THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 41, pp. 34614–34625, October 5, 2012

Multidomain Carbohydrate-binding Proteins Involved in *Bacteroides thetaiotaomicron* Starch Metabolism*[S]

Received for publication, July 2, 2012, and in revised form, August 10, 2012 Published, JBC Papers in Press, August 21, 2012, DOI 10.1074/jbc.M112.397380

Elizabeth A. Cameron[‡1], Mallory A. Maynard[‡], Christopher J. Smith[‡], Thomas J. Smith[§], Nicole M. Koropatkin[‡§2], and Eric C. Martens[‡3]

From the [‡]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109 and the [§]Donald Danforth Plant Science Center, St. Louis, Missouri 63132

Background: *Bacteroides thetaiotaomicron* is a prototype for understanding carbohydrate metabolism by colonic bacteria.
Results: Two nonenzymatic membrane proteins involved in starch metabolism are composed of tandem carbohydrate-binding modules that each bind starch differently.
Conclusion: *B. thetaiotaomicron* has evolved multiple starch-binding modules to compete for different forms of starch.
Significance: Learning how gut bacteria degrade carbohydrates is crucial for understanding their role in nutrition.

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 values 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.

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 starch-utilization system (Sus)[4]-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 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

* This work was supported, in whole or in part, by National Institutes of Health Grant K01DK084214 (to E. C. M), the University of Michigan Biological Sciences Scholars Program, and pilot/feasibility grants from the University of Michigan Gastrointestinal Peptides Research Center (to E. C. M. and N. M. K.).
[S] This article contains supplemental Tables S1–S3 and Figs. S1–S11.
[1] Supported by University of Michigan Genetics Training Grant GM07544.
[2] To whom correspondence may be addressed. E-mail: nkoropat@umich.edu.
[3] To whom correspondence may be addressed: 1150 West Medical Center Dr., Ann Arbor, MI. Tel.: 734-647-5800; Fax: 734-764-3562; E-mail: emartens@umich.edu.

[4] The abbreviations used are: Sus, starch-utilization system; *Bt*, *Bacteroides thetaiotaomicron*; CBM, carbohydrate-binding module; SeMet, selenomethionine; M7, maltoheptaose; αCD, α-cyclodextrin; IgSF, immunoglobulin superfamily; GH, glycoside hydrolase; GM3M3, glucosyl maltotriosyl maltotriose; ITC, isothermal titration calorimetry; rTEV, recombinant tobacco etch virus; Bistris propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; r.m.s., root mean square.

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 toward 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 β2,6-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 noncatalytic 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. Although nonenzymatic CBM-containing proteins have been described as part of cellulosomal complexes (1), nonenzymatic 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 noncatalytic 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 nonenzymatic polypeptides, which is rarely reported, may serve to assist the catalytic function of SusG in this multiprotein system that is present on the cell surface.

EXPERIMENTAL PROCEDURES

Bacterial and Culture Conditions—*Bt* was grown in tryptone/yeast extract/glucose (TYG) media (27) or on brain-heart infusion (BD Biosciences) agar, which 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 with 5 mg/ml of maltose was prepared as described in Ref. 28.

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 using the primers listed in supplemental Table S3 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 counterselection on 5-fluoro-2′-deoxyuridine. Surface expression of SusE and SusF was probed by antibody staining of nonpermeabilized formaldehyde-fixed *Bt* cells grown on minimal media/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 NdeI (SusE) or NheI (SusF) and XhoI 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 multisite-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 1.

The pET28rTEV plasmids containing the allele of interest were transformed into Rosetta (DE3) pLysS cells (EMD Biosciences). Transformed cells were grown at 37 °C for 20 h, and then the plates were scraped to inoculate culture media for protein expression. For native protein expression, the cells were grown in 1 liter of TB, plus kanamycin (50 μg/ml) and chloramphenicol (20 μg/ml) (in 2-liter baffled flasks) at 37 °C until

they reached an OD ~0.4, and the temperature was turned down to 22 °C. Approximately 30 min after lowering the temperature, isopropyl 1-thio-β-D-galactopyranoside was added to a final concentration of 0.5 mM, and the cells continued to grow overnight (16–20 h). Cells were harvested by centrifugation at 6,000 × *g*, and the cell pellets were stored at −80 °C until protein purification. Selenomethioine (SeMet)-substituted protein was produced via the methionine inhibitory pathway (29), as previously described (30).

