fatty acids. *JPEN. J Parenter Enteral Nutr* **14**, 181S-185S (1990).

86 Duncan, S. H., Barcenilla, A., Stewart, C. S., Pryde, S. E. & Flint, H. J. Acetate utilization and butyryl coenzyme A (CoA):acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Appl Environ Microbiol* **68**, 5186-5190 (2002).

87 Duncan, S. H. *et al.* Contribution of acetate to butyrate formation by human faecal bacteria. *Br J Nutr* **91**, 915-923 (2004).

88 Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543-547 (2011).

89 Zeeman, S. C., Kossmann, J. & Smith, A. M. Starch: its metabolism, evolution, and biotechnological modification in plants. *Annu Rev Plant Biol* **61**, 209-234 (2010).

90 Yokoyama, W., Renner-Nantz, J. J. & Shoemaker, C. F. Starch molecular mass and size by size-exclusion chromatography in DMSO-LiBr coupled with multiple angle laser light scattering. *Cereal Chem* **75**, 530-535 (1998).

91 Ball, S. G. & Morell, M. K. From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol* **54**, 207-233 (2003).

92 Buleon, A., Colonna, P., Planchot, V. & Ball, S. Starch granules: structure and biosynthesis. *Int J Biol Macromol* **23**, 85-112 (1998).

93 Wilson, W. A. *et al.* Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiol Rev* **34**, 952-985 (2010).

94 Champ, M. M. *et al.* Small-intestinal digestion of partially resistant cornstarch in healthy subjects. *Am J Clin Nutr* **68**, 705-710 (1998).

95 Vonk, R. J. *et al.* Digestion of so-called resistant starch sources in the human small intestine. *Am J Clin Nutr* **72**, 432-438 (2000).

96 Bird, A. R., Conlon, M. A., Christophersen, C. T. & Topping, D. L. Resistant starch, large bowel fermentation and a broader perspective of prebiotics and probiotics. *Benef Microbes* **1**, 423-431 (2010).

97 Segain, J. P. *et al.* Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease. *Gut* **47**, 397-403 (2000).

98 Luhrs, H. *et al.* Cytokine-activated degradation of inhibitory kappaB protein alpha is inhibited by the short-chain fatty acid butyrate. *Int J Colorectal Dis* **16**, 195-201 (2001).

99 Hamer, H. M. *et al.* Butyrate modulates oxidative stress in the colonic mucosa of healthy humans. *Clin Nutr* **28**, 88-93 (2009).

100 Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S. & Flint, H. J. The microbiology of butyrate formation in the human colon. *FEMS Microbiol Lett* **217**, 133-139 (2002).

101 Avivi-Green, C., Polak-Charcon, S., Madar, Z. & Schwartz, B. Apoptosis cascade proteins are regulated in vivo by high intracolonic butyrate concentration: correlation with colon cancer inhibition. *Oncol Res* **12**, 83-95 (2000).

102 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).

103 Dronamraju, S. S., Coxhead, J. M., Kelly, S. B., Burn, J. & Mathers, J. C. Cell kinetics and gene expression changes in colorectal cancer patients given resistant starch: a randomised controlled trial. *Gut* **58**, 413-420 (2009).

104 Clarke, J. M., Topping, D. L., Bird, A. R., Young, G. P. & Cobiac, L. Effects of high-amylose maize starch and butyrylated high-amylose maize starch on azoxymethane-induced intestinal cancer in rats. *Carcinogenesis* **29**, 2190-2194 (2008).

105 Van Immerseel, F. *et al.* Butyric acid-producing anaerobic bacteria as a novel probiotic treatment approach for inflammatory bowel disease. *J Med Microbiol* **59**, 141-143 (2010).

106 Pan, N. & Imlay, J. A. How does oxygen inhibit central metabolism in the obligate anaerobe Bacteroides thetaiotaomicron. *Mol Microbiol* **39**, 1562-1571 (2001).

107 Sakamoto, M. *et al.* Butyricimonas synergistica gen. nov., sp. nov. and Butyricimonas virosa sp. nov., butyric acid-producing bacteria in the family 'Porphyromonadaceae' isolated from rat faeces. *Int J Syst Evol Microbiol* **59**, 1748-1753 (2009).

108 Walker, A. W. *et al.* Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *Isme J* **5**, 220-230 (2011).

109 McWilliam Leitch, E. C., Walker, A. W., Duncan, S. H., Holtrop, G. & Flint, H. J. Selective colonization of insoluble substrates by human faecal bacteria. *Environ Microbiol* **9**, 667-679 (2007).

110 Macfarlane, S. & Macfarlane, G. T. Composition and metabolic activities of bacterial biofilms colonizing food residues in the human gut. *Appl Environ Microbiol* **72**, 6204-6211 (2006).

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

112 Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233-238 (2009).

113 Reeves, A. R., D'Elia, J. N., Frias, J. & Salyers, A. A. A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J Bacteriol* **178**, 823-830 (1996).

114 D'Elia, J. N. & Salyers, A. A. Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of Bacteroides thetaiotaomicron on starch. *J Bacteriol* **178**, 7173-7179 (1996).

115 Shipman, J. A., Berleman, J. E. & Salyers, A. A. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *J Bacteriol* **182**, 5365-5372 (2000).

116 Cho, K. H., Cho, D., Wang, G. R. & Salyers, A. A. New regulatory gene that contributes to control of Bacteroides thetaiotaomicron starch utilization genes. *J Bacteriol* **183**, 7198-7205 (2001).

117 Cho, K. H. & Salyers, A. A. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **183**, 7224-7230 (2001).

118 Bos, M. P., Robert, V. & Tommassen, J. Biogenesis of the gram-negative bacterial outer membrane. *Annu Rev Microbiol* **61**, 191-214 (2007).

119 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**, 1105-1115 (2008).

120 Cameron, E. A. *et al.* Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism. *J Biol Chem* **287**, 34614-34625 (2012).

121 Ellrott, K., Jaroszewski, L., Li, W., Wooley, J. C. & Godzik, A. Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. *PLoS Comput Biol* **6**, e1000798 (2010).

122 Koropatkin, N. M. & Smith, T. J. SusG: A Unique Cell-Membrane-Associated alpha-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. *Structure* **18**, 200-215 (2010).

123 Kitamura, M. *et al.* Structural and functional analysis of a glycoside hydrolase family 97 enzyme from Bacteroides thetaiotaomicron. *J Biol Chem* **283**, 36328-36337 (2008).

124 D'Elia, J. N. & Salyers, A. A. Effect of regulatory protein levels on utilization of starch by Bacteroides thetaiotaomicron. *J Bacteriol* **178**, 7180-7186 (1996).

125 Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc Natl Acad Sci U S A* **109**, 19786-19791 (2012).

126 Hehemann, J. H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908-912 (2010).

127 Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498-502 (2014).

## Chapter II

## Multi-domain carbohydrate-binding proteins involved in *Bacteroides thetaiotaomicron* starch metabolism


### Abstract

Human colonic bacteria are necessary for the digestion of many dietary polysaccharides. The intestinal symbiont *Bacteroides thetaiotaomicron* uses five outer membrane proteins to bind and degrade starch. Here, we report the x-ray crystallographic structures of SusE and SusF, two outer membrane proteins composed of tandem starch specific carbohydrate-binding modules (CBMs) with no enzymatic activity. Examination of the two CBMs in SusE and three CBMs in SusF reveals subtle differences in the way each binds starch and is reflected in their $K_d$s for both high molecular weight starch and small maltooligosaccharides. Thus, each site seems to have a unique starch preference that may enable these proteins to interact with different regions of starch or its breakdown products. Proteins similar to SusE and SusF are encoded in many other polysaccharide utilization loci that are possessed by human gut bacteria in the phylum Bacteroidetes. Thus, these proteins are likely to play an important role in carbohydrate metabolism in these abundant symbiotic species. Understanding structural changes that diversify and adapt related proteins in the human gut microbial community will be critical to understanding the detailed mechanistic roles that they perform in the complex digestive ecosystem.

## Introduction

Digestion of polysaccharides is one of the major mutualistic roles performed by microorganisms in the human gut[1,2]. Absorption of short chain fatty acids produced by bacterial carbohydrate fermentation contributes up to 10% of our daily calories, depending on the amount and nature of polysaccharides in our diet and the particular assemblage of microbes we each harbor[3,4]. Competition for polysaccharides that enter the gut from both dietary and endogenous mucosal sources is a major factor shaping the relative abundance and physiology of microbial species in the intestinal tract. The high density of microorganisms in the lower gut (over $10^{11}$ per gram of contents), and corresponding competition for nutrients, has driven some species to evolve strategies for scavenging the available polysaccharides.

To compete for polysaccharides, members of the Gram-negative Bacteroidetes, one of a few dominant phyla in the guts of humans and other animals[5,6], have evolved and diversified a series of cell envelope-associated protein systems, termed Sus-like systems[7-10]. Each Sus-like system targets a distinct glycan using substrate-specific enzymes located on the cell surface and in the periplasm. These enzymes function in concert with glycan-binding and transport proteins to assimilate the products of glycan degradation. Sus-like systems are named after the starch-utilization system (Sus) in the human gut symbiont, *Bacteroides thetaiotaomicron* (*Bt*), and defined by the presence of genes encoding homologs of the SusC and SusD proteins. SusC is a predicted outer membrane TonB-dependent transporter that moves starch oligosaccharides into the periplasm[11]. SusD is an outer membrane lipoprotein with a single starch-binding pocket and is essential for *Bt* growth on starch polymers larger than 5 glucose units[12].

SusC and SusD work in concert with three predicted outer membrane lipoproteins, SusE, SusF and SusG[11]. SusG, is an α-amylase essential for growth on high molecular weight starch[13].

Structural analysis of SusG revealed that it contains two starch-binding pockets in addition to the catalytic site, both of which are necessary for efficient degradation of insoluble starch by the purified enzyme[14]. Two additional proteins, SusE and SusF, have poorly defined roles in starch metabolism. Previous phenotypic analyses of mutants lacking expression of the *susE* and *susF* genes reported that they were dispensable for growth on starch *in vitro*[13]; although, they contribute substantially to starch binding by whole cells[11]. Neither SusE nor SusF appears to possess enzymatic activity towards starch, as disruption of the only validated amylase (SusG) is not compensated for by the presence of these proteins. Additional support for the importance of SusE and SusF comes from the presence of similar lipoproteins in most other Sus-like systems with specificity for glycans other than starch[8,9]. Although, only close relatives of these proteins involved in binding starch or similar glycans are currently grouped into the same protein families in the Pfam database: SusE (PF14292, currently 236 members) and PB002941 (currently 88 sequences)[15]. Of note, the former family only corresponds to the first ~125 residues of SusE and does not include SusF; the latter family includes the C-terminal domains of both proteins. Very little sequence level homology exists between these proteins, but some are predicted to adopt carbohydrate binding module (CBM) folds[16,17] and at least one of these proteins with specificity for β-2,6-linked fructan binds polysaccharide in its pure form[9]. Finally, a recent bioinformatics study comparing human gut metagenomic samples to those from non-gut environments found that one of the most abundant human gut-specific microbial protein families includes SusE and SusF[18].

To effectively degrade insoluble glycan structures, many microbial glycoside hydrolases are appended with non-catalytic CBMs. These small β-sheet rich domains, ~100 amino acids, often enhance glycan degradation by tethering the enzyme to the substrate, or by disrupting the

secondary or tertiary structure of the glycan[19-21]. A great number of bacterial amylases contain one or more CBMs, and the removal or mutation of these domains decreases the ability of the enzyme to process insoluble starch[14,22-24]. In some instances, the addition of a starch CBM can impart the ability to degrade raw starch to an amylase that does not otherwise have this capability [25,26]. To date, the carbohydrate active enzymes (CAZy) database recognizes 10 CBM families that bind starch, all of which describe protein domains that are components of amylases. While non-enzymatic CBM-containing proteins have been described as part of cellulosomal complexes[1], non-enzymatic proteins composed of starch-binding CBMs have not been reported.

In this study we investigate the interactions of purified *Bt* SusE and SusF proteins with starch or its oligosaccharides using x-ray crystallographic and biochemical approaches. Structural analyses of SusE and SusF demonstrate that each protein functions as a multivalent starch-binding protein: SusE contains two binding sites and SusF contains three. The C-terminal regions of both proteins encompass two CBMs that are structurally very similar. The extra binding site in SusF is due to the insertion of an additional CBM into the middle of a sequence with otherwise similar topology to SusE. We constructed single and double binding site mutants in SusE and SusF to evaluate the individual contributions of each site to binding starch and various oligosaccharides. Each site displays subtle differences in its starch-binding architecture and binding preference, suggesting that each site is adapted to slightly different starch substrates. Including SusD and SusG, there are a total of eight distinct non-catalytic sites at which Sus proteins bind their substrate. Based on these observations, we speculate that SusE and SusF have evolved to help *Bt* compete for starch in the human intestinal tract, by sequestering starch at the bacterial surface and away from competitors. In addition, the occurrence of CBMs in non-

enzymatic polypeptides, which is rarely reported, may serve to assist the catalytic function of SusG in this multi-protein system that is present on the cell surface.

## Results and Discussion

### *SusE and SusF are surface-exposed lipoproteins*

Both SusE and SusF are predicted to contain an N-terminal signal sequence followed by Cys that should be lipidated after secretion and processing by signal peptidase II. Since a pathway for secreting lipoproteins to the external leaflet of the Gram-negative outer membrane has yet to be defined[27], we examined the cellular location of SusE and SusF by changing the predicted lipidated Cys of each protein to an Ala. This mutation should allow secretion and signal peptide cleavage by signal peptidase I, resulting in a soluble periplasmic form of each protein. Consistent with its predicted location, wild-type (WT) SusE and SusF were detected on the *Bt* cell surface when probed with SusE- or SusF-specific antibodies (**Figure 2.1)**. In contrast, SusE or SusF was not detected on the cell surface of mutant strains producing periplasmic SusE or SusF, although these proteins, in amounts similar to WT, were observed in cell lysates by western blot. Consistent with earlier reports, growth of *Bt* lacking surface expression of SusE and SusF did not result in a significant growth rate defect on maize amylopectin and glycogen (data not shown).

### *SusE and SusF have multiple starch-binding domains*

SusE and SusF were expressed in *E. coli* from constructs that eliminated the N-terminal secretion and lipidation features. Structure determination of both proteins was performed using SAD phasing from crystals obtained from selenomethionine (SeMet)-substituted

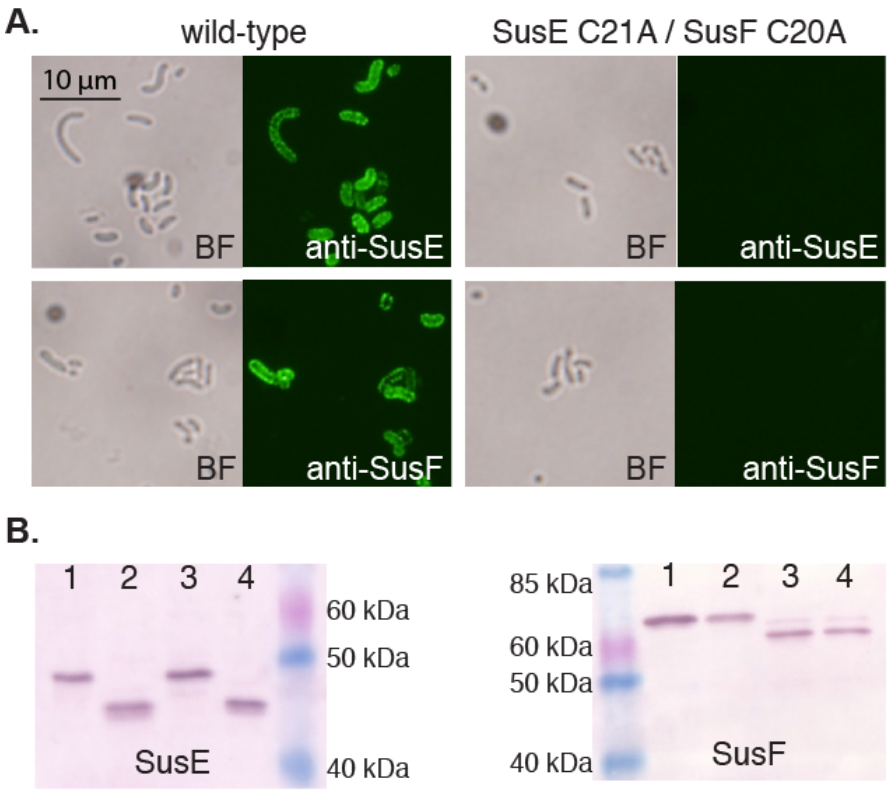



**FIGURE 2.1. SusE and SusF are exposed on the surface of *Bt*.**

Alleles of *susE* and *susF* were created in which the N-terminal cysteine, predicted to be lipidated to tether the proteins to the outer membrane, was mutated to alanine (SusE C21A and SusF C20A). These alleles were recombined into the native *sus* locus. Cells were grown to mid-exponential phase in minimal media (MM) with maltose to induce *sus* expression. **A.** *Bt* staining for SusE and SusF surface expression. Non-permeabilized cells were fixed and probed for SusE and SusF surface expression using polyclonal antisera. Fluorescent images are shown with the corresponding bright field (BF) images. All images are shown on the same scale; bar = 10µm. **B.** Western blot of lysates from whole cells expressing the wild-type and mutant alleles probed in A. Wild-type (1), SusE C21A (2), SusF C20A (3), and SusE C21A SusF C20A (4) *Bt* whole cell lysates were probed for SusE and SusF protein using polyclonal antibodies. Size difference between the wild-type and lipidation signal mutant proteins corresponds to loss of the lipid tail

protein. The initial protein models were built from the SeMet data sets, and then used as models for molecular replacement with the native protein data sets (**Table 2.1, 2.2**). The 2.0Å crystal structure of SusF, the larger of the two proteins, included maltoheptaose (M7) ($R_{work}$ = 19.6%, $R_{free}$ = 24.8%) and encompassed residues 40 – 485. The first 19 residues at the N-terminus of the recombinant SusF were not resolved in electron density, suggesting a flexible linker to the lipidation site. The topology of SusF can be described as three tandem domains (N-terminal, middle and C-terminal) that assume an S-shaped conformation in the crystal structure (**Figure 2.2A**). These domains are packed against each other, although the buried surface area between the N-terminal and middle domain (364 $Å^2$), and middle domain and C-terminal domain (345 $Å^2$) is quite small and includes just a few hydrogen-bonding contacts.

The N-terminal domain (residues 40-160) of SusF consists of a β-sandwich that is similar in overall fold and topology to several immunoglobulin superfamily (IgSF) domains found in cell adhesion proteins including CD28 (1YCD-chainC; RMSD 3.1Å, 8% sequence identity), and CD47 (2JJS-chain A; RMSD 2.7Å, 12% sequence identity). Beyond this N-terminal domain, SusF consists of three β-sandwich CBMs of ~100 amino acids each. We will refer to these as CBMs Fa, Fb, and Fc, using "F" to denote that they are from SusF, and labeling them alphabetically from the N- to C- terminus. The middle domain of SusF (residues 161-274) is composed of CBM Fa, while the C-terminal domain is composed of two distinct CBMs (residues 275-383 as Fb and residues 384-485 as Fc) that are closely packed together via hydrophobic interactions. While each CBM displays unique binding-site features, the overall architecture of each is quite similar and reminiscent of many starch-binding CBMs[21]. Submission of the three individual CBMs of SusF to the DALI server[28] revealed that all share the most structural homology with the X25 domain of the *Bacillus acidopullyticus* glycoside hydrolase (GH) family

**TABLE 2.1. SAD data collection statistics for SusE and SusF**

| | SeMet SusF | SeMet SusE |
|---|---|---|
| **Wavelength (Å)** | 0.97941 | 0.97941 |
| **Resolution (Å)** | 50.0 -2.18<br>(2.22 – 2.18)* | 50.0 -2.65<br>(2.70 – 2.65) |
| **No. independent Reflections** | 40357<br>(2009) | 15067<br>(691) |
| **Completeness** | 99.6<br>(99.3) | 97.7<br>(94.1) |
| **Redundancy** | 5.2<br>(4.4) | 6.9<br>(3.2) |
| **Avg I/Avg σ(I)** | 38.5<br>(11.8) | 22.0<br>(1.75) |
| **$R_{sym}$ (%)** | 6.6<br>(13.3) | 9.7<br>(47.1) |

* Parentheses denote highest resolution shell

**TABLE 2.2. Data collection and refinement statistics**

| **Structure** | SusF / maltoheptaose | SusE / α-cyclodextrin | SusE / maltoheptaose |
|---|---|---|---|
| **PDB Accession** | 4FE9 | 4FEM | 4FCH |
| **Resolution (Å)** | 50.0 – 2.0 (2.03 – 2.00)* | 50.0 – 2.50 (2.54-2.50) | 50.0 – 1.30 (1.32 – 1.30) |
| **Unique reflections** | 46531 (1990) | 17403 (715) | 141909 (6557) |
| **% Completeness** | 89.1 (77.3) | 95.7 (80.9) | 97.4 (91.3) |
| **Redundancy** | 2.5 (1.9) | 9.3 (2.4) | 6.9 (5.7) |
| **Avg, I/ Avg σ (I)** | 16.9 (3.5) | 35.4 (2.0) | 39.5 (4.0) |
| **$R_{sym}$ (%)** | 6.9 (25.3) | 6.6 (36.4) | 8.4 (30.1) |
| **No. proteins atoms** | 3391 | 1709 | 3411 |
| **No. hetero- atoms** | 573 | 111 | 740 |
| **$R_{work}$/# reflections (%)** | 19.9/41493 (26.8/2671) | 20.7/15644 (35.2/1001) | 16.4/127655 (20.0/9221) |
| **$R_{free}$/# reflections (%)** | 24.7/2351 (32.4/166) | 24.1/879 (39.4/57) | 17.8/7098 (21.4/496) |
| **Avg B. factors** | | | |
| **Protein Atoms** | 19.3 | 53.1 | 5.4 |
| **Ligand (sugar)** | 28.2 | 51.4 | 13.2 |
| **Heteroatoms** | 27.8 | 53.1 | 17.3 |
| **RMS deviations:** | | | |
| **Bond length (Å)** | 0.015 | 0.012 | 0.015 |
| **Bond Angles (degrees)** | 2.118 | 0.832 | 2.031 |

* Parentheses denote highest resolution shell

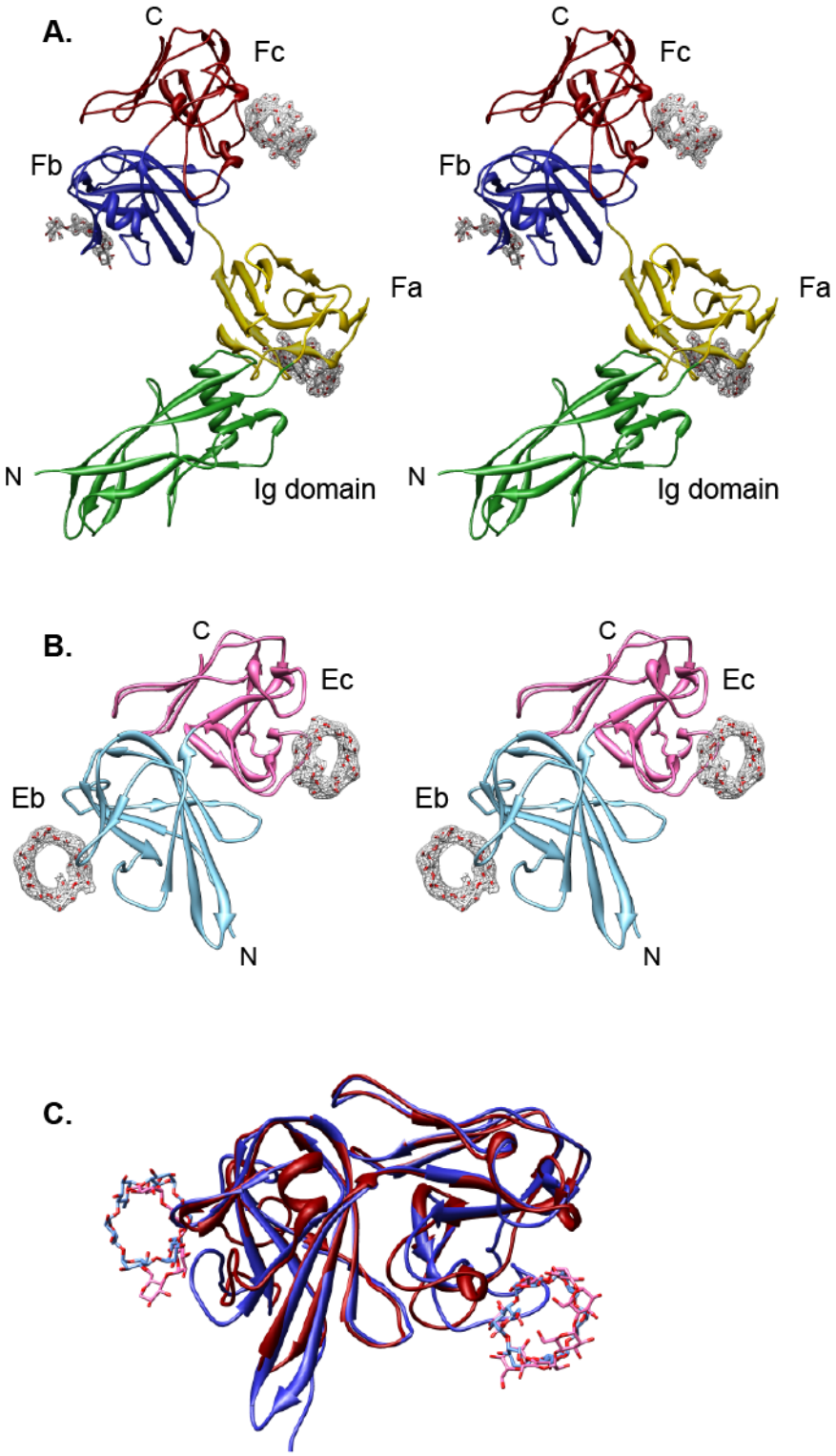

A.
C
Fc
Fb
Fa
N
Ig domain
C
Fc
Fb
Fa
N
Ig domain
B.
C
Ec
Eb
N
C
Ec
Eb
N
C.

**FIGURE 2.2. Ribbon diagram of SusE and SusF structures.**

**A.** Stereo pair cartoon representation of SusF (residues 40-485), with the IgSF domain (residues 40-160) in green, CBM Fa (residues 161-172) in yellow, CBM Fb (residues 275-383) in blue, and CBM Fc (residues 384-485) in red. Bound M7 is displayed as red and white sticks. The electron density from an omit map, contoured at 2.5σ is shown for the ligands. Note that the M7 observed at Fa and Fc is shared across a crystallographic symmetry axis, and therefore the electron density is the same. **B.** Stereo pair cartoon representation of SusE (residues 174-387), with CBM Eb (residues 174-283) colored aqua and CBM Ec (residues 284-385) colored pink. Bound aCD is displayed as red and white sticks. Electron density for aCD from an omit map is displayed and contoured at 2σ. The ligand observed at Eb and Ec is shared across a crystallographic symmetry axis, and therefore the electron density is the same. **C.** Overlay of the SusE CBM Eb and Ec domains (blue) with the SusF CBM Fb and Fc domains (red). The RMSD of the models is 1.3Å for 189 Cα atoms. The ligand aCD bound to SusE is shown as light blue sticks, and the maltotetraose and M7 bound to SusF are shown as pink sticks.

13 pullulanase (pdb 2WAN), with Z-scores of 7.8, 7.3 and 4.9 for the Fa, Fb and Fc CBMs, respectively. While the core β-sandwich structure of the SusE and SusF CBMs are similar to described starch-binding CBMs, the β-strand topology is different which prevented an amino acid sequence based prediction of SusE and SusF as starch-binding CBMs. Therefore, we propose that the five CBMs between SusE and SusF should be added as a novel class of CBMs in the CAZy database[17].

The asymmetric unit of the SusF crystals (C2) contained one molecule of SusF and two molecules of M7, one at CBM Fb and one that adopts a nearly circular conformation and is shared between Fa and Fc of a symmetry related molecule. This packing arrangement does not suggest a dimeric interface, and both size-exclusion chromatography and native PAGE suggest that SusF is a monomer (data not shown). The starch-binding sites of Fb and Fc are oriented nearly 180° away from each other, an arrangement that mimics the orientation of the tandem CBM41 domains of *Streptococcus pneumoniae* SpuA[29]. However, in SusF the additional CBM Fa creates a triangle of binding sites, with each starch-binding site oriented approximately 120° apart (**Figure 2.2A**).

The structure of SusE (residues 35-387) complexed with α-cyclodextrin (aCD) was solved to a resolution of 2.5Å ($R_{work}$ = 20.4%, $R_{free}$ = 24.2%). The final model includes residues 174-387, as the predicted N-terminal domain (residues 38-167) was not observed in the electron density (**Figure 2.2B**). Sufficient space exists in the asymmetric unit for this domain, and both mass spectrometry analysis on SusE prior to crystallization, as well as SDS-PAGE analysis of extensively washed crystals indicated the prominent presence of the full-length (~40 kDa) protein (data not shown). Therefore we conclude that there is a flexible linker between the N- and the C-terminal domain, causing the former to be disordered in the crystal lattice. In the

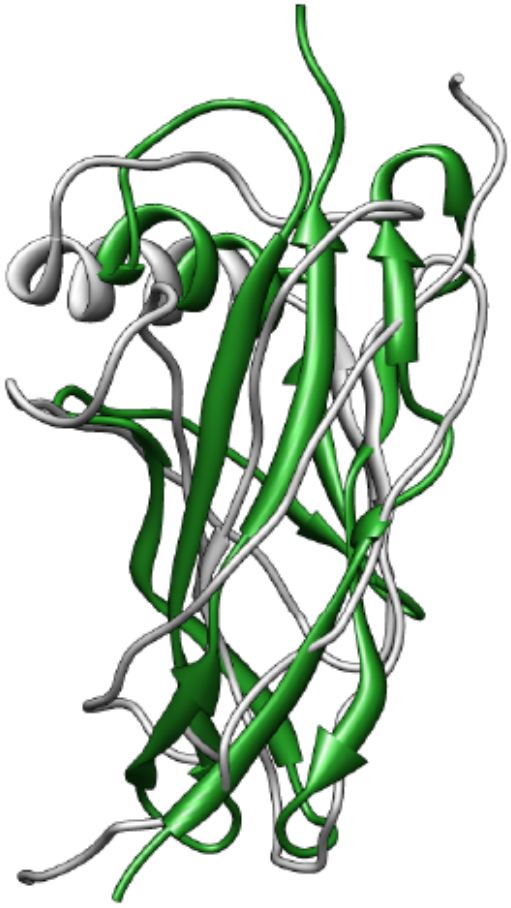


**FIGURE 2.3. Overlay of the N-terminal domains of SusF and SusE (predicted).**

The amino acid sequence of SusE (gray ribbon) from residues 35 through 171 were submitted to the I-TASSER protein structure prediction server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) and compared to the empirically determined structure of SusF (green ribbon). Note that SusF was not deposited in the PDB at the time of the search and therefore not utilized as a template. SusF (residues 40 – 160) and SusE (residues 35-171) were superimposed in COOT, and have a core RMSD of 3.1Å for 91 residues. Amino acid sequence identity for these N-terminal domains is 14.3%.

structure, two symmetry-related molecules of SusE are clustered around a single molecule of aCD. There is very little (285 $Å^2$) buried surface area between the proteins and both size exclusion and native PAGE indicate that SusE is a monomer (data not shown).

The most striking difference between SusE and SusF is that SusE is ~10kDa smaller, due to the absence of a middle domain corresponding to Fa in SusF. While the N-terminal domain of SusE was not resolved in the crystal structure, the predicted structure of residues 38-167 generated using I-TASSER[30,31] suggests a similar IgSF-type fold (**Figure 2.3**). The C-terminal domain of SusE is strikingly similar to the C-terminal domain of SusF and is also composed of two CBMs (residues 174-283 as CBM Eb and residues 284-387 as CBM Ec) packed tightly together. The C-terminal domains of SusE and SusF superimpose with an RMSD of 1.3Å over 189 Cα atoms and share 38.6% sequence identity (**Figure 2.2C**).

***The SusF starch-binding sites coordinate oligosaccharides differently***

Each of the three CBMs in SusF display bound M7 in the crystal structure allowing a comparison of the molecular details of binding at each site. Each site has features universal to many starch-binding proteins: an arc of aromatic amino acids for hydrophobic stacking with glucose and hydrogen-bonding acceptors and donors for interacting with the O-2 and O-3 of glucose. However, each site also displays differences in ligand binding that may impart some specificity regarding which part of a starch molecule is preferred or how tightly it is bound.

A molecule of M7 is shared between the CBMs Fa and Fc of symmetry-related proteins, imposing a circular shape on the linear maltooligosaccharide (**Figure 2.4A**). The ring-like appearance of M7 suggests that the ends of the ligand occur in different places in different molecules and thus an average of these orientations is manifest in the electron density. The Fa

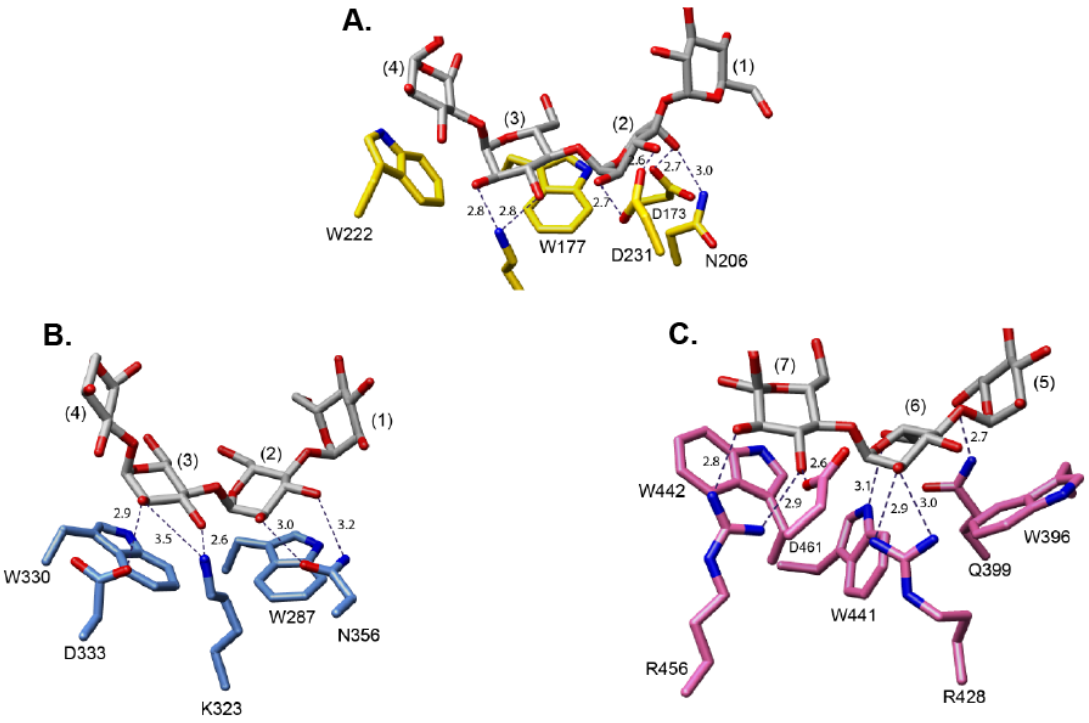



**FIGURE 2.4. Close-up view of the three starch-binding sites in SusF.**

In each panel, M7 is shown as gray and red sticks and the amino acids involved in binding displayed. Dashed lines depict the hydrogen-bonding network between the ligand and protein and distances are shown in Å. Note that in panels A and C, only the area of the ligand bound to protein is displayed. Glucose residues are numbered with glucose (1) indicating the non-reducing end of the maltooligosaccharide. The interactions are shown for **A.** CBM Fa, displaying only glucoses 1 – 4, **B.** CBM Fb (note that only four of the possible seven glucose units were resolved in the electron density), **C.** CBM Fc, displaying only glucoses 5-7.

binding site displays a characteristic aromatic arc (W177 and W222) that stacks against Glc3 and Glc4; however, hydrogen bonding occurs at Glc3 and Glc2. It is more typical in starch-binding sites to observe the same glucose residue anchored in place by both hydrophobic stacking and hydrogen-bonding interactions[14,32-34]. At the Fb site, four of the seven glucose residues of M7 are resolved in the electron density (**Figure 2.4B**). This site, unlike Fa, recognizes only two rather than three glucose moieties, although both monosaccharides at Fb are stabilized via hydrophobic stacking as well as hydrogen-bonding interactions. The Fc binding site is somewhat more extensive than the Fb site. The residues that create the aromatic platform for hydrophobic stacking, W442 and W396, are further apart than those within Fa and Fb, with W441 wedged between these residues, and providing an additional hydrogen-bonding donor to the O-6 of Glc6 (**Figure 2.4C**).

While each of the SusF CBMs displays subtle molecular differences in the binding sites, the orientation of each curved M7 at these surface sites suggests that a long helix of starch could be accommodated with the pitch of the helix lying parallel to the plane of the protein surface. This might allow the protein to recognize and bind the double helical starch structures present in more resistant and insoluble forms of starch (amylose) that transit to the distal intestinal environment.

***CBM Ec has an additional loop that may mediate interactions with single helical starch***

Noting the absence of the N-terminal domain of SusE in our structure of the near full length protein, we decided to pursue a higher resolution structure of the SusE C-terminal domain (residues 172 – 387). A structure of this domain with M7 was solved to a resolution of 1.3Å ($R_{work}$=16.5%, $R_{free}$=17.8%). The space group of this structure was $P2_12_12_1$ with two SusE

molecules per asymmetric unit. These monomers overlay with an RMSD of 0.3Å for all atoms, except one loop (residues 360-365) with a maximum Cα deviation of 2.7Å, likely due to crystal contacts. The C-terminal domain from the SusE structures with aCD and M7 overlay with a RMSD of 0.4Å with no Cα deviations in either starch-binding site.

CBM Eb overlays with CBM Fb with an RMSD of 1.4Å over 93 Cα atoms (33.3% sequence identity). The binding of aCD at Eb is similar to M7 binding at Fb, with adjacent glucose residues bound via both hydrophobic stacking and hydrogen bonding interactions (**Figure 2.5A**). In the SusE structure with M7, no oligosaccharide is bound at Eb, rather a protein-protein crystal contact is made between SusE molecules of adjacent asymmetric units. These crystals were generated using a 2:1 molar ratio of protein to M7, so it is not surprising that one of the starch binding sites was empty. This observation and additional data discussed below suggest that Eb has a weaker binding site relative to Ec.

