Structure of a SusD Homologue, BT1043, Involved in Mucin *O*-Glycan Utilization in a Prominent Human Gut Symbiont[†,‡]

Nicole Koropatkin,[§] Eric C. Martens,[||] Jeffrey I. Gordon,[||] and Thomas J. Smith*[,§]

Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, Missouri 63132, and Center for Genome Sciences, Washington University in St. Louis School of Medicine, St. Louis, Missouri 63108

Received October 16, 2008; Revised Manuscript Received January 5, 2009

ABSTRACT: Mammalian distal gut bacteria have an expanded capacity to utilize glycans. In the absence of dietary sources, some species rely on host-derived mucosal glycans. The ability of *Bacteroides thetaiotaomicron*, a prominent human gut symbiont, to forage host glycans contributes to both its ability to persist within an individual host and its ability to be transmitted naturally to new hosts at birth. The molecular basis of host glycan recognition by this species is still unknown but likely occurs through an expanded suite of outermembrane glycan-binding proteins that are the primary interface between *B. thetaiotaomicron* and its environment. Presented here is the atomic structure of the *B. thetaiotaomicron* protein BT1043, an outer membrane lipoprotein involved in host glycan metabolism. Despite a lack of detectable amino acid sequence similarity, BT1043 is a structural homologue of the *B. thetaiotaomicron* starch-binding protein SusD. Both structures are dominated by tetratrico peptide repeats that may facilitate association with outer membrane β-barrel transporters required for glycan uptake. The structure of BT1043 complexed with *N*-acetyllactosamine reveals that recognition is mediated via hydrogen bonding interactions with the reducing end of β-*N*-acetylglucosamine, suggesting a role in binding glycans liberated from the mucin polypeptide. This is in contrast to CBM 32 family members that target the terminal nonreducing galactose residue of mucin glycans. The highly articulated glycan-binding pocket of BT1043 suggests that binding of ligands to BT1043 relies more upon interactions with the composite sugar residues than upon overall ligand conformation as previously observed for SusD. The diversity in amino acid sequence level likely reflects early divergence from a common ancestor, while the unique and conserved α-helical fold the SusD family suggests a similar function in glycan uptake.

The distal gut microbial community (microbiota) of humans and other mammals is one of the most densely populated ecosystems on Earth. It is dominated by members of Bacteria but also contains members of Archaea and Eukarya (*1, 2*). The microbiota, whose acquisition begins at birth (*3*), plays pivotal roles in a number of facets of human physiology, including development of the immune system (*4, 5*) and acquisition of nutrients (*6*). Despite the essential role of the microbiota, it is not clear how this community retains its functional stability over time with constant changes in diet and in the face of continuous and rapid replacement of the intestinal epithelium and its overlying mucus layer.

While the composition and availability of dietary glycans changes over time, the endogenous pool of host glycans, including glycosaminoglycans, mucin O-linked glycans, and N-linked glycans, offers a more consistent food source for gut microbial species. Members of the bacterial phylum Bacteroidetes, one of two dominant phyla in the human and other mammalian gut microbiotas (*2*), are well-characterized consumers of complex glycans, including those emanating from the host mucosa (*7*). Correspondingly, these species have large collections of glycoside hydrolases and polysaccharide lyases in their genomes (*8*), a feature that likely contributes to this phenotype.

Bacteroidetes thetaiotaomicron is a member of the Bacteroidetes phylum, whose genome has been sequenced (*9*). It has served as a model organism for identifying some of the molecular mechanisms deployed by human gut Bacteroidetes for sensing, recognizing, importing, and processing dietary and host carbohydrates (*10*−*13*). Whole genome transcriptional profiling of *B. thetaiotaomicron* isolated from the intestines of adult gnotobiotic mice fed a diet of simple sugars, as well as gnotobiotic mice surveyed during the suckling period of postnatal development, revealed increased levels of expression of glycoside hydrolases such as hexosaminidases and fucosidases compared to those of gnotobiotic mice consuming a standard polysaccharide-rich chow diet (*12, 14*). These findings suggested an increased level of degradation of host glycans by this bacterium during periods when dietary polysaccharides were scarce. This ability of *B. thetaiotaomicron* to selectively forage on host glycans was shown to be important for fitness in vivo. A

[†] This work was supported by the National Institutes of Health (GM078800) and the Missouri Life Science Trust Fund.

[‡] X-ray coordinates of apo BT1043 and the BT1043−*N*-acetyllactosamine complex have been deposited in the Protein Data Bank (entries 3EHM and 3EHN, respectively) and will be released upon publication.

* To whom correspondence should be addressed: Donald Danforth Plant Science Center, 975 N. Warson Rd., St. Louis, MO 63132, Telephone: (314) 587-1451. Fax: (314) 587-1551. E-mail: tsmith@danforthcenter.org.

[§] Donald Danforth Plant Science Center.

[||] Washington University in St. Louis School of Medicine.

10.1021/bi801942a CCC: $40.75
Published on Web 02/03/2009

mutant strain deficient in mucin *O*-glycan utilization is unable to effectively compete with an isogenic wild-type strain in the intestines of gnotobiotic mice (*13*). This same mutant was also defective in its ability to be transmitted from a colonized mother to newborn pups in the days immediately following their birth. Together, these observations provide evidence that host glycan metabolism contributes to the ability of this species both to persist in a given host and to be transmitted naturally to new hosts.

The ability of Gram-negative, gut *Bacteroides* spp. such as *B. thetaiotaomicron* to forage upon a wide variety of glycans in the gut environment is attributed to the expression of similarly patterned polysaccharide utilization loci (PUL)[1] (*13, 14*). PUL encode cell envelope-associated protein complexes consisting of one or more glycolytic enzymes and homologues of two proteins, SusC and SusD, that are involved in glycan recognition and import (*8*). The starch utilization system (Sus) of *B. thetaiotaomicron* was the first such PUL described (*15*), although 87 additional Sus-like PUL have since been identified in this species (*13*) and several hundred more in other gut *Bacteroides* species (*8*). The Sus system is comprised of eight genes (*susRABCDEFG*) with five of its protein products, SusCDEFG, located in the outer membrane. In response to maltose, the SusR regulator drives expression of the *sus* operon, which is required for growth on amylose, amylopectin, pullulan, and maltooligosaccharides (*16, 17*). Large starch molecules are hydrolyzed by the outer membrane α-amylase SusG and imported into the periplasm through SusC, a predicted TonB-dependent β-barrel porin (*18*−*20*). While SusD has a critical role in starch and maltooligosaccharide binding (*17*), the roles of SusE and SusF are unclear (*19, 20*). Two periplasmic glycoside hydrolases, SusA and SusB, a neopullulanase and an α-glycosidase, respectively, further degrade periplasmic oligosaccharides imported by SusC (*18*).