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 ml 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 repurified on the 5-ml nickel 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 of protein and 2 mM α-cyclodextrin (αCD) against a well solution of 16–20% PEG 6000, 2 M NaCl, 100 mM malonate (pH 5.0). The SusE-αCD crystals were then serially transferred into a cryoprotectant of 22% PEG 6000, 2.3 M NaCl, 50 mM malonate, 2 mM αCD, 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 maltoheptaose (M7) were grown at room temperature from hanging drops against a well solution of 2.5 M ammonium sulfate, 100 mM Bistris propane (pH 7.0). These crystals were flash-frozen in a cryoprotectant containing 2.0 M ammonium sulfate, 80 mM Bistris 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 of protein and 2 mM M7 against a well solution of 6–12% glycerol, 1.5–2 M 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, whereas 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 (31). The structures of SusE and SusF were determined from the SAD data using the AutoSol subroutine within the Phenix software package (32, 33). These initial models of SusE and SusF proteins were then utilized for molecular replacement in Phaser (34) against the native x-ray data sets. Data collection statistics are reported in supplemental Tables S1 and S2. The ramachandran plots for all three structures were generated using the MolProbity structure validation server (35). 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 the favored regions, and the rest in allowed regions. The SusF structure with maltoheptaose had two residues, Glu-89 (55.7, −23.4) and Ser-341 (−29.1, 143.3) that fell just outside the generously allowed region of the ramachandran plot. Glu-89 is part of a left-hand helical turn. A hydrogen bond between the peptidyl O of Leu-87 and the side chain imidazole N of His-91 distort the geometry of this turn. Ser-341 is at the beginning of an α-helix, and a hydrogen bond between the side chain hydroxyl of Ser-341 and the nearby side chain of Glu-379 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 αCD, 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. Isotherms are displayed in supplemental Figs. S3–S11.

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_2O, then once with an excess of PBS. Starch was pelleted and suspended in PBS to make a 100 mg/ml of 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 to 0.1 mg/ml in PBS. Plates were incubated for 2 h at room temperature with agitation. Starch and bound protein was pelleted by centrifugation and the supernatant collected. Unbound

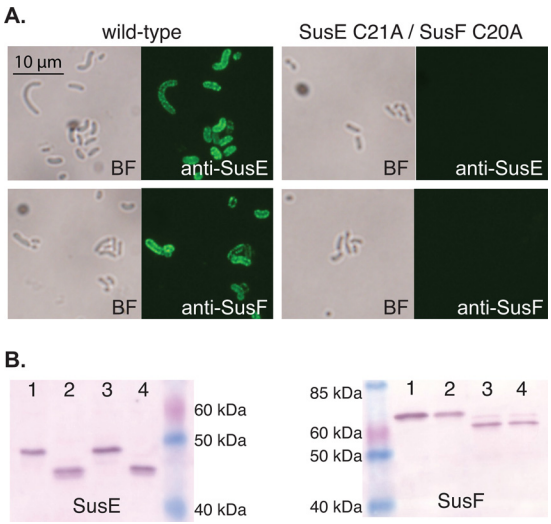

FIGURE 1. **SusE and SusF are exposed on the surface of *B. thetaiotaomicron*.** Alleles of *susE* and *susF* were created in which the N-terminal cysteine, which is lipidated to tether the proteins to the outer membrane, was mutated to alanine (SusE C21A and SusF C20A). These alleles were recombined into the native *sus* locus. Cells were grown to mid-exponential phase in minimal media/maltose to induce expression. *A*, *Bt* staining for SusE and SusF surface expression. Nonpermeabilized 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* = 10 μm. *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 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).

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. Because a pathway for secreting lipoproteins to the external leaflet of the Gram-negative outer membrane has yet to be defined (36), 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 (Fig. 1, *A* and *B*). 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 *Escherichia 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 SeMet-substituted 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 (supplemental Tables S1 and S2).

The 2.0-Å 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 (Fig. 2*A*). 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-barrel that is similar in overall-fold and topology to several immunoglobulin superfamily (IgSF) domains found in cell adhesion proteins including CD28 (1YCD-chainC; r.m.s. deviation 3.1 Å, 8% sequence identity), and CD47 (2JJS-chain A; r.m.s. deviation 2.7 Å, 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, whereas 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. Although 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 (37) revealed that all share the most structural homology with the X25 domain of the *Bacillus acidopullyticus* glycoside hydrolase (GH) family 13 pullulanase (PDB 2WAN), with *Z*-scores of 7.8, 7.3, and 4.9 for the Fa, Fb, and Fc CBMs, respectively. Although 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