The second CBM of SusE (Ec) has the most extensive set of protein-ligand interactions among all five CBM domains contained in SusE and SusF. In the aCD structure Ec contacts 5 out of 6 possible glucose residues, but a different mode of binding was observed in the M7 structure, highlighting the potential for Ec to bind single helical regions of starch (**Figure 2.5**). Tryptophans W336 and W296 of Ec create a hydrophobic arc with W335 wedged between, but not participating in glycan binding. A unique feature of the Ec site is the loop created by residues 353-357 that caps one end of the binding site, with the side chain of I355 centered in front of the aCD ring. This loop provides multiple hydrogen-bonding partners to Glc1, Glc2 Glc3 and Glc6 of aCD, via specific interactions with N353, L354, I355 and D356 (**Figure 2.5B**). This starch-binding loop is unlikely to be flexible, and rather is anchored in place by a network of hydrogen

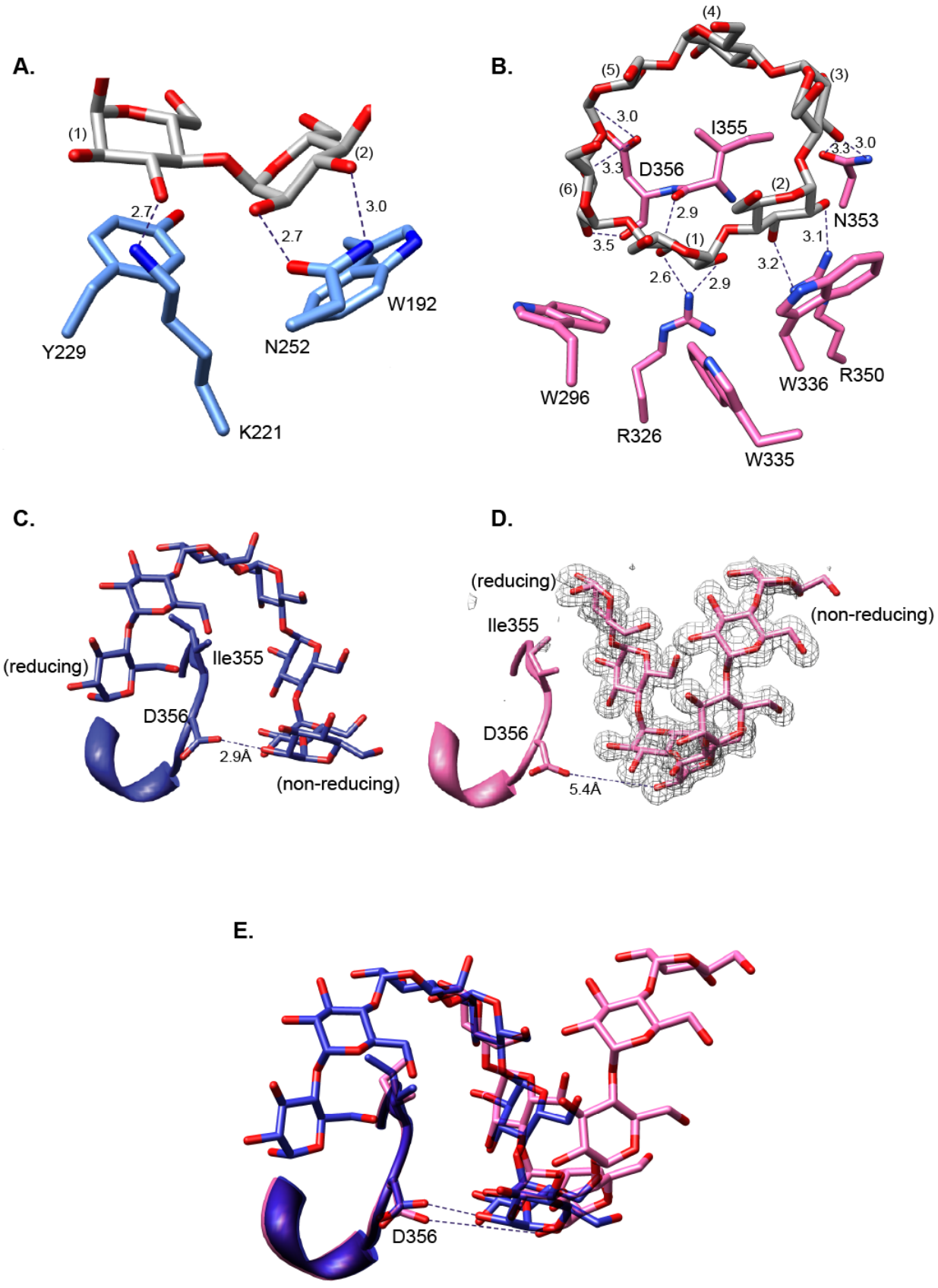

A.
(1)
(2)
2.7
3.0
2.7
W192
Y229
N252
K221
B.
(4)
(5)
(3)
3.0
I355
3.3
3.3
3.0
D356
(6)
(2)
2.9
N353
3.5
(1)
3.1
3.2
2.6
2.9
W296
W336
R350
R326
W335
C.
(reducing)
Ile355
D356
2.9Å
(non-reducing)
D.
(reducing)
(non-reducing)
Ile355
D356
5.4Å
E.
D356

**FIGURE 2.5. Close-up view of the starch-binding sites in SusE.**

Panels **A**. & **B.** depict the structure of SusE with aCD, while Panels **C.-E.** depict the structure of the C-terminal half of SusE with maltoheptaose. **A.** aCD binding at CBM Eb, with the ligand as gray and red sticks, and the amino acids involved in binding displayed. Dashed lines depict the hydrogen-bonding network between the ligand and protein and distances are shown in Å. Note that only the glucose residues involved in binding are displayed. Glucose residues are numbered with glucose (1) indicating the non-reducing end of the maltooligosaccharide. **B.** aCD binding at CBM Ec, as described for panel A. L354 was omitted for clarity. **C.** M7 bound at Ec (chain A) demonstrating the curvature of the ligand and the manner in which it extends over the loop created by residues 353-357. **D.** M7 bound at CBM Ec (chain B). Electron density for maltoheptaose was generated from an omit map, contoured at 3σ. Note that due to crystallographic symmetry the ligand in Panels C and D are the same molecule and thus electron density is only displayed in one panel. **E.** Overlay of M7 bound by chains A (purple) and B (pink) at CBM-Ec, demonstrating the manner in which this site may accommodate a longer molecule of starch.

bonds with an adjacent loop defined by residues 359-362. The topology of this binding site, in particular the centering of the I355 side chain at the ligand, is strikingly similar to the binding of βCD to the glycogen-binding domain of AMP-activated protein kinase[35].

In the structure of SusE with M7, the ligand is shared across a symmetry axis at the CBM Ec, between chain A of one asymmetric unit and chain B of another (**Figure 2.5C, D**). An overlay of these two ligands at chains A and B simulates a model of a 10-glucose long maltooligosaccharide interacting with this extensive binding site (**Figure 2.5E**). In both chains A and B, M7 is anchored to the protein by the same set of hydrophobic stacking interactions with W336 and W296, as well as hydrogen-bonding through R326 and R350. At chain A, the maltooligosaccharide helix, from the non-reducing to reducing end, projects towards the protein against the capping loop (**Figure 2.5C, D**). The peptidyl oxygen atoms of L354 and I355 participate in hydrogen-bonding with hydroxyl groups from adjacent glucose residues as seen in the structure with aCD, but due to the pitch of the oligosaccharide helix, D356 is now 5.4 Å away. However, the same M7 bound by chain B is instead "draped" over this loop, with the maltooligosaccharide from the non-reducing to reducing end extending from the hydrophobic cradle of binding residues and extending up and over the capping loop. Thus, in chain B the non-reducing end of the ligand is nestled closer to the capping loop, such that the glucose at the terminal non-reducing end interacts with D356. In this ligand orientation, I355 intercalates directly into the groove of the M7 helix. As mentioned earlier, the overall atomic structures of chains A and B are nearly identical, with the exception of a helical turn (residues 361 – 365) that is about 15Å from the starch binding site and therefore unlikely to influence binding. The orientation of the starch-binding loop is identical in the structures with M7 and with aCD.

The presence of the starch-binding loop in Ec could govern the forms of starch that bind at this site. A long helix of starch could bind at Eb with the pitch of the helix parallel to the protein surface, similarly to how starch may bind to SusF. At these sites, it is the outer shape of the starch helix that is recognized, and thus single or double helical forms of starch could bind. However, the loop containing I355 that intercalates into one of the grooves of the starch helix at Ec makes interactions with double helical starch unlikely, suggesting this site could be specific for partially unwound single helical forms or small starch breakdown products.

***SusE and SusF display differences in their affinity for starch oligosaccharides***

The chemical and physical structures of starches and related molecules that reach the human colon vary due to a number of features: molecular weight, the pattern and density of α-1,6 branches, the degree to which they have already been degraded by human enzymes and even cooking methods. *Bt* requires the Sus to degrade a variety of different molecules, including amylose, amylopectin and pullulan[36]. While the Sus outer membrane amylase (SusG) will only hydrolyze α-1,4 linkages[14], at least one of the periplasmic amylases (SusB) is promiscuous towards a variety α-glucosidic linkages[37]. Thus, it is possible that SusE and SusF interact with oligosaccharides that contain α-1,6 branches prior to transport across the outer membrane. Moreover, the cyclic maltooligosaccharide aCD mimics the rigid, geometrically constrained curvature of larger amylose molecules, making it possible to probe starch-binding proteins for affinity towards starch secondary structures as opposed to linear oligosaccharides with more flexible helical geometry.

To test the affinity of the various SusE and SusF binding sites for different structures, we performed isothermal titration calorimetry (ITC) using three different starch

**TABLE 2.3. Affinity of SusE and SusF CBMs for maltooligosaccharides determined by ITC**

| Protein - active CBM | Mutations (to Ala) | $K_d$ (μM) aCD | $K_d$ (μM) M7 | $K_d$ (μM) GM3M3 |
|---|---|---|---|---|
| **WT SusE** | none | 86.96 ± 19.7 | 134.2 ± 34.2 | 357.1 ± 12.8 |
| **SusE – B only** | R326 W336, R350 | 386.1 ± 35.8 | 1023.5 ± 36.7 | 3584.2 ± 120.8 |
| **SusE – C only** | W192, K221, Y229, N252 | 97.09 ± 13.2 | 17.04 ± 1.2 | 641.0 ± 90.4 |
| **SusE –no binding** | W192, K221, Y229, N252, R326, W336, R250 | No binding | No binding | Not tested |
| **WT SusF** | none | 769.2 ± 50.9 | 303.0 ± 23.9 | 990.1 ± 107.8 |
| **SusF – A only** | W287, K323, N356, W396, W442, R456 | 775.2 ± 15.6 | 361.0 ± 4.6 | 2710 ± 110.2 |
| **SusF – B only** | W177, K208, W222, D231, W396, W442, W456 | 460.8 ± 51.0 | 309.6 ± 10.5 | 751.9 ± 44.7 |
| **SusF- C only** | W177, K208, W222, D231, W287, K323, N356 | 465.1 ± 34.6 | 97.09 ± 2.9 | 507.6 ± 20.6 |
| **SusF – no binding** | W177, K208, W222, D231, W287, K323, N356, W396, W442, R456 | No binding | No binding | Not tested |

oligosaccharides: aCD, M7 and glucosyl-maltotriosyl-maltotriose (GM3M3), an oligosaccharide of seven glucose units containing two α-1,6 linkages (**Table 2.3**). In addition to examining the overall binding affinities of the two WT proteins, we created a series of site-directed mutants of each protein in which only one ligand-binding site remains active; these proteins are labeled to designate the active CBM remaining (*e.g.*, SusF-A only indicates that the Fa domain is still active while the others have been mutated). For both SusE and SusF, we also created negative controls in which all CBMs were mutated, referred to as SusE-no binding and SusF-no binding. We did not detect any binding with these negative control proteins confirming that the site-directed mutations abolished starch binding. As observed in the crystal structures, it is possible for both proteins to cluster around a single molecule of aCD or M7, and thus it is possible that during the course of the ITC experiment both 1:1 and 2:1 protein:ligand binding events are occurring. Therefore, because we knew the number of binding events to expect approaching saturation, we chose to fit the data to a one-site model and fix N to the number of binding sites in each protein. Thus, our $K_d$ values reflect the relative affinity of each protein for each ligand.

Overall, SusE has a higher affinity for the three ligands compared to SusF. The Eb site displays tighter binding for aCD compared to M7 and GM3M3, likely due to the reduced entropic penalty of binding the geometrically constrained ligand. Many starch-binding sites only recognize 2 or 3 glucose residues and thus the lack of true helical shape in aCD, which is a ring, is compensated for by the fixed geometry of the cyclodextrin[38]. This is not true for ligand binding at CBM Ec. At Ec helical M7 was bound with higher affinity ($K_d$ 17.04 μM) compared to aCD ($K_d$ 97.09 μM); the unique binding site loop in Ec allows the protein to recognize much more of the starch ligand, and thus the pitch of the helix, as seen in M7 in the crystal structure, is required to maximize interactions with the protein. Unexpectedly, all three CBMs of SusF bound

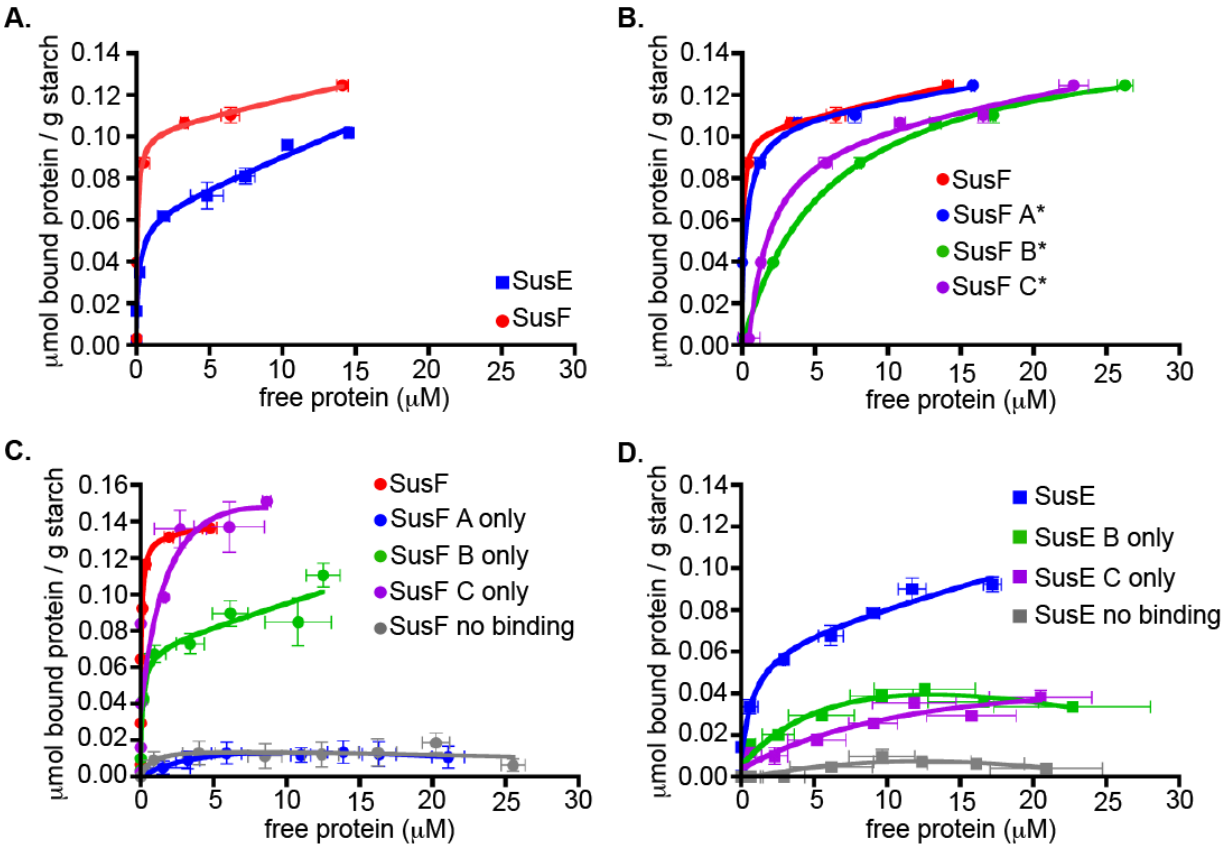



**FIGURE 2.6. Protein binding to insoluble cornstarch**.

Bound protein per gram of starch is plotted as a function of free protein concentration with error bars representing the standard error from three replicates. Data were fit to a one-site total binding equation. **A.** WT SusE and SusF; **B.** WT SusF and mutant forms of SusF where one of the binding sites has been mutated (SusF A*, SusF B* and SusF C*); **C.** WT SusF with mutant forms of SusF where only one binding site remains intact (SusF A only, SusF B only and SusF C only) or where all binding sites were mutated (SusF no binding); **D.** WT SusE and mutant forms of SusE where only one of the binding sites remains intact (SusE B only, SusE C only) or both binding sites were mutated (SusE no binding).

**TABLE 2.4 SusE and SusF binding to high molecular weight insoluble starch**

| **Protein – active CBM** | $K_d$ **(μM)** |
|---|---|
| WT SusE | 0.233 ± 0.092 |
| SusE – B only | 10.62 ± 6.814 |
| SusE – C only | 19.51 ± 51.84 |
| SusE – no binding | [§]Could not be fit |
| WT SusF | 0.106 ± 0.094 |
| SusF – A* | 0.433 ± 0.408 |
| SusF – B* | 5.821 ± 1.363 |
| SusF – C* | 1.002 ± 0.291 |
| SusF – A only | [§]Could not be fit |
| SusF – B only | 0.154 ± 0.095 |
| SusF – C only | 2.182 ± 4.404 |
| SusF – no binding | [§]Could not be fit |

[§] Could not be fit designates curve fits with an R squared value of less than 0.8

M7 with slightly better affinity than aCD, despite our observations from the crystal structure that these sites only recognize 2 or 3 glucose residues. This may suggest that they are more adept at recognizing a flexible helical segment of starch. This preference for partially "unwound" segments of starch may aid in docking the Sus complex to portions of a starch molecule that will be more accessible to the SusG amylase. SusE and SusF bind GM3M3 the weakest of all three ligands, suggesting that while α-1,6-linkages are tolerated, there is unlikely to be a preference for these structures over α-1,4-linked glucose.

***SusE and SusF CBMs contribute differently to binding of insoluble starch***

The presence of multiple starch-binding sites on a single protein introduces the possibility that SusE and SusF bind longer polymers better than small oligosaccharides due to an avidity affect, in which binding at more than one site occurs simultaneously resulting in increased apparent affinity. To determine the binding affinity of WT and binding site mutants of SusE and SusF to insoluble cornstarch we performed adsorption depletion experiments. The error of some of the curve fits are elevated; we attribute this to errors in using the BCA assay near the high and low limits of protein detection, as well as to potential differences in non-specific binding between replicates. We performed this assay many times while refining our final assay conditions (also performed in triplicate) and consistently observed the same binding trends in our data. Interestingly, SusF and SusE bound starch with a $K_d$ of 0.106 ± 0.094 μM and 0.233 ± 0.091 μM respectively, revealing that despite the extra CBM in SusF there was virtually no difference in their affinities for starch (**Figure 2.6A** and **Table 2.4**). In experiments utilizing single CBM mutants of SusF (**Figure 2.6B**), with the mutated CBM designated by *, there is a decrease in the overall affinity for starch when either CBM Fb (SusF B*) or Fc (SusF C*) is

mutated, but no defect when Fa (SusF A*) is alone mutated. Reciprocally, when Fa is left as the only remaining functional starch binding site (SusF – A only; **Figure 2.6C**), the protein has greatly reduced starch binding, and displays a similar isotherm as the SusF no binding mutant. Therefore Fa, the CBM that is unique to SusF, does not contribute to insoluble starch binding, despite its ability to bind smaller maltooligosaccharides. When the CBMs Fb or Fc alone were mutated, the $K_d$ increased by an order of magnitude over WT SusF, suggesting that these sites may work together to bind starch (**Figure 2.6B**). However, the SusF – B only protein displays nearly the same $K_d$ for starch as WT SusF, suggesting that Fb drives binding to insoluble starch even though it displays moderate affinity for maltooligosaccharides compared to Fc. It is possible therefore that the CBMs of SusF are responsible for binding different structural forms of starch, rather than having redundant starch-binding functions that contribute towards the avidity of the protein for starch. When the individual domains of SusE are mutated, there is a substantial loss in insoluble starch-binding, approximately 40- to 80-fold for the SusE – B only and SusE – C only mutants, respectively. Therefore, in terms of insoluble starch binding, the presence of both domains in SusE is critical.

**Prospectus**

In this report, we investigated the biochemical and structural features of SusE and SusF, two cell surface lipoproteins within the *Bt* Sus complex. These proteins are extremely similar in structure, composed of an observed (SusF) or predicted (SusE) N-terminal IgSF domain, followed by two or three tandem starch-binding CBMs. The N-terminal domain of SusE could not be resolved in the crystal structure suggesting inherent flexibility in this domain. This flexibility may allow the predicted N-terminal IgSF domain to dock to SusF or another Sus

protein and still permit mobility of the SusE starch-binding domains to capture starch. Earlier literature suggests that SusE is more susceptible to proteolytic cleavage in a strain lacking SusF, suggesting these proteins may interact[7]. A striking difference between these two proteins is the presence of the additional CBM Fa in SusF, which may impart extra rigidity to the protein because of increased contacts with the flanking domains. While the Fa binding site has moderate affinity for maltooligosaccharides, it is nearly devoid of insoluble starch binding.

CBMs are typically contained within a single glycoside hydrolase polypeptide or associated enzyme complex (*i.e.*, cellulosomes) and enhance accessibility to an insoluble substrate[19]. Tandem CBMs in glycoside hydrolases have been shown to display an avidity affect in binding carbohydrate, whereby relatively low affinity of the individual domains is augmented several fold due to the protein's multivalent interactions with the substrate [20]. For SusF, there is no apparent avidity advantage from the presence of tandem CBMs. Rather, it seemed that each CBM has different starch-binding characteristics, reflected in both the architecture of the starch-binding sites as well as the observed affinities for the ligands tested. In contrast to SusF, both domains of SusE are required for tight binding to insoluble starch, suggesting an avidity affect. The CBM Ec binding site has an additional loop that is likely responsible for its enhanced binding affinity. The structure of SusE with maltoheptaose demonstrates how a longer, single helix of amylose could interact with the Ec site, suggesting that this site, even more so than CBM Fc, may bind relaxed or denatured α-1,4-glucans.

The precise mechanistic role of SusE and SusF in starch metabolism remains unclear, although the data presented here provide a valuable structural and biophysical perspective (**Figure 2.7**). As mentioned above, current protein classification schemes such as Pfam include a narrow range of lipoproteins that are associated with Sus-like systems within the same families

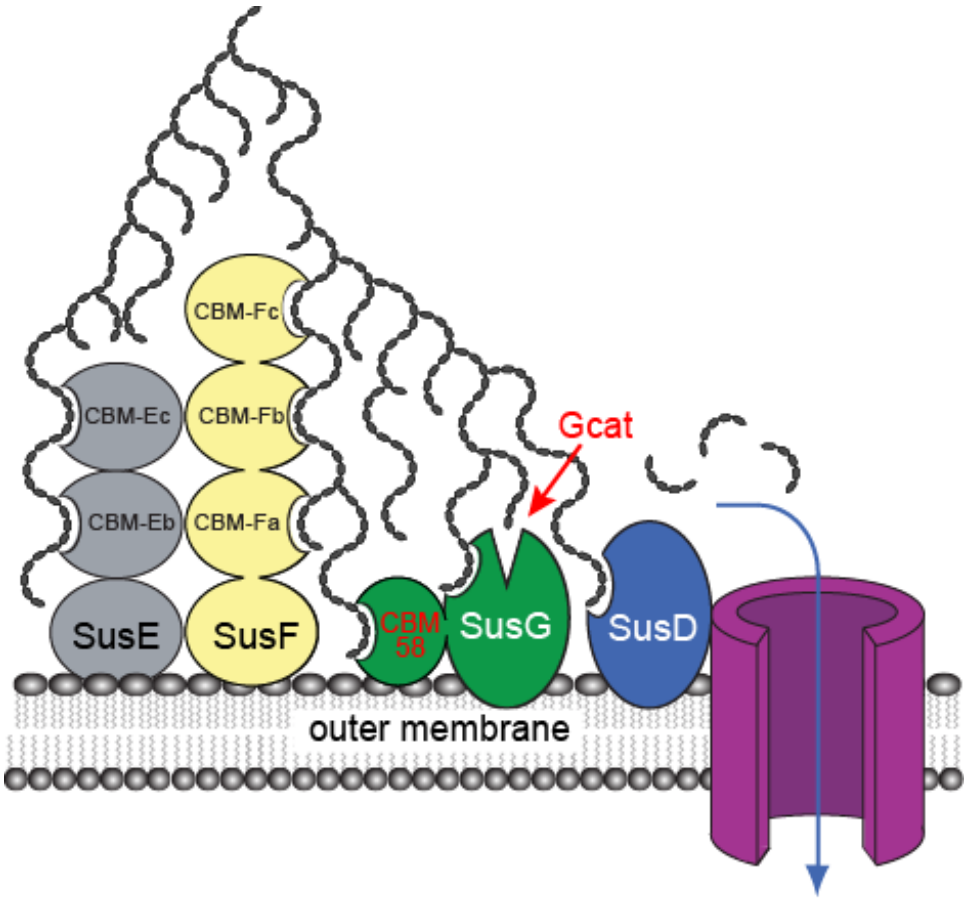



**FIGURE 2.7. Model of Sus outer membrane protein interactions with starch.**

In total, the four outer membrane lipoproteins in the *Bt* starch-utilization system contain at least nine sites that interact with starch or cleaved maltooligosaccharides. Only one of these sites (Gcat) is catalytic and present in the endo-acting amylase, SusG. The remaining eight binding sites are spread across all four lipoproteins. In the model shown, these eight sites make interactions with different regions of a single starch polymer. The nature of potential interactions between individual Sus lipoproteins has not been explored, nor has the stoichiometry of these proteins on the cell surface.

as SusE and SusF. Thus, these groups may exclude many functional or structural homologs that target other glycans, but are missed by primary sequence analysis. Consistent with this idea, one such *Bt* lipoprotein (BT1761) has been shown to bind specifically to β-2,6-linked fructan[9]. Moreover, we have purified two additional proteins (Bacova_04391 and Bacova_02094) from another human gut symbiont, *Bacteroides ovatus*, that have been implicated in metabolism of xylan and β-mannan, respectively. Each of these proteins binds to its predicted target glycan in a gel-retardation assay (data not shown), and the ligand-free crystal structure of Bacova_04391 has been determined by the Joint Center for Structural Genomics (pdb 3ORJ), revealing that it has an N-terminal Ig-like domain followed by two β-sandwich domains resembling CBMs. More work will be needed to establish how these and similar proteins interact with their target glycans, but it is probable that they are part of a diverse group of relatively unexplored glycan-binding proteins that are associated with Bacteroidetes Sus-like systems.

Blocking these two proteins from trafficking to the bacterial surface does not eliminate growth on starch, despite the fact that they contain a total of five starch binding sites. In contrast, SusD has a lower affinity for oligosaccharides and loss of this protein results in complete inability to grow on oligosaccharides greater than five glucose units[12]. Thus, different proteins in Sus-like systems are likely to play different functional roles that are not necessarily dependent on how tightly they bind substrates. Given that two other starch binding sites are present in SusG, including a CBM58 domain[14], it is possible that loss of SusE and SusF is compensated by these additional sites. With structural data in hand for all four of the Sus proteins, we are now in a position to perform this more precise level of mutagenesis and further probe the mechanism of this system. In addition, it is possible that SusE and SusF scavenge starch when it is at low concentrations or sequester it at the cell surface during hydrolysis. Either of these mechanisms

would be valuable to a gut bacterium during competition in the densely populated colonic ecosystem. Regardless of their precise functional role(s), the abundance of proteins related to SusE and SusF in Bacteroidetes Sus-like systems suggests that they are fundamentally important to the fitness and survival of these symbiotic organisms. Our results here shed structural insight into understanding the role of these proteins and provide the basis for future mechanistic studies in live bacteria.

## Materials and Methods

### *Bacterial culture conditions*

*Bt* was grown in tryptone-yeast extract-glucose (TYG) media[39] or on brain-heart infusion (BHI, Beckton Dickinson) agar that included 10% horse blood (Colorado Serum Co.). Antibiotics were added as appropriate including erythromycin (25 μg/ml), gentamicin (200 μg/ml), and 5-fluoro-2'-deoxyuridine (200 μg/ml). Minimal media (MM) with 5 mg/ml maltose was prepared as described in[40].

### *SusE and SusF lipid attachment site mutation*

The *susE* and *susF* genes plus ~700 bp of sequence flanking each gene were amplified from *Bt* strain ATCC 29148 and cloned into the suicide vector pExchange-*tdk*[12]. Mutation of the SusE C21 and SusF C20 codons to alanine was carried out using the QuikChange® site directed mutagenesis kit (Stratagene). The mutated alleles were confirmed by sequencing and introduced into *Bt* by conjugation and counter-selection on 5-fluoro-2'-deoxyuridine. Surface expression of SusE and SusF was probed by antibody staining of non-permeabilized formaldehyde fixed *Bt* cells grown on MM-maltose with rabbit polyclonal antibodies (Cocalico Biologicals) and

detected with an Alexa-Fluor® 488 conjugated goat anti-Rabbit IgG secondary antibody (Molecular Probes). SusE and SusF were detected in *Bt* whole cell lysates by western blot using the rabbit polyclonal primary antibodies mentioned above together with an alkaline phosphatase conjugated goat anti-Rabbit IgG secondary antibody (Sigma).

***Expression of SusE and SusF***

To clone and express the SusE and SusF proteins, the gene fragments corresponding to the soluble domains of SusE (residues 35 – 387 for full length and 172-387 for C-terminal domain) and SusF (residues 21 - 485) were amplified from *Bt* genomic DNA to include *Nde*I (SusE) or *Nhe*I (SusF) and *Xho*I sites at the 5' and 3' ends of the PCR products, respectively. The gene products were ligated into a modified version of pET-28a (EMD Biosciences) containing a recombinant tobacco etch virus (rTEV) protease recognition site. Site directed mutagenesis of the cloned *susE* and *susF* genes was performed using the QuikChange® multi site-directed mutagenesis kit with the *susE*-pET28rTEV or *susF*-pET28rTEV plasmid as the template. Starch-binding residues mutated to alanine in specific CBMs of SusE and SusF are listed in **Table 2.3**.

The pET28rTEV plasmids containing the allele of interest were transformed into Rosetta (DE3) pLysS cells (EMD Biosciences). Transformed cells were grown at 37ºC for 20 hours, and then the plates were scraped to inoculate culture media for protein expression. For native protein expression, the cells were grown in 1L of TB, plus kanamycin (50 μg/ml) and chloramphenicol (20 μg/ml) (in 2L baffled flasks) at 37ºC until they reached an O.D. ~0.4, and the temperature was turned down to 22ºC. Approximately 30 minutes after lowering the temperature, IPTG was added to a final concentration of 0.5mM, and the cells continued to grow overnight (16 – 20 hours). Cells were harvested by centrifugation at 6,000 x g, and the cell pellets were stored at

-80ºC until protein purification. SeMet-substituted protein was produced via the methionine inhibitory pathway[41], as previously described[42].

***Purification of native and SeMet-substituted SusE and SusF***

All SusE and SusF proteins were purified using a 5 ml Hi-Trap metal affinity cartridge (GE Healthcare) according to the manufacturer's instructions. The cell lysate was applied to the column in His Buffer (25 mM $NaH_2PO_4$, 500 mM NaCl, 20 mM imidazole pH 7.4). After sample loading, the column was washed with 40 mls of His Buffer, then proteins were eluted with an imidazole (20 – 300 mM) gradient. The His-tag was removed by incubation with rTEV (1:100 molar ratio of rTEV to protein) at room temperature for 2 h, followed by overnight at 4°C while dialyzing against His Buffer. The cleaved protein was then re-purified on the 5ml Ni column to remove undigested target protein, the cleaved His-tag and His-tagged rTEV. Purified proteins were dialyzed against 20 mM HEPES / 100 mM NaCl (pH 7.0) prior to crystallization, and concentrated using Vivaspin 15 (10,000 MWCO) centrifugal concentrators (Vivaproducts, Inc.).

***Crystallization and Data Collection***

Crystallization conditions were screened via the hanging drop method of vapor diffusion in 96-well plates and using Hampton Screen kits (Hampton Research). Crystals were obtained for the native and SeMet substituted full-length SusE protein at room temperature as hanging drop experiments using 16.5 mg/ml protein and 2 mM aCD against a well solution of 16-20% PEG 6000, 2 M NaCl, 100 mM malonate pH 5.0. The SusE-aCD crystals were then serially transferred into a cryoprotectant of 22% PEG 6000, 2.3 M NaCl, 50 mM malonate, 2 mM aCD, and 19% ethylene glycol and flash-frozen in liquid nitrogen prior to data collection.

Crystals of the SusE C-terminal domain (18 mg/ml) plus 0.5 mM M7 were grown at room temperature from hanging drops against a well solution of 2.5 M ammonium sulfate, 100 mM bis-tris propane pH 7.0. These crystals were flash-frozen in a cryoprotectant containing 2.0 M ammonium sulfate, 80 mM bis tris propane pH 7.0, 1 mM maltoheptaose and 20% ethylene glycol.

Crystals of the native and SeMet-substituted full-length SusF were grown via hanging drop at room temperature using 29.8 mg/ml protein and 2 mM M7 against a well solution of 6-12% glycerol, 1.5-2M Na/$KPO_4$ pH 6.3. The SusF-M7 crystals were then serially transferred into a cryoprotectant of 6-12% glycerol, 1.75-2 M Na/$KPO_4$ pH 6.3, 300 mM NaCl, 2 mM M7, and 16% ethylene glycol and flash-frozen in liquid nitrogen prior to data collection.

SAD x-ray data sets for all SeMet substituted crystals were collected at the Life Sciences Collaborative Access Team (LS-CAT) beamline ID-D at the Advanced Photon Source at Argonne National Labs, Argonne, IL. Native data sets for SusF as well as full-length SusE crystals were also collected at LS-CAT ID-D, while the SusE C-terminal x-ray data were collected at LS-CAT beamline ID-G. X-ray data were processed with HKL3000 and scaled with SCALEPACK [43]. The structures of SusE and SusF were determined from the SAD data using the AutoSol subroutine within the Phenix software package[44,45]. These initial models of SusE and SusF proteins were then utilized for molecular replacement in Phaser[46] against the native X-ray data sets. Data collection statistics are reported in **Tables 2.1-2.2**. The ramachandran plots for all three structures were generated using the MolProbity structure validation server[47]. The structure of the SusE C-terminal domain with M7 had no outliers with 97.4% of residues in favored regions and the rest within the allowed regions of the ramachandran plot. The SusE model with

α-cyclodextrin also had no outliers, and displayed 94.3% of residues within favored regions, and the rest in allowed regions. The SusF structure with maltoheptaose had two residues, E89 (55.7, -23.4) and S341 (-29.1, 143.3) that fell just outside the generously allowed region of the ramachandran plot. E89 is part of a left-hand helical turn. A hydrogen bond between the peptidyl O of L87 and the side chain imidazole N of H91 distort the geometry of this turn. S341 is at the beginning of an α-helix, and a hydrogen bond between the side-chain hydroxyl of S341 and the nearby side chain of E379 may play a role in pulling this residue out of an ideal alignment. For the rest of the SusF model, 96.9% of residues are in the favored regions and the remaining residues in the allowed regions of the ramachandran plot.

***Isothermal Titration Calorimetry***

ITC measurements were carried out using a MicroCal VP-ITC titration calorimeter. Proteins were dialyzed into 50 mM HEPES pH 8.0 and oligosaccharides were prepared using the dialysis buffer. Protein (250 μM) was placed in the sample cell and the reference cell was filled with dialysis buffer. After the temperature was equilibrated to 25°C, a first injection of 2 μl was performed followed by 29 subsequent injections of 10 μl of 20 mM aCD, M7 or glucosyl maltotriosyl maltotriose (GM3M3). The solution was stirred at 305 rpm and the resulting heat of reaction was measured. Data were analyzed using the Origin software package, fixing N to the known number of starch binding sites in the protein of interest. The SusE-C only with GM3M3 isotherm was indicative of two binding events, one being very weak. This weak second binding event is unlikely to be relevant at biological concentrations of starch therefore we included only the first 15 injections in our curve fit to get an approximation of the affinity of the major binding event.

***Adsorption depletion assay***

The affinity of purified SusE and SusF for insoluble cornstarch was determined via adsorption depletion. Cornstarch (Sigma, S4126) was washed twice in an excess of $ddH_20$ then once with an excess of PBS. Starch was pelleted and suspended in PBS to make a 100 mg/mL slurry. 20 mg of starch was pipetted into each well of a microtiter plate, pelleted, and the supernatant discarded. Starch pellets were suspended in 200 µl of protein solution ranging from 1.5 mg/ml to 0.1 mg/mL in PBS. Plates were incubated for two hours at room temperature with agitation. Starch and bound protein was pelleted by centrifugation and the supernatant collected. Unbound protein concentration was determined with the Pierce® Microplate BCA Protein Assay Kit. Bound protein per gram of starch was plotted as a function of free protein from three replicates and fit to a nonlinear regression using the one-site total binding equation (GraphPad Prism).

**Notes**

This work was reprinted and modified with permission from Cameron, E.A., Maynard, M.A., Smith, C.J., Smith, T.J., Koropatkin, N.M., Martens, E.C. (2012) Multidomain Carbohydrate-Binding Proteins Involved in *Bacteroides thetaiotaomicron* Starch Metabolism. *The Journal of Biological Chemistry.* **287**, 34614-34625.

## References


1 Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* **6**, 121-131 (2008).

2 Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* **10**, 323-335 (2012).

3 McNeil, N. I. The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr* **39**, 338-342 (1984).

4 Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031 (2006).

5 Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320**, 1647-1651 (2008).

6 Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65 (2010).

7 Cho, K. H. & Salyers, A. A. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **183**, 7224-7230 (2001).

8 Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284**, 24673-24677 (2009).

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

10 Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).

11 Shipman, J. A., Berleman, J. E. & Salyers, A. A. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *J Bacteriol* **182**, 5365-5372 (2000).

12 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**, 1105-1115 (2008).

13 Reeves, A. R., Wang, G. R. & Salyers, A. A. Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. *J Bacteriol* **179**, 643-649 (1997).

14 Koropatkin, N. M. & Smith, T. J. SusG: A Unique Cell-Membrane-Associated alpha-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. *Structure* **18**, 200-215 (2010).

15 Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res* **40**, D290-301 (2012).

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

17 Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233-238 (2009).

18 Ellrott, K., Jaroszewski, L., Li, W., Wooley, J. C. & Godzik, A. Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. *PLoS Comput Biol* **6**, e1000798 (2010).

19 Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769-781 (2004).

20 Guillen, D., Sanchez, S. & Rodriguez-Sanoja, R. Carbohydrate-binding domains: multiplicity of biological roles. *Appl Microbiol Biotechnol* **85**, 1241-1249 (2010).

21 Machovic, M. & Janecek, S. Starch-binding domains in the post-genome era. *Cell Mol Life Sci* **63**, 2710-2724 (2006).

22 Penninga, D. *et al.* The raw starch binding domain of cyclodextrin glycosyltransferase from Bacillus circulans strain 251. *J Biol Chem* **271**, 32777-32784 (1996).

23 Tan, T. C., Mijts, B. N., Swaminathan, K., Patel, B. K. & Divne, C. Crystal structure of the polyextremophilic alpha-amylase AmyB from Halothermothrix orenii: details of a productive enzyme-substrate complex and an N domain with a role in binding raw starch. *J Mol Biol* **378**, 852-870 (2008).

24 Sumitani, J., Tottori, T., Kawaguchi, T. & Arai, M. New type of starch-binding domain: the direct repeat motif in the C-terminal region of Bacillus sp. no. 195 alpha-amylase contributes to starch binding and raw starch degrading. *Biochem J* **350 Pt 2**, 477-484 (2000).

25 Latorre-Garcia, L., Adam, A. C., Manzanares, P. & Polaina, J. Improving the amylolytic activity of Saccharomyces cerevisiae glucoamylase by the addition of a starch binding domain. *J Biotechnol* **118**, 167-176 (2005).

26 Juge, N. *et al.* The activity of barley alpha-amylase on starch granules is enhanced by fusion of a starch binding domain from Aspergillus niger glucoamylase. *Biochim Biophys Acta* **1764**, 275-284 (2006).

27 Kovacs-Simon, A., Titball, R. W. & Michell, S. L. Lipoproteins of bacterial pathogens. *Infect Immun* **79**, 548-561 (2011).

28 Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem Sci* **20**, 478-480 (1995).

29 Lammerts van Bueren, A. *et al.* The conformation and function of a multimodular glycogen-degrading pneumococcal virulence factor. *Structure* **19**, 640-651 (2011).

30 Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738 (2010).

31 Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* **9** (2008).

32 Boraston, A. B. *et al.* A structural and functional analysis of alpha-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *J Biol Chem* **281**, 587-598 (2006).

33 Lammerts van Bueren, A. & Boraston, A. B. The structural basis of alpha-glucan recognition by a family 41 carbohydrate-binding module from Thermotoga maritima. *J Mol Biol* **365**, 555-560 (2007).

34 van Bueren, A. L., Higgins, M., Wang, D., Burke, R. D. & Boraston, A. B. Identification and structural basis of binding to host lung glycogen by streptococcal virulence factors. *Nat Struct Mol Biol* **14**, 76-84 (2007).

35 Polekhina, G. *et al.* Structural basis for glycogen recognition by AMP-activated protein kinase. *Structure* **13**, 1453-1462 (2005).

36 Anderson, K. L. & Salyers, A. A. Genetic evidence that outer membrane binding of starch is required for starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **171**, 3199-3204 (1989).

37 Kitamura, M. *et al.* Structural and functional analysis of a glycoside hydrolase family 97 enzyme from Bacteroides thetaiotaomicron. *J Biol Chem* **283**, 36328-36337 (2008).