SusD is a starch-binding protein with a novel α-helical fold, unlike any other carbohydrate binding module (CBM) studied to date (*17*). SusD preferentially binds helical amylose as suggested by the atomic structure of SusD complexed with α-cyclodextrin and the imposition of a similar geometry on bound maltoheptaose. Isothermal titration calorimetry confirmed that SusD binds cyclic oligosaccharides with higher affinity than linear forms, suggesting that recognition is dominated by the three-dimensional conformation of oligosaccharide ligands, rather than specific interactions with the composite sugars. Finally, a deletion mutant that lacked only production of SusD was unable to grow on starch molecules composed of at least six glucose units, indicating an essential role for this protein in starch uptake. On the basis of recent analysis of SusD and the notable expansion of SusD homologues among *Bacteroides* spp. (*8*), it is likely that other members of this protein family have similar roles in glycan recognition and uptake.

BT1043 from *B. thetaiotaomicron* is believed to play a role similar to that of SusD in a PUL that likely targets mucin *O*-glycans. Its expression is dramatically upregulated in vivo, ∼190-fold, in mice consuming glycan-deficient diets (*12, 14*) and in vitro, ∼101-fold, in host glycan fractions enriched for mucin *O*-glycans (*13*). In contrast to SusD, which recognizes the α-1,4-glucosidic linkages contained in starch, it is hypothesized that BT1043 recognizes linkages that are specific to mucin *O*-glycans. To elucidate the molecular determinants of host glycan recognition by BT1043, we determined its atomic structure in the presence and absence of *N*-acetyllactosamine (LacNAc), a common disaccharide component of host glycans. While BT1043 does not share any direct sequence homology with SusD, it has a remarkably similar protein fold, including four conserved tetratrico peptide repeats (TPRs). The TPR units provide a scaffold for the rest of the protein structure and may mediate protein−protein interactions with other members of the outer membrane glycan-binding complex. Extensive hydrogen bonding interactions with the reducing end of the β-*N*-acetylglucosamine moiety of LacNAc suggest that BT1043 recognizes free mucin *O*-glycans, although its cognate glycan remains to be determined. The glycan-binding pocket in BT1043 has a shape and location homologous to those found in SusD. However, while the amylose-binding site of SusD presents a relatively smooth, shallow cleft on the protein surface for interactions with the glycan backbone, the binding pocket of BT1043 is more articulated and likely makes more specific interactions with the composite sugar moieties themselves. To assess the generality of these findings, the structures of SusD and BT1043 are compared to the crystal structure of another putative mucin *O*-glycan-binding SusD homologue, BT3984. Despite diversity at the amino acid sequence level, a likely consequence of early divergence from a common ancestor, SusD-like proteins have a conserved α-helical fold, suggesting a homologous function in glycan uptake.

MATERIALS AND METHODS

Heterologous Protein Expression. The *BT1043* gene (residues 18−546) was amplified by PCR from genomic DNA prepared from *B. thetaiotaomicron* ATCC 29148 (also known as VPI-5482). The amplicon was cloned into pET28rTEV where the thrombin cleavage site of pET-28a (Novagen) has been modified to a tobacco etch virus (TEV) protease cleavage site. pET28rTEV-*bt1043* was transformed into Rosetta (DE3) pLysS (Novagen) for protein expression. Cells were grown in TB medium at 37 °C with shaking (225 rpm) until they reached an OD of ∼0.4, at which time the temperature was adjusted to 22 °C. Once the cultures reached an OD of ∼0.8, cells were treated with 0.2 mM ITPG to induce BT1043 expression and allowed to grow for 16 h at 22 °C. Cells were subsequently harvested by centrifugation, frozen in liquid nitrogen, and stored at 80 °C. Selenomethionine-substituted protein was produced via the methionine inhibitory pathway (*21*), as previously described (*22*).

Purification of Native and Selenomethionine-Substituted BT1043. Both native and selenomethionine-substituted BT1043 were purified using a 5 mL Hi-Trap metal affinity cartridge according to the manufacturer's recommendations (GE Healthcare). The cell lysate was loaded onto the column in His buffer [25 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole (pH 8.0)], and BT1043 was eluted using an imidazole (10 to 300 mM) gradient. The His tag was removed by incubation with rTEV (1:100 molar ratio relative to BT1043) at room temperature for 16 h. The cleaved protein was then dialyzed against His buffer and passed over

[1] Abbreviations: LacNAc, *N*-acetyllactosamine; PUL, polysaccharide utilization loci; rmsd, root-mean-square deviation.

Table 1: Data Collection Statistics[a]

	peak	inflection	remote	*N*-acetyllactosamine
space group	$P2_12_12_1$			$P2_13$
unit cell (Å)	$a = 60.28$, $b = 102.23$, $c = 175.75$			$a = b = c = 155.99$
wavelength (Å)	0.97934	0.97951	0.97167	0.97167
resolution (Å)	50.0−1.96 (2.03−1.96)	50.0−1.96 (2.03−1.96)	50.0−1.94 (2.01−1.94)	50.0−2.75 (2.80−2.75)
no. of independent reflections	77053 (7003)	77151 (6885)	78668 (7058)	32650 (1478)
completeness (%)	97.3 (90.1)	97.1 (88.3)	96.9 (88.1)	98.4 (90.6)
redundancy	4.1 (3.7)	4.1 (3.6)	4.1 (3.5)	8.3 (4.2)
avg I/avg $\sigma(I)$	35.4 (14.8)	35.0 (12.7)	32.2 (8.5)	20.5 (3.4)
R_{sym} (%)	8.6 (12.7)	6.8 (11.4)	6.2 (13.9)	7.8 (25.2)

[a] Values in parentheses represent values in the highest-resolution shells.

the affinity column to remove the His-tagged rTEV and undigested BT1043. Purified BT1043 was dialyzed against 20 mM HEPES and 100 mM NaCl (pH 7.0) and concentrated to ~9.5 mg/mL for crystallization.