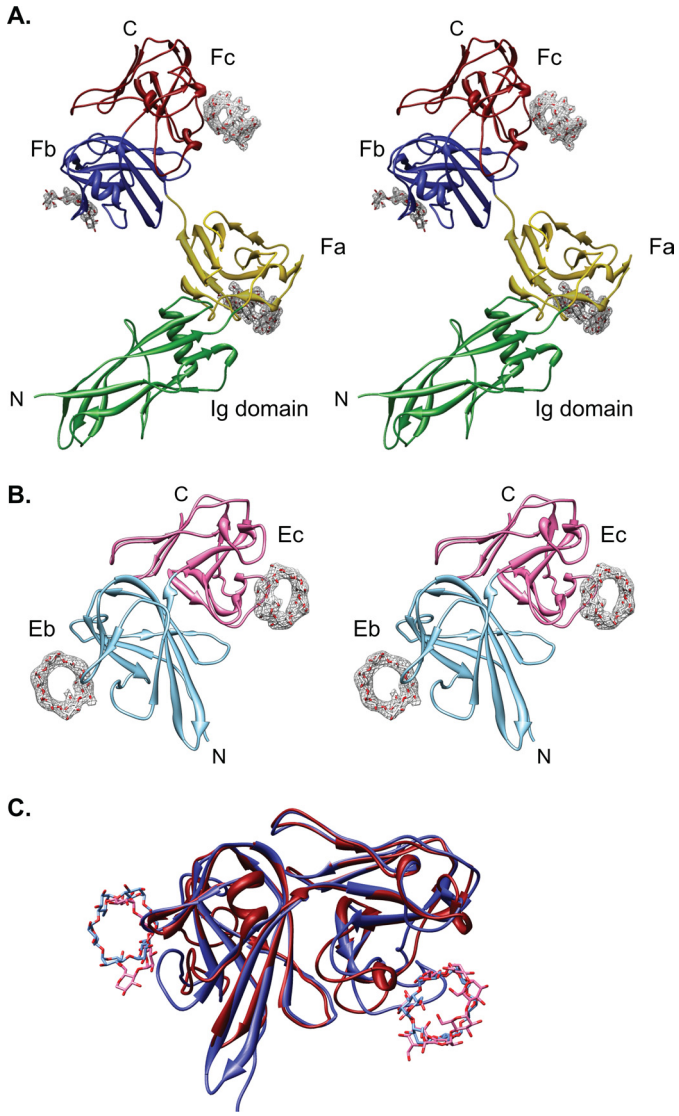

FIGURE 2. **Ribbon diagram of SusE and SusF structures.** *A,* schematic representation of SusF (residues 40–485), with the IgSF domain (residues 40–160) in *green*, CBM Fa (residues 161–172) in *yellow*, CBM Fb (residues 275–383) in *blue*, and CBM Fc (residues 384–485) in *red*. Bound M7 is displayed as *red* and *white sticks*. The electron density from an omit map, contoured at 2.5 σ 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,* schematic representation of SusE (residues 174–387), with CBM Eb (residues 174–283) colored *aqua* and CBM Ec (residues 284–385) colored *pink*. Bound αCD is displayed as *red* and *gray sticks*. Electron density for αCD from an omit map is displayed and contoured at 2 σ. 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 r.m.s. deviation of the models is 1.3 Å for 189 Cα atoms. The ligand αCD bound to SusE is shown as *light blue sticks*, and the maltotetraose and M7 bound to SusF are shown as *pink sticks*.

CBM41 domains of *Streptococcus pneumoniae* SpuA (38). However, in SusF the additional CBM Fa creates a triangle of binding sites, with each starch-binding site oriented ~120° apart (Fig. 2*A*).

The structure of SusE (residues 35–387) complexed with αCD was solved to a resolution of 2.5 Å (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 (Fig. 2*B*). 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 C-terminal domains, causing the former to be disordered in the crystal lattice. In the structure, two symmetry-related molecules of SusE are clustered around a single molecule of αCD. 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 ~10 kDa smaller, due to the absence of a middle domain corresponding to Fa in SusF. Although the N-terminal domain of SusE was not resolved in the crystal structure, the predicted structure of residues 38–167 generated using I-TASSER (39, 40) suggests a similar IgSF-type-fold (supplemental Fig. S1). 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 r.m.s. deviation of 1.3 Å over 189 Cα atoms and share 38.6% sequence identity (Fig. 2*C*).

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 (Fig. 3*A*). 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 binding site displays a characteristic aromatic arc (Trp-177 and Trp-222) 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, 41–43). At the Fb site, four of the seven glucose residues of M7 are resolved in the electron density (Fig. 3*B*). 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, Trp-442 and Trp-396, are further apart than those within Fa and Fb, with Trp-441 wedged between these residues, and providing an additional hydrogen-bonding donor to the O-6 of Glc6 (Fig. 3*C*).

Although each of the SusF CBMs displays subtle molecular differences in the binding sites, the orientation of each