38 Janecek, S., Svensson, B. & MacGregor, E. A. Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. *Enzyme Microb Technol* **49**, 429-440 (2011).

39 Holdeman, L. V., Cato, E. D. & Moore, W. E. C. *Anaerobe Laboratory Manual (Blacksburg, Va.: Virginia Polytechnic Institute and State University Anaerobe Laboratory).* (1977).

40 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**, 447-457 (2008).

41 Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229**, 105-124 (1993).

42 Koropatkin, N. M., Koppenaal, D. W., Pakrasi, H. B. & Smith, T. J. The structure of a cyanobacterial bicarbonate transport protein, CmpA. *J Biol Chem* **282**, 2606-2614 (2007).

43 Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 *Macromolecular Crystallography, part A* (eds C. W. Jr. Carter & R. M. R.M. Sweet) 307-326 (Academic Press, 1997).

44 Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta crystallographica. Section D, Biological crystallography* **58**, 1948-1954 (2002).

45 Zwart, P. H. *et al.* Automated structure solution with the PHENIX suite. *Methods Mol Biol* **426**, 419-435 (2008).

46 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674 (2007).

47 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* **66**, 12-21 (2010).

## Chapter III

## Multi-functional nutrient binding proteins adapt human symbiotic bacteria for glycan competition in the gut


### Abstract

To compete for the dynamic stream of nutrients flowing into their ecosystem, colonic bacteria must respond rapidly to new resources then catabolize them efficiently once detected. The *Bacteroides thetaiotaomicron* starch utilization system (Sus) is a model for nutrient acquisition by symbiotic gut bacteria, which harbor thousands of related Sus-like systems. Four different Sus outer-membrane proteins (SusD,E,F,G) contain a total of eight starch-binding sites that we demonstrate using genetic and biochemical approaches play distinct roles in starch metabolism. SusD enhances starch sensing, allowing *sus* transcriptional activation at much lower concentrations than without this function. Seven additional binding sites across SusE,F,G are less critical for *sus* activation but optimize starch growth rate in a polysaccharide capsule-dependent manner *in vitro* and *in vivo* in gnotobiotic mice. This study reveals how numerically dominant families of carbohydrate binding proteins in the human microbiome fulfill separate and cooperative roles to optimize gut commensal bacteria for glycan acquisition.


### Introduction

A critical symbiotic function of the dense community of bacteria (microbiota) that inhabit the human gut is to break down complex carbohydrates (glycans) that our own digestive

enzymes cannot degrade. Short chain fatty acids and other products from bacterial glycan fermentation are a significant source of nutrition for the host, improving the health of intestinal tissue and directly modulating leukocyte development[1-3]. The collection of carbohydrates available in the colon and the ability of particular bacteria to degrade them shape the membership and abundance of the microbial community[4-6]. Since alterations in the microbiota have been linked to a number of health conditions including inflammatory bowel diseases[7,8], colon cancer[9,10] and susceptibility to pathogens[11,12], the ability to manipulate the composition and physiology of this ecosystem through non-invasive routes like diet or prebiotics is attractive to promote or restore health. For such interventions to be applied, the rules governing diet-microbiota interactions must first be elucidated in mechanistic detail.

Complex carbohydrates may be the most abundant class of nutrients flowing into the colonic ecosystem, but the precise identities and amounts of these molecules change from meal-to-meal and wane in between feedings. Not surprisingly, competition for glycans in the densely populated colon has driven some gut bacteria to evolve complex systems to sense and scavenge available forms of these nutrients[13]. Individual members of the Bacteroidetes, one of the most abundant bacterial taxa in the human gut[14], encode numerous multi-protein complexes with machinery to bind, degrade and import glycans[15]. These complexes are termed Sus-like systems after the prototypic starch utilization system (Sus) expressed by *Bacteroides thetaiotaomicron* (*Bt*), and the gene clusters encoding them are known as polysaccharide utilization loci (PULs). Sus-like systems have so far been found in all sequenced gut Bacteroidetes, constituting as much as 20% of the genome[16,17]. The *Bt* Sus was the first such system described[18] and has become a model for studying glycan acquisition by the many homologous Sus-like systems that degrade a wide variety of chemically diverse dietary and host-derived glycans[4,19-21].

A major unresolved aspect of the function of Sus-like systems is how the component proteins function during various stages of the catabolic process. Sus is required for *Bt* to utilize various forms of starch, a highly abundant component of the human diet composed solely of α−1,4 and α−1,6 linked glucose but exhibits tremendous variability in secondary/tertiary structure based on the relative positions of these linkages. The *Bt* Sus is able to target different forms of starch from several different plant sources[22], demonstrating that it can accommodate this inherent structural diversity. Moreover, when *Bt* is first exposed to high molecular weight starch, transcription of *sus* is rapidly activated and reaches maximum levels within just 5 minutes[23]. This system remains highly responsive to starch at concentrations as low as 0.01 mg/ml revealing exquisite sensitivity to low substrate levels.

We have previously solved x-ray crystallographic structures of the four peripheral Sus outer-membrane proteins (OMPs)[24-26]. These studies enigmatically revealed that there are a total of eight distinct starch-binding sites spread across these four OMPs, in addition to a single catalytic site in the enzyme SusG (**Figure 3.1)**. SusD, a conserved component of all Sus-like systems, contains a single binding site and deletion of *susD* eliminates *Bt* growth on starches longer than 5 glucose units[24]. In addition to its catalytic site, the amylase SusG contains two non-enzymatic starch-binding sites[25]. A SusG surface-binding site is contained within the catalytic domain and is distinct from the active site. The second non-catalytic binding site in SusG is contained within a separate carbohydrate-binding module (CBM) domain, with the binding surface oriented ~180° away from the catalytic and surface binding sites. *In vitro* the SusG CBM domain decreases enzymatic activity on soluble starch, but enhances degradation of insoluble starch[25]. The final two Sus OMPs, SusE and SusF, are multi-domain binding proteins with SusE containing two tandem CBMs and SusF containing three[26]. These two proteins are dispensable

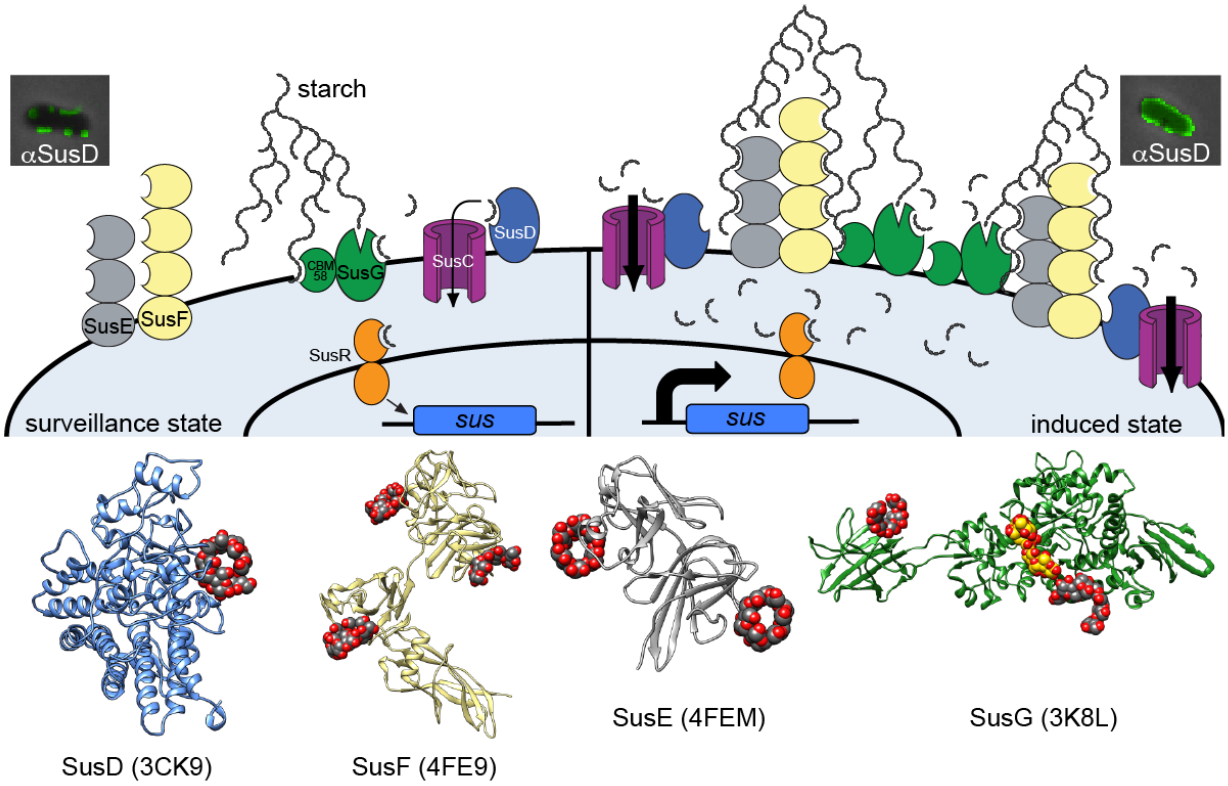



**FIGURE 3.1. Model of the *Bt* Starch utilization system (Sus) incorporating findings in this study**.

Four outer-membrane starch-binding lipoproteins SusD (blue), SusE (gray), SusF (yellow) and SusG (green) contribute to starch binding, along with a TonB-dependent transporter SusC (purple). Ribbon diagrams of the crystal structures of SusD, E, F, and G are displayed at the bottom to highlight the position and number of binding sites in each protein (note that an N-terminal domain that is uninvolved in starch binding is missing from the SusE structure). Ligands bound to each of the eight binding sites are shown in red/gray space fill; a ligand in the catalytic site of SusG is shown in red/yellow space fill. The SusG CBM58 and surf binding sites are labeled. All corresponding binding sites are schematized as open crescents in the cartoon and the SusG catalytic site as an open "V". In the uninduced “surveillance state” there is a low level of Sus machinery on the cell surface (SusD levels by antibody staining are shown in micrograph at top left). When starch is introduced SusG cleaves it to release maltoOS. The SusD binding activity is critical for binding and import of this maltoOS signal, which is then sensed by SusR, leading to increased expression of the *sus* locus (all seven functional genes are schematized as a blue box), initiating the “induced state”. During the induced state, the cell surface is flooded with Sus machinery (SusD levels by antibody staining shown at upper right) and the SusE, F and G binding sites become cooperatively important for binding starch molecules that have penetrated the *Bt* capsule (not shown in illustration). In this phase SusD binding activity is no longer essential (shown by lack of ligand occupancy in SusD) but another SusD function is required based on site-directed mutagenesis.

for growth and, despite a high level of structural similarity between the SusE and SusF CBMs, display distinct substrate preferences and affinities at individual binding sites[26]. The differences in binding site specificity and/or preference, along with variable phenotypes when genetically eliminated, suggest that these sites may play different roles in starch metabolism. On the other hand, the general binding preferences of all sites for unbranched α–1,4 maltoOS could suggest that the functions of these eight sites are largely redundant.

In this study we systematically explore how these eight different Sus OMP binding sites contribute to *Bt* starch metabolism. We show that the binding site contained in SusD is critical for *Bt* to sense available starch and respond transcriptionally and that this protein has an additional function that is separable from its binding site. In contrast, we show that SusE and SusF, which are alone dispensable for starch utilization, are required in combination with either of the SusG binding sites in a starch substrate-dependent fashion. Most dramatically, a mutant lacking SusE, F and the SusG surface site is completely unable to grow on high molecular weight corn starch, and this and other phenotypes can be compensated by loss of this symbiont's surface polysaccharide capsule. Thus, it appears that one role of these carbohydrate-binding proteins is to assist with capture of external carbohydrates in spite of the diffusion barrier that imposed by these species' ubiquitous protective capsule[27]. Our results provide an important new layer of mechanistic insight into the function of this family of glycan acquisition systems that is abundantly represented in the human microbiome. By demonstrating separable functions for the apparently redundant Sus OMPs, we provide insight into how gut commensal bacteria have evolved to become more competitive for nutrients in their densely populated habitat.

**Results**

***The SusD binding site is required for growth on large starch molecules***

Homologues of SusD can be identified in most if not all sequenced gut Bacteroidetes[16,28], so we reasoned that these proteins must play a critical role(s). Based on structural and biochemical analysis, SusD contains a single starch binding site and has no enzymatic activity, but when deleted is required for growth on starch molecules longer than 5 glucose units (**Figure 3.2**)[24]. To address whether the binding activity of SusD *per se* is required for starch growth, we constructed a strain carrying a mutant allele of *susD* where three critical binding residues (W98, N101, W320)[24] were mutated to alanine, abolishing all measurable affinity for starch (**Figure 3.3A**). Identical to the *susD* deletion strain (Δ*susD*), the binding deficient SusD mutant (SusD*) is unable to grow on amylopectin (AP) from either maize or potato (**Figure 3.4A, Figure 3.3B**), despite being trafficked to the cell surface similarly to native SusD (**Figure 3.3C**). Thus, we conclude that the binding capacity of SusD is directly required for growth on very large starch molecules.

Interestingly, we found that SusD* growth could be restored if a small amount of maltose, which alone is too low to support substantial growth (**Figure 3.3B**), is added to high molecular weight starch cultures (**Figure 3.4B**). This observation was initially made with commercially available starches, some of which contained minor amounts of contaminating maltoOS (**Table 3.1**), which we hypothesized could bypass the SusD* defect. Consistent with this idea, when these starches are dialyzed to remove contaminants, the SusD* mutant exhibited a complete defect **(Figure 3.3B**). We were also surprised to observe that addition of a small amount of maltose was insufficient to restore growth to the Δ*susD* mutant, indicating that the binding site of SusD and the presence of the protein itself (irrespective of binding capacity) are likely to play separable roles (**Figure 3.4B**). This difference in response to maltose is unlikely to

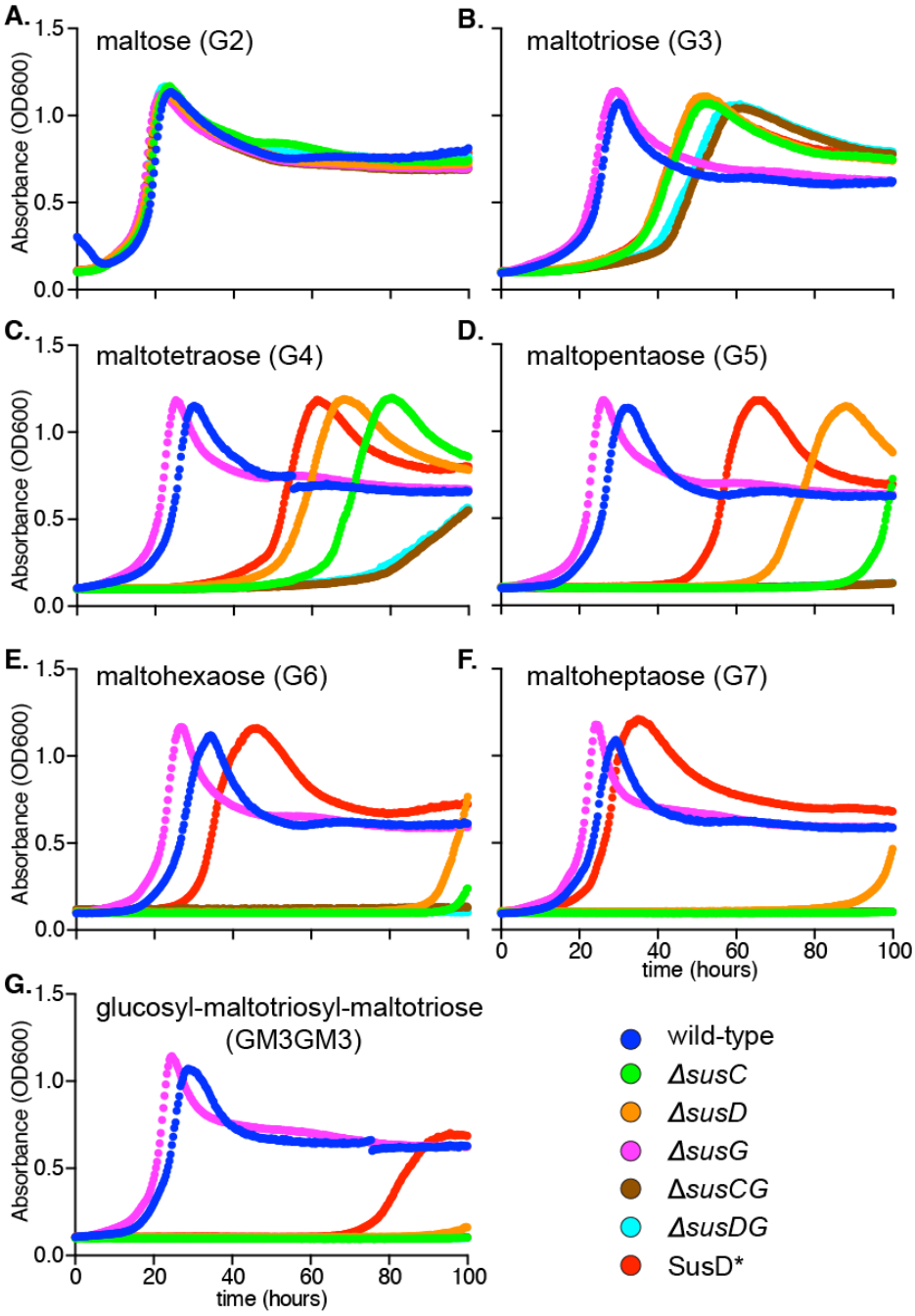
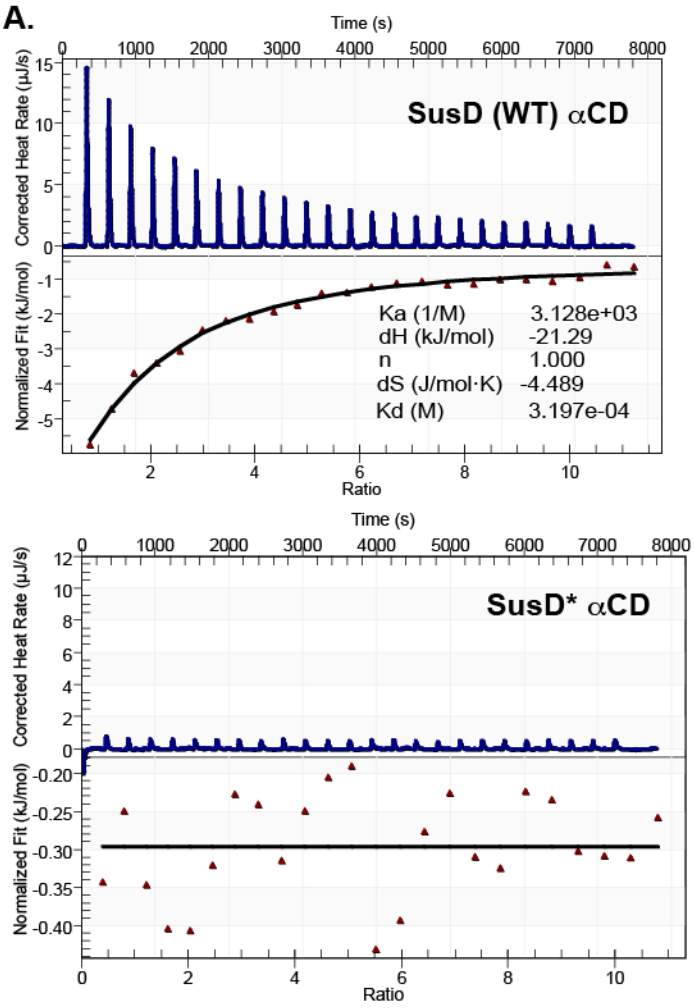
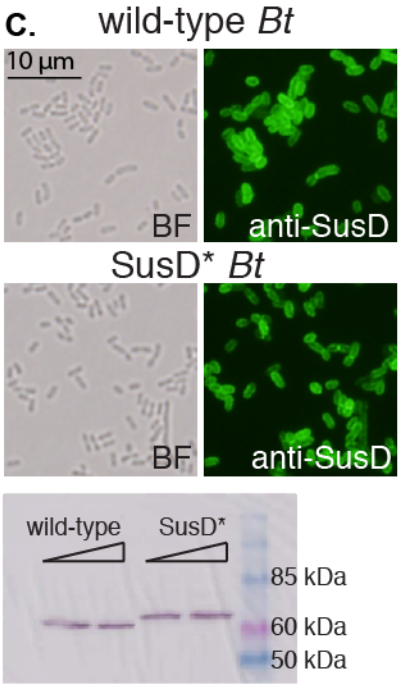



**FIGURE 3.2. SusC, D and G mutant growth curves on maltoOS**

Representative growth curves of various SusC, SusD and/or SusG *Bt* mutants on minimal media with 5mg/mL **A.** maltose **B.** maltotriose **C.** maltotetraose **D.** maltopentaose **E.**maltohexaose **F.** maltoheptaose **G.** glucosyl-maltotriosyl-maltotriose (GM3GM3), a branched glucose heptamer with two α-1,6 linkages. As observed previously with Δ*susD Bt*[24], after 100 hours we began to see the emergence of suppressor mutants that, despite lacking required Sus components, had regained the ability to grow on longer substrates. Thus, we only analyzed growth up to 100 hours.


A.
Time (s)
SusD (WT) αCD
Corrected Heat Rate (µJ/s)
Normalized Fit (kJ/mol)
Ka (1/M) 3.128e+03
dH (kJ/mol) -21.29
n 1.000
dS (J/mol·K) -4.489
Kd (M) 3.197e-04
Ratio
SusD* αCD


C.
wild-type Bt
10 µm
BF
anti-SusD
SusD* Bt
wild-type
SusD*
85 kDa
60 kDa
50 kDa

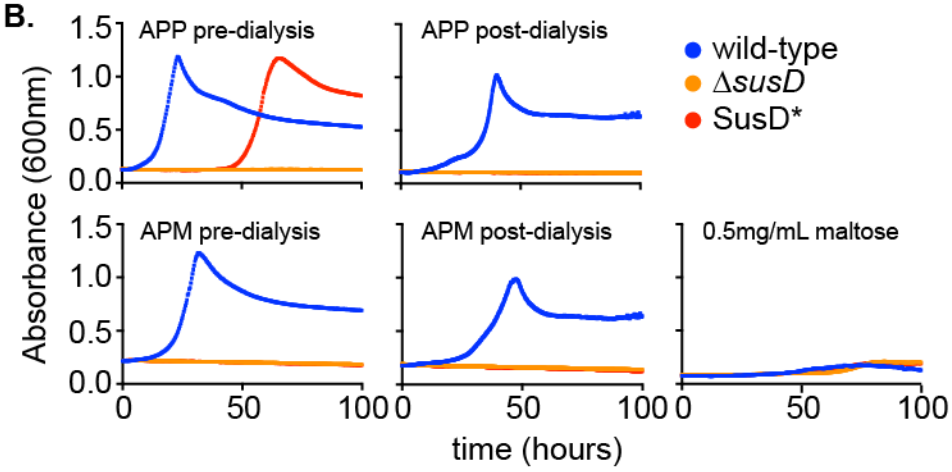

B.
APP pre-dialysis
APP post-dialysis
APM pre-dialysis
APM post-dialysis
0.5mg/mL maltose
wild-type
ΔsusD
SusD*
Absorbance (600nm)
time (hours)

**FIGURE 3.3. The SusD* protein is unable to bind starch, stable on the *Bt* surface, and leads to a phenotype distinct from wild-type or Δ*susD Bt***

**A.** Isothermal titration calorimetry was performed with α-cyclodextrin (αCD) to verify that the binding mutant SusD* fails to bind starch. Top panel: SusD wild-type protein (positive control, for reference) Bottom panel: SusD*. The SusD wild-type data were fit to an independent one-site binding model, fixing n=1 due to the known stoichiometry of binding. The data for SusD* and was fit to a blank constant (no binding) model. **B.** Representative growth curves with 5mg/mL AP-potato (APP) or AP-maize (APM) either pre- or post-dialysis or 0.5mg/mL maltose as the sole carbon source **C.** Wild-type and SusD* *Bt* were stained (green) using anti-SusD antibodies. Similar levels of surface SusD expression were observed. Anti-SusD western blots were also performed on whole cell lysates from wild-type and SusD* cells, again similar SusD levels were observed.

**TABLE 3.1. Maltooligosaccharide concentration in AP- pre and post dialysis detected by high pH anion exchange chromatography**
**(ng maltoOS per µg total carbohydrate)**

| Starch | G1 | G2 | G3 | G4 | G5 | G6 | G7 | Total |
|---|---|---|---|---|---|---|---|---|
| AP-maize (pre dialysis) | .392 | ND | ND | ND | ND | ND | ND | **0.392** |
| AP-potato (pre dialysis) | .228 | .17 | .286 | .676 | ND | ND | ND | **1.36** |
| AP-maize (post dialysis) | .174 | .17 | ND | ND | ND | ND | ND | **0.344** |
| AP-potato (post dialysis) | .282 | .12 | ND | ND | ND | ND | ND | **0.402** |

ND = not detected

be due to effects of the Δ*susD* in-frame deletion on up- or downstream *sus* genes since we have previously shown that the remaining genes are expressed to wild-type levels in maltose[24].

Quantification of two aspects of *Bt* growth revealed that the exponential doubling time of SusD* in AP-maize + maltose is indistinguishable from wild-type (**Figure 3.4C**), but the lag phase of this strain was significantly lengthened (**Figure 3.4D**). Based on this, we hypothesized that SusD binding is important for efficient import of maltoOS − even as small as maltose − which in turn act as the signal to induce *sus* transcription. In the absence of normal SusD function, these signals must still enter the cell because both *susD* mutants, as well as a *susC* transporter mutant, are able to grow normally on maltose alone (**Figure 3.2A**). To test if SusD binding was important for growth on a particular size range of maltoOS we tested growth of the two SusD mutants on a panel of maltoOS ranging from two to seven glucose units (**Figure 3.4E**). We observed that without the entire SusD protein (Δ*susD*) lag time continued to increase with substrate size. In contrast, the lag time of SusD* sequentially increased on maltoOS up to 5 units long, and then began to decrease on longer chains. Based on this, we conclude that SusD binding is optimized for utilization of mid range maltoOS (G3-G5) and less important on shorter or longer maltoOS. Interestingly, SusD binding also plays a critical role during growth on the double-branched oligosaccharide glucosyl-maltotriosyl-maltotriose (GM3GM3), which is likely to be a product of enzymatic processing of α1,6 branch regions, and this defect is absent in the isomeric linear maltoheptaose (**Figure 3.2F,G**). Although it remains to be tested, our observations suggest that other components of the Sus machinery are able to compensate for lack of SusD binding in the presence of longer, linear substrates in a fashion that is dependent on the presence of SusD protein but not its binding site.

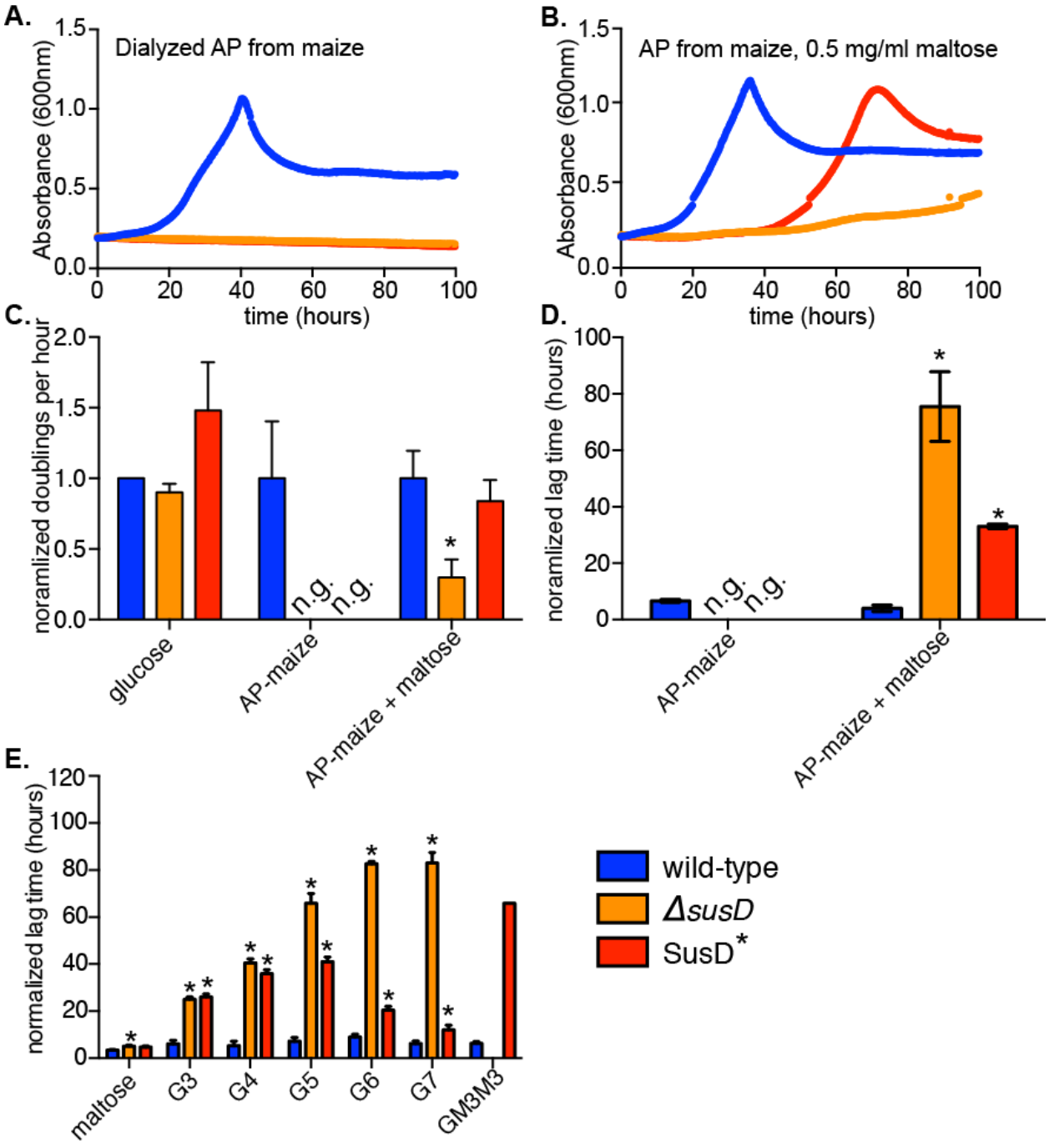

A.
Dialyzed AP from maize
Absorbance (600nm)
time (hours)
B.
AP from maize, 0.5 mg/ml maltose
Absorbance (600nm)
time (hours)
C.
noramlized doublings per hour
n.g.
n.g.
glucose
AP-maize
AP-maize + maltose
D.
noramlized lag time (hours)
n.g.
n.g.
AP-maize
AP-maize + maltose
E.
normalized lag time (hours)
maltose
G3
G4
G5
G6
G7
GM3M3
wild-type
ΔsusD
SusD*

**FIGURE 3.4. Requirement for SusD binding for growth on large starches is overcome with low levels of maltose.**

Representative growth curves of wild-type, Δ*susD* and SusD* *Bt* with **A.** 5 mg/mL dialyzed AP-maize or **B.** 5 mg/mL dialyzed AP-maize + 0.5 mg/mL maltose as the sole carbon source. **C.** The exponential rates from three replicate growth curves (including those shown in A. and B.) were measured and normalized to glucose rate then wild-type rate (n.g. denotes no growth observed up to 100 hours). **D.** Lag times (time for absorbance to reach or surpass 0.35) were measured for three biological replicates and normalized to the corresponding glucose growth for each replicate. **E.** Normalized lag times for *Bt* grown on linear maltoOS ranging from two glucose units (maltose) to seven glucose units (G7) or GM3M3, a branched glucose heptamer containing two α-1,6 linkages. In all panels, error bars indicate standard error across three replicates. Statistically significant differences from wild-type (*$p < 0.05$) were determined using a one-tailed unpaired student's t test. Note that statistics could not performed on the SusD* GM3M3 growth as only one replicate reached sufficient growth density.

***SusD binding enhances transcriptional sensitivity to starch***

In the above experiments, we reasoned that increased lag time equates to decreased ability to sense or process available substrates. To directly test this idea and determine if the SusD binding site is important for efficient sensing of maltoOS, we examined the transcriptional responses of SusD* to concentrations of starch or maltoOS spanning several orders of magnitude. Naïve *Bt* cells, grown to mid exponential phase in glucose (a condition that does not activate *sus* expression), were washed anaerobically and introduced into medium containing maltotriose (G3), maltoheptaose (G7) or AP-maize in concentrations ranging from 250ng/mL to 2.5mg/mL. Samples were collected 30 minutes after the media change, when *sus* transcription reaches its maximum[23], and *susC* message levels were quantified (**Figure 3.5**).

Consistent with our hypothesis that SusD enhances sensory responses to starch and maltoOS during the surveillance-to-active degradation transition (**Figure 3.1** inset micrographs), *sus* expression in response to both oligosaccharides and starch was significantly attenuated in the SusD* mutant compared to wild-type, sometimes requiring ~$10^3$ higher substrate concentration to elicit similar responses as wild-type (**Figure 3.5**). Even at the highest tested concentration of 2.5mg/mL, the SusD* mutant showed lowered levels of *sus* expression on G3, which, although not statistically significant, could explain the increased lag time compared to wild-type when the mutant is grown on this substrate.

In our growth experiments we observed that while SusD* had an increased lag on G3, it did not on G7. Indeed, we observed that the SusD* mutant had a more robust transcriptional response to G7 than to G3, though it was still significantly attenuated compared to wild-type (**Figure 3.5B**). In SusD*, *sus* expression was ~10-fold higher at 250μg/mL G7 than at the same concentration of G3. Additionally, at the highest tested concentration, the SusD* mutant

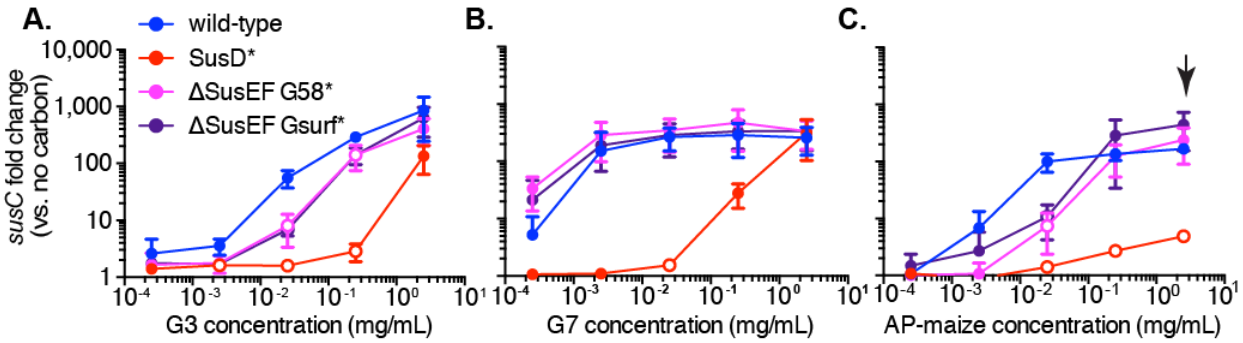



**FIGURE 3.5. SusD binding is critical for sensing starch**

Expression of *susC* transcript in wild-type and mutant *Bt* cells exposed to maltoOS or starch for 30 minutes: **A.** maltotriose (G3) **B.** maltoheptaose (G7) or **C.** AP-maize. Fold-change was calculated relative to glucose-grown, washed cells used to inoculate starch/maltoOS cultures. The average and standard deviation of three individual replicates is shown. P values were calculated using a one-tailed unpaired student's t test and statistically significant differences compared to wild-type ($p < 0.05$) are shown as open circles. Arrow in C. highlights an AP-maize concentration where ΔSusEF Gsurf* displays no growth despite having *sus* transcript levels equal to wild-type.

exhibited *sus* levels that were indistinguishable from wild-type, which likely explains the lack of defect observed when SusD* is grown on a comparable (5mg/mL) G7 concentration. These data demonstrate that SusD binding is critical for *Bt* to mount an optimal transcriptional response to available maltoOS and allows *Bt* to sense nutrients at concentrations several orders of magnitude lower than without SusD function. As expected, given the lack of growth of SusD* on AP-maize and this strain's defect in sensing small maltoOS, this mutant exhibited a severe defect in *sus* expression in response to AP-maize (**Figure 3.5C**). Since SusD homologs are a defining feature of Bacteroidetes Sus-like systems that target dozens of other glycans[15], this feature may be a fundamental and conserved aspect of this molecular mechanism.

***The SusE, F, and G binding sites work together to enhance Bt growth rate in a substrate dependent manner***

SusD represents just one of the eight starch binding sites contained within the Sus outer-membrane proteins. Across SusE, F and G there are seven additional non-catalytic starch binding sites (**Figure 3.1**). To investigate the role of these additional binding sites we created a series of *Bt* mutants lacking one or multiple binding sites in SusE, F, and G. To abolish surface expression of SusE and SusF we used a strain containing two mutant alleles (ΔSusEF), in which the N-terminal cysteine residue of each protein was mutated to prevent lipidation and trafficking to the outer-membrane[26]. To abrogate binding activity of the SusG CBM58, three critical binding residues (W287, W299, N330) were mutated to alanine (SusG58*), which abolished binding ability of the CBM (**Figure 3.6**). Similarly, we mutated three critical residues of the SusG surface site (W460, Y469, D473) to alanine to create the SusGsurf* binding deficient allele. Note that we could not directly verify that this mutation eliminated binding at this site because

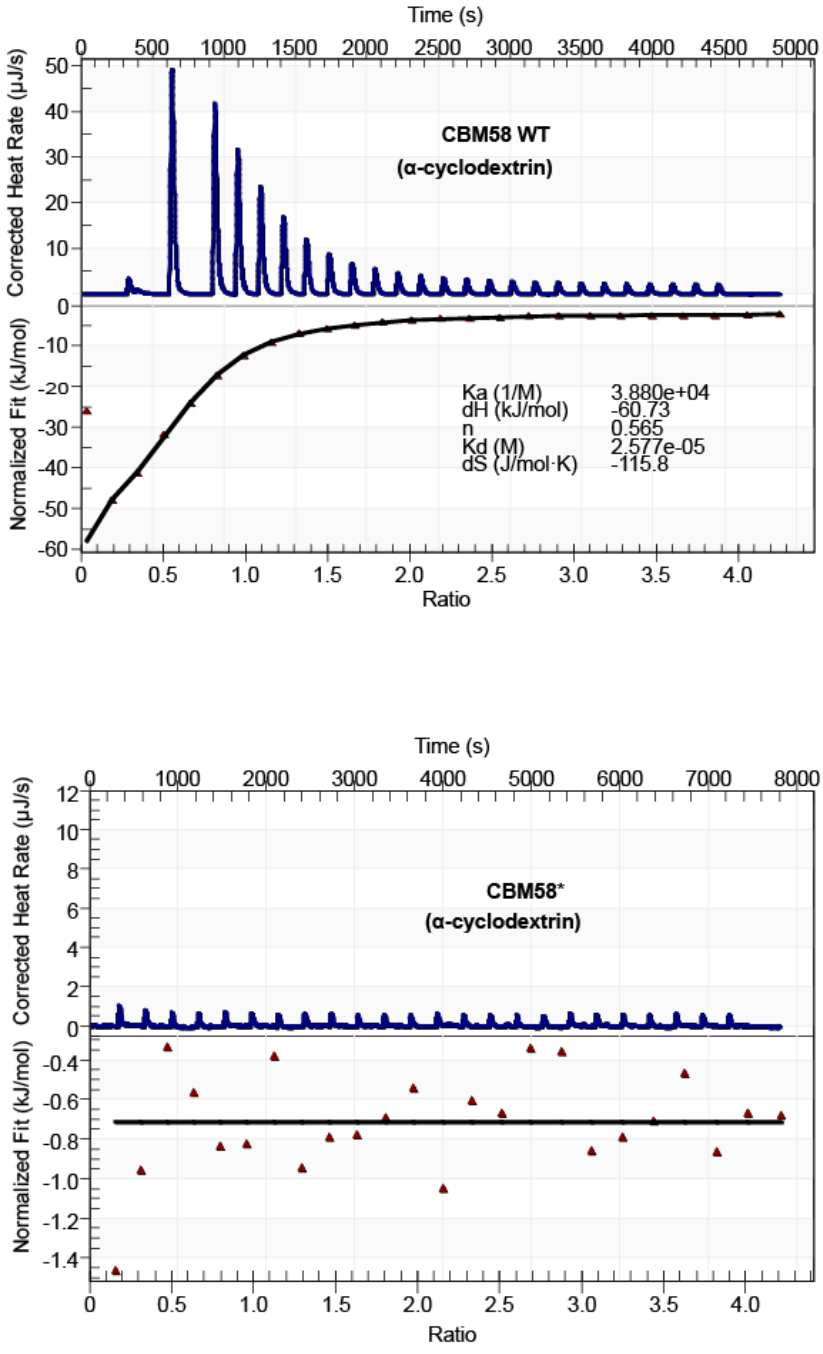



**FIGURE 3.6. SusG CBM58* is deficient in binding**

Isothermal titration calorimetry was performed with α-cyclodextrin to verify that the binding mutant SusG-CBM58* fails to bind starch. Top panel: wild-type CBM58, Bottom Panel: CBM58*. The wild-type CBM58 data were fit to an independent one-site binding model, fixing n=1 due to the known stoichiometry of binding. The data for CBM58* were fit to a blank constant (no binding) model.

the catalytic site is still present and, if mutated to become non-catalytic would itself become a binding site. However, previous *in vitro* studies with SusG revealed that this mutation decreases both binding and degradation of insoluble corn starch[25] and, as discussed below, we show that loss of this surface binding site reduces growth in intact bacteria.