Crystallization and Data Collection. Initial crystallization conditions for apo-BT1043 were determined via the hanging drop method using the Hampton Screen (Hampton Research). Large single crystals of SeMet BT1043 were grown at 4 °C in batch plates by seeding small crystals into mother liquor that contained ~5 mg/mL BT1043, 6−7% polyethylene glycol 6000, 50 mM NaCl, and 50 mM 2-(cyclohexylamino)ethanesulfonic acid (CHES) (pH 9.0). Crystals were of space group $P2_12_12_1$ with two molecules in the asymmetric unit and the following unit cell dimensions: $a = 60.28$ Å, $b = 102.23$ Å, and $c = 175.75$ Å. Crystals of native BT1043 complexed with *N*-acetyllactosamine were grown at room temperature via the hanging drop method. BT1043 (OD_{280} ~ 25) was mixed with solid *N*-acetyllactosamine (LacNAc) to a final concentration of 120 mM and equilibrated against mother liquor containing 12−14% polyethylene glycol 4000, 50 mM Na/K phosphate, and 100 mM succinate (pH 5.0). Crystals of BT1043 with LacNAc were in cubic space group $P2_13$ with two molecules per asymmetric unit and the following unit cell dimensions: $a = b = c = 159.06$ Å.

Both crystal forms were serially transferred to final cryoprotectant solutions containing 20% ethylene glycol and 15−21% of the appropriate PEG in synthetic mother liquor prior to being flash-frozen with liquid nitrogen. SeMet BT1043 diffraction maxima were collected on a 3 × 3 tiled "SBC3" CCD detector on Structural Biology Center beamline 19-BM, and the BT1043−LacNAc data were collected similarly on Structural Biology Center beamline 19-ID (Advanced Photon Source, Argonne National Laboratory, Argonne, IL). X-ray data were processed with HKL3000 and scaled with SCALEPACK (*23*). Data collection statistics are listed in Table 1.

X-ray Structure Determination. The structure of BT1043 was determined via MAD phasing from the X-ray data collected using the selenomethionine-substituted crystals. The program SOLVE (*24*) was used to determine and refine the initial positions of the selenomethionines, and RESOLVE (*25*) was then applied for solvent flattening and initial model building. Alternate cycles of manual model building in O (*26*), with maximum-likelihood refinement with CNS (*27*), was used to build and refine the 2.0 Å selenomethionine-substituted BT1043 ($R_{work} = 17.1\%$; $R_{free} = 20.7\%$). The BT1043−LacNAc structure was determined via molecular replacement using AMoRe (*28*) from the CCP4 suite of programs (*29*) with the SeMet BT1043 structure as a search model. Alternate cycles of manual model building in O and refinement using CNS were combined to complete the model. Relevant refinement statistics are presented in Table 2. Ramachandran plot analysis using PROCHECK revealed that the selenomethionine-substituted 2.0 Å apo structure of BT1043 had 90.9% of its residues in the most favored regions, 8.5% in the additionally allowed regions, 0.3% in the generously allowed regions, and 0.2% in the disallowed regions (*30*). In chains A and B, T59 assumes ϕ and ψ angles of 40° and −117°, respectively, as part of a turn between a 3_{10}-helix and α2, which lines part of the glycan-binding pocket. The electron density in this region is unambiguous. The 2.8 Å structure of BT1043 with *N*-acetyllactosamine had 84.4% of its residues in the most favored regions, 15% in the additionally allowed regions, and 0.6% in the generously allowed regions with no residues in the disallowed regions. In the structure of BT1043 with *N*-acetyllactosamine, T59 is in the proximity of the ligand, and the ϕ and ψ angles of this loop are slightly altered as a result of glycan binding.

Table 2: Refinement Statistics[a]

	apo, SeMet	*N*-acetyllactosamine
PDB entry	3EHM	3EHN
resolution (Å)	50.0−2.0 (2.01−2.0)	50.0−2.80 (2.82−2.80)
no. of protein atoms	8065	8106
no. of heteroatoms	670	178
R_{work}	17.1 (18.2)	19.6 (28.7)
R_{free}	20.7 (21.8)	24.6 (35.6)
no. of reflections	65281 (1226)	27651 (3080)
avg B values (Å^2)		
protein atoms	18.9	35.6
ligand	not applicable	45.4
solvent	25.445	22.1
rmsd		
bond lengths (Å)	0.0052	0.0069
bond angles (deg)	1.26	1.33

[a] Values in parentheses represent values in the highest-resolution shells.

Isothermal Titration Calorimetry. ITC measurements were carried out using a MicroCal VP-ITC titration calorimeter (MicroCal, Inc.). BT1043 was dialyzed overnight against a solution containing 20 mM HEPES (pH 7.0) and 100 mM NaCl, prior to the experiment. A solution of *N*-acetyllactosamine (Sigma-Aldrich) was prepared using dialysis buffer. BT1043 concentrated to 0.247 mM was placed in the reaction cell, and the reference cell was filled with deionized water. After the temperature was equilibrated to 25 °C, 25 successive 10 μL injections of 20 mM *N*-acetyllactosamine were made while the mixture was being stirred at 460 rpm, and the resulting heat of reaction was measured. The observed isotherm suggested weak binding, with a K_d of ~12 mM for one binding site.