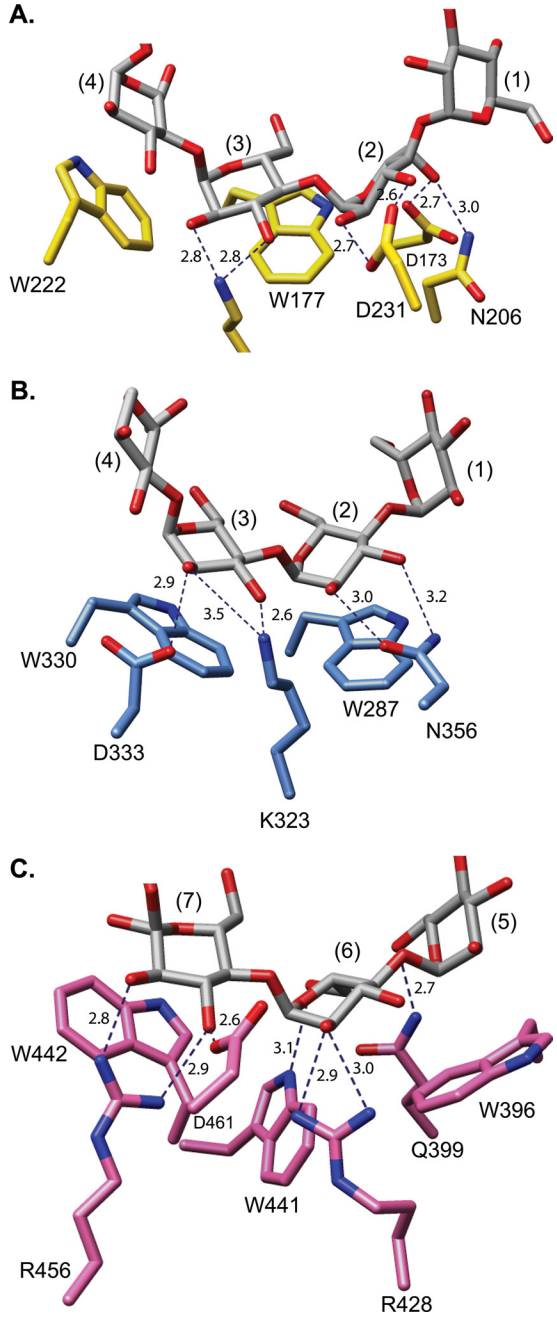

FIGURE 3. **Close-up view of the three starch-binding sites in SusF.** In each panel, M7 is shown as *gray* and *red sticks* and the amino acids involved in binding displayed. *Dashed lines* depict the hydrogen-bonding network between the ligand and protein and distances are shown in Å. Note that in *panels A* and *C*, only the portion of the ligand involved in protein binding is displayed. Glucose residues are numbered with glucose (*1*) indicating the nonreducing 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.

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 1.3 Å (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 r.m.s. deviation of 0.3 Å for all atoms, except one loop (residues 360–365) with a maximum Cα deviation of 2.7 Å, likely due to crystal contacts. The C-terminal domain from the SusE structures with αCD and M7 overlay with a r.m.s. deviation of 0.4 Å with no Cα deviations in either starch-binding site.

CBM Eb overlays with CBM Fb with an r.m.s. deviation of 1.4 Å over 93 Cα atoms (33.3% sequence identity). The binding of αCD at Eb is similar to M7 binding at Fb, with adjacent glucose residues bound via both hydrophobic stacking and hydrogen bonding interactions (Fig. 4*A*). 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 αCD structure Ec contacts 5 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 (Fig. 4). Tryptophans Trp-336 and Trp-296 of Ec create a hydrophobic arc with Trp-335 wedged between, but not participating in glycan binding. A unique feature of the Ec site is the loop created by residues 353–357 that cap one end of the binding site, with the side chain of Ile-355 centered in front of the αCD ring. This loop provides multiple hydrogen-bonding partners to Glc1, Glc2, Glc3, and Glc6 of αCD, via specific interactions with Asn-353, Leu-354, Ile-355, and Asp-356 (Fig. 4*B*). This starch-binding loop is unlikely to be flexible, and rather is anchored in place by a network of hydrogen bonds with an adjacent loop defined by residues 359–362. The topology of this binding site, in particular the centering of the Ile-355 side chain at the ligand is strikingly similar to the binding of βCD to the glycogen-binding domain of AMP-activated protein kinase (44).

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 (Fig. 4, *C* and *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 (Fig. 4*E*). In both chains A and B, M7 is anchored to the protein by the same set of hydrophobic stacking interactions with Trp-336 and Trp-296, as well as hydrogen-bonding through Arg-326 and Arg-350. At chain A, the maltooligosaccharide helix, from the nonreducing to reducing end, projects toward the protein against the capping loop (Fig. 4, *C* and *D*). The peptidyl oxygen atoms of Leu-354 and Ile-355