Growth experiments with dialyzed starches as the sole carbon source revealed that while ΔSusEF and SusG58* mutants did not demonstrate any growth attenuation compared to wild-type, loss of all six Sus CBMs in ΔSusEF G58* resulted in a decreased exponential growth rate (doublings per hour) on both substrates (**Figure 3.7A**). Due to variability between replicates this was not statistically significant however, ΔSusEF G58* did display a significantly increased lag time on AP-maize compared to wild-type (**Figure 3.7B**).

Loss of the SusG surface site alone (SusGsurf*) resulted in a significant growth defect (**Figure 3.7C,D**). In fact, this mutation had greater impact on growth than loss of the other six CBM binding sites, showing a significant decrease in exponential rate on AP-maize and a significantly longer lag on both substrates. Interestingly, the SusGsurf* defect was further exacerbated – resulting in complete loss of growth on AP-maize – by the loss of SusE and SusF. This suggests that the SusG surface site is the most critical of the seven SusE, F, G binding sites for growth on high molecular weight starch, but works cooperatively with the sites contained in SusE, F. Control staining of whole cells with antisera specific for SusG suggested that the observed defects were not associated with decreased trafficking of the mutant proteins to the cell surface (**Figure 3.8**). Representative growth curves from the experiments described above are displayed in **Figure 3.9**.

Since our results above suggest that SusD has a critical role(s) independent of its binding function (**Figure 3.4B**), we tested if the binding sites in SusE and SusF are solely responsible for

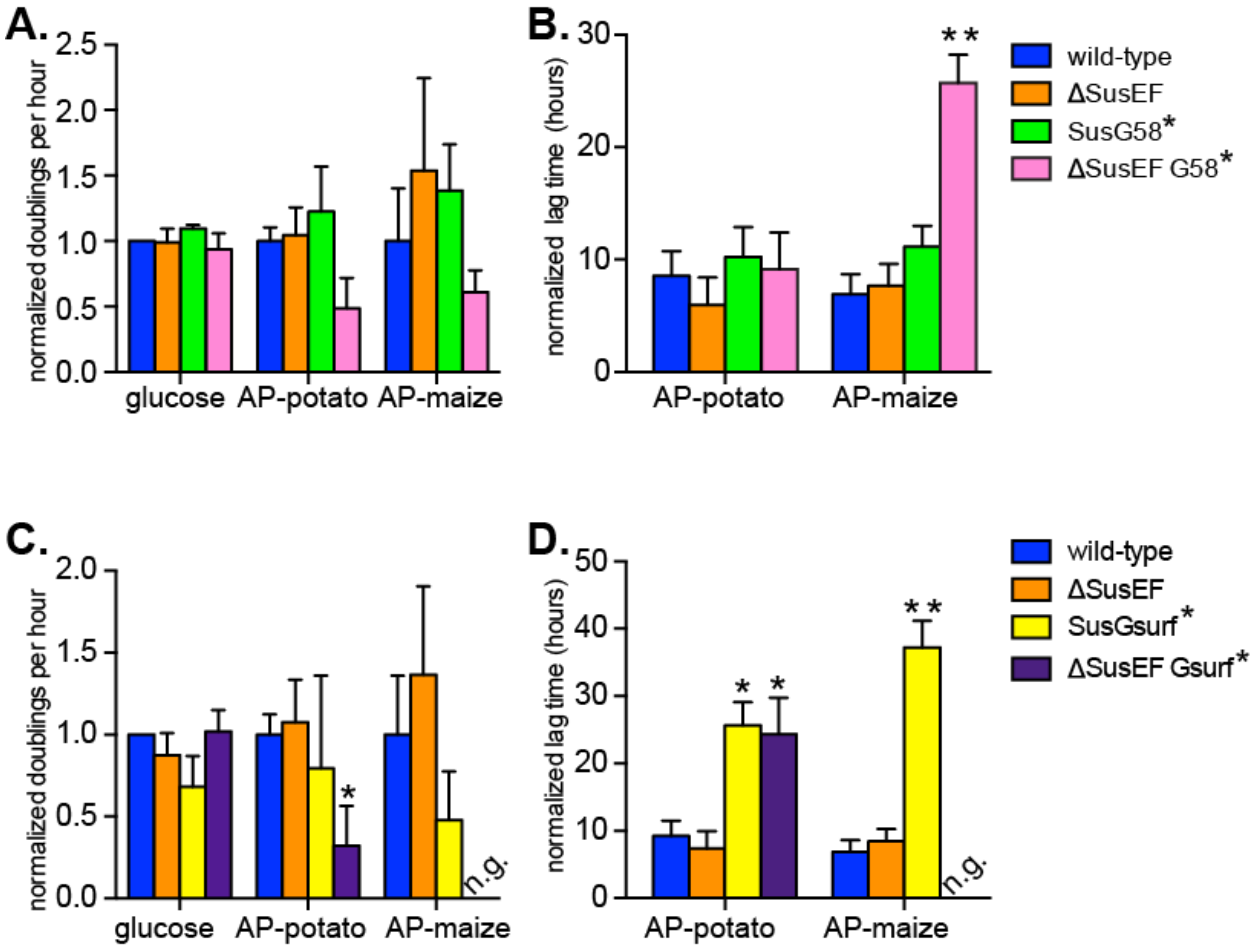


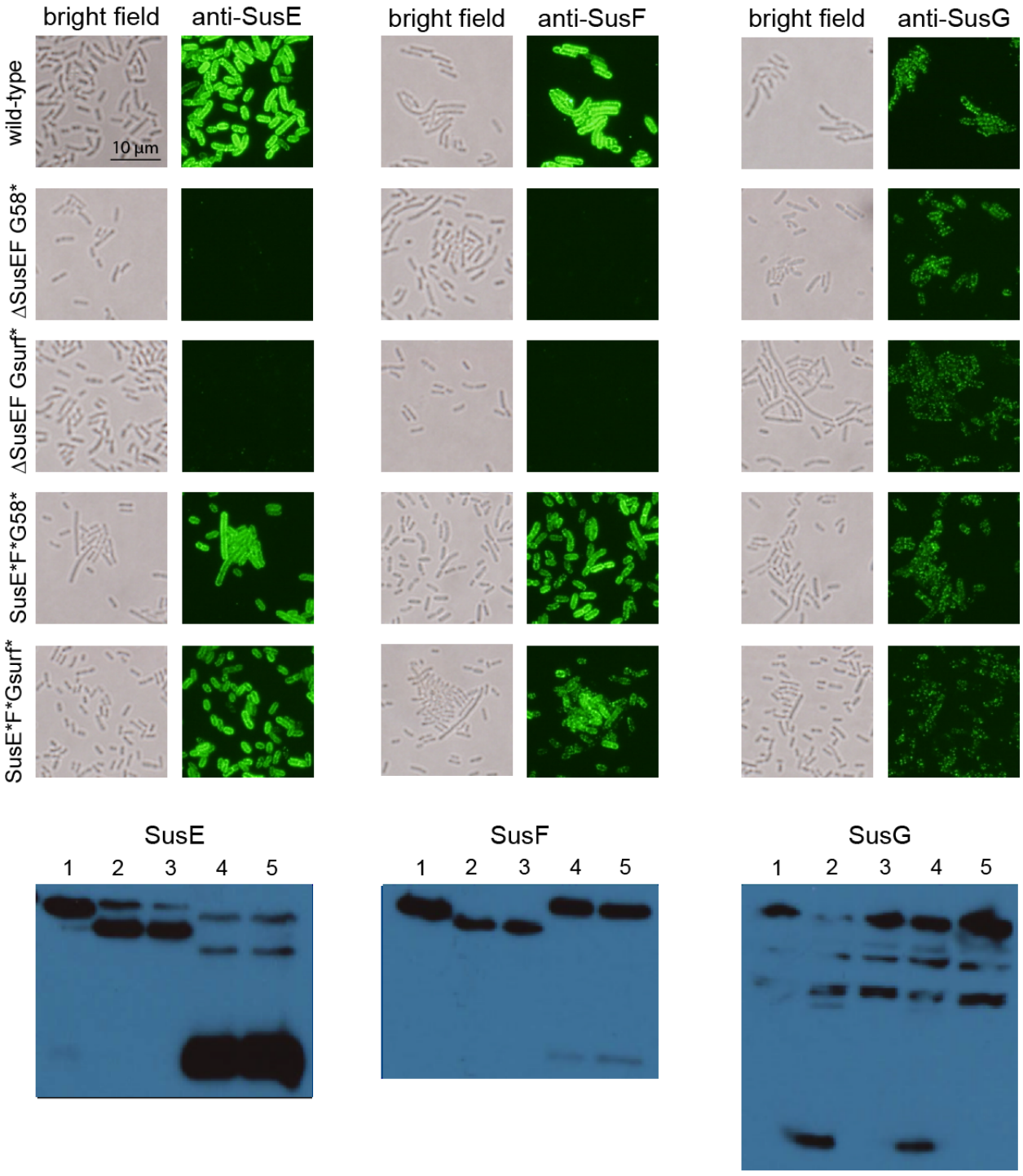

**FIGURE 3.7. SusEFG binding sites enhance *Bt* starch growth via overlapping roles**

Growth assays were performed with wild-type and mutant *Bt* strains on the sole carbon sources denoted. Normalized doublings per hour and normalized lag times were calculated for strains lacking one or more Sus CBMs (**A., B.**) or strains lacking SusE, SusF and/or the SusG surface site (**C., D.**) as described in the legend for Figure 3.4. Averages and standard errors across three replicates are shown. Statistically significant differences vs wild-type were calculated using a one-tailed unpaired student's t test (*$p < 0.05$; **$p < 0.01$).

their contribution to starch growth, or if restoring the presence of binding deficient proteins to the cell surface would recover some of the growth defect. To this end we created binding deficient alleles of SusE and SusF using site directed mutagenesis to target critical binding residues in each of the SusE and SusF CBMs. The resulting alleles, *susE** and *susF**, encode proteins that show no measurable binding to starch[26], but are also expressed on the surface of the cell to levels similar as wild-type (**Figure 3.8**). The resulting mutant, SusE*F*G58*, displayed a growth profile nearly identical to that of ΔSusEF G58* (**Figure 3.10**), suggesting that – unlike for SusD – it is loss of the SusE and SusF binding functions that contribute to their function. This conclusion is further corroborated by growth of a SusE*F*Gsurf*, which was nearly identical to the ΔSusEF Gsurf* strain, including the most severe loss of growth phenotype on AP-maize (**Figure 3.10B**).

***SusE, F, G binding sites are important for growth on high molecular weight starch***

Interestingly, loss of the SusE, F, G binding sites generally resulted in more severe defects on AP-maize than on AP-potato (**Figure 3.7**). Starch structure can differ significantly between plant sources, so we hypothesized that the SusE, F, G binding sites were required for specific structural aspects of starch that are more prevalent in AP-maize, which tends to have a higher degree of α-1,6 branching[29] and a higher molecular weight[30] relative to AP-potato. To investigate which of these structural aspects required the SusE, F, G binding sites, we performed growths on a series of enzyme-treated substrates in which the molecular weight and branching density were reciprocally varied. Waxy cornstarch (WCS), a high amylopectin starch similar to the AP-maize previously used, was treated with β-amylase (BA) and/or branching enzyme (BE). Both of these enzymes modify starch to increase the number of α-1,6 linkages while decreasing




1 wild-type
2 ΔSusEF G58*
3 ΔSusEF Gsurf*
4 SusE*F*G58*
5 SusE*F*Gsurf*

**FIGURE 3.8. Mutated SusE, SusF and SusG alleles are appropriately expressed in *Bt* cells**

(Upper panels) Wild-type, ΔSusEF G58*, ΔSusEF Gsurf*, SusE*F*G58* and SusE*F*Gsurf* *Bt* cells were fixed and stained for SusE, SusF or SusG using the appropriate antibodies. Bright-field and corresponding fluorescent images are shown side by side. Consistent with mutation of their lipidation site, the ΔSusE and ΔSusF alleles were not detected on the cell surface but SusE*, SusF*, Sus G58* and SusGsurf* alleles were detected on the *Bt* cell surface at levels similar to wild-type. (Lower panels) Whole cell lysates from strains listed above were collected and probed for expression of SusE, SusF and SusG using western blotting (see numbered key below blots). Consistent with mutation of their lipidation site only the ΔSusE and ΔSusF proteins are still expressed, but their molecular weight is slightly lower than the wild-type proteins, consistent with loss of the lipid tail. Although degradation of some of the proteins is observed (most notably SusE*, SusG58* and SusGsurf*) we believe this can be tolerated as levels similar to wild-type are observed in surface staining of these proteins.

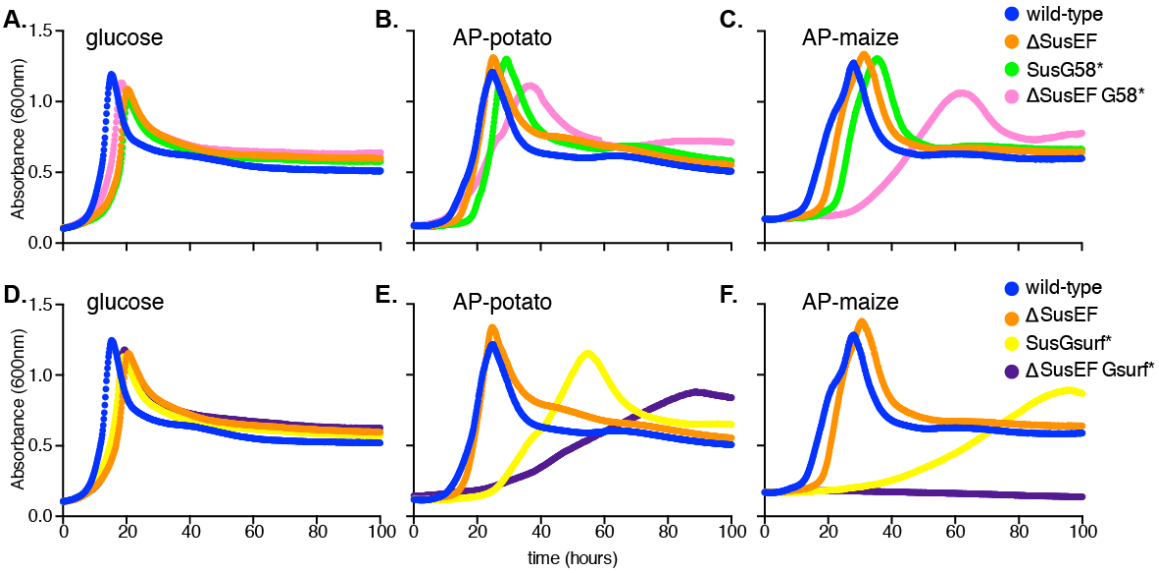



**FIGURE 3.9. Growth curves of SusE, F and G mutants on starch**

Growth assays were performed with *Bt* strains grown on minimal media with the sole carbon sources denoted. We only observed growth up to 100 hours to exclude appearance of suppressor mutants as described above.

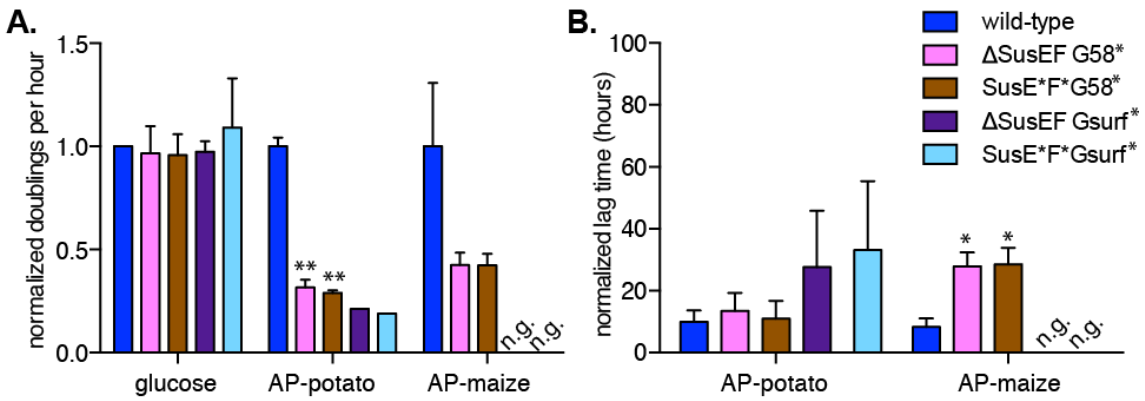



**FIGURE 3.10. The binding ability of SusE and SusF is solely responsible for their contribution to *Bt* starch growth**

Growth assays were performed with *Bt* strains either not expressing SusE and SusF (ΔSusEF G58* and ΔSusEF Gsurf*) or expressing binding deficient SusE and SusF (SusE*F*G58* and SusE*F*Gsurf*) on the sole carbon sources denoted. Normalized doublings per hour (**A.**) and normalized lag times (**B.**) were calculated as described in Figure 3.4. Averages and standard errors across three replicates are shown. Statistically significant differences vs. wild-type were calculated using a one-tailed unpaired student's t test (*$p < 0.05$; **$p < 0.01$).

the average molecular weight[31] (**Figure 3.11A**). Growth rate of the ΔSusEF G58* mutant improved and lag time was significantly decreased as molecular weight decreased by ~$10^3$, but branch density nearly doubled (**Figure 3.11B**). Thus, we conclude that this constellation of binding sites is more important for adapting the cell to higher molecular weight substrates as opposed to those with more branches. As seen with AP-maize, the ΔSusEF Gsurf* strain had a severe defect on the WCS substrate, in fact in only one of three replicates was any growth observed. Growth of this strain improved on both enzyme treated starch preparations, which is seen most notably as a decrease in lag time. Again we conclude that because the ΔSusEF Gsurf* strain is more adept at growth on lower molecular weight, yet more highly branched starches, that this collection of binding sites is suited to aiding growth on high molecular weight starches.

***The combined SusE, F, and G binding sites play little role in starch sensing***

Since we determined that the SusD binding site is primarily involved in sensing available maltoOS and is not required for growth rate on starch when this blockade is bypassed with maltose (**Figure 3.4B**), we sought to determine if the binding sites in SusE, F and G play a similar role or if they confer a mostly separate downstream role. To test this, we performed the same starch, G3 and G7 exposure experiments done previously with SusD* with the ΔSusEF G58* and ΔSusEF Gsurf* strains. Compared to susD*, both strains were substantially more responsive to limited concentrations of maltoOS, and both mutants only showed significant defects relative to wild-type on G3, albeit this defect was much less severe than SusD* (**Figure 3.5A,B**). These data suggest that the SusE, F, G binding sites play a far less critical role, relative to SusD, in enhancing bacterial ability to sense and respond to maltoOS. Because substantial growth defects were observed for the ΔSusEF G58* and ΔSusEF Gsurf* mutants on larger

**A.**

| | % α-1,6 | MW (Da) |
|---|---|---|
| WCS | 5.3 | $2x10^8$ |
| BE WCS | 7.1 | $3x10^5$ |
| BEBA WCS | 12.9 | $2x10^5$ |

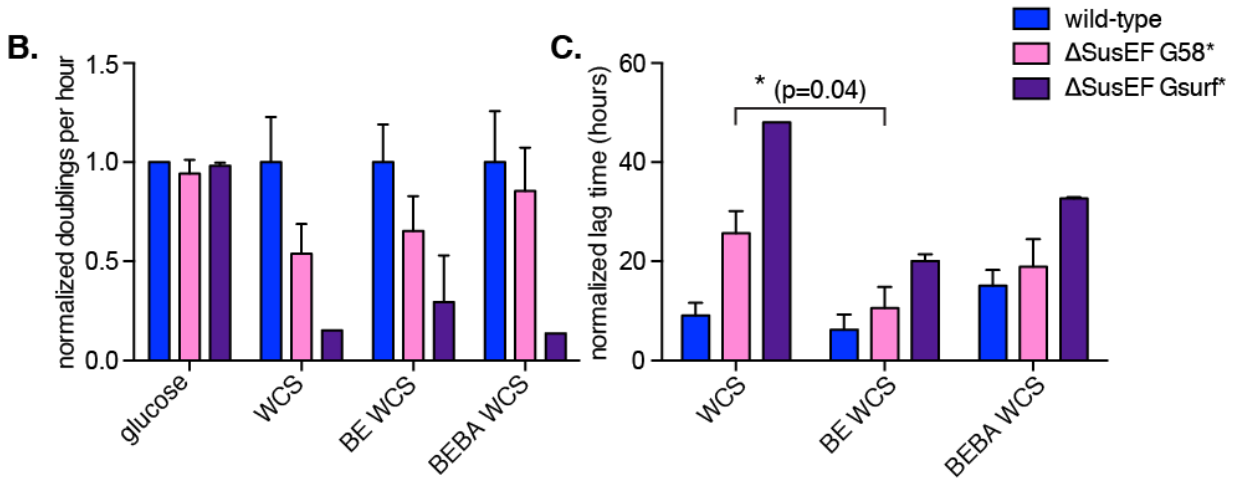



**FIGURE 3.11. The SusEFG binding sites enhance *Bt* growth on high molecular weight starch**

Growth assays were performed with indicated *Bt* strains on untreated waxy cornstarch (WCS), branching enzyme treated WCS (BE WCS) or branching enzyme + β-amylase treated WCS (BEBA WCS). Total α-1,6 linkage percentage and average molecular weight is shown for each substrate (**A.**) Normalized doublings per hour (**B.**) and normalized lag times (**C.**) were calculated as described in Figure 3.4. Averages and standard errors across three replicates are shown. Statistically significant differences vs. wild-type were calculated using a two-tailed unpaired student's t test. Statistically significant differences were calculated using a one-tailed unpaired student's t test.

starches, we also tested the transcriptional response of these mutants to dialyzed AP-maize. At the highest AP-maize concentration, wild-type *sus* transcription was over 30-fold higher than that of SusD*, but both combined SusE, F, G mutants were indistinguishable from wild-type. Although, both SusE, F, G combined mutants displayed significantly lower *sus* expression at a single intermediate concentration of 250μg/mL (**Figure 3.5C**). Normal *sus* expression levels in both of the combined SusE, F, G mutants at the highest starch concentration (black arrow in **Figure 3.5C**) is particularly striking since at comparable AP-maize concentrations of 5mg/ml (2x that used here), the ΔSusEF G58* mutant had a significant growth defect and the ΔSusEF Gsurf* mutant was completely unable to grow. Thus, we conclude from these experiments that the SusE, F, G binding sites serve a function(s) that is largely distinct from that of the SusD binding site and that they optimize *Bt* starch growth (*e.g.*, by enhancing growth rate) independent of enhancing the transcriptional response.

***Loss of SusE, F, and G binding sites does not increase maltoOS release during growth on starch***

We hypothesized that multiple binding sites may serve to sequester starch breakdown products after they were degraded by SusG to prevent cross-feeding of surrounding species or enhance substrate channeling. Therefore we sought to test whether loss of the SusE, F, G binding sites increased non cell-bound maltoOS in the supernatant during growth on starch. The various *Bt* strains were grown to mid-log phase on either AP-maize or maltose as a positive control, and supernatants were collected. Glycans in the supernatants were labeled with 2-Aminobenzamide and maltoOS (defined as seven glucose units or shorter) were individually quantified using high-pH anion exchange chromatography (**Table 3.2**). Only a very low level of maltoOS release,

primarily glucose and maltose, could be detected in the supernatants of starch grown wild-type *Bt* cells. Neither the ΔSusEF G58* nor the ΔSusEF Gsurf* strain displayed increased maltoOS release and in fact had slightly decreased maltoOS levels compared to wild-type. These data do not support a role for the SusE, F, G binding sites promoting growth by virtue of their ability to sequester starch breakdown products. Rather, they highlight the exquisite efficiency of this system to scavenge catalyzed starch products, even in the absence of these functions. It is worth noting that SusD is still present in the strains tested and, given the complete loss of growth on AP-maize, we could not perform a parallel experiment using the SusD* mutant without adding maltose to the culture. Given our results connecting SusD to enhanced maltoOS sensing, it is plausible that SusD plays the prominent role in sequestering released maltoOS, even during active catalysis.

***Requirement of the SusE, F, and G binding sites is dependent on polysaccharide capsule***

Among human associated members of the phylum Bacteroidetes, the ability to produce a polysaccharide capsule is enriched specifically in gut species compared to oral isolates[27], suggesting the capsule provides a competitive advantage specifically in the intestinal environment. This capsule layer can be up to several hundred nanometers thick, homogenously covers the cell surface[32], and may represent a significant barrier for large extracellular carbohydrates to penetrate and reach the cell surface Sus machinery. We hypothesized that the multiple SusE, F, G binding sites may have evolved to offset this barrier and increase overall affinity of the *Bt* surface for starch, for example by holding on to starch chains as they are being degraded. To test whether loss of capsule would reduce the growth defects observed in the SusE, F, G binding mutants, we created ΔSusEF G58* and ΔSusEF Gsurf* strains in a *Bt* mutant that

**TABLE 3.2. Maltooligosaccharide concentration in *Bt* supernatants detected by high pH anion exchange chromatography (ng maltoOS per µl)**

| Sample | G1 | G2 | G3 | G4 | G5 | G6 | G7 | Total |
|---|---|---|---|---|---|---|---|---|
| **wild-type maltose grown** | 34.04 | 1149.36 | 0.74 | ND | ND | ND | ND | **1184.14** |
| **wild-type APM grown** | 63.104 | 37.896 | 3.884 | 1.62 | 1.932 | 1.22 | 2.32 | **111.98** |
| **ΔSusEF G58* APM grown** | 31.936 | 21.944 | 15.528 | 5.044 | 3.036 | 3.376 | 4.296 | **85.16** |
| **ΔSusEF Gsurf* APM grown** | 13.772 | 13.564 | 6.376 | 0.76 | 0.528 | 0.636 | 0.8 | **36.44** |
| **ΔSusEF G58* Δcps-all APM grown** | 26.904 | 16.468 | 3.776 | 2.8 | 0.88 | 0.88 | 1.252 | **52.96** |

ND = not detected

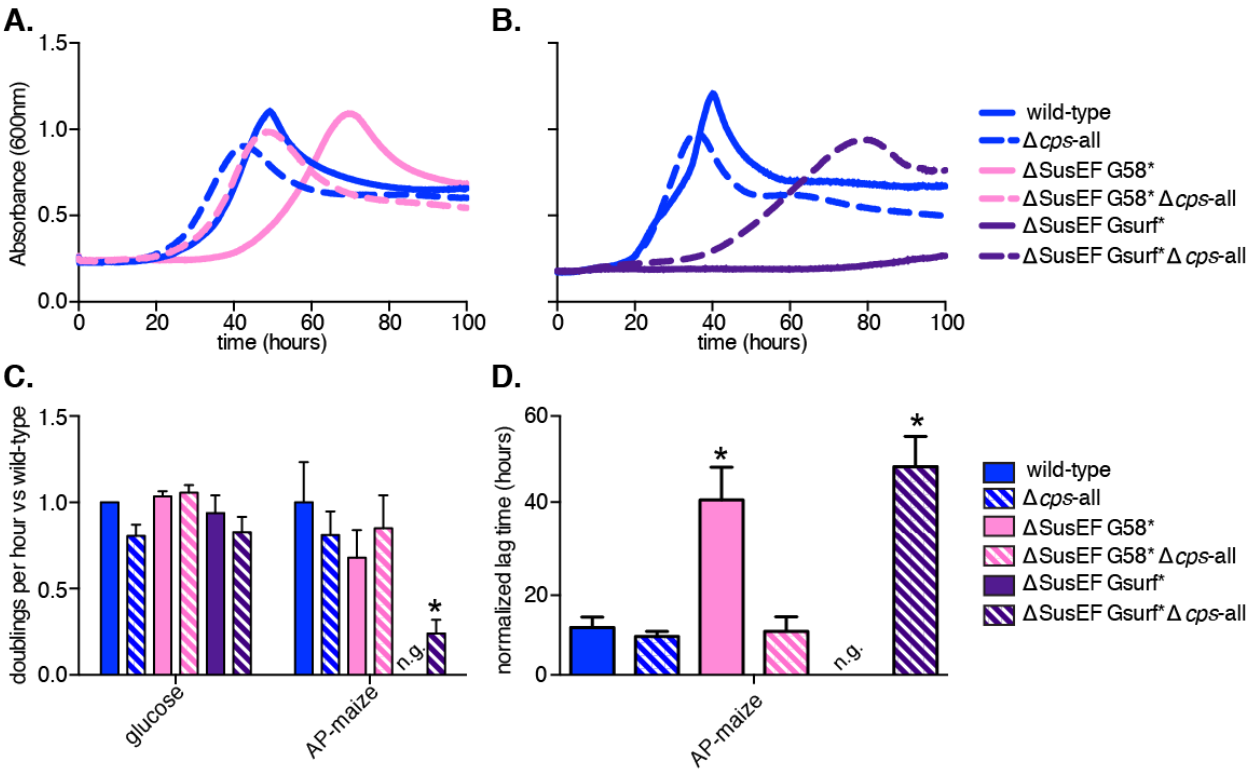



**FIGURE 3.12. Contribution of the SusEFG binding sites is capsule dependent**

**A., B.** Representative growth curves of *Bt* strains grown with AP-maize as the sole carbon source. **C.** Normalized doublings per hour and **D.** normalized lag times were calculated for three replicates (including those shown in **A.** and **B.**) as described in Figure 3.4. Average and standard error are depicted for three biological replicates. Statistically significant differences vs wild-type were calculated using a one-tailed unpaired student's t test (*$p < 0.05$).

does not express a polysaccharide capsule (ΔSusEF G58* Δcps-all and ΔSusEF Gsurf* Δcps-all).

Strikingly, we observed that the growth defect associated with loss of the Sus CBMs (ΔSusEF G58*) was abolished in the acapsular strain (ΔSusEF G58* Δcps-all). The significantly increased lag time associated with ΔSusEF G58* was restored to wild-type levels in the acapsular strain, and the growth curves of wild-type and ΔSusEF G58* Δcps-all were nearly identical (**Figure 3.12A,D**). When the ΔSusEF Gsurf* mutant, which was completely unable to grow on AP-maize, was tested in an acapsular background (ΔSusEF Gsurf*Δcps-all) growth was substantially restored, albeit to less than wild-type levels (**Figure 3.12B,C**). This data suggests that the SusE, F, G binding sites play a measurable role in the presence of polysaccharide surface capsule and may have evolved to counter-act the diffusion barrier created by the *Bt* capsule. This data is in agreement with our data suggesting the SusE, F, G binding sites are particularly important for growth on high molecular weight starch, as one would expect that it would be more difficult for starch molecules with a high degree of polymerization to penetrate the capsule layer compared to smaller substrates. Consistent with the SusD binding site playing a role distinct from that of the SusE, F, G binding sites, loss of the *Bt* capsule did not restore growth of the SusD* mutant (data not shown).

***Sus binding sites confer a fitness advantage in vivo on a starch-rich diet***

The Sus has evolved in the context of selective pressures encountered in the gut; therefore, we expect that the functions of the Sus binding sites are particularly important in the intestinal environment. To test this, we performed a competition experiment in gnotobiotic mice. Groups of germ-free C57Bl/6 mice were inoculated with an equal amount of each of three *Bt* strains: wild-type, Δ*susC* (control that cannot use starch) and either SusD* or ΔSusEF Gsurf*, as

these two binding site mutants had the most severe growth defects. To ensure that a significant amount of starch escaped host digestion and reached the colon we used a diet high in resistant starch (RS), which is not easily degraded by mammalian amylases. Previous *in vitro* studies reported that *Bt* cannot degrade this type of starch but upon co-culture with *Ruminococcus bromii*, a species that degrades RS very well, *Bt*'s growth on RS was significantly enhanced[33]. Therefore, we colonized half of the groups with *R. bromii*, to investigate whether the addition of *R. bromii* would increase the amount and types of starch available to *Bt* and perhaps exacerbate competitive defects of the mutants.

Mice were fed a starch free diet until colonization with the appropriate strains was established. Mice were then switched to a sequential feeding regimen of two diets that each contained 50% of two different high amylose RS preparations. DNA was extracted from fecal samples over time and the relative abundance of each *Bt* strain was quantified by qPCR directed at a unique genomic tags inserted into each mutant[16], the relative abundance of each strain was normalized to its abundance on the day of the diet switch (Day 0).

In the mice colonized with wild-type, Δ*susC* and SusD* the fitness of wild-type *Bt* was enhanced on the RS rich diets with its normalized abundance increasing over 10-fold over the course of the experiment (**Figure 3.13A,B**). The abundance of SusD* and Δ*susC* stayed relatively stable throughout the experiment with modest fold-change decreases that were not statistically significant. We did not see any perturbations in strain abundance associated with switching from one type of RS to another. In contrast to *in vitro* growth studies[33], the presence of *R. bromii* did not appear to significantly alter *Bt* fitness on RS *in vivo* (compare blue curves in **Figure 3.13A,B**). It is important to note that while starch is the only dietary carbohydrate present in this experiment, host mucosal glycans, for which *Bt* encodes several Sus-like degradative

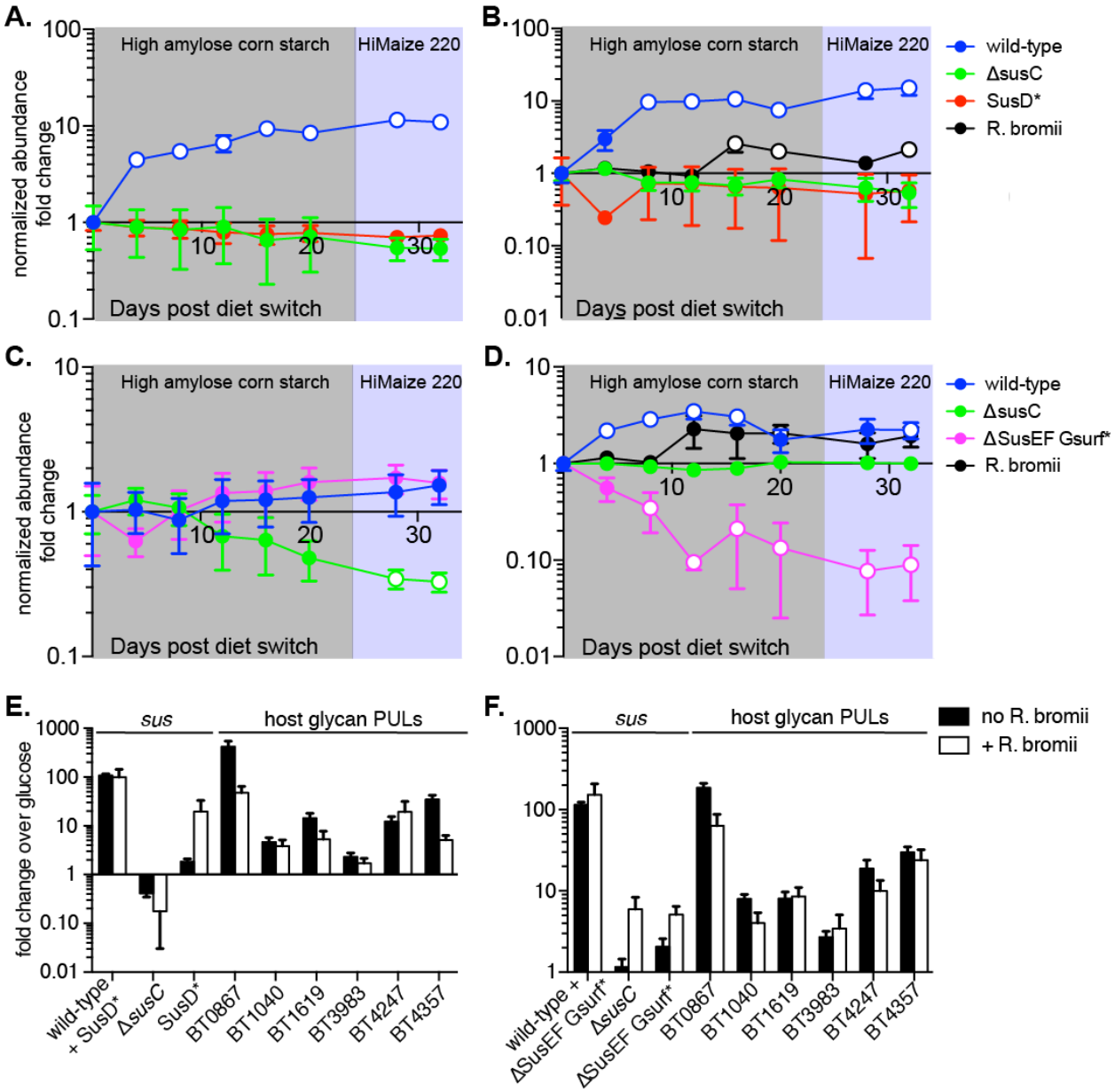

A.
B.
C.
D.
E.
F.
High amylose corn starch
HiMaize 220
normalized abundance
fold change
Days post diet switch
wild-type
ΔsusC
SusD*
R. bromii
ΔSusEF Gsurf*
sus
host glycan PULs
fold change over glucose
no R. bromii
+ R. bromii
wild-type + SusD*
ΔsusC
SusD*
BT0867
BT1040
BT1619
BT3983
BT4247
BT4357
wild-type + ΔSusEF Gsurf*
ΔsusC
ΔSusEF Gsurf*

**FIGURE 3.13. The Sus binding sites enhance wild-type *Bt* fitness *in vivo* on a starch-rich diet**

Germ-free mice fed a starch-free diet were colonized with wild-type, Δ*susC* and either SusD* (**A.** and **B.**) or ΔSusEF Gsurf* (**C.** and **D.**) *Bt*. Half of the groups (**B.** and **D.**) were also colonized with the keystone starch degrading species *Ruminococcus bromii*. Once colonization was established, mice were switched to a diet rich in high amylose resistant cornstarch for 24 days (shaded in gray) then switched to a diet rich in HiMaize 220 (shaded in purple) for the remainder of the experiment. The relative abundance of each strain was determined by quantifying unique genomic tags using qPCR and normalized to their abundance on the day of the initial diet switch (Day 0). The average and standard error across five mice is shown. Open circles represent a significant change ($p < 0.05$) in normalized abundance vs Day 0, calculated using two-tailed student's t test. **E.** and **F.** Transcript levels from cecal contents collected at the end of the experiment were probed using qPCR and fold change calculated over *in vitro* glucose grown *Bt*. (**E.**) displays transcripts from groups represented in panels A. and B. and (**F.**) displays data from groups represented in panels C. and D. To probe for strain-specific *sus* expression primers were designed such that only *sus* transcript from a subset of the *Bt* strains present (denoted on the x-axis) would be amplified. *sus* levels were normalized to the relative abundance of the strain from which they were amplified. Transcript levels of PULs targeting host mucosal glycans were probed as well.

systems[16], are a constant nutrient source. On both starch-free and RS diets, it is likely that these host glycans provide a constant alternative for *Bt*, explaining why the Δ*susC* and SusD* strains are able to maintain colonization (this is explored in more detail below).