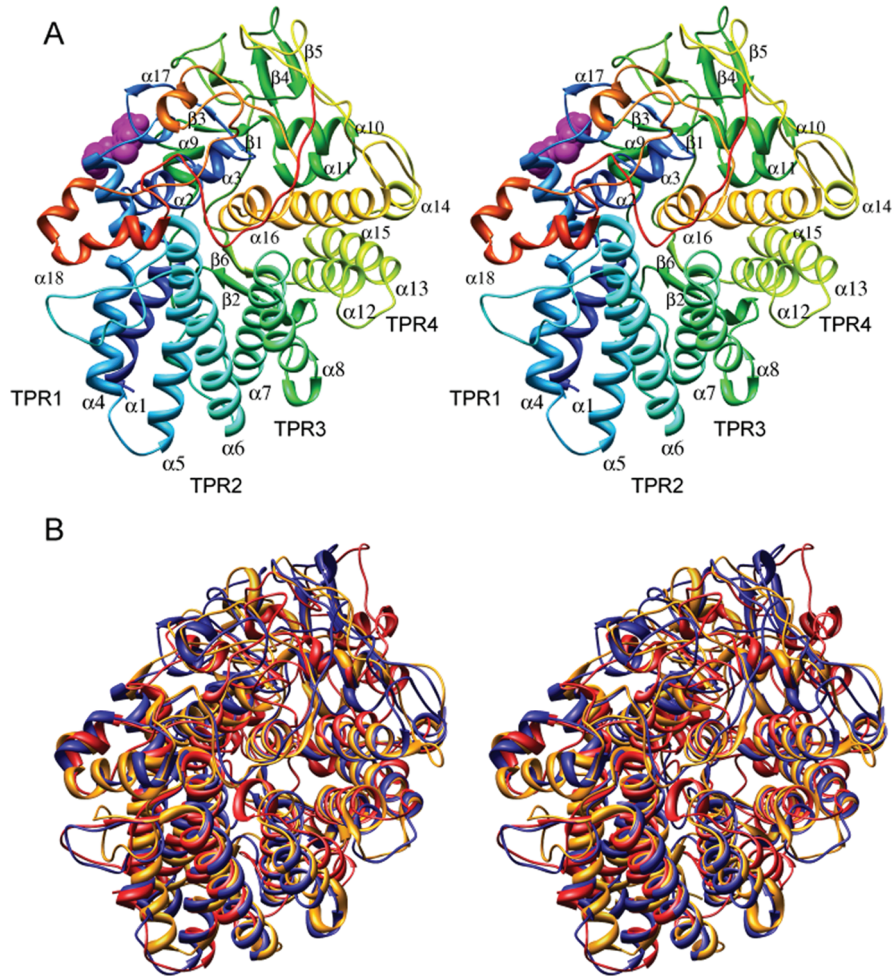

FIGURE 1: Structure of selenomethionine-substituted BT1043 at 2.0 Å resolution. (A) Cartoon representation of apo BT1043, colored from blue to red from the N-terminus to the C-terminus, respectively. The *N*-acetyllactosamine ligand (magenta spheres) has been modeled to show the location of the glycan-binding pocket. The 18 α-helices and six *β*-strands are labeled as shown; 3_{10}-helices are not labeled. Four helix−turn−helix pairs, or tetratrico peptide repeats (TPRs), are labeled with α1 and α4 as TPR1, α5 and α6 as TPR2, α7 and α8 as TPR3, and α12 and α13 as TPR4. (B) Overlay of BT1043 (blue), SusD (green; PDB entry 3CK7), and BT3984 (red; PDB entry 3CGH) in a view similar to that in panel A to highlight to similarities in the TPR domain.

Glycan Array Screening. BT1043 was fluorescently labeled using the Alexa Fluor 488 Protein labeling kit (Molecular Probes, Inc.) according to the manufacturer's instructions. The protein was then sent to the Consortium for Functional Glycomics Protein-Glycan Interaction Core (H) facility at Emory University for glycan array screening (www.functionalglycomics.org). Glycan binding was assessed on a printed array (version 3.1) of 377 mammalian glycans using labeled BT1043 at 0.2 mg/mL. No significant binding was detected above background. Because binding may have been affected by the fluorophore that reacts with exposed lysine residues, screening was attempted a second time using His-tagged BT1043 (0.2 mg/mL) with fluorescently labeled anti-His antibodies, provided by the facility. Again, glycan binding was not detected. Failure to detect binding may have been due to the absence of ligands specific for BT1043 or K_d values greater than the limit of detection (high micromolar to low millimolar).

RESULTS

The X-ray crystal structure of the soluble portion of BT1043 (residues 18−546) was determined using MAD phasing methods with selenomethionine-substituted protein crystals. The resulting 2.0 Å structure (R_{work} = 17.1%; R_{free} = 20.7%) was refined using the remote wavelength data. BT1043 is a monomer that has an α-helical fold consisting of 18 α-helices, nine 3_{10}-helices, three two-stranded anti-parallel *β*-sheets, and multiple reverse turns (Figure 1A). Most striking is the presence of four tetratrico peptide repeats (TPR), each consisting of a helix−turn−helix motif (Figure 1). These helix−turn−helix pairs are defined by α1 (residues 39−50) and α4 (residues 99−125) as TPR1, α5 (residues 130−154) and α6 (residues 174−193) as TPR2, α7 (residues 211−230) and α8 (residues 234−245) as TPR3, and α12 (residues 358−371) and α13 (residues 378−393) as TPR4. Together, these TPR units create a right-handed superhelix along one side of the protein and cradle the rest of the structure, with α15 (residues 443−459) and α16 (residues 463−473) packing against these TPR units. BT1043 shares significant structural homology with SusD as well as BT3984 from *B. thetaiotaomicron* (PDB entry 3CGH), a predicted mucin glycan-binding SusD homologue (*13*). Each of these proteins contains four TPR units in the same relative positions, creating a right-handed superhelix along one side of the structure (Figure 1B). Beyond the TPR domain, these structures deviate substantially, although they are nearly

entirely α-helical. BT3984 can be superimposed onto BT1043 with an rmsd of 1.37 Å over 373 Cα atoms (34% identical sequence by structural alignment over 373 Cα atoms). Interestingly, SusD can be superimposed onto BT1043 with an rmsd of 1.78 Å over 207 Cα atoms (14.5% identical sequence by structural alignment over 207 Cα atoms) even though BLASTP failed to detect any amino acid sequence homology between BT1043 and SusD. This suggests that proteins in the SusD family may display a wide range of sequence variability despite having similar three-dimensional folds.