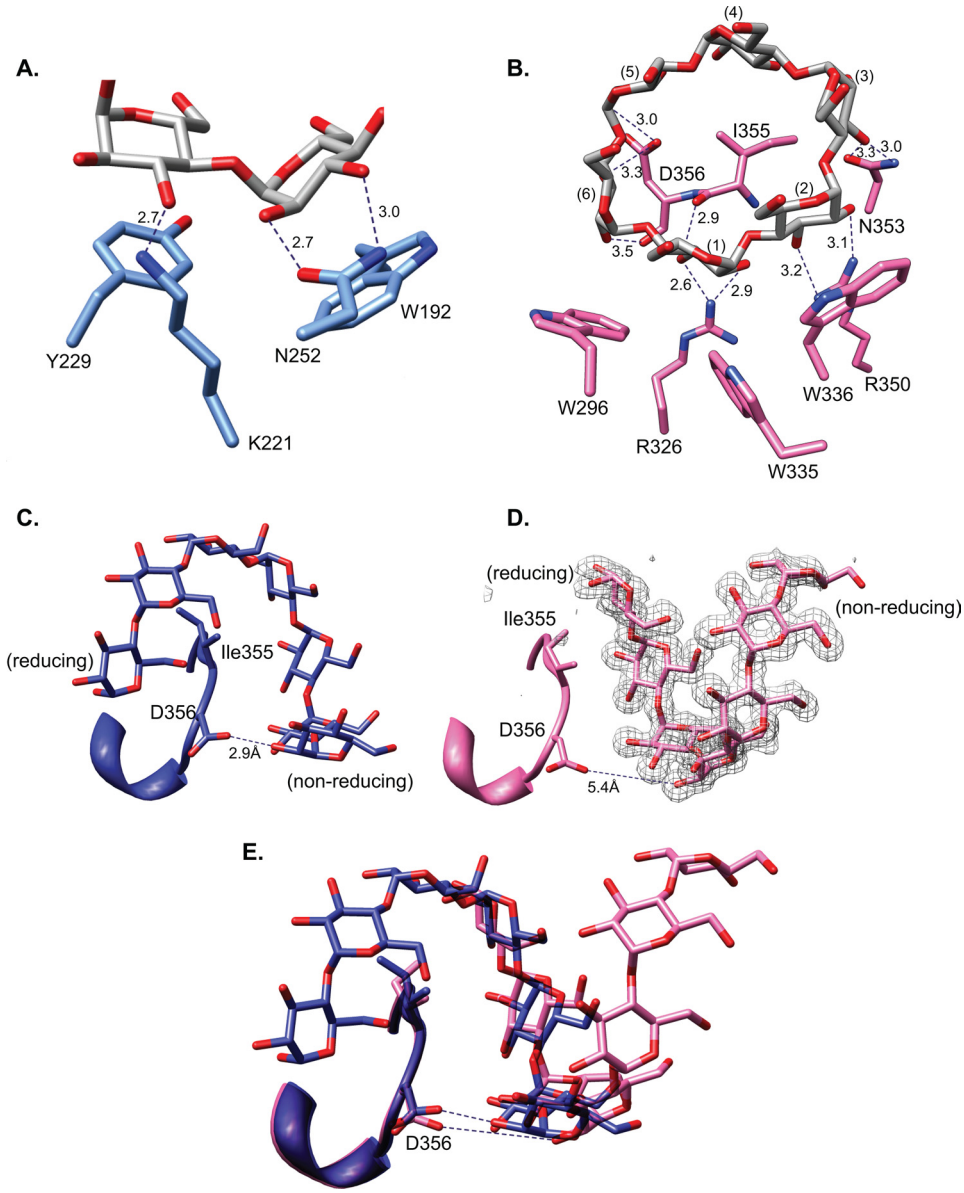

FIGURE 4. **Close-up view of the starch-binding sites in SusE.** *Panels A* and *B* depict the structure of SusE with αCD, whereas *panels C–E* depict the structure of the C-terminal half of SusE with maltoheptaose. *A,* αCD binding at CBM Eb, with the ligand as *gray* and *red sticks*, and the amino 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 nonreducing end of the maltooligosaccharide. *B,* αCD binding at CBM Ec, as described for *panel A*. Leu-354 was omitted for clarity. For a stereo view of this site, see supplemental Fig. S2. *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 3 σ. 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.

participate in hydrogen bonding with hydroxyl groups from adjacent glucose residues as seen in the structure with αCD, but due to the pitch of the oligosaccharide helix, Asp-356 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 nonreducing to reducing end extending from the hydrophobic cradle of binding residues and extending up and over the capping loop. Thus, in chain B the nonreducing end of the ligand is nestled closer to the capping loop, such that the glucose at the terminal nonreducing end interacts with Asp-356. In this ligand orientation, Ile-355 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 15 Å 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 αCD.

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 Ile-355 that intercalates into one

TABLE 1
Affinity of SusE and SusF CBMs for maltooligosaccharides determined by ITC

Protein–active CBM	Mutations (to Ala)	K_d αCD	K_d M7	K_d GM3M3
			μM	
WT SusE	None	86.96 ± 19.7	134.2 ± 34.2	357.1 ± 12.8
SusE–B only	Arg-326 Trp-336, Arg-350	386.1 ± 35.8	1023.5 ± 36.7	3584.2 ± 120.8
SusE–C only	Trp-192, Lys-221, Tyr-229, Asn-252	97.09 ± 13.2	17.04 ± 1.2	641.0 ± 90.4
SusE–no binding	Trp-192, Lys-221, Tyr-229, Asn-252, Arg326, Trp-336, Arg-250	No binding	No binding	Not tested
WT SusF	None	769.2 ± 50.9	303.0 ± 23.9	990.1 ± 107.8
SusF–A only	Trp-287, Lys-323, Asn-356, Trp-396, Trp-442, Arg-456	775.2 ± 15.6	361.0 ± 4.6	2710 ± 110.2
SusF–B only	Trp-177, Lys-208, Trp-222, Asp-231, Trp-396, Trp-442, Trp-456	460.8 ± 51.0	309.6 ± 10.5	751.9 ± 44.7
SusF–C only	Trp-177, Lys-208, Trp-222, Asp-231, Trp-287, Lys-323, Asn-356	465.1 ± 34.6	97.09 ± 2.9	507.6 ± 20.6
SusF–no binding	Trp-177, Lys-208, Trp-222, Asp-231, Trp-287, Lys-323, Asn-356, Trp-396, Trp-442, Arg-456	No binding	No binding	Not tested