Interestingly, in the groups colonized with ΔSusEF Gsurf* we did observe differences in strain abundance between the *Bt* only and *R. bromii* colonized group, most notably in the abundance of the ΔSusEF Gsurf* strain itself. In the absence of *R. bromii,* ΔSusEF Gsurf* behaved very similarly to wild-type, with both strains making modest, non-significant fold increases in normalized abundance over the course of the experiment (**Figure 3.13C**). However, in the presence of *R. bromii* the normalized abundance of ΔSusEF Gsurf* decreased significantly over the course of the experiment, while the wild-type abundance increased (**Figure 3.13D**).

At the end of the experiment, cecal contents were collected from all mice, total RNA extracted and corresponding cDNA probed for *sus* transcript levels as well as expression of PULs directed at host mucosal glycans. Because each of the *Bt* strains differed in the sequence of their *sus* locus due to the *susD, E, F* and *G* mutations we designed primers that allowed us to examine strain-specific *sus* levels. The level of *sus* transcription in mice colonized with the SusD* mutant indicated that most transcript was produced by wild-type *Bt*, regardless of the presence of *R. bromii* (**Figure 3.13E**). Although, the level of sus expression by the SusD* mutant increased ~18-fold in the presence of *R. bromii*, suggesting that this species liberates small maltoOS to induce *sus* expression in this mutant. As expected, the combined *Bt* community exhibited expression of several PULs previously associated with degradation of host glycans[16], confirming that this alternative nutrient pool is being accessed in these conditions. Similar results were observed in mice colonized with the ΔSusEF Gsurf* strain, suggesting that most expression

of specific *sus* transcript also derived from wild-type bacteria in this condition and that host glycans are targeted as alternatives.

Taken together, these data suggest that degradation of dietary starch by host amylases and *R. bromii* liberate different forms of this nutrient that require distinct Sus binding proteins. SusD likely contributes to utilization of saccharides, possibly smaller maltoOS, released by host digestion. In contrast, the combined presence of SusE, F and Gsurf contribute to utilization of starch released in the presence of *R. bromii*, which may correspond to longer pieces of starch that these functions are essential to metabolize *in vitro*.

**Discussion**

Microorganisms that thrive in the densely colonized and competitive gut ecosystem undoubtedly have evolved features to enhance their ability to recognize and scavenge nutrients. In this study, we demonstrate that the abundant gut symbiont, *Bt*, has evolved multiple starch binding proteins that, via unique and sometimes cooperative roles, optimize this bacterium for starch acquisition. We present a model where the SusD starch-binding site is critical for initial sensing of starch by enhancing utilization of medium length maltoOS, leading to efficient and rapid induction of the *sus* locus. The seven remaining binding sites spread across SusE, F, G contribute far less to the transcriptional response to starch but instead optimize *Bt* growth rate on starch in a capsule dependent manner, suggesting they act to offset the loss of affinity created by this barrier. We hypothesize that the SusE, F, G binding sites are most critical once the Sus machinery is highly expressed and function to keep local concentrations of starch surrounding the *Bt* cell high so catalysis can occur with maximum efficiency (**Figure 3.1**).

We have demonstrated that despite the absolute requirement for SusD in *Bt* starch utilization, the need for its binding ability *per se* can be circumvented with small maltoOS, demonstrating a critical function for SusD independent from binding. A potential binding-independent role for SusD is in formation and/or stabilization of the putative Sus complex. Indeed, previous cross-linking evidence suggests physical interaction between SusC and SusD[34]. SusD may promote additional interaction between SusE, F and G, which are needed to efficiently acquire and degrade substrate, and SusC, through which maltoOS are imported. This hypothesis is supported by the presence of a tetratricopeptide repeat domain on SusD[24], a motif associated with protein-protein interactions, although these additional roles for SusD remain to be explored in detail.

The amounts and forms of starch that reach the colonic microbiota are difficult to predict as, unlike other plant polysaccharides, human enzymes in the upper digestive tract degrade a significant portion. However, studies monitoring starch digestion as it passes through the digestive tract suggest that approximately 20% of digestible starch and 50% of resistant starch reaches the human colon[35,36]. Furthermore, starch-degrading enzymes are among the most common carbohydrate active enzymes in the human microbiome[13], suggesting it is indeed an important nutrient source for the gut microbiota. We found that the SusD binding site increased *Bt*'s ability to sense available starch, by allowing *sus* expression at starch concentrations several orders of magnitude lower than without SusD binding. The ability to sense and respond rapidly to available nutrients is critical in the gut where there is intense competition for nutrients and also a constantly changing carbohydrate landscape due to meal-to-meal variation.

SusD, unlike SusE, F, or G, is a conserved component of all Sus-like systems[16]. This is reflected in the fact that the SusD binding site serves a unique function that cannot be

compensated by binding sites in the other proteins. We hypothesize that SusD homologs in other systems serve a similar function in enhancing sensitivity to their cognate substrates, but this remains to be tested.

In contrast to the dramatic phenotype associated with loss of SusD binding, eliminating between one and five of the CBMs in SusE, F and/or G does not significantly effect *Bt* starch growth. However, loss of all six CBMs contained in these three proteins does cause a significant growth defect. Additionally, the defect associated with loss of the SusG surface-binding site, which is substantial by itself, is exacerbated by the further loss of SusE and SusF. Unlike for SusD, there must be overlapping roles for these Sus binding sites, as they appear to be able to compensate for loss of others. SusE, F and G are not conserved members of Sus-like systems. However, emerging studies suggest that there is functional conservation of cell surface carbohydrate binding proteins in other systems[4,19]. These divergent carbohydrate-binding proteins may fill similar roles as their functional counterparts in Sus. However, it is unlikely that binding proteins in other systems will universally exhibit cooperation or overlap with binding sites contained in their accompanying surface enzymes, because in a recent study on xyloglucan degradation the x-ray crystallographic structure of an essential endo-acting xyloglucanase failed to reveal any additional carbohydrate binding sites associated with this enzyme[19].

Experiments with an acapsular *Bt* strain suggest that the SusE, F, G binding sites have evolved redundant roles to offset a loss of affinity that is imposed by production of protective surface polysaccharides. The capsule-dependent role of these binding proteins may be particularly critical for starch acquisition since it is a large and potentially highly branched plant polysaccharide (on average $10^7$-$10^9$ Da in unprocessed corn starch). However, the forms of other host and dietary polysaccharides that are attacked by living *Bacteroides* cells *in vivo* may be

similar or greater in complexity, owing to their incorporation in plant cell wall particles or high molecular weight secreted mucin glycoproteins. One can imagine that these larger nutrient scaffolds will be more difficult to interact with through the thick capsular polysaccharide mesh. In light of many emerging studies on gut and environmental Sus-like systems required for degradation of other polysaccharides, it will be interesting in the future to determine if binding functions akin to those contributed by SusE, F, G play similar or different roles in other systems.

This study provides another layer of mechanistic understanding to a polysaccharide degradation paradigm that has been markedly expanded in bacterial members of the human gut microbiota and for which the *Bt* Sus is the best understood example. We demonstrate that individual binding proteins, with similar biochemical specificities when analyzed in pure form *in vitro*, play unique roles in the context of a multi-protein complex expressed on the surface of a symbiotic gut bacterium. Investigating these molecular mechanisms in great detail not only contributes to our understanding of the fundamental physiology of our gut microbial symbionts, but may also offer clues about how to intervene in their biology and the food webs in which they participate. The latter is the goal of pre- and probiotic approaches that aim to alter or stabilize the function of the gut microbial community and its potential contribution to inflammation and colorectal cancer, various metabolic diseases (obesity, diabetes) and invasion by outside pathogens.

## Materials and Methods

### *Bacterial strains and culture conditions*

*Bt* ATCC 29148 (VPI-5482) strains were routinely grown in tryptone-yeast extract-glucose (TYG) media[37] or on brain-heart infusion (BHI, Beckton Dickinson) agar that included

10% horse blood (Colorado Serum Co.). *R. bromii* L2-63 was grown in Hungate tubes containing YCFA media[33] supplemented with 2mg/mL each glucose, cellobiose and soluble starch (YCFA-GCS). Antibiotics were supplemented to medias as appropriate including erythromycin (25 μg/ml), gentamicin (200 μg/ml), tetracycline (2μg/mL) and 5-fluoro-2'-deoxyuridine (FUdr, 200 μg/ml). Minimal media (MM) was prepared as described previously[16]. Cultures were grown at 37°C in an anaerobic chamber (Coy Manufacturing, Grass Lake, MI; 10% $H_2$, 5% $CO_2$ and 85% $N_2$). Mono and oligosaccharides were filter sterilized and stored at 4°C until use, higher molecular weight starches were sterilized by autoclaving, and stored at 4°C for at least 6 days before use to allow starches to fully retrograde. Carbon sources were used at a final concentration of 5mg/ml unless specified otherwise.

***Bt growth curves***

Strains were grown overnight in TYG then passaged into MM + glucose and allowed to grow to stationary phase. Anaerobically, MM+glucose grown cells were washed and resuspended in an equal volume of 2X MM (no carbon) then diluted 1:50 into 2X MM (no carbon). 100μl inoculated 2X MM was added to 96 well plates containing 100μl desired carbon source (10mg/mL) that had equilibrated overnight in the anaerobic chamber. Absorbance at 600nm was measured every ten minutes on an automated plate reading device as described previously[17]. To calculate *normalized doublings per hour* we used the portion of the graph corresponding to absorbance readings between 0.6-0.8 for all data except those in figure 3.12; here 0.4-0.6 was used to account for the lower maximum absorbance observed in the Δcps-all mutant. Data points were fit to an exponential growth equation and doublings per hour were normalized to wild-type glucose; because some mutants could exhibit defects in glucose growth, this allowed visualization of variation for each strain on glucose. Growth of each strain was then

normalized to wild-type for each substrate. *Lag time* was defined as the time required for the absorbance reading for a particular strain to reach or exceed 0.35 at 600nm. To account for variation in inoculum size or environment between experiments the lag time on glucose for that individual experiment was subtracted from the lag time for the substrate of interest; in all cases cultures on glucose had shorter lag time than on starch or maltoOS. Statistical significance was assessed using a one-tailed unpaired student's t-test. Note that there is a relatively significant amount of variability between independent *Bt* growth replicates. We hypothesize this is due to variations in the anaerobic chamber environment (i.e. oxygen, hydrogen levels) as well as small variations in starch/maltoOS concentration as we have demonstrated the exquisite sensitivity of the Sus. As observed previously with Δ*susD*, suppressor mutants were observed after ~100 hours of growth that, despite still lacking the required Sus components, had regained the ability to grow on starch. To exclude these from our analysis only the first 100 hours of growth was analyzed.

***Genetic manipulation of Bt***

Lipidation deficient alleles of *susE* and *susF* were constructed by mutating the C-terminal cysteine residues of each protein as described previously[26]. The *susG* 58* *susG* surf* and *susD** alleles were constructed by amplifying the appropriate gene plus flanking sequence from the *Bt* ATCC 29148 genome and cloning the fragment into the pExchange-tdk suicide vector[24]. Critical binding residues (**SusG58*:** W287, W299 and N330 **SusGsurf***: W460, Y469 and D473 **SusD*:** W98, N101 and W320) were mutated to alanine using the Quikchange® multi-site directed mutagenesis kit (Stratagene) using pExchange-tdk construct containing the appropriate gene fragment as template. Desired mutations were screened for via PCR using a primer complementary to mutated residues at the 3' end, and confirmed by sequencing. The *susE** and

*sus*F* alleles had been created previously in the protein expression plasmid, pET28rTEV[26]. These alleles were amplified from the pET28rTEV constructs, 700 bp of flanking sequence was added via soe PCR and this insert was cloned into pExchange-tdk. Constructs were introduced into the appropriate *Bt* background by conjugation[24]. Merodiploids were selected for on plates containing erythromycin (25μg/mL) and plated on BHI-blood containing 5-fluoro-2'-deoxyuridine (200μg/mL) to select for recombinants. Clones were screened by PCR and confirmed with sequencing.

***Confirmation of SusG-CBM58 and SusD-binding deficient mutants***

We previously validated that recombinantly expressed SusG with mutations in the surface (surf) binding site displays defects in the ability access or bind granular starch, although there is no defect in the enzyme's ability to digest soluble starch or maltooligosaccharides[25]. To confirm that mutations made in the starch-binding residues of the CBM58 of SusG and those in SusD eliminated starch-binding, we used site-directed mutagenesis to mutate our original pET28rTEV constructs of both the isolated CBM58 domain of SusG[25] and SusD[24], described in previous studies. The recombinant CBM58* and SusD* proteins were expressed in E. coli, purified via Ni affinity chromatography, and the His-tag removed using TEV protease, as described for the WT version of these proteins. Isothermal titration calorimetry on a TA instruments NanoITC SV was performed to confirm the lack of binding by both proteins to α-cyclodextrin (aCD). Both proteins were dialyzed extensively against 20mM HEPES 100mM NaCl pH 7.0 prior to experiments and the dialysate was used to prepare ligand solutions for the injections. For each protein, 0.15 – 0.2 mM protein was loaded into the 1.3 ml sample cell, and 24 x 10ul injections of 3 – 6 mM α-cyclodextrin were made every 300s while stirring at a rate of

350rpm. All titrations were performed at 25° C. The data were plotted and fit with NanoAnalyze (TA instruments).

***Monitoring transcriptional response to maltoOS and starch***

*Bt* strains were grown to mid exponential phase (OD 0.6-0.85) in MM+glucose, washed in MM (no carbon) and resuspended in an equal volume MM (no carbon). 1mL of this 'no carbon' sample was taken and added immediately to RNA protect (Qiagen). The remaining cells were then diluted 1:1 with MM + appropriate concentration of desired carbon source and time noted precisely at the addition. 1mL samples were collected at 30 minutes post spike-in into RNA protect reagent. RNA was isolated with RNAeasy mini kit (Qiagen) according to manufacturer instructions. DNA digestion was performed with TURBO™ DNAseI (Ambion) followed by an additional RNA clean-up with RNAeasy mini kit. Reverse transcription was performed with SuperScript III (Invitrogen) reverse transcriptase using random primers. cDNA quantification was performed with a Mastercycler® ep realplex (Eppendorf), using KAPA SYBR® FAST qPCR Master Mix and 400 nM *susC* primers, or 62.5 nM Bacteroides 16S rRNA primers, for 40 cycles of 95°C for 3 sec, 55°C for 8 sec, 72°C for 20 sec. *susC* levels were normalized to 16S levels then converted to fold-change over the no carbon condition.

***Analysis of maltoOS concentrations by HPAEC***

Analysis of 2-Aminobenzamide labeled starch stocks and culture supernatants was performed by high-pH anion exchange chromatography (HPAEC) at the UCSD Glycotechnology Core. MaltoOS levels were quantified using known standards of maltoOS ranging from glucose to maltoheptaose (G1-G7).

To quantify maltoOS levels in starch stocks pre and post dialysis starch solutions of known concentrations were made in water and labeled with 2-Aminobenzamide prior to analysis by HPAEC.

To quantify maltoOS release in culture supernatants, *Bt* cultures were grown in MM+AP-maize. At mid-log phase culture was collected, spun and the supernatant collected into a fresh tube. Supernatant was immediately boiled for 20 minutes then frozen at -20°C until analysis. Samples were labeled with 2-Aminobenzamide prior to analysis by HPAEC

***Immunofluorescence and Western-blotting***

Surface expression of Sus proteins was probed by antibody staining of non-permeabilized formaldehyde fixed *Bt* cells grown on MM-maltose with rabbit polyclonal antibodies (Cocalico Biologicals) and detected with an Alexa-Fluor® 488 conjugated goat anti-Rabbit IgG secondary antibody (Molecular Probes). Sus proteins were detected in *Bt* whole cell lysates by western blot using the rabbit polyclonal primary antibodies mentioned above together with either an alkaline phosphatase (Figure 3.3) or horseradish peroxidase (Figure 3.8) conjugated goat anti-Rabbit IgG secondary antibody (Sigma).

***Gnotobiotic mouse experiment***

Six week old male C57Bl/6 germ-free mice were pre-fed a starch-free diet (Harlan-Teklad TD.130280) for seven days before bacterial colonization. Mice were randomly divided into four groups of five mice and caged separately by group. *Bt* strains containing previously published genomic tags[16] were grown in TYG and *R. bromii* was grown in YCFA-GCS media then mixed 1:1:1 (wild-type *Bt* : Δ*susC Bt* : SusD*/ΔSusEF Gsurf* *Bt*) or 1:1:1:2 (wild-type *Bt* : Δ*susC Bt* : SusD*/ΔSusEF Gsurf* *Bt* : *R. bromii*) by volume. Bacterial mixtures were kept anaerobic until just before inoculation of mice, which each received 100µl by oral gavage on

Day 0. Groups receiving *R. bromii* were given three additional gavages on Days 4, 8 and 12 of 100µl overnight *R. bromii* culture only. On Day 14 all mice were switched to Harlan-Teklad Low Glycemic Control Diet (TD.120455), containing 50% (w/w) High Amylose Resistant Corn Starch. DNA was extracted from fecal pellets and *Bt* strains enumerated as previously described[16], *R. bromii* levels were quantified using a similar strategy but with primers to Rb_05420, a dockerin type I repeat that has no sequence homology to genes from other organisms. On Day 38 all groups were switched to a different Harlan-Teklad Low Glycemic Control Diet (TD.08810) containing 50% (w/w) Hi Maize 220 Resistant Starch.

On Day 53 mice were sacrificed via cervical dislocation and cecal contents was collected, flash frozen and stored at -80°C. RNA was purified via phenol chloroform extraction, ethanol precipitated and cleaned up using the Quick-RNA™ mini-prep kit (Zymo Research) with optional on column DNAse step. DNA was further removed using TURBO™ DNAseI (Ambion), followed by a final clean up using the Quick-RNA™ mini prep kit. cDNA quantification was performed on the Mastercycler® ep realplex (Eppendorf) using the conditions described above. For strain specific *sus* quantification *sus* transcripts were normalized to 16S levels that had been adjusted to reflect the abundance of the strain(s) in question. Host glycan PUL genes were normalized to total 16S levels. Transcript levels were then converted to fold change over cDNA obtained from the appropriate strain of *in vitro* MM+glucose grown *Bt*.

**Notes**

This work was reprinted and modified with permission from Cameron, E.A., Kwiatkowski, K.J., Lee B.H., Hamaker B.R., Koropatkin N.M., Martens, E.C. Multi-functional nutrient binding proteins adapt human symbiotic bacteria for glycan competition in the gut. *Manuscript submitted for publication May, 2014*.

## References


1 McNeil, N. I. The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr* **39**, 338-342 (1984).

2 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).

3 Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* **341**, 569-573 (2013).

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

5 Martinez, I., Kim, J., Duffy, P. R., Schlegel, V. L. & Walter, J. Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects. *PLoS One* **5**, e15046 (2010).

6 Walker, A. W. *et al.* Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J* **5**, 220-230 (2011).

7 Kostic, A. D., Xavier, R. J. & Gevers, D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* **146**, 1489-1499 (2014).

8 Garrett, W. S. *et al.* Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* **8**, 292-300 (2010).

9 Kostic, A. D. *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207-215 (2013).

10 Sears, C. L. & Garrett, W. S. Microbes, Microbiota, and Colon Cancer. *Cell Host Microbe* **15**, 317-328 (2014).

11 Sekirov, I. *et al.* Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. *Infect Immun* **76**, 4726-4736 (2008).

12 Theriot, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. *Nature Commun* **5**, 3114 (2014).

13 El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Rev Microbiol* **11**, 497-504 (2013).

14 Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638 (2005).

15 Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284**, 24673-24677 (2009).

16 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**, 447-457 (2008).

17 Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).

18 Shipman, J. A., Berleman, J. E. & Salyers, A. A. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *J Bacteriol* **182**, 5365-5372 (2000).

19 Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498-502 (2014).

20 Dodd, D., Moon, Y. H., Swaminathan, K., Mackie, R. I. & Cann, I. K. Transcriptomic analyses of xylan degradation by Prevotella bryantii and insights into energy acquisition by xylanolytic bacteroidetes. *J Biol Chem* **285**, 30261-30273 (2010).

21 Renzi, F. *et al.* The N-glycan Glycoprotein Deglycosylation Complex (Gpd) from Capnocytophaga canimorsus Deglycosylates Human IgG. *PLoS Path* **7** (2011).

22 Anderson, K. L. & Salyers, A. A. Biochemical evidence that starch breakdown by Bacteroides thetaiotaomicron involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *J Bacteriol* **171**, 3192-3198 (1989).

23 Rogers, T. E. *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol Microbiol* (2013).

24 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**, 1105-1115 (2008).

25 Koropatkin, N. M. & Smith, T. J. SusG: A Unique Cell-Membrane-Associated alpha-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. *Structure* **18**, 200-215 (2010).

26 Cameron, E. A. *et al.* Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism. *J Biol Chem* **287**, 34614-34625 (2012).

27 Coyne, M. J. & Comstock, L. E. Niche-specific features of the intestinal bacteroidales. *J Bacteriol* **190**, 736-742 (2008).

28 Ellrott, K., Jaroszewski, L., Li, W., Wooley, J. C. & Godzik, A. Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. *PLoS Comput Biol* **6**, e1000798 (2010).

29 Morell, M. K., Samuel, M. S. & O'Shea, M. G. Analysis of starch structure using fluorophore-assisted carbohydrate electrophoresis. *Electrophoresis* **19**, 2603-2611 (1998).

30 Yokoyama, W., Renner-Nantz, J. J. & Shoemaker, C. F. Starch molecular mass and size by size-exclusion chromatography in DMSO-LiBr coupled with multiple angle laser light scattering. *Cereal Chem* **75**, 530-535 (1998).

31 Lee, B. H. *et al.* Enzyme-Synthesized Highly Branched Maltodextrins Have Slow Glucose Generation at the Mucosal alpha-Glucosidase Level and Are Slowly Digestible In Vivo. *PLoS One* **8**, e59745 (2013).

32 Martens, E. C., Roth, R., Heuser, J. E. & Gordon, J. I. Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J Biol Chem* **284**, 18445-18457 (2009).

33 Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J* **6**, 1535-1543 (2012).

34 Cho, K. H. & Salyers, A. A. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **183**, 7224-7230 (2001).

35 Vonk, R. J. *et al.* Digestion of so-called resistant starch sources in the human small intestine. *Am J Clin Nutr* **72**, 432-438 (2000).

36 Champ, M. M. *et al.* Small-intestinal digestion of partially resistant cornstarch in healthy subjects. *Am J Clin Nutr* **68**, 705-710 (1998).

37 Holdeman, L. V., Cato, E. D. & Moore, W. E. C. *Anaerobe Laboratory Manual.*, (Virginia Polytechnic Institute and State University Anaerobe Laboratory, 1977).

# Chapter IV

## Discussion

Glycan degradation is an important function of the human gut microbiota that directly affects host health. Short chain fatty acids produced as bacterial fermentation products serve as nutrients for our intestinal cells[1] and improve colonic health[2-5]. The availability of carbohydrates and the ability of specific species to metabolize them profoundly influences the composition of the bacterial community, which can in turn have effects on the health of the host[6]. The Bacteroidetes, one of the two predominant phyla in the human gut[7], have an expanded capacity for glycan degradation compared to other bacterial taxa[8-10], making them key players in this process. The Bacteroidetes degrade a wide variety of plant and host-derived carbohydrates via multi-protein complexes termed Sus-like systems. Work presented in this dissertation provides new insight into the structure and function of the Starch Utilization System (Sus) in *Bacteroides thetaiotaomicron* (*Bt*), the first of these systems discovered and a model for glycan acquisition by the Bacteroidetes[11-15].

### Chapter Summary

The work described in this dissertation investigated the molecular mechanisms governing starch acquisition by *Bt,* an abundant human gut symbiont. X-ray crystallographic structures of two outer-membrane lipoproteins in the *Bt* Sus were solved, and these proteins were characterized biochemically. This work allowed us to probe the function of the four outer-membrane Sus proteins in mechanistic detail. This work has significantly advanced our

understanding of the molecular mechanisms of carbohydrate acquisition in an important group of human gut symbionts.

The Sus is an eight-protein system required for *Bt* to utilize starch as a carbon source[11-15]. SusE and SusF are two outer-membrane lipoproteins encoded by the *Bt sus* locus that, previous to this work, were of undefined function. These proteins were previously shown to be dispensable for *Bt* starch growth *in vitro* although they contributed to the total ability of *Bt* cells to bind starch[15]. SusEF-positioned proteins are found in almost all Sus-like systems[16], and related proteins are enriched in the human microbiome vs. related environmental species[17] suggesting they indeed play an important role specifically in the gut environment.

In Chapter II both SusE and SusF were co-crystallized with small starch ligands, and the x-ray crystal structures resolved. SusE and SusF were found to both be composed of multiple carbohydrate-binding modules (CBMs), with SusE containing two and SusF three (**Figure 2.2)**. The C-termini of these two proteins display very similar folds, which is consistent with the ~38% sequence identity observed specifically at the C-termini. Each CBM of SusE and SusF contains a single starch-binding site that all display a similar mode of binding. Aromatic residues form a hydrophobic platform that interacts with the glucose rings of starch via stacking interactions, and hydrogen bonding between polar amino acid side chains and the hydroxyl groups of glucose residues stabilizes the binding (**Figure 2.4, 2.5**). However, differences in the number and orientation of binding contacts between the sites suggested these sites might display differences in their strength of binding. Indeed, distinct affinities and preferences for particular starch structures were observed across the five SusE and SusF CBMs (**Table 2.3, Figure 2.6**). We hypothesized this may help *Bt* to accommodate the structural heterogeneity between starches it encounters in the gut.

Work described in chapter II completed a series of studies determining the structures of the four Sus outer-membrane lipoproteins[18,19]. This work revealed that across these four lipoproteins (SusD,E,F,G) there is a total of eight non-enzymatic starch binding sites. This immediately raises the question of the utility of encoding so many binding sites: are they redundant or do they serve unique functions during *Bt* starch acquisition? Interestingly SusD, which has a single starch binding site and no enzymatic activity, is required for *Bt* starch growth[18] while SusE and SusF, containing five starch binding sites between them, are dispensable for growth on starch[15]. This, along with biochemical data showing differences in substrate preference and affinity between the eight SusDEFG binding sites, led us to hypothesize that these binding proteins serve unique functions in the Sus complex.

In chapter III we show that the SusD,E,F,G binding sites are not redundant and indeed have distinct roles during different phases of *Bt* starch acquisition. We demonstrate that while SusD binding ability is required for growth on large starch molecules, addition of a low level of maltooligosaccharides (maltoOS), which stimulates induction of the *sus* locus, will rescue growth of the SusD binding mutant on starch. However, addition of maltoOS does not rescue growth of a *susD* deletion mutant, demonstrating that SusD has a critical function independent of its binding ability (**Figure 3.4**). Although this binding independent function is not entirely clear there is evidence from previous studies as well as work discussed in Appendix I that suggests SusD interacts with SusC[15], perhaps stabilizing this transporter as well as potentially facilitating interactions with other Sus components.

We go on to show that SusD binding activity plays an important role in import of the initial maltoOS signal, leading to expression of the *sus* locus so that rapid degradation of large starches can begin. The SusD binding site enhances *Bt*'s ability to sense low levels of starch,

allowing *sus* expression at concentrations several orders of magnitude lower than without this function (**Figure 3.5**). These data show that the SusD binding site is critical in the initial 'surveillance phase' and allows *Bt* to rapidly and efficiently sense and respond to available starch.

We show that while loss of SusE and SusF does not lead to a significant growth defect, when lost in combination with either the SusG CBM58 or the SusG surface site, a significant growth defect is observed (**Figure 3.7**). This suggests that these binding sites have partially redundant roles, as remaining binding sites appear to compensate for the loss of others. Additionally, we show that unlike SusD, SusE and SusF do not have a crucial function independent of their binding ability, as expression of binding-deficient proteins results in an identical phenotype to complete lack of expression of SusE and SusF (**Figure 3.10**). In contrast to SusD, the SusE,F,G binding sites did not significantly enhance *sus* locus transcription (**Figure 3.5**). This supports a model in which SusEFG are more critical once the *sus* locus has been induced, during active catalysis and growth on starch.

Interestingly, we show that the growth defect observed in the SusE,F,G binding site mutants was partially complemented by loss of the *Bt* polysaccharide capsule (**Figure 3.12**). The *Bt* capsule is a barrier that may make it difficult for large carbohydrates (like starch) to reach the cell surface, where the Sus machinery is located. Therefore, we hypothesize that the SusEFG binding sites have evolved to offset this obstacle.

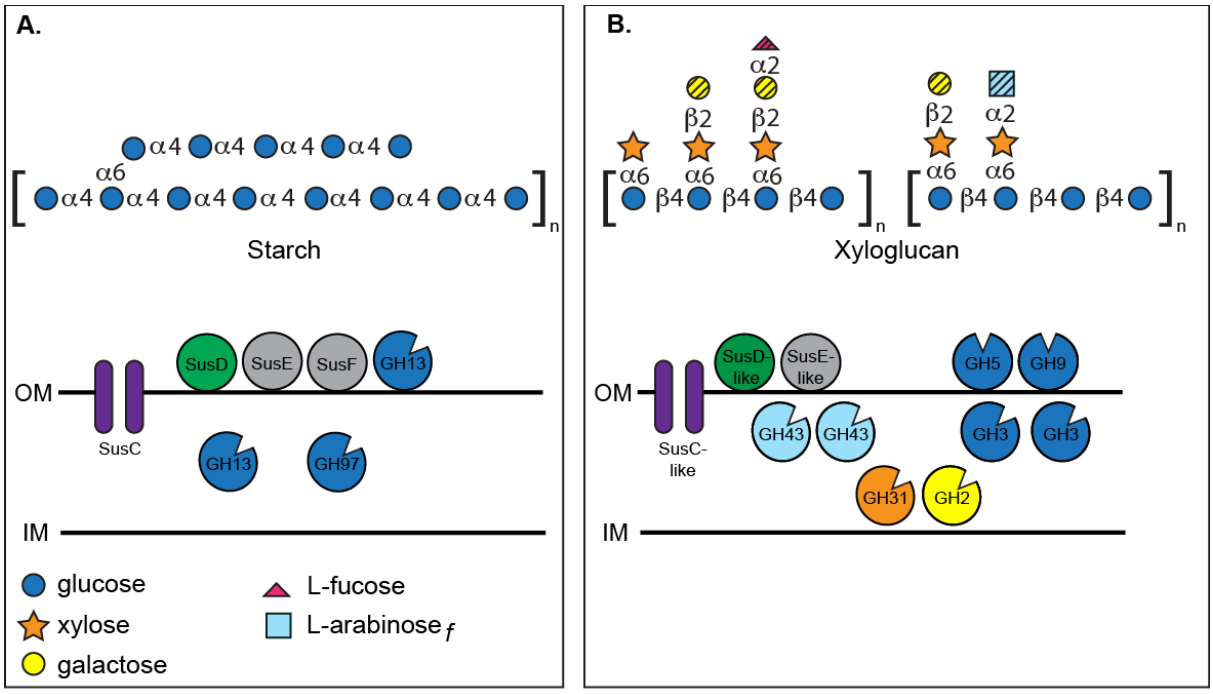
Finally, germ-free mice were colonized with wild-type and mutant *Bt* and fed a resistant starch rich diet. We observed that *in vivo* the Sus binding sites did confer a fitness advantage to the wild-type strain, but this advantage was dependent on the particular mutants included in the competition and also the colonization status with or without another symbiont (*Ruminococcus*

*bromii*) that directly targets resistant starch. Furthermore, we observed that the wild-type strain exhibited higher levels of *sus* transcription compared to the Δ*susC*, SusD* or ΔSusEF Gsurf* *Bt* (**Figure 3.13**). We hypothesize that in this mouse model wild-type *Bt* is efficiently utilizing the starch available to proliferate, whereas the mutant strains are not efficiently catabolizing starch, and instead target host mucosal glycans, as PULs directed toward these structures were highly expressed in the mouse cecum. These data demonstrate that expressing a complete Sus was beneficial in a mouse model of colonization on a starch-rich diet.

The work presented here significantly enhances our understanding of the structure and function of the *Bt* Sus on a molecular level. This is significant as the Sus is a prototype for numerous similar systems encoded by the Bacteroidetes, which, because of their extended ability to degrade numerous glycans, are major players in carbohydrate degradation in the human gut.

**Similarities between the *Bt* Sus and other Sus-like systems**

The ability to easily culture and genetically manipulate *Bt* makes it an excellent model for studying glycan acquisition by the Bacteroidetes. Additionally, compared to other PULs the *Bt* Sus is relatively simple, containing only eight proteins and three enzymes. Sus-like systems that target carbohydrates with multiple sugar monomers and glycosidic linkages can be significantly more complex (**Figure 4.1**). For example a single *Bt* PUL targeting the plant pectin rhamnogalacturonan II contains 32 enzymes[20]. Therefore, while the *Bt* Sus is a simple and convenient model, it is important to assess whether the molecular lessons learned from the *Bt* Sus can be applied more broadly to other Sus-like systems.




C.

| Glycan | # of unique linkages | Species | # of PULs in species | # of enzymes |
|---|---|---|---|---|
| **Starch** | 2 | *B. theta* | 1 | 3 |
| **Xyloglucan** | 5 | *B. ovatus* | 1 | 8 |
| **Yeast α-mannan** | 4 | *B. theta* | 3 | 12 |
| **Mucin O-glycans** | 12 | *B. theta* | 15 | 17 |
| **Rhamnogalacturonan II** | 22 | *B. theta* | 1 | 32 |

**FIGURE 4.1. Variations in functional complexity among Sus-like systems**

(**A.** and **B.**) Simplified models of two Sus-like systems in human gut *Bacteroides* are depicted with their cognate substrates. Enzymes are shown in their cellular locations (known or predicted): above the outer membrane (OM) are extracellular lipoproteins and between the OM and inner membrane (IM) are periplasmic enzymes. Enzymes are color-coded according the sugar residues they cleave and are denoted by their glycoside hydrolase (GH) family. **A**. Starch contains only two unique glycosidic linkages and accordingly the *B. theta* Sus contains only three enzymes, relatively few compared to other systems (see **C.**). A GH13 (SusG) performs the initial degradation of starch at the cell surface. In the periplasm maltooligosaccharides are degraded by another GH13 (SusA) and a GH97 (SusB) **B**. Depiction of a *Bacteroides ovatus* PUL that targets the hemicellulose xyloglucan, a heteropolymer with multiple monosaccharides and glycosidic linkages[21] (variable sugar residues are shown with black hatch marks). Xyloglycan structure differs between plant sources and thus two varieties are shown. The increased complexity of xyloglucan compared to starch is reflected in the greater number and diversity of enzymes in this PUL. **C.** A select number of *Bacteroides* PULs and the substrates they target are listed. The complexity of the PULs (number of enzymes) increases as the complexity of the substrate (# of unique linkages) increases.

***The SusEF-like proteins***

The structure of SusE and SusF were elucidated in Chapter II, revealing that they are structural homologs of each other and both composed of multiple CBMs. The majority of Sus-like systems identified contain at least one SusEF-positioned protein[22] but there is little or no conserved sequence homology of these genes across different systems[9]. However, as other Sus-like systems are being studied more extensively it appears there is some functional conservation of the SusEF-positioned proteins as non-enzymatic, outer-membrane carbohydrate-binding proteins.

*Bt* encodes a Sus-like system that targets the β-2,6 linked fructose polymer levan, a component of plant cell walls. However, this system does not confer the ability to grow on the related fructan inulin, where a β-2,1 bond links the fructose residues. It was found that the hybrid two-component system responsible for upregulation of this PUL recognized monomeric fructose and thus did not lend any linkage specificity to the system[23]. Additionally, only one of three glycoside hydrolases in this PUL demonstrates specificity for the β-2,6 linkage. However, both the SusD homolog and a SusE-like protein in this system bound specifically to β-2,6 linked fructans, displaying no affinity for β-2,1 linked fructans[23]. In this system it appears that the non-enzymatic binding proteins, including the SusE-like protein, are responsible for much of the specific recognition of levan. Understanding the basis of this specificity can lead to a greater understanding of how and why certain dietary glycans affect particular bacterial populations and how we might alter diet in a deliberate way to shape the bacterial community in the gut.

Starch and levan are both relatively simple carbohydrates, however SusEF-like proteins in Sus-like systems targeting complex glycans have also been found to display similar function. In the related species *Bacteroides ovatus* (*Bo*) a PUL was identified targeting xyloglucan[21], a

family of cell wall glycans found in plants like lettuce, onions and tomatoes. Xyloglucans have a β-1,4 linked glucose backbone with sidechains containing several different sugar residues and glycosidic linkages that give this family of glycans a high level of structural heterogeneity[24]. Again, this PUL encodes a single SusE-like protein that binds xyloglucan oligosaccharides but has no detectable enzymatic activity[21]. Here again we see further evidence for functional conservation of this SusEF-like family as carbohydrate-binding proteins, even in a system targeting a significantly more complicated substrate.

Functional conservation of the SusEF-like proteins has even been observed in Bacteroidetes that are distantly related to the gut *Bacteroides*, such as *Capnocytophaga canimorsus*. *C. canomorsus* is a facultative anaerobe that is a common member of the oral microbiota of dogs and cats[25] and can cause infections in humans following a bite from these animals[26]. This organism contains a Sus-like system that targets N-linked glycans on the surface of host cells[27]. This system is strikingly similar to the *Bt* Sus and contains a SusE-like (GpdE) and a SusF-like protein (GpdF). GpdE and GpdF are both outer-membrane lipoproteins that have sequence homology with annotated glycan binding proteins and are not predicted to have any enzymatic activity[27]. Interestingly in this system, GpdE was required for deglycosylation of host cells whereas GpdF was not, suggesting that unlike their cognate proteins in the *Bt* Sus, GpdE and GpdF may have distinct roles.

It appears that despite having no sequence homology across systems, the SusEF-like proteins display functional conservation as glycan binding proteins. The ability of SusEF-like proteins to bind carbohydrates has been observed in close relatives of *Bt* (*Bacteroides ovatus*) as well as distant relatives (*Capnocytophaga canimorsus*), and we expect this will be a theme

across many Sus-like systems in other species. However, the precise role of the SusEF-like proteins in the context of their respective multi-protein complexes may differ across systems.

In chapter III we presented evidence of a role for SusE and SusF in offsetting a loss of affinity caused by the *Bt* polysaccharide capsule. We hypothesize that this may be a starch-specific function as starch has one of the highest degrees of polymerization among plant polysaccharides and thus may be more inhibited from penetrating the capsule layer. While sheer size of the molecule may be one of the most significant challenges during starch utilization, degradation of other glycans likely pose other challenges. Most plant and host derived carbohydrates are smaller than starch but many have much more complex structures, containing multiple sugar residues and linkages that SusEF-like proteins may be involved in accommodating. Interestingly in the *Bt* levan PUL we again see a substrate-specific role for the SusE-like protein. Here a SusE-like protein, plus a linkage specific surface enzyme, adapts *Bt* to specifically use levan over inulin. The ability to utilize inulin is more widespread in the *Bacteroides* compared to levan; therefore this levan-specific PUL may give *Bt* access to a nutritional niche where it encounters less competition.

The lack of sequence homology between the SusEF-like proteins[9] supports a model in which these proteins have diverged and evolved functions that are specific to the system's target glycan. An interesting line of future research will be to investigate whether the function of SusEF-like proteins is substrate specific and adapt the system to specific challenges associated with acquiring or degrading its target. For example, *Bt* encodes a number of Sus-like systems targeted toward host mucosal glycans[16], which have an incredible level of diversity in the sugar residues they contain as well as the linkages connecting them[28]. SusEF-like proteins may help *Bt* accommodate this diversity, perhaps by containing multiple binding sites with specificity for

different sugar residues, expanding *Bt*'s ability to recognize different structures found within the mucosal layer.

***SusD homologs***

In chapter III we described a previously unknown function for the SusD binding site, enhancing *Bt*'s ability to sense available starch. We found that a small amount of maltoOS could partially compensate for loss of the SusD binding function but not for loss of the entire protein, suggesting SusD has a critical function independent of binding. We hypothesize that this role may be in formation of the Sus complex and that SusD facilitates interactions between other Sus components. This is supported by evidence of a physical interaction between SusD and SusC[15], as well as the presence of four tetratricopeptide repeat (TPR) units, a motif associated with making protein-protein interactions[29], in the SusD structure[18]. Along with SusC, SusD is one of the two conserved components of all Sus-like systems[16]; therefore, SusD homologs in other systems may fulfill similar roles that are separable from their binding capacity.