Like SusD, BT1043 was not predicted to have a TPR domain since it lacks the traditional TPR consensus sequence (W4-L7-G8-Y11-A20-F24-A27-P32) (*31*). TPR domains are typically involved in mediating protein−protein interactions. Most often, TPR-containing proteins recognize a peptide from another protein via the concave interior surface created by the right-handed superhelical twist of several TPR units. In BT1043, two antiparallel β-strands, β2 (residues 260−262) and β6 (residues 353−365), are recognized in this concave area via hydrophobic interactions with the N-terminal halves of α5 and α7, while α16 makes mostly hydrogen bonding and electrostatic interactions with the C-terminal portion of these same α-helices. In a similar manner, α15 has some hydrophobic and some hydrogen bonding interactions with α12 and α13 of TPR4. Thus, in the structure of BT1043 as well as other SusD family proteins, this concave surface acts as a scaffold by cradling the rest of the protein structure. There has been speculation that SusD homologues interact with their predicted TonB-dependent porins via these conserved TPR units, perhaps via the solvent-exposed loops connecting adjacent helices and TPR units. This edge-on mode of binding between a TPR domain and target protein was observed between the TPR domain of $p67^{phox}$ and Rac, two subunits of NAPDH oxidase (*32*). In this complex, two loops that connect TPR units 1 and 3 in $p67^{phox}$ create the binding surface for Rac-GTP, resulting in recognition at the edge of the TPR bundle rather than the concave surface. While the crystal structures of some TPR proteins display large movements of the TPR helices and/or disruption of individual TPR bundles when crystallized under difference conditions (*31, 33*), the TPR domains of SusD and BT1043 are invariant even when the proteins are crystallized in different space groups. The stability of the TPR arrangement in the SusD family of proteins is likely a result of the multiple interactions these units make with the rest of the protein, resulting in much less flexibility compared to that seen in other TPR-containing proteins.

Previous transcriptional analysis revealed that the PUL operon encoding BT1043 (*BT1042−45*) is upregulated ∼101-fold during growth in vitro on porcine mucin glycan fractions enriched for mucin *O*-glycans, as opposed to glycosaminoglycans or *N*-glycans (*13*). This pool of glycans is estimated to have as many as 79 different oligosaccharides species of varying lengths (*34*). To identify the repertoire of oligosaccharides recognized by BT1043, glycan binding was examined by the Consortium for Functional Glycomics Protein-Glycan Interaction Core (H) facility at Emory University. Glycan binding to 377 natural and synthetic mammalian glycans on a printed array (version 3.1) was assessed using both fluorescently labeled BT1043 and fluorescently labeled antibodies to the N-terminal His tag of BT1043. Neither attempt identified a potential glycan. This was not altogether surprising since even SusD had a weak binding affinity for even a closely related analogue of its natural ligand. To define the structural elements involved in glycan recognition, the atomic structure of BT1043 was determined in the presence of *N*-acetyllactosamine (LacNAc), a common disaccharide component (β-D-galactosyl-1,4-β-D-*N*-acetylglucosamine) of mucin *O*-glycans that could be obtained in reasonable quantities for biochemical studies (*35*). Transcriptional analysis indicates that the *BT1042−1045* locus of *B. thetaiotaomicron* is upregulated ∼4-fold in response to LacNAc, suggesting that LacNAc may represent a part of a larger glycan more readily recognized by BT1043 (*13*). In addition, isothermal titration calorimetry (ITC) revealed that BT1043 binds LacNAc with a K_d of ∼12 mM (data not shown).

While LacNAc may not be an ideal analogue of BT1043's cognate glycan, the results with SusD lend support to using these small oligosaccharides to identify at least some of the key elements involved in recognition of the larger glycans (*17*). ITC studies demonstrated that SusD binds linear maltoheptaose with a K_d of ∼1 mM but binds the circular analogues α-, β-, and γ-cyclodextrin with K_d values of 0.089, 0.065, and 0.146 mM, respectively. Cocrystal structures of SusD with various maltooligosaccharides revealed that the mode of binding was the same in each case. Indeed, even though ITC failed to detect binding of maltotriose to SusD, it was possible to cocrystallize maltotriose complexed with SusD, and this small oligosaccharide bound in a manner identical to that of the trisaccharide component of the larger oligosaccharides. A similar binding trend is emerging on work with the galactomannan/glucomannan-binding SusD homologue from *Bacteroides ovatus* (unpublished results).

There is an interesting parallel between this family of glycan binding proteins and the capsid proteins of the *Calicviruses.* Norwalk virus is able to bind both A and H blood group antigens, and this interaction is thought to greatly enhance subsequent interactions with the viral receptor. The structure of a portion of the capsid protein complexed with both polysaccharides was made possible only by the addition of very large amounts of glycans during crystallization (*36*), as in these studies. They found that while the H blood group polysaccharide interacts with the terminal α-fucose, β-galactose moieties, the A type blood group antigen interacts only via the terminal α-GalNAc moiety. Nevertheless, the protein residues involved in hydrogen binding and van der Waals contacts were common to both glycans. Therefore, the protein contacts observed in the BT1043−LacNAc complex are more than likely to be biologically relevant, but it also seems likely that, in vivo, additional contacts are made with longer polysaccharides. Also, as was proposed with SusD (*17*), the authors suggested that the relatively weak protein−glycan interactions might be overcome by multivalent interactions.

To ensure saturation of BT1043 with ligand, ∼120 mM LacNAc was added to BT1043 during crystallization and freezing; smaller amounts of ligand were not attempted. The 2.8 Å structure of BT1043 with LacNAc (R_{work} = 19.6%; R_{free} = 24.6%) revealed one molecule of LacNAc clearly bound in the glycan-binding site predicted by the structure of SusD (Figure 2A). The apo and LacNAc-bound structures of BT1043 can be superimposed with an rmsd of 0.4 Å,