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 α1,6-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 (45). Although the Sus outer membrane amylase (SusG) will only hydrolyze α1,4-linkages (14), at least one of the periplasmic amylases (SusB) is promiscuous toward a variety α-glucosidic linkages (46). Thus, it is possible that SusE and SusF interact with oligosaccharides that contain α1,6-branches prior to transport across the outer membrane. Moreover, the cyclic maltooligosaccharide αCD mimics the rigid, geometrically constrained curvature of larger amylose molecules, making it possible to probe starch-binding proteins for affinity toward 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 oligosaccharides: αCD, M7, and glucosyl-maltotriosyl-maltotriose (GM3M3), an oligosaccharide of seven glucose units containing two α1,6-linkages (Table 1). 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, whereas 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 αCD 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 with SusF. The Eb site displays tighter binding for αCD compared with 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 a true helical shape in αCD, which is a ring, is compensated for by the fixed geometry of the cyclodextrin (47). 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 with αCD (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 M7 with slightly better affinity than αCD, 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 whereas α1,6-linkages are tolerated, there is unlikely to be a preference for these structures over α1,4-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. We performed adsorption depletion experiments to determine the binding affinity of WT SusE and SusF, as well as binding site mutants of SusE and SusF, to insoluble cornstarch. 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 potential differences in nonspecific binding between replicates. We performed this assay many times while refining our final assay conditions (also performed in triplicate)

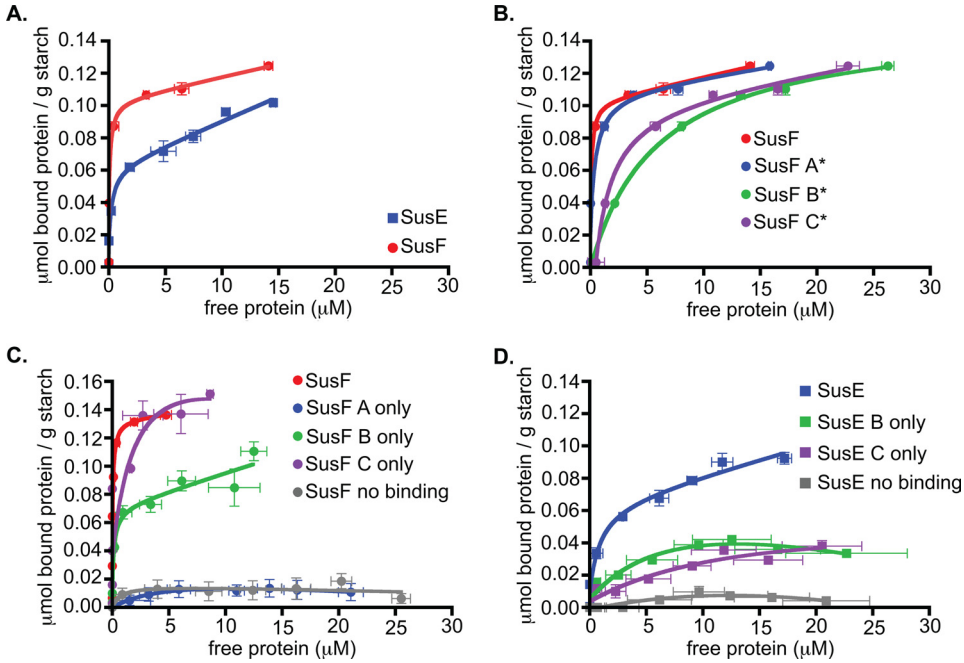

FIGURE 5. **Protein binding to insoluble cornstarch.** Bound protein per gram of starch is plotted as a function of free protein concentration with error bars representing the S.E. from three replicates. Data were fit to a one-site total binding equation. *A,* WT SusE and SusF; *B,* WT SusF and mutant 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*).

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 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 (Fig. 5*A* and Table 2). In experiments utilizing single CBM mutants of SusF (Fig. 5*B*), with the mutated CBM designated by an asterisk, 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*) alone is mutated. Reciprocally, when Fa is left as the only remaining functional starch-binding site (SusF-A only; Fig. 5*C*), 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 (Fig. 5*B*). 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 with 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 toward the avidity of the protein for starch. When the individual domains of SusE are mutated, there is a substantial loss in insoluble starch binding, ~40–80-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.

TABLE 2
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[a]
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[a]
SusF–B only	0.154 ± 0.095
SusF–C only	2.182 ± 4.404
SusF–no binding	Could not be fit[a]

[a] Could not be fit, designates curve fits with an *R* squared value of less than 0.8.