Indeed all SusD homologs whose structures have been deposited in the Protein Data Bank are dominated by alpha helices and contain TPR motifs[30], supporting a conserved function for this group of proteins. This structural homology cannot be predicted by primary sequence as evidenced by BT1043, a SusD-like protein involved in host glycan utilization that, despite having no sequence homology, displays the same alpha-helical fold as SusD and is dominated by TPR units[31].

Aside from their potential role in complex formation, the glycan binding function of the SusD protein family appears to be conserved as well. SusD homologs from distantly related organisms (*Bt*, *Bo*, *C. canimorsus*) and targeting very different substrates (starch[18], levan[23], O-

glycans[31], N-glycans[27], xyloglucan[21]) have all been shown to be non-enzymatic glycan binding proteins. To date, all studies where the SusD homolog was deleted have led to significant growth attenuation on the cognate substrate[18,23,27]. The apparent critical role of SusD homologs provides an interesting contrast with the SusEF-like family of proteins, which are also glycan binding proteins but upon deletion often result in no apparent phenotype[15], or one that is less severe than loss of the SusD-like protein[27]. In the *Bt* starch system the SusD binding site is uniquely required for the early stages of starch sensing and *sus* locus induction. It will be an interesting area of future study to investigate whether SusD homologs also enhance the ability to sense and respond to their respective substrates, and if this explains why these glycan proteins are more essential than their SusEF-like counterparts.

The *Bt* Sus is relatively simple compared to Sus-like systems targeting more complex substrates. However, it appears that there is sufficient functional homology among the components of Sus-like systems that we can draw meaningful conclusions about Bacteroidetes glycan acquisition from studying the *Bt* Sus. Even in systems targeting complex substrates (xyloglucan, host mucosal glycans), and systems found in distantly related species (*C. canimorus*) we observe functional similarities with the *Bt* Sus. We believe that there is sufficient evidence that molecular mechanisms discerned from the *Bt* Sus are useful in understanding glycan acquisition as a whole by this important group of human commensals. Testing this hypothesis will be an important focus of future work and could reveal important modifications to our mechanistic model based on the *Bt* Sus.

### ***Bacteroides*** **glycan catabolism is involved in intestinal disease**

The work detailed in this dissertation strengthens our understanding of the mechanisms used by an abundant group of gut bacteria to access carbohydrates. It has been known for quite some time that carbohydrate metabolism is a major factor that determines which commensal species are present and their abundance in the gut, which can in turn influence the health of the host. However, recent studies are uncovering that carbohydrate metabolism, specifically by gut *Bacteroides*, can directly influence the behavior of intestinal pathogens and can either contribute to or prevent intestinal disease.

***Bacteroides glycan catabolism can enhance infection by intestinal pathogens***

To colonize the gut and cause disease, pathogenic bacteria must compete with the trillions of bacteria that reside there. Several recent studies have revealed that many intestinal pathogens have evolved to take advantage of the catabolic activities of the gut microbiota. In a mouse model of infection the ability to catabolize sialic acid, a sugar found in the host mucosal layer, enhanced colonization levels of both *Clostridium difficile* and *Salmonella typhimurium.* However, neither of these organisms expresses the sialidases necessary to liberate this sugar from mucin glycoproteins, suggesting they scavenge sialic acid that has been cleaved by other species. *Bt* has sialidase activity, and it was shown that the presence of *Bt* enhanced colonization of *C. difficile* and *S. typhimurium* in the mouse gut in a manner that was dependent on expression of the *Bt* sialidase[32]. Here we see two distantly related pathogens that both take advantage of the expanded saccharolytic ability of the *Bacteroides*, using the nutrients that *Bt* liberates (but does not use itself) to achieve sufficient titers to cause disease.

In addition to using *Bacteroides*-liberated sugars as nutrients, some intestinal pathogens use them as environmental cues to regulate expression of virulence genes. Enterohaemorrhagic

*E. coli* (EHEC) has a complex signal cascade that uses multiple signals to time expression of its locus of enterocyte effacement (LEE). Fucose, a sugar abundant in the host mucus layer, is one of the signals feeding into this cascade to repress expression of the LEE[33]. LEE repression in the presence of fucose ensures that these virulence genes are not expressed in the mucus layer but only when EHEC has penetrated this layer and reached the intestinal epithelium. Again, this intestinal pathogen does not encode the necessary enzymes to cleave fucose from mucosal glycoproteins and relies on resident gut microbes to perform this action. The presence of *Bt*, which encodes multiple fucosidases, can indeed modulate LEE expression levels when co-cultured with EHEC in mucin media[33]. Here we see that an intestinal pathogen uses microbiota-liberated sugar signals to assess its precise location in the gut, leading to a more effective infection and increased disease severity.

In both of the cases discussed above, pathogenic microbes use the glycosidic abilities of commensal *Bacteroides* to enhance their ability to colonize and cause disease in the mammalian intestine. Both of these studies showed that *Bt* was capable of modulating the behavior of the intestinal pathogens by cleaving either sialic acid or fucose. Although *Bt* is not the only species capable of liberating these mucosal sugars, it is likely that the expanded saccharolytic abilities of the Bacteroidetes[10,34] makes them particularly important at this pathogen-commensal interface. The cross-talk between the Bacteroidetes and intestinal pathogens via glycosidic activity represents an interesting future target for preventing and/or treating intestinal disease; however, more studies are needed to fully understand the mechanisms governing this interaction.

***Host glycan catabolism may play a role in Bacteroides related colitis***

Dysregulation of the immune system in the intestine can lead to a chronically inflamed state known as inflammatory bowel disease (IBD). The microbiota plays a critical role in progression of this disease and is in fact required for development of IBD in most animal models[35]. The protective mucus layer that overlies the epithelium is a critical barrier that separates the gut microbiota from the host tissue. Penetration of bacteria into the inner mucus layer is associated with intestinal inflammation and colitis[36-38]. The host mucus layer is composed of proteins that are heavily glycosylated with a wide variety of sugars. Some microbial species have evolved strategies to scavenge host mucosal sugars as a nutrient source. The *Bacteroides* specifically appear to be enriched in this capacity[16,34,39,40] although certain species of *Ruminococcus*[41], *Akkermansia*[42] even *E. coli* (cleaves the protein backbone)[43] have also been shown to do this.

A recent study published by Thaddeus Stappenbeck's lab sought to identify specific bacterial populations responsible for initiating inflammation in a genetically susceptible mouse model of IBD. Antibiotic treated susceptible mice were gavaged with different bacterial populations cultured using selective media. Mice that were given the gavage enriched for *Bacteroides* species developed the most severe inflammation, and colonization with either *Bt* or *Bacteroides vulgatus* alone was sufficient to induce colitis in a germ-free susceptible mouse. Interestingly, gavage with commensal Enterobacteriaceae did not induce inflammation despite the fact that this group was enriched in the microbiota of diseased mice[44]. The *Bacteroides* as a genus are enriched in their capacity to degrade host mucosal glycans, a trait that may present an interesting potential mechanism for *Bacteroides* induced colitis. The inability of the Enterobacteriaceae to induce this colitis fits with this model as this bacterial group is largely

unable to degrade complex glycans and instead scavenges mono and disaccharides liberated by other species[45].

Furthermore, Sus-like systems have recently been shown to contribute to long-term colonization of the mucus layer in certain *Bacteroides* species. In a recent study by Sarkis Mazmanian's lab it was found that pre-colonization of germ-free mice with a particular *Bacteroides* species precluded later colonization with the same species, but not with a different species of *Bacteroides*[46], suggesting that these bacteria are filling species-specific niches in the gut. A transposon mutant screen in *Bacteroides fragilis* identified a single genetic locus that, when disrupted, allowed for subsequent colonization by wild-type *B. fragilis*. This genetic locus was identified as encoding a Sus-like system and conferred the ability to colonize intestinal crypts (dubbed the crypt colonization factor locus*, ccf*). It was hypothesized that the *ccf* Sus-like system likely targets, or interacts with, a host glycan found within crypts, giving the bacteria the ability to occupy this microenvironment. Finally, it was shown that the *ccf* locus allowed *B. fragilis* to recover from various perturbations of the gut environment (antibiotic treatment, infection with a pathogen) and maintain a stable and long-term colonization of the host[46]. Here we see that a Sus-like system confers the ability to colonize a microenvironment that is very closely associated with the host epithelium, the intestinal crypt. Colonization of this environment may lead to activation of certain immune pathways and could trigger inflammation in susceptible individuals. Furthermore, crypt colonization provides a certain level of protection from gut perturbations, such as antibiotic treatment, and leads to stable, long-term colonization of the host.

Evidence is emerging that host glycan utilization by the *Bacteroides* may be involved in the development of IBD. Certain species of *Bacteroides* are sufficient to cause disease in a genetically susceptible mouse model of IBD, and a Sus-like system confers *B. fragilis* with the

ability to colonize intestinal crypts. These host glycan systems therefore represent a potential "virulence factor" that is highly context specific, as the majority of the time these organisms are innocuous, even beneficial to the host. Future work is needed to determine the link between *Bacteroides* and IBD, and what role host glycan utilization plays in this process.

## Future Research

### *Sus complex assembly and structure*

The work detailed in this dissertation, along with much of the research on the *Bt* Sus, has focused on the structure and function of individual Sus proteins. An area that remains understudied is how the Sus outer-membrane proteins (OMPs) interact and work together to bind, degrade and import starch. Early work from Abigail Salyers' lab uncovered evidence that at least some of the Sus proteins form a physical complex with one another. When lysates from Sus-expressing formaldehyde cross-linked *Bt* cells were run on a native gel, the same high molecular weight bands appeared when blotted with either anti-SusC or anti-SusD antibodies, suggesting formation of a SusC-SusD complex[15]. SusC also exhibited higher sensitivity to proteolytic cleavage on the surface of *Bt* cells not expressing SusE or SusF compared to wild-type *Bt*, suggesting that SusC may also interact with these Sus OMPs, resulting in its protection from proteolysis. Similarly, SusE exhibited higher proteolytic sensitivity when SusF was not present, suggesting an interaction between these two structural homologs[15]. These seminal studies provide preliminary evidence of interactions between Sus OMPs. However, we now have more refined techniques for genetic manipulation of *Bt* that will allow us to more closely examine how the loss of one or more Sus OMPs, or even mutations of specific amino acids, affects formation of the putative Sus complex.

Evidence of direct protein interactions has been observed in other Sus-like systems as well. In *C. canimorsus* a streptavidin tag was added to the SusC-like GpdC and used to capture this protein from membrane preparations. Proteins that co-purified with GpdC were identified by western blot and mass spectrometry. The two enzymes in this system, GpdG and SiaC, could be detected by western blot, however the other components of this system (GpdD, GpdE and GpdF) could only be detected by mass spectrometry[27], suggesting that particular components of this system form more stable interactions while others may be transient. Studies described in Appendix I support a model where *Bt* Sus complex formation is dynamic; therefore, a similar approach to that used in the *C. canimorsus* study could be useful for identifying interaction partners. Protein precipitation coupled to mass spectrometry could identify Sus OMP interactions that are transient and therefore unable to be captured by less sensitive methods. Determining which Sus OMPs form physical interactions is an important part of understanding how these proteins work together. However, for the complete picture we must also examine the dynamics of Sus complex assembly in a live *Bt* cell.

Developing imaging technology is allowing for the visualization of dynamic protein interactions in real time in live cells. Single molecule fluorescent imaging was recently used to visualize interactions between Sus OMPs on the surface of live *Bt* cells revealing interesting aspects of Sus complex formation shown in Appendix I. Fluorophore labeled SusG was found to co-localize with antibody labeled SusD, SusE, SusF and SusG. The level of co-localization increased in the presence of starch compared to glucose or maltose (**Figure A.6**), suggesting that formation of the Sus complex is starch-induced. Interestingly, SusG was found to exist primarily as monomers or dimers when *Bt* is grown in glucose but formed clusters in the presence of starch (**Figure A.11**), suggesting that the starch-induced Sus complex may contain several of each of

the Sus OMPs. SusG proteins were found to separate into two distinct populations distinguished by their diffusion rates on the cell surface. The $D_{fast}$ population is hypothesized to be freely diffusing SusG and the $D_{slow}$ population hypothesized to be SusG proteins whose movement is confined by interaction with starch or one or more Sus OMPs. Loss of SusD or loss of SusE and SusF increased the percentage of $D_{fast}$ SusG molecules (**Table A.2**), suggesting that loss of these proteins destabilizes the Sus complex, increasing the amount of non-complex associated SusG. In addition, loss of SusD or SusE and SusF increased average diffusion rate of the $D_{slow}$ population (**Table A.2**), demonstrating that the remaining complex-associated SusG was more motile, presumably because it is no longer confined by its interactions with SusD or SusE,F.

The work in Appendix I sheds new light on Sus complex assembly and supports a model where Sus OMP interaction is highly dynamic. The putative fluidity of Sus-like systems offers an explanation for how multiple enzymes (required to degrade complex substrates) can all interact with the substrate and SusCD-like proteins, which would be difficult to account for in a static model. However, the molecular details of how these proteins come together are still unknown. In chapter III we show that SusD has a critical binding-independent function that we hypothesize is involved in SusC interaction and/or complex assembly. The TPR motif of SusD[18] is an intriguing candidate for facilitating Sus OMP interaction, however its ability to mediate protein-protein interactions remains to be proven. In chapter III targeted amino acid changes to the SusD binding site were made to study the function of this specific portion of SusD. A similar strategy could be used in combination with the imaging techniques detailed in Appendix I to identify specific residues that are critical for complex assembly. This approach would allow us to separate the different roles of this multifunctional protein and discern which functions are critical during different stages of *Bt* starch growth. Facilitating Sus complex assembly would be a novel

function for SusD and may provide insight into why SusD homologs are often more essential than the SusEF-like proteins, despite both being non-enzymatic glycan binding proteins.

In contrast to SusD, the other four Sus OMPs do not contain any domains or motifs that are typically associated with facilitating protein-protein interactions. Identification of how these proteins interact molecularly would not only contribute to our understanding of Sus-like systems but may lead to discovery of novel modes of protein-protein interaction.

**Final Conclusions**

The work presented in this dissertation has allowed us to build a detailed, mechanistic model of Sus structure and function. This system is just one of the 88 total Sus-like systems expressed by *Bt*[16] and targets just one of the dozens of glycans this species is capable of utilizing. The ability of this organism to sense available glycans and rapidly express the appropriate Sus-like system(s) is vital to its success in the gut. Here we demonstrate that the Sus binding sites play critical roles in sensing starch, allowing expression of the *sus*, and ensuring starch catalysis occurs at a maximal rate. This model provides valuable molecular insight into degradation of starch, but we hypothesize it can also be applied to the vast number of Sus-like systems in the Bacteroidetes that target dozens of different glycans. The ability to degrade so many different glycans and effectively take advantage of the ever-changing carbohydrate landscape certainly contributes to the success of the Bacteroidetes as gut commensals. These studies lay the groundwork for further investigation of Sus-like systems and are therefore an important first step towards a complete understanding of the mechanism of glycan degradation by a significant portion of the human gut microbiota.

The function of Sus-like systems affects host health both directly (enhancing infection by intestinal pathogens) and indirectly (shaping microbiota composition). Investigation into the role of the microbiota in human health has revealed that this bacterial community performs numerous symbiotic functions while also playing an important role in a number of diseases. Recent studies have shown that modulating the gut microbiota is an effective way to maintain health and prevent or treat intestinal disease. Diet represents an attractive non-invasive way to shape the gut community. However, to use a dietary intervention effectively we must first understand how certain species will respond to particular carbohydrates and why. This work provides insight into how gut bacteria respond to and degrade dietary carbohydrates and paves the way for future studies that will enhance our understanding of how diet affects the composition and metabolic activities of the microbiota. In the future this may lead to design of dietary treatments to treat or prevent intestinal disease and harness the maximum benefits from our microbial gut community.

**References**


1 McNeil, N. I. The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr* **39**, 338-342 (1984).

2 Hamer, H. M. *et al.* Review article: the role of butyrate on colonic function. *Aliment Pharmacol Ther* **27**, 104-119 (2008).

3 Rombeau, J. L. & Kripke, S. A. Metabolic and intestinal effects of short-chain fatty acids. *JPEN. J Parenter Enteral Nutr* **14**, 181S-185S (1990).

4 Duncan, S. H., Barcenilla, A., Stewart, C. S., Pryde, S. E. & Flint, H. J. Acetate utilization and butyryl coenzyme A (CoA):acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Appl Environ Microbiol* **68**, 5186-5190 (2002).

5 Duncan, S. H. *et al.* Contribution of acetate to butyrate formation by human faecal bacteria. *Br J Nutr* **91**, 915-923 (2004).

6 Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nature Rev Microbiol* **10**, 323-335 (2012).

7 Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638 (2005).

8 Salyers, A. A., West, S. E., Vercellotti, J. R. & Wilkins, T. D. Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Appl Environ Microbiol* **34**, 529-533 (1977).

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

10 El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Rev Microbiol* **11**, 497-504 (2013).

11 Reeves, A. R., D'Elia, J. N., Frias, J. & Salyers, A. A. A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J Bacteriol* **178**, 823-830 (1996).

12 D'Elia, J. N. & Salyers, A. A. Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of Bacteroides thetaiotaomicron on starch. *J Bacteriol* **178**, 7173-7179 (1996).

13 Shipman, J. A., Berleman, J. E. & Salyers, A. A. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *J Bacteriol* **182**, 5365-5372 (2000).

14 Cho, K. H., Cho, D., Wang, G. R. & Salyers, A. A. New regulatory gene that contributes to control of Bacteroides thetaiotaomicron starch utilization genes. *J Bacteriol* **183**, 7198-7205 (2001).

15 Cho, K. H. & Salyers, A. A. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **183**, 7224-7230 (2001).

16 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**, 447-457 (2008).

17 Ellrott, K., Jaroszewski, L., Li, W., Wooley, J. C. & Godzik, A. Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. *PLoS Comput Biol* **6**, e1000798 (2010).

18 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**, 1105-1115 (2008).

19 Koropatkin, N. M. & Smith, T. J. SusG: A Unique Cell-Membrane-Associated alpha-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. *Structure* **18**, 200-215 (2010).

20 Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).

21 Larsbrink, J. *et al.* A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506**, 498-502 (2014).

22 Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284**, 24673-24677 (2009).

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

24 Hoffman, M. *et al.* Structural analysis of xyloglucans in the primary cell walls of plants in the subclass Asteridae. *Carbohydr Res* **340**, 1826-1840 (2005).

25 Blanche, P., Bloch, E. & Sicard, D. Capnocytophaga canimorsus in the oral flora of dogs and cats. *J Infect* **36**, 134 (1998).

26 Le Moal, G., Landron, C., Grollier, G., Robert, R. & Burucoa, C. Meningitis due to Capnocytophaga canimorsus after receipt of a dog bite: case report and review of the literature. *Clin Infect Dis* **36**, e42-46 (2003).

27 Renzi, F. *et al.* The N-glycan Glycoprotein Deglycosylation Complex (Gpd) from Capnocytophaga canimorsus Deglycosylates Human IgG. *Plos Pathog* **7** (2011).

28 Larsson, J. M., Karlsson, H., Sjovall, H. & Hansson, G. C. A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19**, 756-766 (2009).

29 D'Andrea, L. D. & Regan, L. TPR proteins: the versatile helix. *Trends Biochem Sci* **28**, 655-662 (2003).

30 Bolam, D. N. & Koropatkin, N. M. Glycan recognition by the Bacteroidetes Sus-like systems. *Curr Opin Struct Biol* **22**, 563-569 (2012).

31 Koropatkin, N., Martens, E. C., Gordon, J. I. & Smith, T. J. Structure of a SusD homologue, BT1043, involved in mucin O-glycan utilization in a prominent human gut symbiont. *Biochemistry* **48**, 1532-1542 (2009).

32 Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96-99 (2013).

33 Pacheco, A. R. *et al.* Fucose sensing regulates bacterial intestinal colonization. *Nature* **492**, 113-117 (2012).

34 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 Environ Microbiol* **33**, 319-322 (1977).

35 Peloquin, J. M. & Nguyen, D. D. The microbiota and inflammatory bowel disease: insights from animal models. *Anaerobe* **24**, 102-106 (2013).

36 Van der Sluis, M. *et al.* Muc2-deficient mice spontaneously develop colitis, indicating that Muc2 is critical for colonic protection. *Gastroenterology* **131**, 117-129 (2006).

37 Johansson, M. E. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc Natl Acad Sci U S A* **105**, 15064-15069 (2008).

38 Johansson, M. E. *et al.* Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model. *PLoS One* **5**, e12238 (2010).

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

40 Marcobal, A., Southwick, A. M., Earle, K. A. & Sonnenburg, J. L. A refined palate: bacterial consumption of host glycans in the gut. *Glycobiology* **23**, 1038-1046 (2013).

41 Hoskins, L. C. *et al.* Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *J Clin Invest* **75**, 944-953 (1985).

42 Derrien, M., Collado, M. C., Ben-Amor, K., Salminen, S. & de Vos, W. M. The Mucin degrader Akkermansia muciniphila is an abundant resident of the human intestinal tract. *Appl Environ Microbiol* **74**, 1646-1648 (2008).

43 Grys, T. E., Siegel, M. B., Lathem, W. W. & Welch, R. A. The StcE protease contributes to intimate adherence of enterohemorrhagic Escherichia coli O157:H7 to host cells. *Infect Immun* **73**, 1295-1303 (2005).

44 Bloom, S. M. *et al.* Commensal bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease. *Cell Host Microbe* **9**, 390-403 (2011).

45 Winter, S. E., Lopez, C. A. & Baumler, A. J. The dynamics of gut-associated microbial communities during inflammation. *EMBO Rep* **14**, 319-327 (2013).

46 Lee, S. M. *et al.* Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501**, 426-429 (2013).

# Appendix

## Super-resolution imaging captures carbohydrate utilization dynamics in human gut symbionts


### Abstract

Gut microbes play a key role in human health and nutrition by catabolizing a wide variety of glycans via enzymatic activities that are not encoded in our genomes. The ability to recognize and process carbohydrates strongly influences the structure of the gut microbial community. While the effects of diet on the microbiota are well documented, little is known about the molecular processes driving metabolism. To provide mechanistic insight into carbohydrate catabolism in gut symbionts, we studied starch processing in real time in the model *Bacteroides thetaiotaomicron* Starch Utilization System (*Bt* Sus) by single-molecule fluorescence. Although previous studies have explored Sus protein structure and function, the transient interactions, assembly and collaboration of these outer membrane proteins have not yet been elucidated in live cells. Our live-cell super-resolution imaging reveals that the polymeric starch substrate dynamically recruits Sus proteins, serving as an external scaffold for bacterial membrane assembly of the Sus complex, which in turn efficiently captures and degrades starch. Furthermore, by simultaneously localizing multiple Sus outer membrane proteins on the *Bt* cell surface, we have characterized the dynamics and stoichiometry of starch-induced Sus complex assembly on the molecular scale. Finally, based on Sus protein knockout strains, we have discerned the mechanism of starch-induced Sus complex assembly in live anaerobic cells with nanometer-scale resolution. Our insights into this conserved nutrient uptake mechanism pave the

way for the development of dietary or pharmaceutical therapies to control Bacteroidetes in the intestinal tract to enhance human health and treat disease.

## Introduction

The human gut contains trillions of densely colonized bacteria that directly influence our health[1]. The majority of these symbionts have a beneficial relationship with humans and promote the degradation of host-indigestible complex glycans, producing short-chain fatty acids that can be utilized by both microbes and humans[2,3]. To efficiently compete for both dietary and host-derived mucosal glycans, gut microbes have evolved a number of different strategies that allow them to scavenge nutrients in the densely populated human gut[1,4].

*Bacteroides thetaiotaomicron (Bt)*, a prominent Gram-negative anaerobic bacterial symbiont in the human gut, metabolizes over a dozen different glycans[5,6]. The Starch Utilization System (Sus) is a well-known multi-protein system that is essential for *Bt* to catabolize starch, a large glucose polymer that is abundant in the human diet. Sus consists of eight proteins, SusRABCDEFG (**Figure A.1**)[7], five of which are involved in starch acquisition at the cell surface. Based on previous biochemical, structural and genetic analysis, the outer membrane-associated SusCDEF assist starch binding to the cell surface[8-12], while SusG, an α-amylase, degrades starch into smaller oligosaccharides[13,14]. SusC, a TonB-dependent transporter, imports these oligosaccharides to the periplasm for further degradation into mono- and disaccharides by SusA and SusB[5,15]. The transcriptional regulator SusR activates Sus expression in the presence of starch or starch derivatives such as the disaccharide maltose[16]. Similarly patterned protein systems, termed Sus-like systems, comprise ~18% of the *Bt* genome and have been identified in the majority of sequenced gut Bacteroidetes[17], making the *Bt* Sus an important model for studying glycan acquisition by gut bacteria.

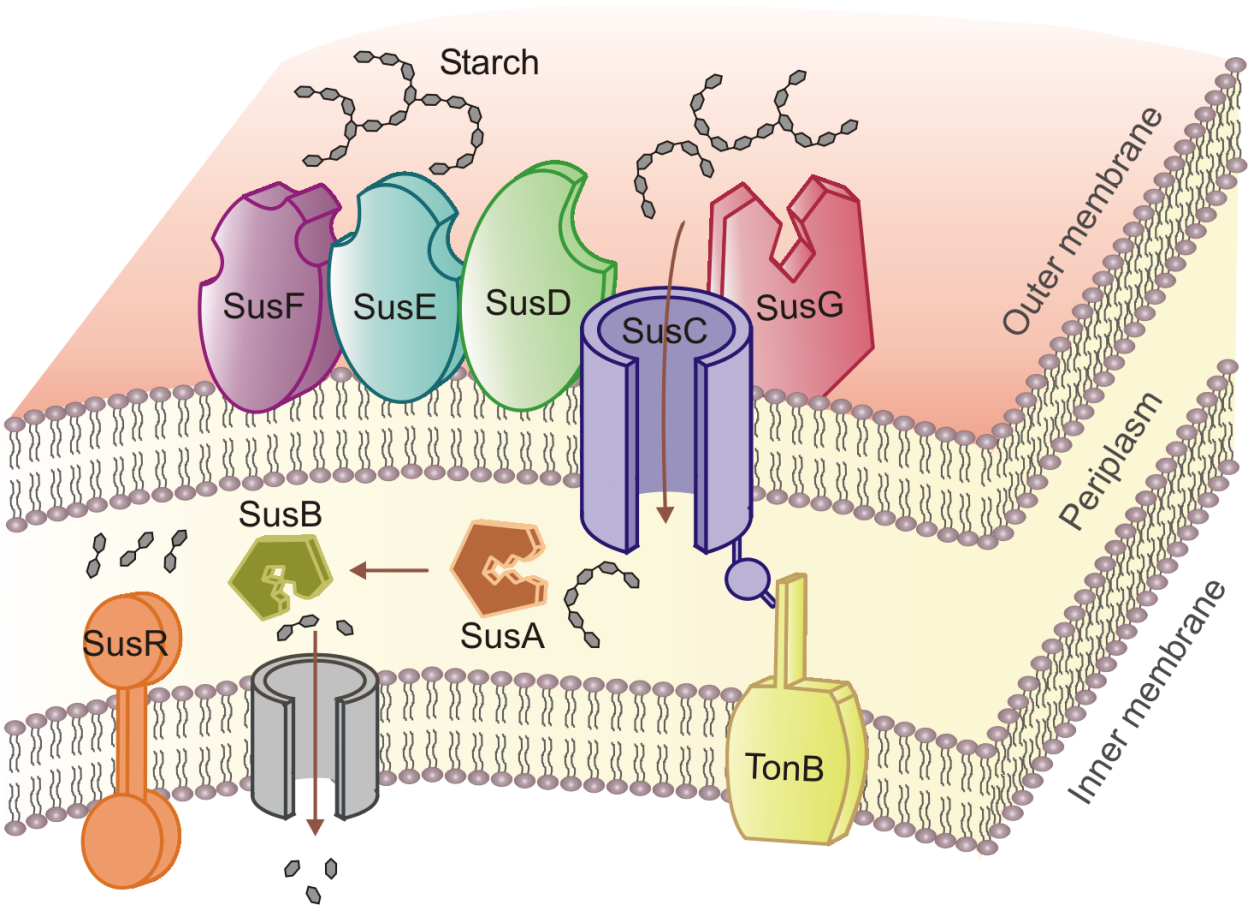



**FIGURE A.1. Model for starch catabolism by *Bt* Sus**

Sus consists of eight proteins (SusRABCDEFG), including five outer membrane-associated proteins that promote starch binding, degradation and import. The exact interactions among these proteins and their stoichiometry have not been elucidated with conventional techniques, but nine sites that interact with starch have been discovered by protein structure determination[12].

Although previous studies have explored Sus protein structure and function, the interactions and assembly of these outer membrane proteins (OMPs) have not yet been elucidated in live cells. Formaldehyde cross-linking and non-denaturing gel electrophoresis studies have shown evidence for SusC/SusD interactions[9]. Furthermore, SusE appears to interact with both SusF and SusCD, forming an OMP complex[9]. Together, these ensemble studies provide a static picture of putative protein associations, but do not reveal the transient interactions that occur during starch catabolism in cells. Therefore, to reveal the precise mechanisms of Sus protein assembly and collaboration during starch processing, we have monitored Sus proteins and their dynamic interactions in real time in live microbes.

Fluorescent labeling of proteins is invaluable for studying intracellular biology[18,19]. Despite the power of fluorescence imaging to explore complex biological systems, standard optical microscopy is unable to fully resolve dynamics and biomolecular interactions on length scales smaller than the ~0.5-$\mu$m diffraction limit[20,21]. To overcome the resolution barrier and to reveal assembly and real-time Sus OMP dynamics under anaerobic conditions, we applied single-molecule super-resolution imaging to fluorophore-labeled Sus proteins[20]. Two-color single-molecule imaging of fluorescently-tagged starch substrates and SusG, an enzyme required for starch catabolism[13], enabled the direct observation of Sus-mediated starch degradation in live *Bt.* Furthermore, by simultaneously localizing multiple Sus OMPs in the presence of starch on the *Bt* cell surface, we characterized starch-induced Sus complex assembly with nanometer-scale resolution. Finally, based on Sus protein knockout strains, the mechanism of starch-induced Sus complex assembly was discerned.

## Results

### *Live-cell imaging of SusG*

Fluorescent labeling of proteins presents unique challenges in live-cell imaging of anaerobic bacteria. Most fluorescent proteins (FPs) require oxygen for maturation[22], precluding their use under anaerobic conditions. Recent advances in covalent labeling of proteins with small fluorescent molecules using a fusion partner, such as the HaloTag® (HT) protein, provide promising alternatives to FPs[23,24]. We applied the HaloTag enzymatic labeling technique to monitor SusG in an oxygen-free environment in live *Bt.* To generate the SusG-HT fusion protein, SusG was fused to HT, a modified haloalkane dehalogenase protein (**Figure A.2A**)[14,25]. Comparable growth rates in starch of *Bt* containing wild-type SusG (SusG-WT) and *Bt* with SusG-HT indicate that this SusG modification has minimal effect on Sus complex-mediated starch degradation (**Figure A.2B-D**).

To determine the positions of SusG on the cell membrane, SusG-HT was fluorescently-labeled using a tetramethyl rhodamine (TMR)-HT ligand (L). Super-resolution imaging of fluorophore-labeled SusG-HT (SusG-HTL) in fixed *Bt* cells revealed the random distribution of stationary SusG proteins at discrete places on the cell membrane (**Figure A.3A-C**). To monitor the dynamic behavior of SusG in live cells, it is essential to maintain an oxygen-free environment throughout the detection time. To overcome this challenge, we assembled live bacterial cells on 2% agarose pads containing minimal media, a carbohydrate source and a reducing agent between two tightly sealed coverslips (**Figure A.4A**) in an anaerobic chamber[26]. Cell division was apparent at 37°C in cells assembled as described above (**Figure A.4B**), providing an opportunity to track SusG on the membrane in real time in live anaerobes (**Figure A.3D-F**). **Figure A.3E** and **A.3F** show that SusG-HTL is membrane-localized: the increased

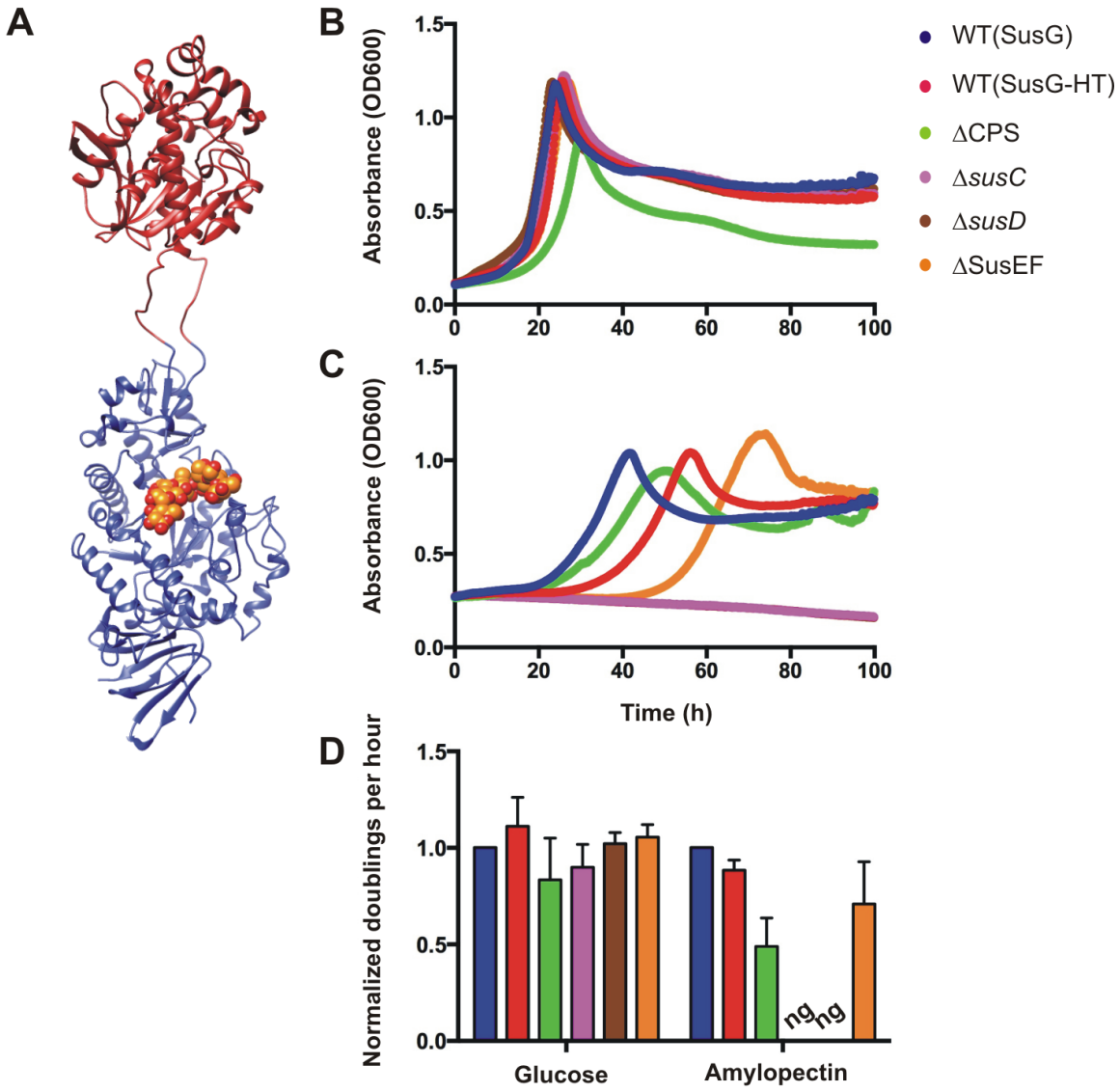



**FIGURE A.2. Structure of SusG-HT fusion protein and comparison of *Bt* growth rates**

**A.** Structure of HaloTag protein (modified haloalkane dehalogenase, red) fused with SusG (blue) containing a bound maltoheptaose molecule (red and yellow spheres) in the active site [14]. To generate the SusG-HT fusion, the carbohydrate binding module (CBM) 58 of SusG, which is dispensable for SusG catalytic activity, was replaced by HT protein. Growth curves of *Bt* strains **B.** in glucose and **C.** in amylopectin. **D.** Normalized doublings per hour showing the effects of mutations on bacterial growth in media containing glucose or amylopectin. Growth curves were obtained by averaging six replicate curves performed at 37°C in an anaerobic chamber. Doubling times are calculated from the exponential growth phase (OD 0.6-0.8) of three separate experiments with the wild-type, WT(SusG), rate normalized to 1.0 (ng denotes no growth). WT(SusG-HT) refers to *Bt* cells containing HaloTag protein-fused SusG. ΔCPS indicates the polysaccharide capsule-free *Bt* cells. ΔSusEF, Δ*susC* and Δ*susD* are the SusEF, *susC* and *susD* knockout strains, respectively. Except for WT(SusG), all other strains include SusG fused to the HaloTag protein.

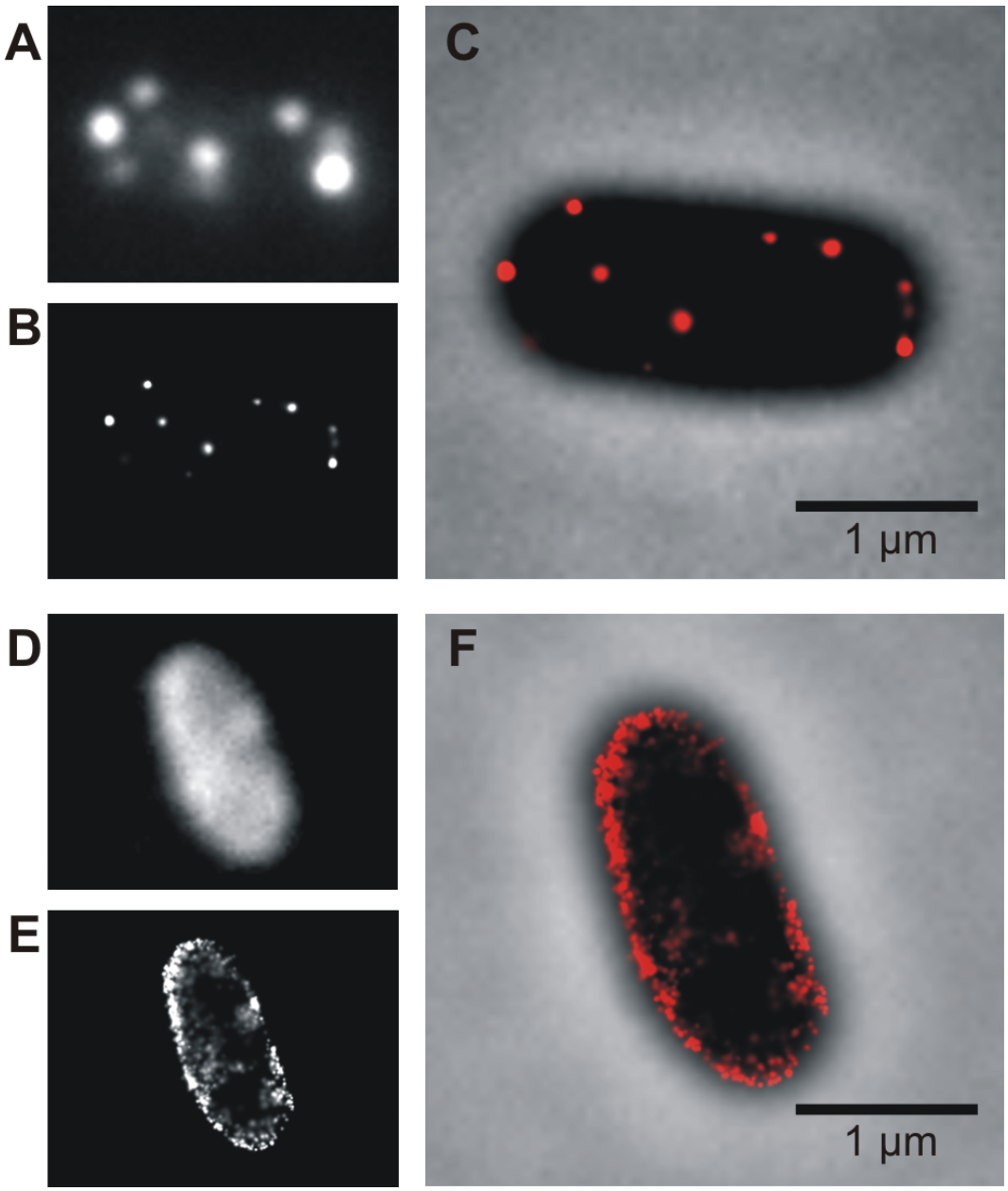



**FIGURE A.3. Single-molecule imaging of HaloTag-labeled SusG in glucose-grown *Bt***

Diffraction-limited (**A.** and **D.**) and reconstructed super-resolution localization (**B.** and **E.**) images of TMR-HaloTag-labeled SusG (SusG-HTL) in fixed and live-cells, respectively. (**C.** and **F.**) Merged phase-contrast cell (black) and super-resolution localization images of SusG-HTL (red) in fixed and live *Bt*, respectively. Fixed-cell images show a few discrete spots representing stationary SusG molecules, while the live-cell images include many spots that correspond to SusG moving on the membrane over time.