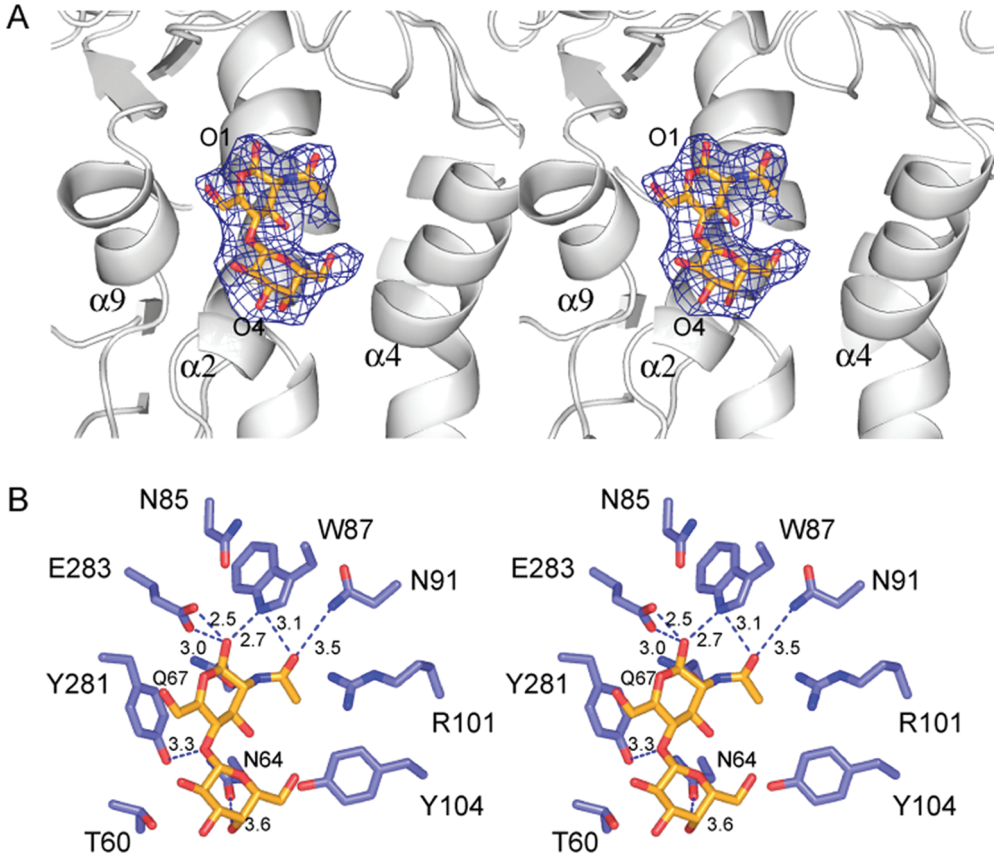

FIGURE 2: *N*-Acetyllactosamine (LacNAc) binding in the 2.8 Å structure of BT1043. (A) Omit map (3σ) of native BT1043 complexed with LacNAc. The reducing O1 atom of *N*-acetylglucosamine (GlcNAc) and the nonreducing O4 atom of galactose are labeled, along with the α-helices that shape the glycan-binding site. (B) Close-up view of LacNAc binding to BT1043. Coordinating and nearby residues are displayed, with distances labeled in angstroms.

revealing virtually no structural changes upon glycan binding. The shallow glycan-binding pocket is shaped by α2 and α9, in the proximity of though not interacting with residues at the N-terminus of α4 that comprises TPR1. LacNAc is anchored to the binding pocket through interactions at the reducing end of *β*-*N*-acetylglucosamine (GlcNAc) (Figure 2B). O1 of the GlcNAc anomeric carbon is positioned within ~3.0 and 2.5 Å of the side chain carboxylic oxygens of E283 and 2.7 Å from the $N^{\varepsilon 1}$ atom of W87. Similarly, the acetyl O atom of GlcNAc is positioned 3.1 Å from the $N^{\varepsilon 1}$ atom of W87 and 3.5 Å from the carboxamide N atom of N91. The amide N atom of GlcNAc points toward the side chain carboxamide oxygen of Q67, 3.3 Å away, and the hexose ring is stacked against the phenolic side chain of Y281 at a distance of ~4 Å. The galactose moiety does not have any significant interactions with the protein, with a single hydrogen bond between O4 and the carboxamide O atom of N64. The extensive interactions between GlcNAc and BT1043 suggest that the cognate ligand for BT1043 contains GlcNAc, although it may be preceded by one or more different monosaccharides and/or by different linkages (i.e., *β*1−3 versus *β*1−4). It is likely that the cognate ligand of BT1043 is much longer than a disaccharide with sugars toward the nonreducing end anchoring it to the protein since the porcine *O*-glycan fraction that stimulated the BT1042−1045 locus by 101-fold was comprised of longer (9−14 sugars) oligosaccharides (*34*). In addition, the glycan-binding pocket of BT1043 contains a number of possible hydrogen bonding acceptors and donors at the reducing end of LacNAc that may play a role in binding a longer glycan. Since SusD prefers maltooligosaccharides with at least six glucose units which allows the ligand to adopt a helical conformation, it is likely that BT1043 prefers a somewhat longer oligosaccharide with an overall shape that can better complement the binding pocket. The fact that *N*-acetylglucosamine is anchored to BT1043 via its reducing end suggests that the protein recognizes *O*-glycans liberated from the mucin protein, as opposed to those attached to the polypeptide and therefore lacking an exposed reducing sugar (*37*).

On the basis of on the structural alignment of BT1043 and SusD, the general features of the glycan-binding sites of BT1043 and SusD were compared (Figure 3A). The SusD glycan binding site has similar size and shape. However, the residues lining the glycan-binding pocket are quite different. The SusD-binding pocket is unique in that the hydrophobic residues that dominate starch binding (W98, W320, and Y296) form an arc that complements the shape of helical α-amylose. Shorter to medium-length oligosaccharides, including maltoheptaose, bind to SusD with much less affinity than the cyclodextrins that have a fixed, curved geometry (*38*). Since BT1043 binds to different oligosaccharides, it is not surprising that BT1043 does not have this arc of hydrophobic residues. The most conservative difference between the two proteins is the substitution of Y281 in BT1043 for W320 in SusD. The conservation of R118 in SusD with R101 in BT1043 and W121 in SusD with Y104 in BT1043 is interesting (Figure 3A), although neither of these residues was observed to interact with glycans in the various SusD complexes (*38*).