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. Although 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.* cel-

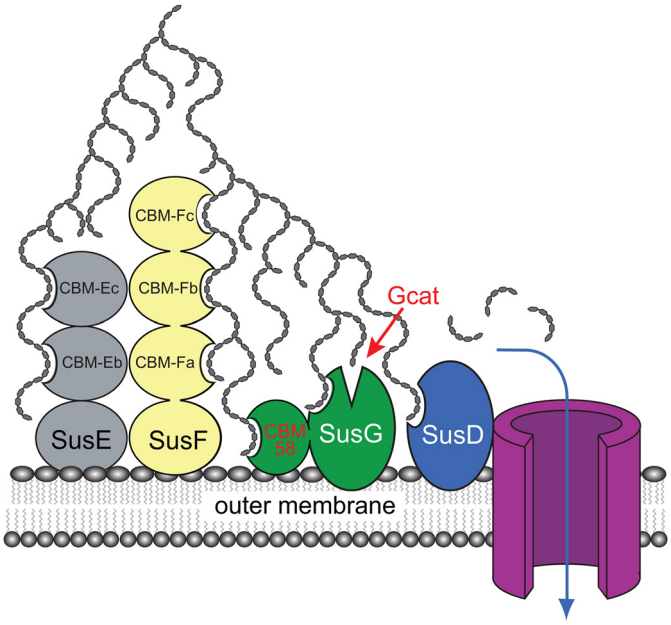

FIGURE 6. **Model of Sus outer membrane protein interactions with starch.** In total, the four outer membrane lipoproteins in the *Bt* starch-utilization system contain at least nine sites that interact with starch or cleaved maltooligosaccharides. Only one of these sites (*Gcat*) is catalytic and present 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 liproteins has not been explored, nor has the stoichiometry of these proteins on the cell surface.

lulosomes) 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 severalfold due to the multivalent interactions of the protein 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 α1,4-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 (Fig. 6). 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 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 β2,6-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 a complete inability to grow on oligosaccharides greater than 5 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.

Acknowledgments—We thank the United States Department of Energy, Office of Science, Office of Basic Energy Sciences for use of the Advanced Photon Source under Contract DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and Michigan Technology Tri-Corridor program Grant 085P1000817.

REFERENCES

1. Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R., and White, B. A. (2008) Polysaccharide utilization by gut bacteria. Potential for new insights from genomic analysis. *Nat. Rev. Microbiol.* **6,** 121–131
2. Koropatkin, N. M., Cameron, E. A., and Martens, E. C. (2012) How glycans shape the human gut microbiota. *Nat. Rev. Microbiol.* **10,** 323–335
3. McNeil, N. I. (1984) The contribution of the large intestine to energy supplies in man. *Am. J. Clin. Nutr.* **39,** 338–342
4. Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031
5. Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., Schlegel, M. L., Tucker, T. A., Schrenzel, M. D., Knight, R.,