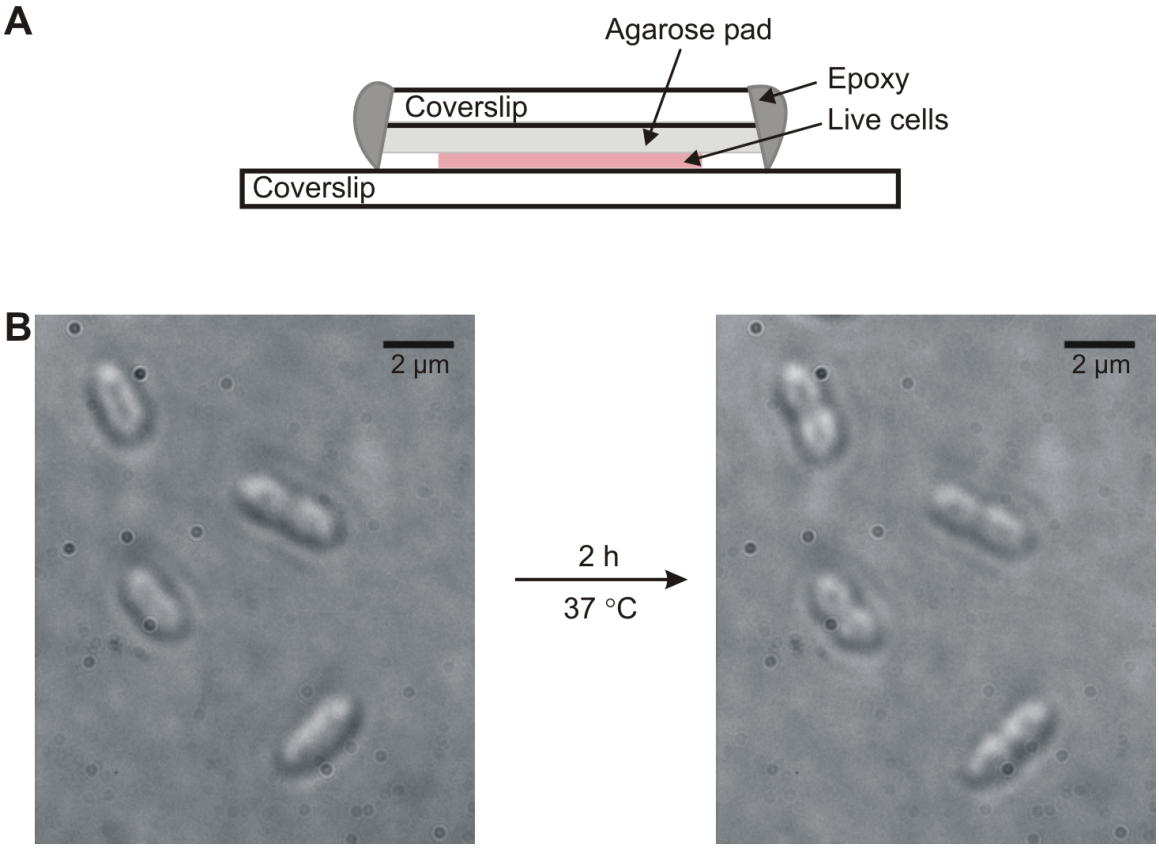



**FIGURE A.4. Live-cell imaging of anaerobic bacteria**.

**A.** Schematic representation illustrating the assembly of *Bt* cells on 2% agarose pads containing minimal media, a reducing agent and a sugar source as explained in the methods. Coverslip edges were sealed with epoxy to maintain an oxygen-free environment for live-cell imaging[26]. **B.** White-light image of *Bt* cells on a slide assembled as in 'A' showing cellular division after incubation for 2 h at 37°C using an objective heater.

concentration of fluorescent spots along the cell edges is as expected for the 2D projection of a cell membrane.

***Sus proteins assemble to process starch***

A key feature of Sus-like systems is the collective action of multiple proteins during glycan binding and degradation[5]. To understand the precise coordinated roles of these proteins during glycan catabolism, we compared the pairwise assembly of Sus OMPs on the cell surface in glucose, which is the monomeric subunit of starch, and in the presence of maize amylopectin (AP), a common plant starch. In addition to HaloTag labeling of SusG, Sus protein-specific antibodies (Abs) were used to concurrently demarcate other Sus proteins on the membrane (**Figure A.5A-E**). Comparison of Alexa 488-conjugated Ab-labeled SusG-WT and SusG-HT (**Figure A.5A and A.5B**) reveals a similar number of Alexa 488 foci per cell, indicating that the introduction of the HT protein does not impede antibody labeling of SusG under these conditions.

As a first step toward understanding Sus complex assembly, SusG-HTL and a second Ab-labeled Sus protein (SusD-, SusE-, or SusF-Ab) were simultaneously monitored in fixed-cells. The Sus protein positions on the membrane were detected with < 20 nm accuracy by fitting individual molecules to a 2D Gaussian function[27]. To accurately measure protein co-localization, super-resolution images of SusG-HTL and Sus-Ab proteins were reconstructed from these positions. Merged images of reconstructed SusG-HTL (red) and Sus-Ab (green) localizations qualitatively indicate protein assembly (**Figure A.6A-D**), although a robust quantitative method is necessary to distinguish differences in Sus protein co-localization with respect to various carbohydrates.

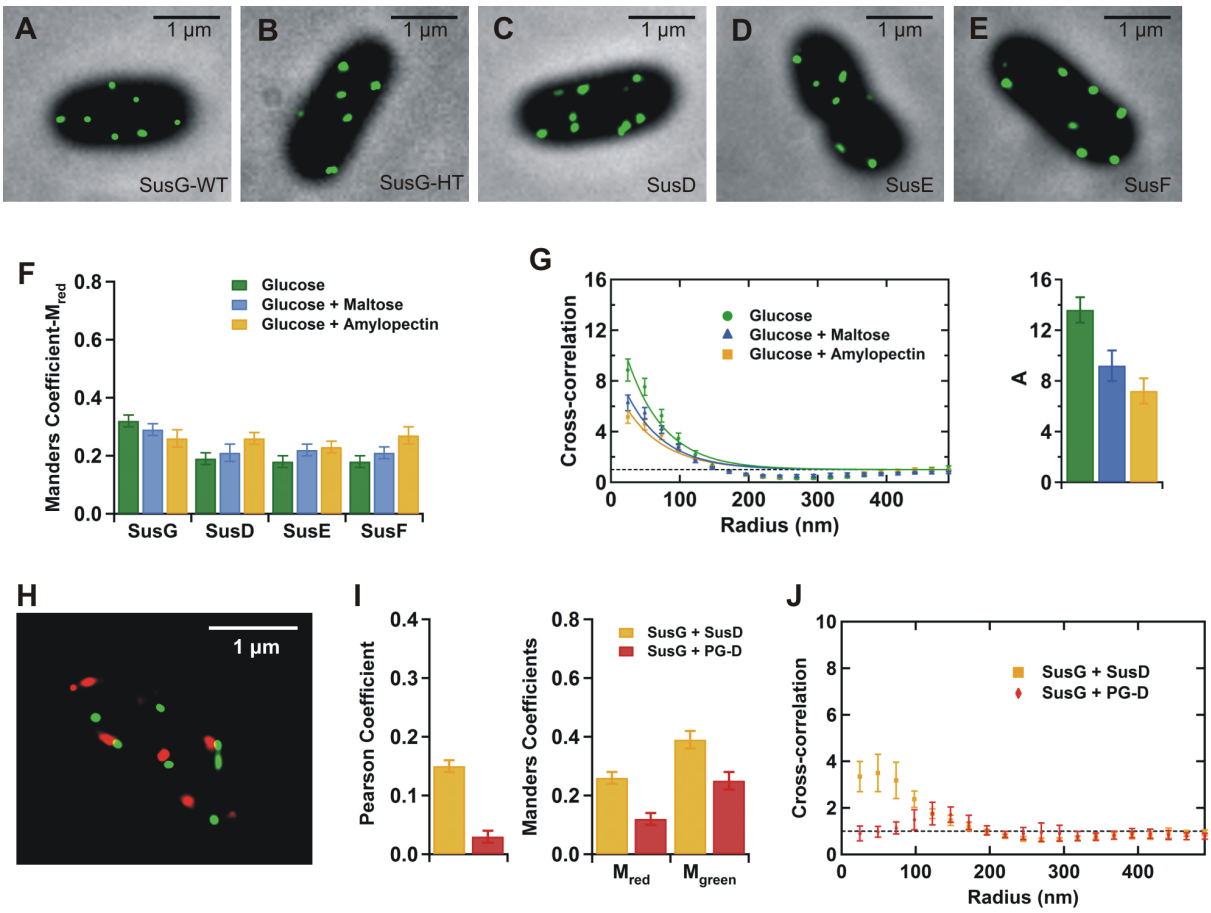



**FIGURE A.5. Antibody labeling and pairwise imaging of Sus proteins**

**A.** Antibody-labeled wild-type SusG (SusG-WT), and **B.** antibody-labeled HaloTag protein-fused SusG (SusG-HT), in cells. (**C.-E.**) *Bt* cells with antibody-labeled SusD, E and F, respectively. All Sus proteins in **A.-E.** were labeled with Alexa 488-congugated antibodies (green). **F.** Manders coefficient ($M_{red}$) showing the quantitative analysis of protein co-localization between SusG-HTL and antibody-labeled other Sus proteins, as indicated. **G.** Cross-correlation functions, $c(r)$, between SusG-HTL and SusG-Ab (left) and cross-correlation amplitude (A) obtained from the fits (right). Protein co-localization experiments were performed in three different sugars as indicated. **H.** Merged images of super-resolution reconstructed localizations of SusG-HTL (red) and antibody-labeled pectic galactan SusD-like protein (PG-D, green) in a representative single *Bt* cell. **I.** Pearson and Manders coefficients comparing the co-localization of SusG with SusD (yellow) and PG-D (red). **J.** Cross-correlation of SusG and SusD (yellow) or SusG and PG-D (red) in *Bt* cells. $c(r) = 1$ for random protein localization (black, dashed line). Error bars represent the standard error of the mean between cells.

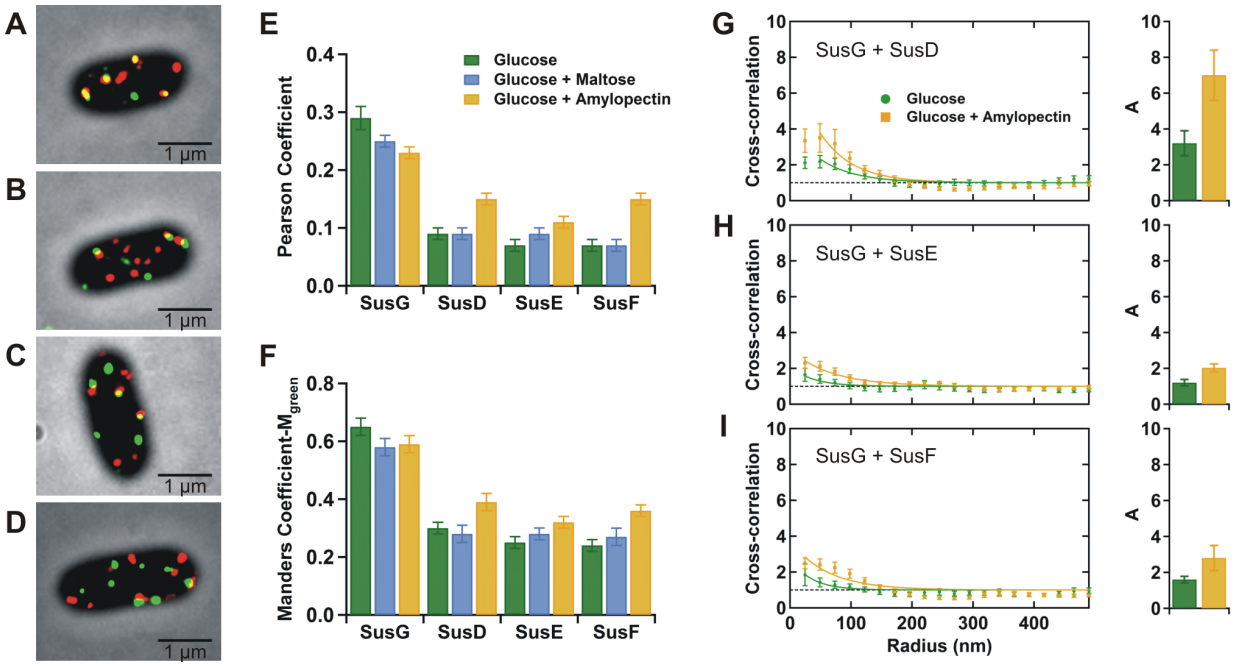



**FIGURE A.6. Co-localization of Sus proteins in fixed-cells**

**A.-D.** Representative merged super-resolution and cell images showing simultaneous localization of SusG-HTL (red) and antibody-labeled SusG, D, E or F (green) respectively, in glucose-grown *Bt*. (**E.** and **F.**) Quantitative analysis of protein co-localization between SusG-HTL and antibody-labeled Sus proteins by Pearson and Manders ($M_{green}$) coefficients. (**G.-I.**) Cross-correlation functions, c(r), (left) and the cross-correlation amplitude (A) obtained from the fit (right) for Sus protein pairs indicated. c(r) = 1 for random protein localization (black, dashed lines). Error bars: standard error of the mean between cells.

Pearson's correlation coefficient (PC) and Manders coefficients ($M_{red}$ and $M_{green}$) quantify biomolecular co-localization based on pixel intensities[28,29]. PC measures the linear correlation between two channels and provides values ranging from –1 (negative correlation) to +1 (positive correlation). Manders coefficients describe co-localization of molecules with respect to an individual channel and increase from 0 to 1 with rising co-localization. Antibody labeling was less efficient than HaloTag labeling due to the stringent protocol that we used to prevent non-specific labeling; consequently the $M_{green}$ coefficient more accurately represents Sus protein co-localization. The Pearson and Manders coefficients indicate higher co-localization levels between SusG and SusD, E or F in the presence of amylopectin than in glucose (**Figures A.6E**, **A.6F**, and **A.5F**). Interestingly, the disaccharide maltose, which enhances Sus protein expression[16] but does not require digestion prior to import and is not highly polymeric, did not enhance protein co-localization as much as amylopectin, which must be degraded to enter the bacterial cell. This suggests that the observed protein co-localization in amylopectin is not due to random Sus protein localization, but rather, this clustering is specifically due to starch-induced complex assembly. As expected, co-localization between SusG-HTL and SusG-Ab was the greatest, irrespective of the sugar source.

Sus OMP assembly was further evaluated by analyzing the cross-correlation of SusG-HTL and Sus-Ab. By fitting each cross-correlation curve to an exponential decay, the degree of co-localization between each pair of variables and the size of the co-localized clusters were defined by an amplitude (A) and a correlation length ($\xi$), respectively[30,31]. Consistent with the Pearson and Manders coefficients, we observed low amplitudes that indicate only moderate co-localization of SusG with SusD, E and F in glucose. This implies that SusG transiently interacts with other Sus proteins in the absence of starch, which may expedite processing when starch

| Cell Type | Protein Pair | Sugar Source | A | ξ (nm) |
|---|---|---|---|---|
| **WT (SusG-HT)** | SusG-HTL + SusG-Ab | Glu (n = 29) | 13.6 ± 1.5 | 54 ± 6 |
| | | Mal (n = 21) | 9.2 ± 1.2 | 55 ± 9 |
| | | AP (n = 22) | 7.2 ± 1.0 | 58 ± 10 |
| | SusG-HTL + SusD-Ab | Glu (n = 30) | 3.2 ± 0.7 | 57 ± 10 |
| | | AP (n = 25) | 7.0 ± 1.4 | 54 ± 9 |
| | SusG-HTL + SusE-Ab | Glu (n = 22) | 1.2 ± 0.2 | 59 ± 11 |
| | | AP (n = 27) | 2.0 ± 0.2 | 68 ± 9 |
| | SusG-HTL + SusF-Ab | Glu (n = 21) | 1.6 ± 0.2 | 56 ± 10 |
| | | AP (n = 22) | 2.8 ± 0.7 | 60 ± 15 |
| **ΔSusEF** | SusG-HTL + SusD-Ab | Glu (n = 10) | 4.3 ± 1.2 | 77 ± 19 |
| | | AP (n = 12) | 8.6 ± 1.5 | 59 ± 7 |

**TABLE A.1. Cross-correlation analysis**

Amplitude of cross-correlation (A) and correlation cluster length (ξ) were obtained after fitting cross-correlation curves for HaloTag-labeled SusG (SusG-HTL) and antibody-labeled Sus proteins (Sus-Ab) to an exponential function, $c(r) = 1 + A \exp(-r/\xi)$, as in Figures A.5, A.6, and A.13. The cross-correlation function $c(r)$, quantifies the increased probability of finding a signal a distance $r$ away from another given signal. Cross-correlation was observed in media containing glucose (Glu), maltose (Mal) or amylopectin (AP), as indicated

becomes available. However, amylopectin enhanced Sus protein co-localization for all three pairs as implied by higher cross-correlation amplitudes (**Figure A.6G-I** and **Table A.1**). This effect was most prominent between SusG and D. Also in agreement with the correlation coefficients, the amplitude for SusG labeled with both HaloTag and antibodies was not enhanced in amylopectin or maltose compared to glucose (**Figure A.5G**). Interestingly, all Sus protein pairs exhibited ~50-nm cluster lengths between SusG-HTL and Sus-Ab proteins regardless of the sugar source or the Ab-labeled Sus protein (**Table A.1**).

*Bt* can express 88 different gene clusters to process various glycans by forming Sus-like systems[17,32]. To verify that the observed Sus OMP co-localization is the result of specific interactions among Sus proteins, SusG-HTL and an Ab-labeled SusD-like protein specific for pectic galactan (PG-D) were simultaneously monitored using amylopectin and pectic galactan as sugar sources[33]. Although fluorescence imaging indicates that both starch and pectic galactan utilization systems can be expressed simultaneously in *Bt*, we observed no significant co-localization between SusG and PG-D compared to SusG and SusD (**Figure A.5H-J**).

To further test our hypothesis that the Sus proteins cluster in the presence of starch, random membrane protein localizations were simulated in MATLAB, generating red and green foci corresponding to SusG-HTL and Sus-Ab, respectively. These Monte Carlo simulations confirm random co-localization contributes minimally to the Manders coefficients and is not detected in cross-correlation analysis. Furthermore, $M_{red}$ and $M_{green}$ depend only weakly on the number of proteins within the experimentally observed range (~10-15 protein foci/cell; **Figure A.7A and A.7B**). Next, we investigated the effect of microscope focus region on random protein localization. Except for the middle of the cell, which is experimentally unattainable as our ~1-$\mu$m focus depth is similar to the *Bt* cell diameter, apparent protein co-localization was not

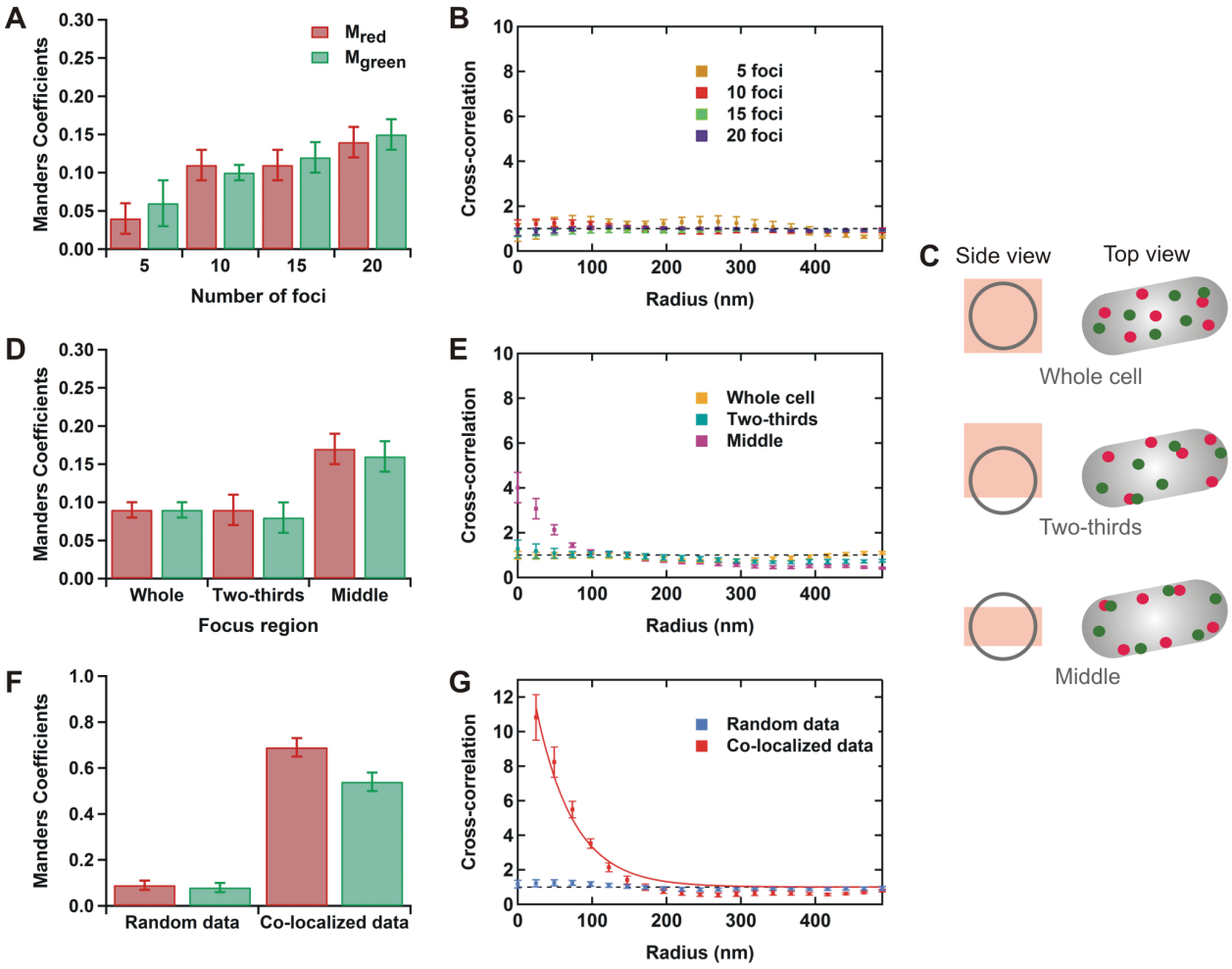



**FIGURE A.7. Simulations of membrane protein localizations**

**A.** Manders coefficients ($M_{red}$ and $M_{green}$) and **B.** cross-correlation demonstrating the effect of number of foci on random protein localization. **C.** Schematic representation of the focus region, i.e., the portion of cell illuminated by the ~1-$\mu$m focus depth of the high-NA microscope (left) and the top view of the corresponding cell (right). Red and green foci represent the HaloTag-labeled SusG and antibody-labeled Sus proteins, respectively. (**D.** and **E.**) Manders coefficients and cross-correlation showing the effect of focus region on random protein co-localization. **F.** Comparison of simulated random versus simulated co-localized (within 50 nm) protein distributions by Manders coefficients and **G.** corresponding comparison of cross-correlation. Error bars represent the standard error of the mean obtained from 20 simulated cells.

significantly affected by the microscope focus (**Figure A.7C-E**). In contrast to random protein localization, we were able to reproduce the experimentally observed protein co-localization using simulated co-localized data with ~50-nm cluster lengths (**Figure A.7F and A.7G**). These simulations support our conclusion that the measured co-localization is a consequence of starch-induced Sus OMP assembly.

***Starch confines SusG motion***

The protein diffusion rate is inversely proportional to the size of the protein or protein complex. Accordingly, changes in mobility can provide insight into how an individual protein associates with other proteins in cells. To provide a baseline for interactions between SusG and other Sus OMPs, the dynamic behavior of SusG was characterized by live-cell imaging of SusG-HTL in glucose. Single-molecule trajectories demonstrating the movement of individual proteins on the membrane were obtained by tracking localized molecules (**Figure A.8A**)[34]. The observed Mean Square Displacement (MSD) slopes of individual trajectories revealed the presence of at least two distinct SusG populations (**Figure A.8B**): mobile (red) and confined (blue). In glucose, the mobile population predominated, suggesting that SusG tends to diffuse freely along the cell membrane during growth in this simple sugar.

To explore SusG/starch interactions during carbohydrate degradation, two-color single-molecule experiments were performed using Alexa 488-labeled maltoheptaose (MH-Alexa488) or amylopectin (AP-Alexa488) in live *Bt* (**Figure A.9A-C**)[35,36]. Single-molecule SusG-HTL trajectories clearly show dynamic interactions between SusG and starch molecules (**Figure A.8E**, **Figure A.9E)**, and single-step analysis of SusG-HTL shows a preponderance of very small steps at the AP-Alexa 488 location (**Figure A.8G**). For detailed analysis of these

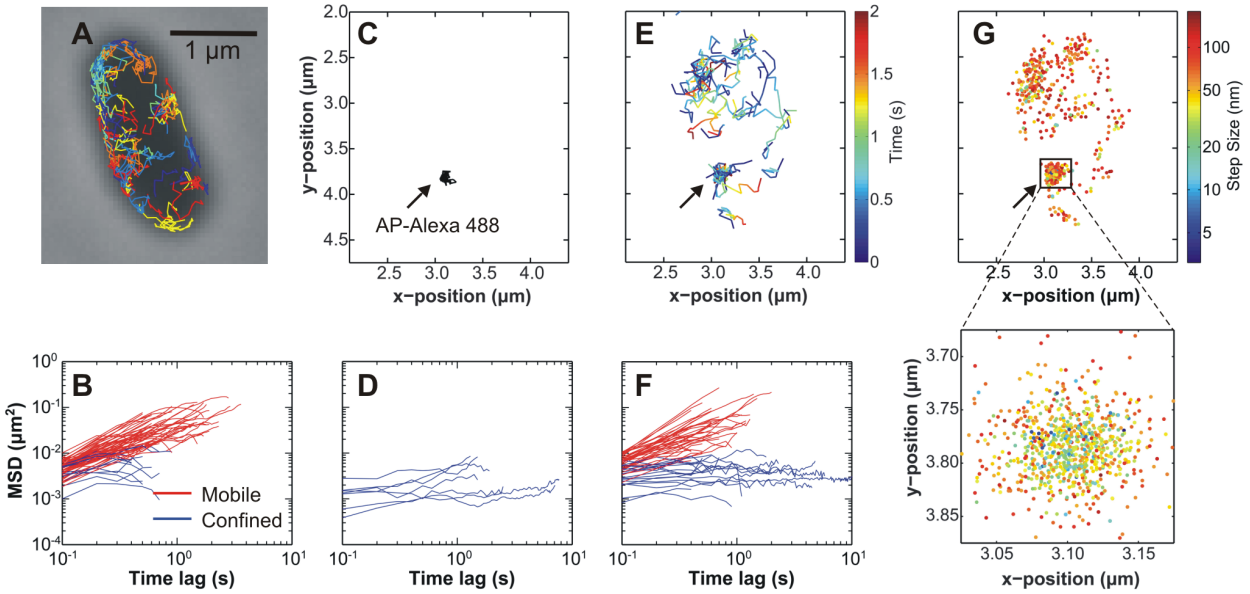



**FIGURE A.8. SusG diffuses heterogeneously but is confined in the presence of starch**

**A.** Single-molecule trajectories of SusG-HTL in glucose (random colors). **B.** Mean square displacement (MSD) vs. time lag for tracks observed on the cell in A. Based on diffusion coefficients ($D$), trajectories were categorized into two subpopulations: mobile (red, $D > 0.01$ $\mu m^2/s$) and confined (blue, $D < 0.01$ $\mu m^2/s$). (**C.** and **D.**) Tracks and MSD plot showing confined movement of Alexa488-labeled amylopectin (AP-Alexa488, black) bound to a cell. (**E.** and **F.**) Time-dependent tracks and MSD plot of SusG-HTL in the presence of AP-Alexa488 (position denoted by arrow). **G.** Spatial distribution of SusG-HTL step-sizes obtained from the tracks in E.

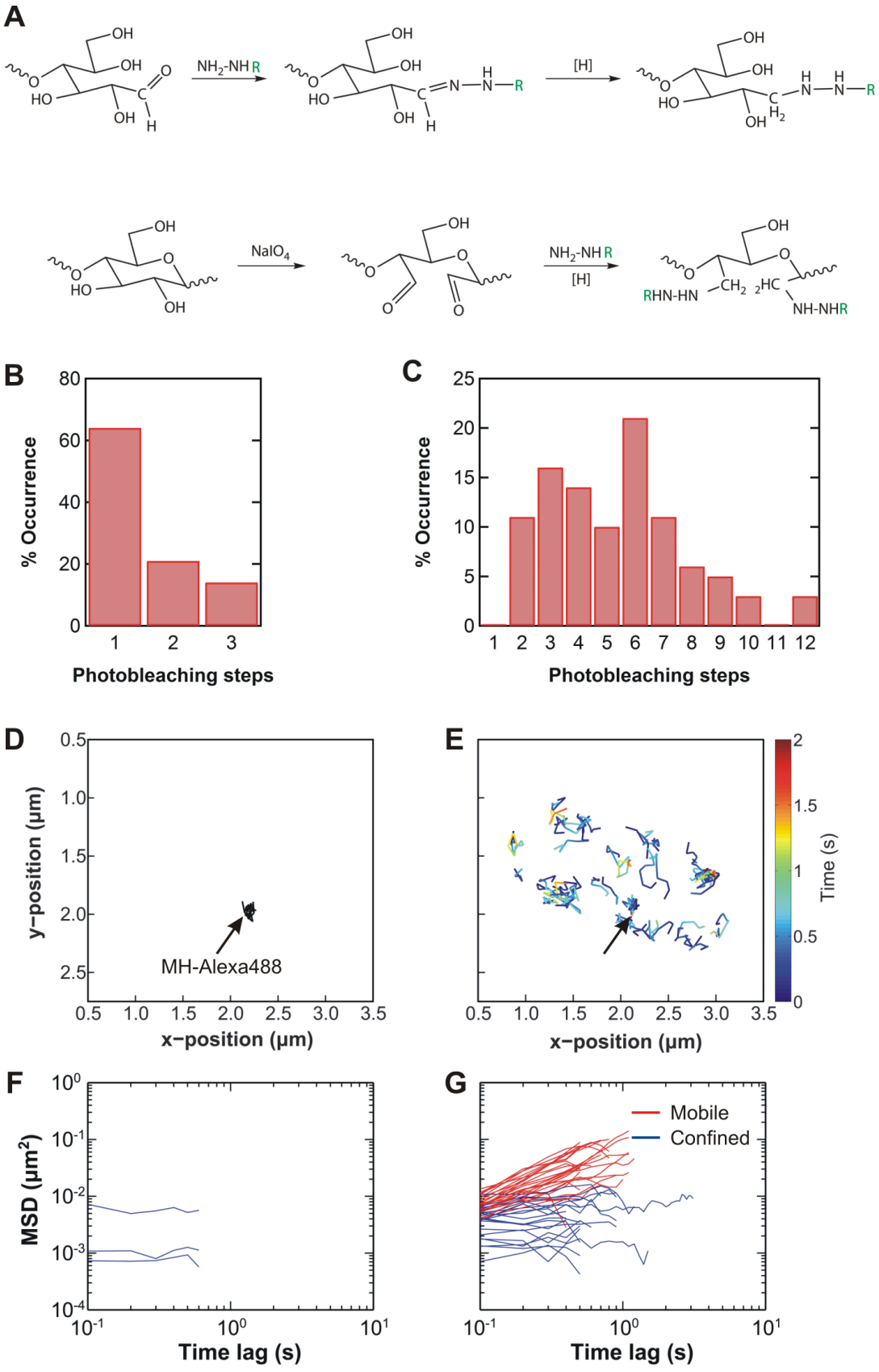

A
NH2-NH R
[H]
NaIO4
RHN-HN
NH-NHR
B
% Occurrence
Photobleaching steps
C
% Occurrence
Photobleaching steps
D
y−position (μm)
x−position (μm)
MH-Alexa488
E
x−position (μm)
Time (s)
F
MSD (μm²)
Time lag (s)
G
Mobile
Confined
Time lag (s)

**FIGURE A.9. Fluorophore labeling of sugars and interactions between SusG and fluorophore-labeled maltoheptaose (MH)**

**A.** Reaction schemes for reductive fluorophore labeling of maltoheptaose (top) and oxidative amylopectin labeling (bottom) using Alexa 488 hydrazide as described in the methods section. (**B.** and **C.**) Occurrence of photobleaching steps detected per fluorescent spot with fluorophore-labeled maltoheptaose (MH-Alexa488) or amylopectin (AP-Alexa488), respectively. Number of photobleaching steps corresponds to the number of fluorescent tags detected per sugar molecule. **D.** Molecular tracks (black) and **F.** corresponding MSD plot of the confined motion of Alexa 488-labeled MH. **E.** Time-dependent molecular tracks of SusG-HTL (colored lines) showing interactions between SusG and MH-Alexa488 in live cells. Black arrow indicates the position of fluorophore-labeled MH observed in 'D'. **G.** MSD plot for SusG diffusion in the presence of MH-Alexa488. Trajectories with diffusion coefficients less than 0.01 $\mu m^2/s$ are categorized as confined (blue) and those with larger coefficients are mobile (red) SusG molecules.

interactions, MSDs were obtained from SusG-HTL molecular tracks in the presence of fluorophore-labeled amylopectin or maltoheptaose. In contrast to the predominantly freely diffusing SusG observed in glucose, the presence of starch increased the proportion of confined SusG molecules (blue curves in **Figure A.8F** and **Figure A.9G**). Since both fluorophore-labeled sugars attached to *Bt* did not show any detectable movements within the experimental observation time (**Figure A.8C** and **A.8D**, **Figure A.9D** and **A.9F**), we attribute the confined SusG population to direct interactions between sugars and SusG, either alone or complexed with other Sus OMPs. The large amylopectin was multiply labeled and easily observed (**Figure A.9C**). Therefore, AP-Alexa488 was used to characterize dynamic interactions between SusG and starch for subsequent analysis.

***SusG exhibits multiple diffusion modes***

Heterogeneous motion of SusG-HTL implies the presence of multiple diffusion modes, even within the trajectory of a single SusG protein (**Figure A.10**). To extract these diffusion coefficients ($D$), single-step analysis was performed by fitting the Cumulative Probability Distribution (CPD) of the squared step sizes to a three-term exponential function that best describes the data (**Figure A.11A-D**)[34,37].

In glucose, mobile SusG-HTL predominantly (61%) diffused rapidly ($D_{fast}$ = 0.020 $\mu m^2/s$); we conclude that this fast movement represents the dynamic behavior of individual, freely diffusing SusG molecules. Less frequently (39%), SusG-HTL diffused slowly ($D_{slow}$ = 0.0050 $\mu m^2/s$), possibly due to interactions with one or more other Sus OMPs (**Figure A.11A-D**, **Figure A.12A** and **A.12B**, and **Table A.2**). In starch, most SusG-HTL (58%) moved slowly ($D_{slow}$ = 0.0015 $\mu m^2/s$), in contrast to the less frequently observed (42%) fast-moving SusG-HTL

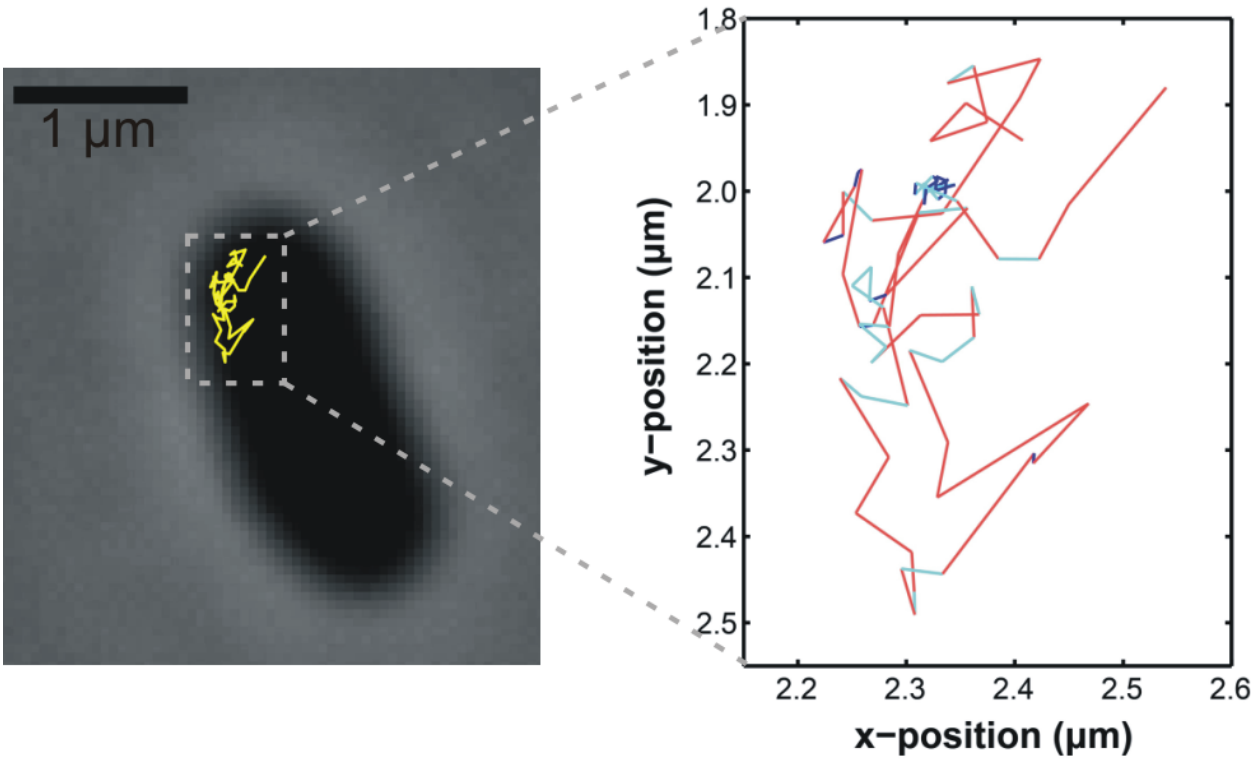



**FIGURE A.10. SusG diffuses heterogeneously in live *Bt***

A typical SusG-HTL single-molecule trajectory superimposed on the phase-contrast cell image (left) and the enlarged trajectory showing different step sizes of SusG-HTL in glucose (right). Red indicates large steps (> 45 nm), corresponding to the fast diffusion mode, cyan indicates small steps (20 – 45 nm), corresponding to the slow diffusion mode and blue indicates SusG-HTL that appears immobile within the localization accuracy (< 20 nm).

molecules ($D_{fast}$ = 0.008 $\mu m^2/s$). The increased proportion of slow-moving SusG-HTL together with the decreased $D_{slow}$ in starch further supports our model of starch-induced Sus OMP assembly. Consistent with the decreased $D_{fast}$ in starch relative to glucose, stoichiometry determined from the number of photobleaching steps[38] indicates that SusG primarily exist as monomers or dimers in glucose, but it tends to form clusters in the presence of starch (**Figure A.11E** and **A.11F**).

In addition to the mobile SusG populations, SusG appeared immobile on the cell membrane ~6-7% of the time, both in glucose and in starch. To test for the possibility that this immobile population resulted from interactions between SusG-HTL and the *Bt* polysaccharide capsule[39], we monitored SusG-HTL dynamics in capsule-free *Bt* cells (ΔCPS). SusG-HTL in ΔCPS behaved similarly to the wild-type in both glucose and amylopectin (**Figure A.12A-D** and **Table A.2**). Thus, we attribute the immobile populations to interactions between SusG and other components on the membrane, as well as to artifacts from imaging inherently 3D motion in 2D. Since the fraction of immobile population remains unchanged in all further analysis, we omit it from future discussion.