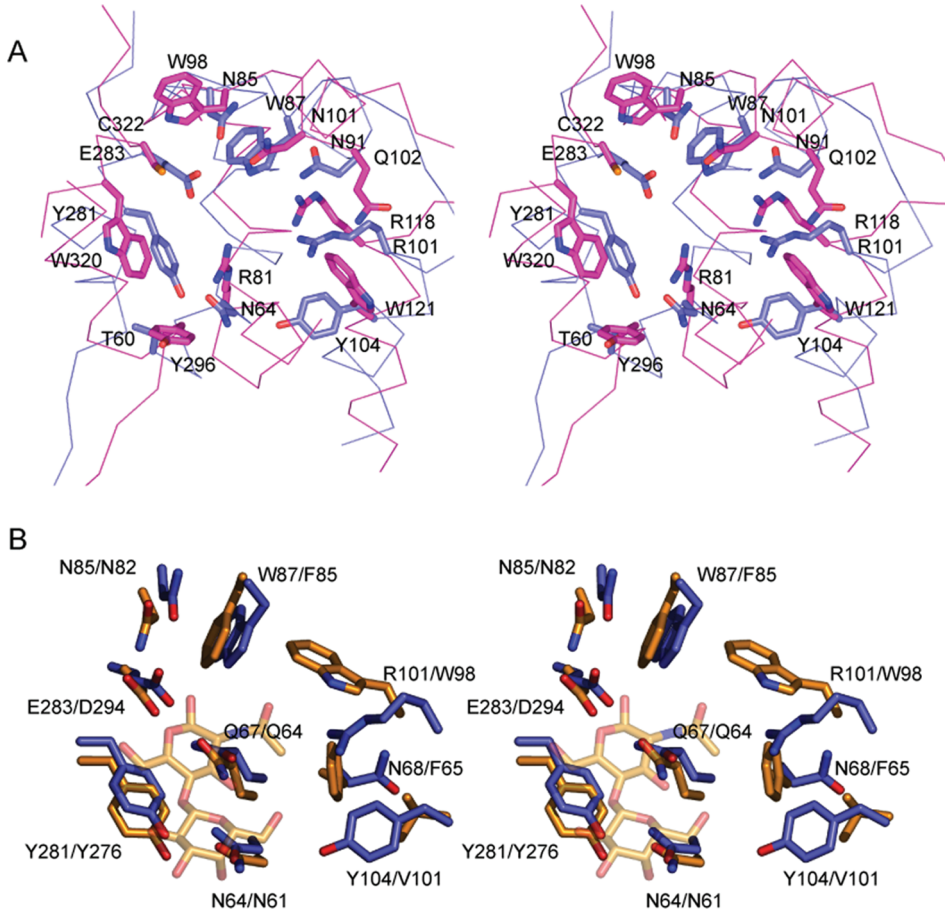

FIGURE 3: Overlay of the SusD and BT1043 glycan-binding pockets. (A) Close-up view of the amino acids lining the glycan-binding pocket of BT1043 (blue) with that of SusD (pink). Starch-binding residues in SusD are overlaid with the equivalent residues in BT1043. The ribbon representation of the glycan-binding pocket is shown to highlight the similarities in the overall shape and size of the site, despite differences in the amino acid sequence. (B) Close-up view of the BT1043 (blue-purple) LacNAc binding site with the equivalent glycan-binding site in BT3984 (yellow-orange).

The structure of SusD homologue BT3984 (PDB entry 3CGH) from *B. thetaiotaomicron* was recently deposited by the Joint Center for Structural Genomics (JCSG). Like BT1043, BT3984 is also highly upregulated in the mouse distal gut when the bacterial food source is limited to host mucin glycans (*13*). Interestingly, expression of these two proteins is stimulated differently during in vitro growth on host glycan fractions that are enriched for *O*-glycans, suggesting that they recognize and respond to different host glycan cues. Unlike the locus containing *BT1043*, the PUL carrying *BT3984* (*BT3983*−*88*) is slightly upregulated (∼10-fold) during growth on mucin core 1. A comparison of the BT1043 glycan-binding site with the equivalent site in BT3984 shows that nearly half of the residues lining this shallow pocket are conserved, including Y281 and Q67 (BT1043 numbering) that help coordinate GlcNAc in BT1043 (Figure 3B). In BT3984, F85 replaces W87 in BT1043, eliminating the hydrogen bonding donor provided by the indole N atom. In BT1043, LacNAc binds on one side of the glycan-binding pocket and is not within reasonable hydrogen bonding or hydrophobic stacking distance of R101, N68, or Y104. Interestingly, it is this side of the binding pocket that shows the most variability between BT1043 and BT3984, with R101, N68, and Y104 in BT1043 being replaced with W98, F65, and V101 in BT3984, respectively (Figure 3B). While the significance of these differences is unknown, it is certainly a possibility that these residues aid in binding branched versus linear *O*-glycan structures (*37*).

Although there are similarities in the overall fold and glycan-binding pocket architecture, electrostatic surface renderings of SusD, BT1043, and BT3984 highlight less obvious yet important differences among these proteins (Figure 4). For binding maltooligosaccharides, the SusD glycan-binding pocket features a shallow, flatter surface, comprised of an arc of aromatic residues that complements the shape of helical starch. Thus, in SusD, glycan recognition appears to rely more on the shape (e.g., van der Waals interactions) of the substrate, and less upon electrostatic or hydrogen bonding interactions between the protein and glucose residues. In contrast, the binding pockets of BT1043 and BT3984 are more articulated and dominated by polar residues that can provide a network of hydrogen bonding interactions. This may suggest that oligosaccharide recognition by *O*-glycan targeting SusD family members is dominated by electrostatic and hydrogen bonding interactions with individual monosaccharide components, as opposed to interacting mainly with the backbone of the glycan.