and Gordon, J. I. (2008) Evolution of mammals and their gut microbes. *Science* **320,** 1647–1651
6. Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J. M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S. D., and Wang, J. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65
7. Cho, K. H., and Salyers, A. A. (2001) Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **183,** 7224–7230
8. Martens, E. C., Koropatkin, N. M., Smith, T. J., and Gordon, J. I. (2009) Complex glycan catabolism by the human gut microbiota. The *Bacteroidetes* Sus-like paradigm. *J. Biol. Chem.* **284,** 24673–24677
9. Sonnenburg, E. D., Zheng, H., Joglekar, P., Higginbottom, S. K., Firbank, S. J., Bolam, D. N., and Sonnenburg, J. L. (2010) Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations. *Cell* **141,** 1241–1252
10. Martens, E. C., Lowe, E. C., Chiang, H., Pudlo, N. A., Wu, M., McNulty, N. P., Abbott, D. W., Henrissat, B., Gilbert, H. J., Bolam, D. N., and Gordon, J. I. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9,** e1001221
11. Shipman, J. A., Berleman, J. E., and Salyers, A. A. (2000) Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **182,** 5365–5372
12. Koropatkin, N. M., Martens, E. C., Gordon, J. I., and Smith, T. J. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16,** 1105–1115
13. Reeves, A. R., Wang, G. R., and Salyers, A. A. (1997) Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **179,** 643–649
14. Koropatkin, N. M., and Smith, T. J. (2010) SusG. A unique cell membrane-associated α-amylase from a prominent human gut symbiont targets complex starch molecules. *Structure* **18,** 200–215
15. Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L., Eddy, S. R., Bateman, A., and Finn, R. D. (2012) The Pfam protein families database. *Nucleic Acids Res.* **40,** D290–301
16. Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V., and Gordon, J. I. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* **299,** 2074–2076
17. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The carbohydrate-active EnZymes database (CAZy). An expert resource for glycogenomics. *Nucleic Acids Res.* **37,** D233–238
18. Ellrott, K., Jaroszewski, L., Li, W., Wooley, J. C., and Godzik, A. (2010) Expansion of the protein repertoire in newly explored environments. Human gut microbiome specific protein families. *PLoS Comput. Biol.* **6,** e1000798
19. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules. Fine-tuning polysaccharide recognition. *Biochem. J.* **382,** 769–781
20. Guillen, D., Sanchez, S., and Rodriguez-Sanoja, R. (2010) Carbohydrate-binding domains. Multiplicity of biological roles. *Appl. Microbiol. Biotech.* **85,** 1241–1249
21. Machovic, M., and Janecek, S. (2006) Starch-binding domains in the post-genome era. *Cell Mol. Life Sci.* **63,** 2710–2724
22. Penninga, D., van der Veen, B. A., Knegtel, R. M., van Hijum, S. A., Rozeboom, H. J., Kalk, K. H., Dijkstra, B. W., and Dijkhuizen, L. (1996) The raw starch-binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. *J. Biol. Chem.* **271,** 32777–32784
23. Tan, T. C., Mijts, B. N., Swaminathan, K., Patel, B. K., and Divne, C. (2008) 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
24. Sumitani, J., Tottori, T., Kawaguchi, T., and Arai, M. (2000) New type of starch-binding domain. The direct repeat motif in the C-terminal region of *Bacillus* sp. no. 195 α-amylase contributes to starch binding and raw starch degrading. *Biochem. J.* **350,** 477–484
25. Latorre-García, L., Adam, A. C., Manzanares, P., and Polaina, J. (2005) Improving the amylolytic activity of *Saccharomyces cerevisiae* glucoamylase by the addition of a starch-binding domain. *J. Biotechnol.* **118,** 167–176
26. Juge, N., Nøhr, J., Le Gal-Coëffet, M. F., Kramhøft, B., Furniss, C. S., Planchot, V., Archer, D. B., Williamson, G., and Svensson, B. (2006) The activity of barley α-amylase on starch granules is enhanced by fusion of a starch-binding domain from *Aspergillus niger* glucoamylase. *Biochim. Biophys. Acta* **1764,** 275–284
27. Holdeman, L. V., Cato, E. D., and Moore, W. E. C. (1977) *Anaerobe Laboratory Manual,* Virginia Polytechnic Institute and State University Anaerobe Laboratory, Blacksburg, VA
28. Martens, E. C., Chiang, H. C., and Gordon, J. I. (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host. Microbe.* **4,** 447–457
29. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229,** 105–124
30. Koropatkin, N. M., Koppenaal, D. W., Pakrasi, H. B., and Smith, T. J. (2007) The structure of a cyanobacterial bicarbonate transport protein, CmpA. *J. Biol. Chem.* **282,** 2606–2614
31. Otwinowski, Z., and Minor, W. (1997) in *Methods in Enzymology* (Carter, C. W. J., and Sweet, R. M., eds) pp. 307–326, Academic Press, New York
32. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX. Building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954
33. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., and Adams, P. D. (2008) Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426,** 419–435
34. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674
35. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity. All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21
36. Kovacs-Simon, A., Titball, R. W., and Michell, S. L. (2011) Lipoproteins of bacterial pathogens. *Infect. Immun.* **79,** 548–561
37. Holm, L., and Sander, C. (1995) Dali. A network tool for protein structure comparison. *Trends Biochem. Sci.* **20,** 478–480
38. Lammerts van Bueren, A., Ficko-Blean, E., Pluvinage, B., Hehemann, J. H., Higgins, M. A., Deng, L., Ogunniyi, A. D., Stroeher, U. H., El Warry, N., Burke, R. D., Czjzek, M., Paton, J. C., Vocadlo, D. J., and Boraston, A. B. (2011) The conformation and function of a multimodular glycogen-degrading pneumococcal virulence factor. *Structure* **19,** 640–651
39. Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER. A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5,** 725–738
40. Zhang, Y. (2008) I-TASSER server for protein three-dimensional structure prediction. *BMC Bioinformatics* **9,** 40
41. Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A., and Law, V. (2006) A structural and functional analysis of α-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *J. Biol. Chem.* **281,** 587–598
42. van Bueren, A. L., and Boraston, A. B. (2007) The structural basis of α-glucan recognition by a family 41 carbohydrate-binding module from *Thermotoga maritima*. *J. Mol. Biol.* **365,** 555–560

43. van Bueren, A. L., Higgins, M., Wang, D., Burke, R. D., and Boraston, A. B. (2007) Identification and structural basis of binding to host lung glycogen by streptococcal virulence factors. *Nat. Struct. Mol. Biol.* **14,** 76–84
44. Polekhina, G., Gupta, A., van Denderen, B. J., Feil, S. C., Kemp, B. E., Stapleton, D., and Parker, M. W. (2005) Structural basis for glycogen recognition by AMP-activated protein kinase. *Structure* **13,** 1453–1462
45. Anderson, K. L., and Salyers, A. A. (1989) Genetic evidence that outer membrane binding of starch is required for starch utilization by *Bacteroides thetaiotaomicron*. *J. Bacteriol.* **171,** 3199–3204
46. Kitamura, M., Okuyama, M., Tanzawa, F., Mori, H., Kitago, Y., Watanabe, N., Kimura, A., Tanaka, I., and Yao, M. (2008) Structural and functional analysis of a glycoside hydrolase family 97 enzyme from *Bacteroides thetaiotaomicron*. *J. Biol. Chem.* **283,** 36328–36337
47. Janeček, Š., Svensson, B., and MacGregor, E. A. (2011) Structural and evolutionary aspects of two families of noncatalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. *Enzyme Microb. Technol.* **49,** 429–440

Multidomain Carbohydrate-binding Proteins Involved in *Bacteroides thetaiotaomicron* Starch Metabolism

Elizabeth A. Cameron, Mallory A. Maynard, Christopher J. Smith, Thomas J. Smith, Nicole M. Koropatkin and Eric C. Martens

J. Biol. Chem. 2012, 287:34614-34625.
doi: 10.1074/jbc.M112.397380 originally published online August 21, 2012