***Dynamic interactions among Sus proteins***

To further elucidate the starch-induced Sus complex assembly mechanism, SusG-HTL diffusion was characterized in Sus protein knockout strains in glucose and starch. First, to reveal interactions between SusG and SusC during starch processing, SusG-HTL dynamics were monitored in *sus*C gene knockout cells (Δ*susC*). Although SusC is essential for starch catabolism in *Bt*, the absence of SusC did not affect the fast ($D_{fast}$ = 0.022 $\mu m^2/s$, 61%) or the slow ($D_{slow}$ = 0.0055 $\mu m^2/s$, 39%) SusG-HTL diffusion rates in glucose (**Figure A.12I** and **Table A.2**).

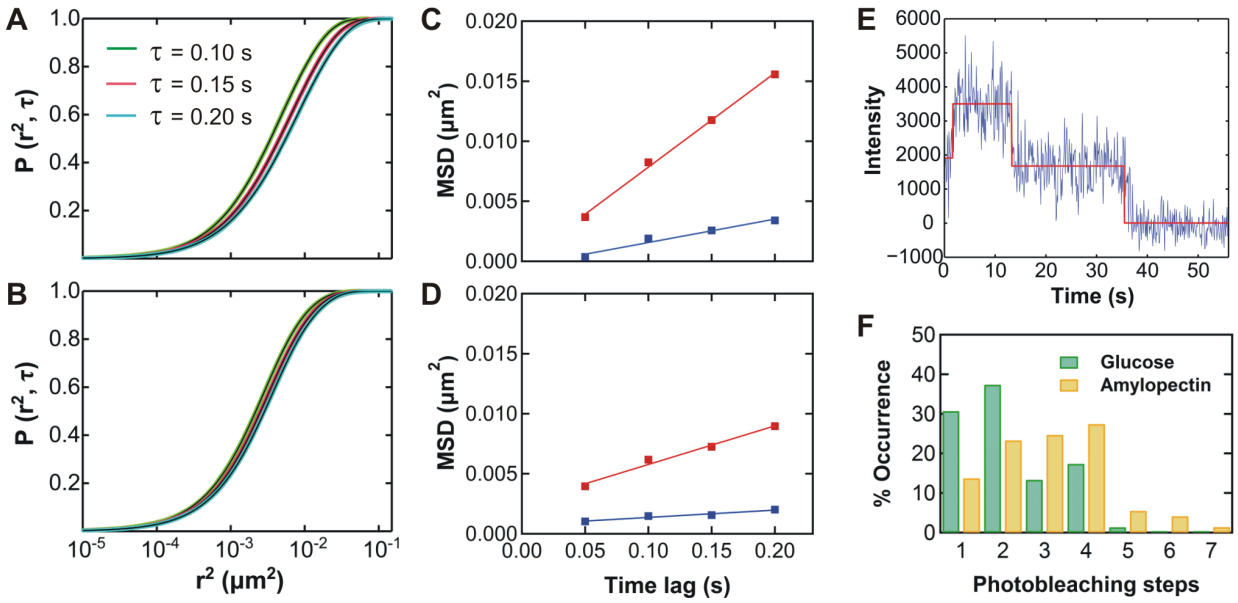



**FIGURE A.11. Cumulative Probability Distribution (CPD) analysis of SusG dynamics and bleaching analysis of SusG stoichiometry**

(**A.** and **B.**) Distributions of squared step sizes ($r^2$) of SusG-HTL fit to a three-term CPD function in glucose and amylopectin, respectively. Raw data (colored lines) and corresponding fits (black lines) were obtained for three different time lags ($\tau$), as indicated. (**C.** and **D.**) MSD vs. $\tau$ obtained from fitting the CPD curves of SusG-HTL in glucose and amylopectin, respectively. The MSD plot slopes reveal fast (red) and slow (blue) diffusion modes for SusG in live cells. **E.** Typical fluorescent intensity trace for TMR-HaloTag-labeled SusG showing multiple photobleaching steps corresponding to several fluorophore-labeled SusG molecules detected in a given foci. The red line fit was obtained from a change-point finding algorithm [38]. **F.** Occurrence of the number of photobleaching steps, revealing the approximate cluster size of SusG molecules in glucose (green) and amylopectin (yellow).

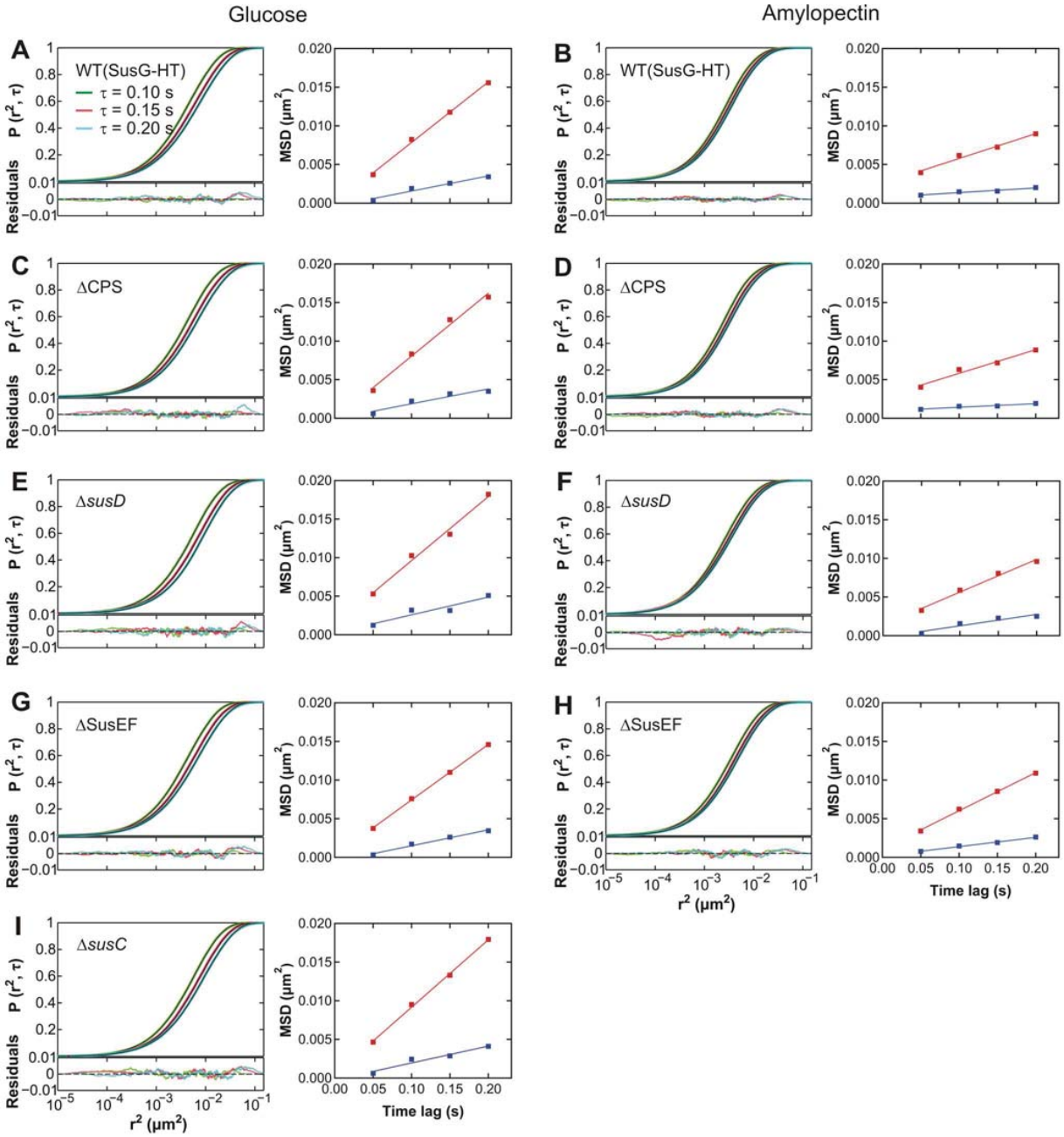

Glucose
Amylopectin
A
WT(SusG-HT)
τ = 0.10 s
τ = 0.15 s
τ = 0.20 s
B
WT(SusG-HT)
C
ΔCPS
D
ΔCPS
E
ΔsusD
F
ΔsusD
G
ΔSusEF
H
ΔSusEF
I
ΔsusC
P (r², τ)
Residuals
MSD (μm²)
r² (μm²)
Time lag (s)

**FIGURE A.12. Cumulative Probability Distribution (CPD) analysis of SusG dynamics**

(**A.** and **B.**) Distribution of squared step sizes of *Bt* SusG-HTL fit to a three-term CPD function (left) and corresponding MSD plot (right) in glucose and amylopectin, respectively. Raw data (colored lines) and corresponding fits (black lines) including residuals are shown for three different time lags ($\tau$) as mentioned in the figure legend. MSD plot slopes corresponding to the average diffusion coefficients for the fast (red) and slow (blue) diffusion modes of SusG. CPD plots (left) and corresponding MSD vs. time lags (right) obtained for SusG-HTL in capsule-free cells (ΔCPS) in **C.** glucose and **D.** amylopectin. CPD plots (left) and corresponding MSD vs. time lags (right) obtained for SusG-HTL in *susD* gene knockout cells (Δ*susD*) in **E.** glucose and **F.** amylopectin. CPD plots (left) and corresponding MSD vs. time lags (right) obtained for SusG-HTL in SusE and SusF proteins knockout cells (ΔSusEF) in **G.** glucose and **H.** amylopectin. **I.** CPD (left) and MSD plot (right) for SusG-HTL diffusion in *susC* gene knockout cells (Δ*susC*) in glucose. All fitting results are tabulated in Table A.2.

| Cell Type | Sugar Source | $D_{fast}$ (μm2/s) | Fast mode (%) | $D_{slow}$ (μm2/s) | Slow mode (%) |
|---|---|---|---|---|---|
| **WT (SusG-HT)** | Glu (n = 22) | 0.020 ± 0.001 | 60.7 ± 1.5 | 0.0050 ± 0.0008 | 39. 3 ± 1.4 |
| | AP (n = 14) | 0.008 ± 0.001 | 41.7 ± 1.9 | 0.0015 ± 0.0003 | 58.3 ± 1.8 |
| **ΔCPS** | Glu (n = 13) | 0.021 ± 0.002 | 52.9 ± 2.6 | 0.0048 ± 0.0010 | 47.1 ± 2.8 |
| | AP (n = 6) | 0.008 ± 0.001 | 36.9 ± 3.8 | 0.0012 ± 0.0003 | 63.1 ± 3.4 |
| **Δ*susD*** | Glu (n = 19) | 0.021 ± 0.002 | 57.4 ± 5.4 | 0.0058 ± 0.0013 | 42.6 ± 5.5 |
| | AP (n = 12) | 0.011 ± 0.001 | 47.6 ± 3.8 | 0.0037 ± 0.0008 | 52.4 ± 4.1 |
| **ΔSusEF** | Glu (n = 18) | 0.018 ± 0.001 | 59.4 ± 3.1 | 0.0053 ± 0.0005 | 40.6 ± 2.8 |
| | AP (n = 16) | 0.012 ± 0.001 | 50.7 ± 4.8 | 0.0030 ± 0.0003 | 49.3 ± 5.4 |
| **Δ*susC*** | Glu (n = 16) | 0.022 ± 0.001 | 61.0 ± 2.5 | 0.0055 ± 0.0010 | 39.0 ± 2.0 |

**TABLE A.2. Summary of CPD analysis of SusG dynamics**

Diffusion coefficients ($D_{fast}$ and $D_{slow}$) and the percentages of molecules in each diffusion mode obtained by CPD analysis as in Figures A.11A-D and A.12. The percentages of mobile SusG molecules in fast or slow modes were calculated by the fraction of molecules in fast ($\alpha$) and slow ($\beta$) diffusion modes at a given time obtained from CPD fitting results. Diffusion coefficients were obtained after analyzing cells in glucose (Glu) or amylopectin (AP). WT (SusG-HT) refers to *Bt* cells containing HaloTag protein-fused SusG. ΔCPS indicates capsule-free cells, ΔSusEF, Δ*susD* and Δ*susC* correspond to SusEF, *susD* and *susC* knockout cells, respectively.

The roles of SusE and SusF in Sus complex assembly were explored by monitoring SusG-HTL dynamics in *Bt* lacking SusE and SusF surface expression (ΔSusEF; **Figure A.12G** and **A.12H**, and **Table A.2**). The loss of SusEF did not alter SusG-HTL diffusion in glucose. Conversely, relative to wild-type cells, in starch, ΔSusEF gave rise to an increased fast mode population ($D_{fast}$ = 0.012 $\mu m^2/s$; 51% for ΔSusEF vs. 42% for WT). This suggests that the Sus complex is destabilized without SusEF. Furthermore, in starch, ΔSusEF *Bt* displayed a two-fold increase in SusG-HTL $D_{slow}$ ($D_{slow}$ = 0.0030 $\mu m^2/s$, 49%). This increase in $D_{slow}$, the diffusion coefficient assigned to the motion of SusG associated with other Sus OMPs, supports the presence of SusE and/or F in the wild-type complex. Regardless of the observed SusG diffusion differences, co-localization between SusG and SusD was not affected by the absence of SusEF (**Figure A.13**). This suggests that SusG interacts with SusD independently of SusEF, either by direct interactions or by mutual interactions with starch.

Finally, to further probe the interaction of SusG and SusD in the presence of starch, SusG-HTL dynamics were monitored in *susD* gene knockout cells (Δ*susD*). SusG-HTL in Δ*susD* showed similar dynamics to wild-type *Bt* in glucose (**Figure A.12E** and **Table A.2**). Furthermore, Δ*susD* had a similar effect as ΔSusEF on SusG-HTL dynamics in starch (**Figure A.12F**): relative to wild-type cells in starch, the absence of SusD in starch increased $D_{slow}$ for SusG-HTL ($D_{slow}$ = 0.0037 $\mu m^2/s$, 52%) and increased the population of fast-moving molecules ($D_{fast}$ = 0.011 $\mu m^2/s$, 48%). The enhanced $D_{slow}$ in Δ*susD* provides evidence that SusD is also a member of the starch-induced Sus complex. Taken together, the absence of any one or several other Sus OMPs did not influence the overall SusG dynamics in glucose, but clearly affected $D_{slow}$ in starch, indicating dynamic associations between SusG and other Sus proteins during starch degradation.

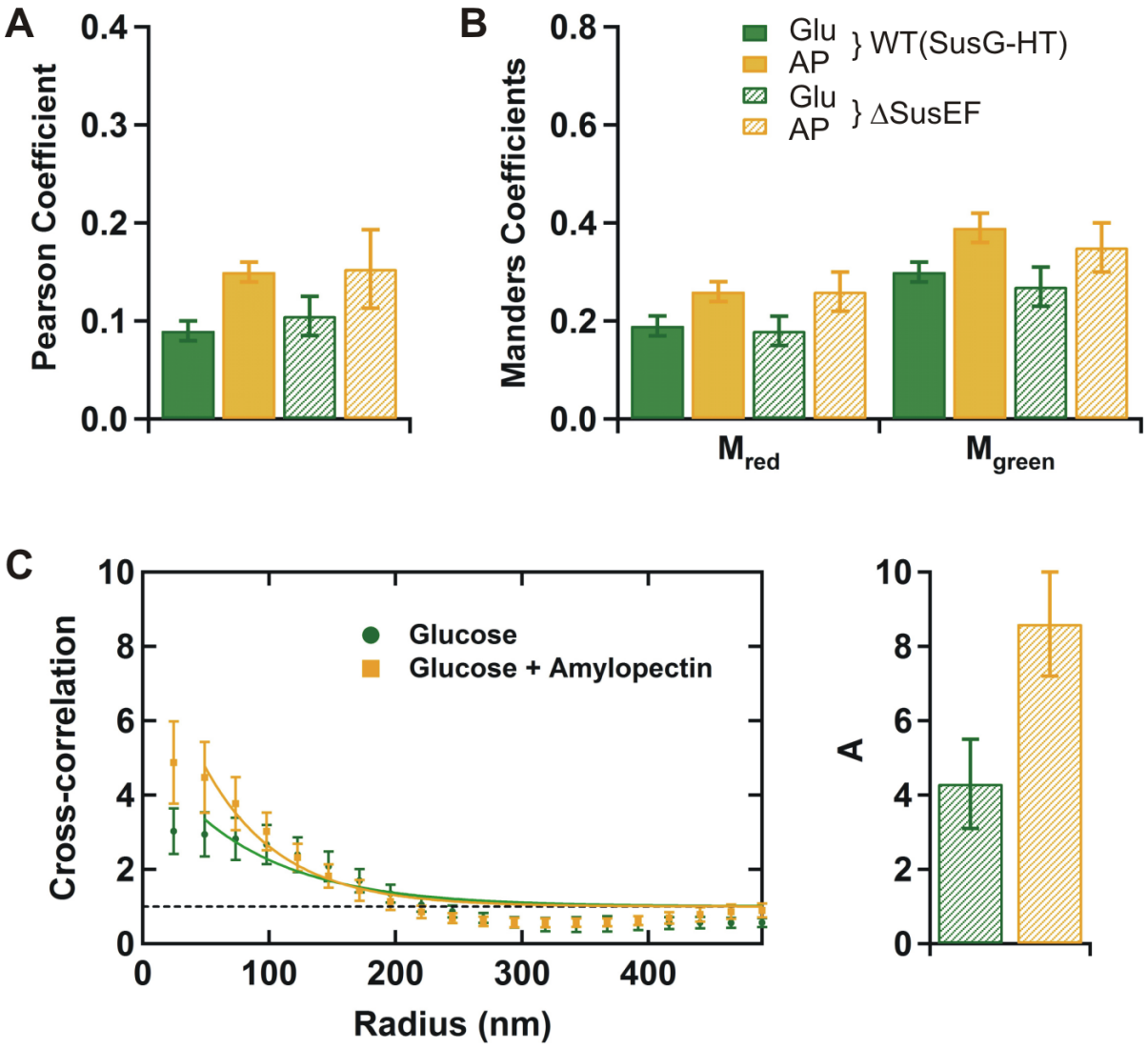



**FIGURE A.13. Simultaneous imaging of SusG and SusD in ΔSusEF cells**

**A.** Pearson and **B.** Manders coefficients showing protein co-localization among HaloTag-labeled SusG (SusG-HTL) and antibody-labeled SusD in SusE and SusF knockout strain (ΔSusEF, shaded bars). Pearson and Manders coefficients obtained for SusG-SusD pair in *Bt* cells containing all Sus proteins (WT(SusG-HT), solid bars) were included for comparison. **C.** Cross-correlation between localized SusG-HTL and antibody-labeled SusD (left), and the cross-correlation amplitude (*A*) obtained from the fit (right) in ΔSusEF. Error bars represent the standard error of the mean.

**Discussion**

The human gut Bacteroidetes promote complex glycan digestion in the gut by coordinated actions of membrane-associated protein complexes. The ability of this bacterial group to target a wide variety of polysaccharides makes them key players in this important symbiotic process. Despite their importance to human health, the precise mechanisms by which these proteins perform their functions are still obscure[1,5]. Using the *Bt* Sus as a model, we have characterized the assembly and real-time dynamics of these OMPs in live cells. Although Sus proteins were predicted to assemble to process starch, this phenomenon had not been directly observed in live bacteria. To reveal interactions among Sus proteins during starch catabolism with high resolution, we employed single-molecule super-resolution imaging in live *Bt* to detect fluorophore-labeled Sus proteins in real-time.

Protein correlation studies performed in fixed cells collectively revealed that simple, non-polymeric sugars such as glucose or maltose do not induce Sus complex assembly. On the other hand, the presence of large starch molecules enhanced Sus protein co-localization in *Bt*, suggesting the collaborative degradation of starch by a multi-component Sus complex. SusG diffusion was slowed in starch compared to glucose, partly due to direct contact with starch itself. However, the loss of one or more Sus OMPs further altered the SusG diffusion rate, suggesting interactions between Sus OMPs in the presence of starch. Taken together, we propose a model in which starch-induced Sus OMP complex assembly promotes starch processing in live *Bt* (**Figure A.14**). Our dynamic model fits well with existing knowledge of other Bacteroidetes Sus-like systems, which exhibit increasing numbers of OMPs (both enzymes and binding proteins) as polysaccharide linkage complexity increases[40]. Since protein complex formation is primarily linked to substrate, more complex Sus-like systems can evolve to incorporate

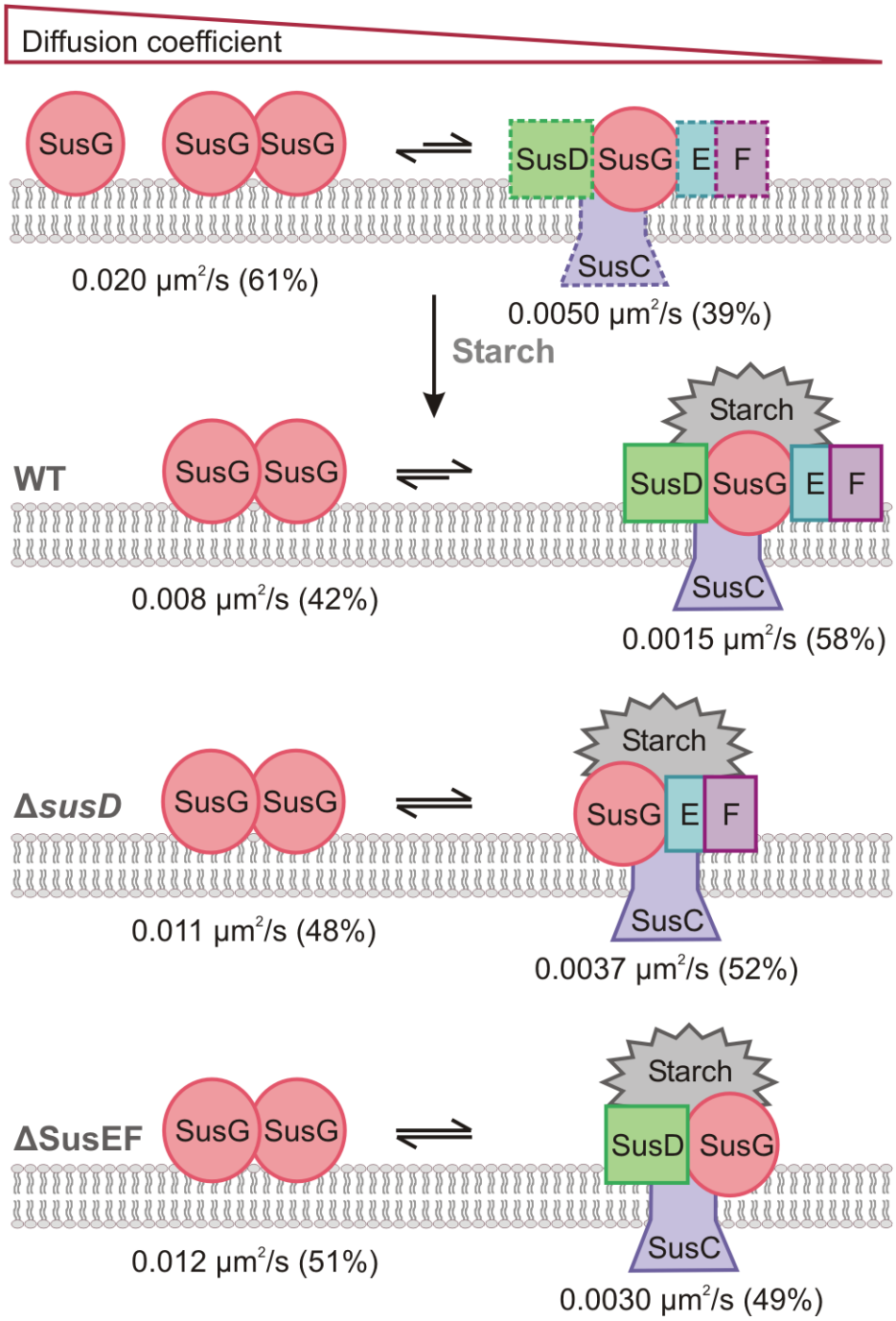



**FIGURE A.14. Proposed model for starch-induced assembly of the Sus complex**

In the absence of starch, SusG predominantly exists as fast diffusing free molecules rather than slow-moving SusG that are in complex with one or more other Sus partners as shown in dashed lines. In the presence of starch, interactions between starch and Sus proteins increase the slow-moving SusG population due to starch-induced Sus complex assembly. This complex diffuses faster in Δ*susD* and ΔSusEF, supporting the presence of SusD and of SusE and/or SusF in the Sus OMP complex.

additional OMP functions in the form of freely diffusing OM lipoproteins that need not fit into a more precisely arranged protein complex.

Interestingly, none of our Sus protein knockouts affected the SusG diffusion rates in glucose. Consistent with our proposed model, these data suggest that the observed moderate protein co-localization in glucose results from transient interactions among Sus proteins on the membrane. In starch, the absence of SusD or SusEF increased the overall diffusion rate of SusG, suggesting that at least a single copy of these proteins plays a role in starch-induced Sus OMP complex. In addition to the change in diffusion rates, a decrease in the proportion of slow-moving SusG in starch implies that the absence of one or more Sus proteins may decrease the overall complex stability.

The carbohydrate environment in the gut is constantly changing, making it critical for gut bacteria to rapidly sense and respond to available glycans. Starch induced-assembly of the Sus complex on the membrane is an apt approach for efficient starch processing in *Bt*. This dynamic process allows Sus OMPs to transition from a rapidly diffusing "surveillance" state in the absence of starch to a complex that can efficiently capture, degrade and import glycans into the cell from a single locus when the target substrate becomes available. SusG partitioning between slow and fast modes suggests that even in the presence of starch, the interaction of SusG and other Sus proteins is dynamic. Perhaps SusG disengagement from both starch and other Sus proteins provides both the substrate and the other Sus proteins additional degrees of freedom to facilitate maltooligosaccharide import. The dynamic assembly we observed for *Bt* Sus OMP-mediated starch degradation suggests a general mechanism by which many other Sus-like systems may operate in gut bacteria. If so, our real-time observations of Sus OMP dynamics during starch catabolism and our protein correlation analysis not only provide insight into how

this multi-protein system works in live *Bt*, but also will pave the way to understanding myriad Sus-like systems in other human gut symbionts.

## Materials and Methods

### *Bacterial growth conditions and genetic manipulation.*

*Bt* was grown at 37˚C under anaerobic conditions in media containing tryptone-yeast extract-glucose (TYG) and diluted into minimal media containing a 0.5% w/v of a carbohydrate source[17]. To monitor protein expression in the presence of maltose or maize amylopectin, cells at mid-late log phase in minimal media containing glucose were incubated for 10 min in fresh, pre-reduced media containing glucose and the appropriate sugar (glucose:maltose or glucose:amylopectin ~ 5:1)[33]. Genetic manipulation of *Bt* was achieved using a counter selectable allelic exchange method as previously described[10].

### *HaloTag-labeling of SusG.*

A construct with *Bt* SusG fused to a HaloTag® protein (SusG-HT) was made by replacing CBM58 of SusG (residues 219-336) with the HaloTag® protein (inactive haloalkane dehalogenase)[14,25]. *Bt* expressing SusG-HT at the native promoter was labeled with TMR-HaloTag® (5 $\mu$M, Promega) by incubation for 10-15 min at 37˚C in the dark as recommended by the manufacturer. To remove excess dye, cells were washed once with PBS buffer (pH 7.5) followed by two 10-min incubations in PBS at 37˚C. Cells were then incubated in 1x minimal media/PBS for 30 min at 37˚C, followed by re-suspension in fresh minimal media containing the appropriate sugar for live-cell imaging. For fixed-cell imaging, cells were further incubated in

4% paraformaldehyde for 15 min and washed twice with PBS at room temperature before re-suspending in fresh minimal media/PBS.

***Antibody labeling of Sus proteins.***

To monitor Sus proteins on the cell surface, formaldehyde-fixed non-permeabilized *Bt* cells were blocked in PBS containing 2% goat serum and incubated with rabbit polyclonal antibodies (Cocalico Biologicals) specific to individual Sus proteins. After washing with PBS, the cells were incubated in Alexa 488-conjugated goat anti-rabbit IgG secondary antibodies (Molecular Probes). Antibody-labeled cells were rinsed several times with PBS and re-suspended in minimal media for cellular imaging. For protein co-localization experiments, cells were labeled with TMR-HaloTag® and fixed with formaldehyde prior to antibody labeling.

***Super-resolution imaging of live bacterial cells.***

For live-cell imaging, fluorophore-labeled cells were incubated in fresh media for 30 min at 37˚C in the anaerobic chamber before imaging. Both fixed and live fluorophore-labeled *Bt* cells were deposited in minimal media containing a carbohydrate source and a reducing agent onto pads of 2% agarose in the same media for super-resolution imaging. The coverslip edges were sealed with 5 Minute® Epoxy (Devcon) to maintain an oxygen-free environment (Figure S2)[26]. *Bt* cells were imaged on an Olympus IX71 inverted fluorescence microscope equipped with a 1.40 NA, 100x oil-immersion wide-field/phase-contrast objective or a 1.49 NA, 100x oil-immersion TIRF objective (Olympus). *Bt* containing SusG labeled with TMR-HaloTag® (SusG-HTL) and Alexa 488-conjugated antibody-labeled Sus proteins were excited with 561-nm (Coherent Sapphire 561-50) and 488-nm (Coherent Sapphire 488-50) lasers, respectively.

Fluorescence emission intensities were detected on a 512 x 512 pixel Photometrics Evolve EMCCD at 10-20 frames per second with appropriate filters[41]. To monitor protein dynamics under native conditions, live bacterial cells were imaged at 37˚C in the presence of glucose or fluorophore-labeled carbohydrates using an objective heater (Bioptechs).

***Fluorophore labeling of carbohydrates.***

Maltoheptaose (Sigma Aldrich) was fluorophore-labeled at the reducing end using a 1:0.2:1.2 molar ratio of maltoheptaose, Alexa 488 hydrazide (Molecular Probes) and 2-picoline borane (Sigma Aldrich)[35]. During the labeling reaction, maltoheptaose was first dissolved in a 1:3 water:methanol mixture containing 2.5% (v/v) acetic acid and incubated with Alexa 488 hydrazide at 65˚C in the dark. After 30 min of incubation, 2-picoline borane was added to the reaction mixture and further incubated at 65˚C for 45 min. To remove unbound fluorophores, the reaction mixture was purified using a sephadex G-10 column (PD MiniTrap G-10, GE Healthcare) followed by HPLC with a C18 reversed-phase column.

To label amylopectin with fluorescent probes, 100 $\mu$L of 10 mg/mL amylopectin from maize (Sigma Aldrich) was oxidized by 1 $\mu$L of 25 mM sodium periodate (Sigma Aldrich) for 60 min at room temperature[36]. The reaction was stopped by addition of 5 $\mu$L ethylene glycol. Oxidized amylopectin was fluorophore-labeled using 2 $\mu$L of 17.5 mM Alexa 488 hydrazide by incubation at 65˚C for 30 min. After addition of 2 $\mu$L of 100 mM 2-picoline borane, the reaction mixture was further incubated for 60 min at 65˚C to perform reductive amination [35]. The excess dye was removed with a sephadex G-25 column (PD SpinTrap G-25, GE Healthcare).

***Image processing.***

Stacked images were analyzed by fitting the point-spread function of isolated single-molecules in each imaging frame to a 2D symmetric Gaussian function to localize the emitter positions using the MATLAB nonlinear least squares regression function *nlinfit*[27]. Super-resolution images were reconstructed from these positions by plotting each localized fit as a 2D Gaussian with constant intensity and with standard deviation equal to the statistical localization accuracy (95% confidence interval on the position). Sus protein co-localization was analyzed by computing two pixel intensity-based quantities: Pearson's correlation coefficient and the Manders coefficients[29]. These coefficients were obtained using standard ImageJ plugins to analyze reconstructed super-resolution images of the red (TMR) and green (Alexa 488) channels corresponding to SusG-HTL and antibody-labeled Sus proteins, respectively. Cross-correlation functions between different protein pairs were analyzed using fast Fourier transforms in MATLAB[30]. Cross-correlation was performed on a whole bacterial cell mask determined from the reconstructed images.

Single-molecule tracking was performed using a custom MATLAB code that determines molecular trajectories as a function of time. Single-molecule tracks were constructed by connecting molecules that are localized in consecutive frames within 150 nm for a minimum of 0.7 s. Ensemble MSDs were found for every time lag (time interval between positions, $\tau$) by fitting the cumulative probability distribution of the squared step sizes to a three-term exponential function consisting of one immobile and two mobile terms that best describe the data [34,37].

$$P(U,\tau) = 1 - \left[ \alpha . \exp\left( \frac{-U}{\left\langle r_1^2(\tau) \right\rangle + \sigma^2} \right) + \beta . \exp\left( \frac{-U}{\left\langle r_2^2(\tau) \right\rangle + \sigma^2} \right) + \gamma . \exp\left( \frac{-U}{\sigma^2} \right) \right]$$

where $P(U,\tau)$ denotes the probability that the squared displacement ($r^2$) for a given time ($\tau$) does not exceed the specific value $U$. The coefficients $\alpha$, $\beta$ and $\gamma$ indicate the fraction of molecules in the fast, slow and immobile modes, respectively, at any given time within the localization accuracy ($\sigma$) and $\alpha + \beta + \gamma = 1$. Average SusG diffusion coefficients were determined from the linear slope of MSD vs. $\tau$ for the first four $\tau$ values[42].

***Monte Carlo Simulations.***

To support protein co-localization results, membrane protein localization was further studied using data generated in MATLAB to simulate random cell surface localizations. The *Bt* cell was modeled as a cylinder (length: 1.5 $\mu$m, radius: 0.5 $\mu$m) with 0.5-$\mu$m radius hemispheric caps. The MATLAB function *random* was used to generate a specified number (5-20) of random simulated localizations of each color (red and green) according to a uniform distribution on the cell surface. The intensity and width of each data point was randomly selected according to a normal distribution about the experimental averages. The effect of focal plane position was investigated by constraining the localizations in the axial direction to the whole cell, to the top 0.67 $\mu$m of the cell and to the middle 0.5 $\mu$m of the cell. For simulated co-localized data, the MATLAB function *randsample* was used to generate red points according to the MATLAB weighting function *rowweight*, which was used to specify a normal distribution ($\sigma$ = 50 nm) about randomly generated green points.

**Notes**

This work was reprinted and modified with permission from Karunatilaka K.S., Cameron E.A., Martens E.C., Koropatkin N.M., Biteen J.S. Super-Resolution Imaging Captures Carbohhydrate Utilization Dynamics in Human Gut Symbionts. *Submitted for Publication*.

Both KKS and EAC contributed to the design of the experiments described in Appendix I. EAC constructed bacterial strains and performed bacterial growths while KKS performed imaging studies and analyzed the resulting data.

**References**


1 Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nature Rev Microbiol* **10**, 323-335 (2012).

2 Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920 (2005).

3 McNeil, N. I. The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr* **39**, 338-342 (1984).

4 Collaborative computational project, n. The CCP4 suite: Programs for protein crystallography. *Acta Crystallographica* **D50**, 760-763 (1994).

5 Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284**, 24673-24677 (2009).

6 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 Environ Microbiol* **33**, 319-322 (1977).

7 Tancula, E., Feldhaus, M. J., Bedzyk, L. A. & Salyers, A. A. Location and characterization of genes involved in binding of starch to the surface of Bacteroides thetaiotaomicron. *J Bacteriol* **174**, 5609-5616 (1992).

8 Anderson, K. L. & Salyers, A. A. Genetic evidence that outer membrane binding of starch is required for starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **171**, 3199-3204 (1989).

9 Cho, K. H. & Salyers, A. A. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **183**, 7224-7230 (2001).

10 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**, 1105-1115 (2008).

11 Shipman, J. A., Berleman, J. E. & Salyers, A. A. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *J Bacteriol* **182**, 5365-5372 (2000).

12 Cameron, E. A. *et al.* Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism. *J Biol Chem* **287**, 34614-34625 (2012).

13 Shipman, J. A., Cho, K. H., Siegel, H. A. & Salyers, A. A. Physiological characterization of SusG, an outer membrane protein essential for starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **181**, 7206-7211 (1999).

14 Koropatkin, N. M. & Smith, T. J. SusG: A Unique Cell-Membrane-Associated alpha-Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. *Structure* **18**, 200-215 (2010).

15 D'Elia, J. N. & Salyers, A. A. Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of Bacteroides thetaiotaomicron on starch. *J Bacteriol* **178**, 7173-7179 (1996).

16 D'Elia, J. N. & Salyers, A. A. Effect of regulatory protein levels on utilization of starch by Bacteroides thetaiotaomicron. *J Bacteriol* **178**, 7180-7186 (1996).

17 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**, 447-457 (2008).

18 Fernandez-Suarez, M. & Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nature Rev Mol Cell Biol* **9**, 929-943 (2008).

19 Chudakov, D. M., Matz, M. V., Lukyanov, S. & Lukyanov, K. A. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev* **90**, 1103-1163 (2010).

20 Biteen, J. S. & Moerner, W. E. Single-molecule and superresolution imaging in live bacteria cells. *Cold Spring Harbor perspectives in biology* **2**, a000448 (2010).

21 Huang, B., Babcock, H. & Zhuang, X. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell* **143**, 1047-1058 (2010).

22 Remington, S. J. Fluorescent proteins: maturation, photochemistry and photophysics. *Curr Opin Struct Biol* **16**, 714-721 (2006).

23 Los, G. V. & Wood, K. The HaloTag: a novel technology for cell imaging and protein analysis. *Methods Mol Biol* **356**, 195-208 (2007).

24 Wombacher, R. & Cornish, V. W. Chemical tags: applications in live cell fluorescence imaging. *J Biophotonics* **4**, 391-402 (2011).

25 Los, G. V. *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **3**, 373-382 (2008).

26 Karunatilaka, K. S. *et al.* Single-Molecule Imaging Can Be Achieved in Live Obligate Anaerobic Bacteria. *Single Molecule Spectroscopy and Superresolution Imaging Vi* **8590** (2013).

27 Biteen, J. S. *et al.* Super-resolution imaging in live Caulobacter crescentus cells using photoswitchable EYFP. *Nat Methods* **5**, 947-949 (2008).

28 Manders, E. M., Stap, J., Brakenhoff, G. J., van Driel, R. & Aten, J. A. Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J Cell Sci* **103 ( Pt 3)**, 857-862 (1992).

29 Bolte, S. & Cordelieres, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc* **224**, 213-232 (2006).

30 Veatch, S. L. *et al.* Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PloS One* **7**, e31457 (2012).

31 Sengupta, P. *et al.* Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nat Methods* **8**, 969-975 (2011).

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

33 Rogers, T. E. *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol Microbiol* (2013).

34 Liao, Y., Yang, S. K., Koh, K., Matzger, A. J. & Biteen, J. S. Heterogeneous single-molecule diffusion in one-, two-, and three-dimensional microporous coordination polymers: directional, trapped, and immobile guests. *Nano Lett* **12**, 3080-3085 (2012).

35 Ruhaak, L. R., Steenvoorden, E., Koeleman, C. A., Deelder, A. M. & Wuhrer, M. 2-picoline-borane: a non-toxic reducing agent for oligosaccharide labeling by reductive amination. *Proteomics* **10**, 2330-2336 (2010).

36 Stephen, T. L. *et al.* Transport of Streptococcus pneumoniae capsular polysaccharide in MHC Class II tubules. *Plos Pathog* **3**, e32 (2007).

37 Sonnleitner, A., Schutz, G. J. & Schmidt, T. Free Brownian motion of individual lipid molecules in biomembranes. *Biophys J* **77**, 2638-2642 (1999).

38 Watkins, L. P. & Yang, H. Detection of intensity change points in time-resolved single-molecule measurements. *J Phys Chem B* **109**, 617-628 (2005).

39 Martens, E. C., Roth, R., Heuser, J. E. & Gordon, J. I. Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J Biol Chem* **284**, 18445-18457 (2009).

40 Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).

41 Moerner, W. E. & Fromm, D. P. Methods of single-molecule fluorescence spectroscopy and microscopy. *Rev Sci Instrum* **74**, 3597-3619 (2003).

42 Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium (vol 82, 041914, 2010). *Phys Rev E* **83** (2011).