DISCUSSION

The atomic structures of the three SusD family proteins determined to date (BT1043, SusD, and BT3984) display a unique α-helical fold. By comparison, the carbohydrate-

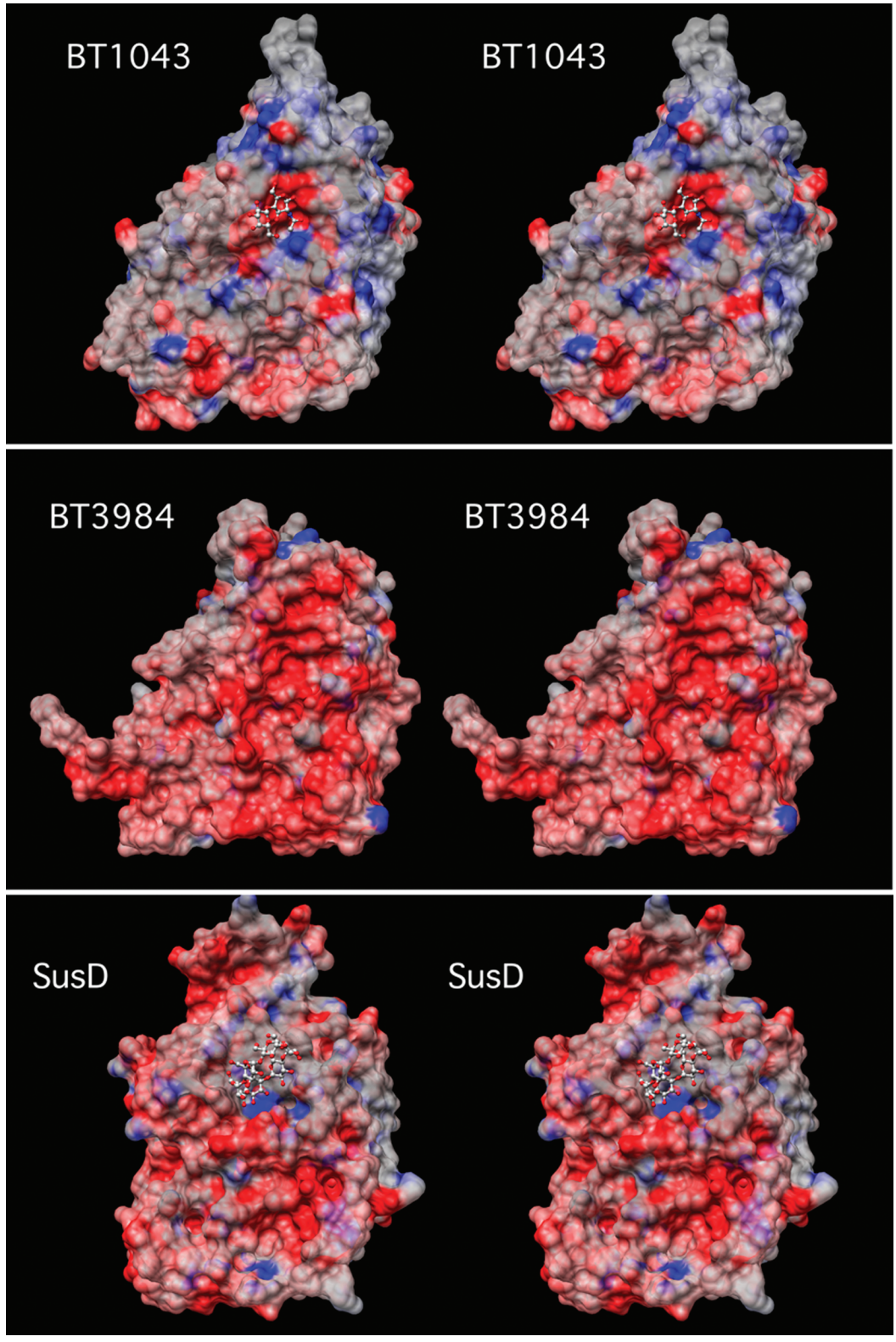

FIGURE 4: Comparison of glycan-binding pockets of BT1043, BT3984, and SusD. Shown here are stereopairs of surface representations generated in CHIMERA (*46*) of the glycan binding surfaces of these three polysaccaride binding proteins colored according to the electrostatic potential (red for negative and blue for positive) calculated with GRASP (*47*) using the DelPhi (*48*) algorithm. The ball-and-stick figure in the BT1043 model represents the structure of the bound LacNAc and α-cyclodextrin in the case of the SusD surface representation.

binding modules (CBMs) associated with cellulases and other glycoside hydrolases and lectin proteins are dominated by β-sheets (*39*). The most conserved feature of the SusD family is the presence of four TPR units that form a right-handed superhelix along one side of the protein and provide a scaffold upon which the rest of the α-helical structure is packed. BLAST searches with SusD, BT3984, and BT1043 amino acid sequences failed to detect the presence of the TPR domain since they all lack the archetypal consensus sequence, W4-L7-G8-Y11-A20-F24-A27-P32 (*31*). Proteins containing TPR units are typically involved in protein−protein interactions (*31*), and it is possible that the conserved TPR

units of the SusD family are required for associating with the other members of the extracellular protein complex, particularly SusC, a predicted TonB-dependent β-barrel porin that is a defining feature of the Sus-like PUL. Previous studies demonstrated that both SusC and SusD are required for starch binding to the cell surface. Further, since they were found to copurify, they likely form a functional oligomer that recognizes starch (*19, 20*). The most variable part of the SusD family structures contains a series of loops and α-helices that lie opposite the TPR superhelix (Figure 1B). The biological significance is unknown but could have implications for the way different SusD family proteins interact with other accessory proteins or glycolytic enzymes in a potential complex. Note that while a few residues from these variable regions span one side of the glycan-binding pockets in these proteins, the size and shape of this site are relatively conserved in all three structures. In the crystal structure of SusD complexed with α-cyclodextrin, two molecules of SusD were wrapped around the oligosaccharide, suggesting that tighter binding of the glycan may be mediated through avidity effects (*17*). On the basis of the close structural similarity of BT1043 and BT3984 with SusD, it is possible that these proteins also associate with their respective SusC homologues to create multivalent, high-affinity binding sites for mucin glycans.

All gut *Bacteroides* PUL defined to date contain *susC*/*susD* homologues that are usually found in combination with one or more genes encoding glycolytic enzymes. These clusters are often linked to genes encoding extracytoplasmic function sigma factors (ECF-σ)/anti-σ factor pairs, hybrid two-component systems, or other types of sensors or regulators (*8, 13*). BT1043 was initially identified as a SusD homologue based upon an iterative BLAST search within the *B. thetaiotaomicron* genome, using each low-scoring hit as a query sequence to identify more divergent SusC/SusD paralogs. Genes included as *susD* homologues in this search also had to meet the criteria of inclusion in a gene pair downstream of an independently identified *susC* homologue (*8*). The striking expansion of genes related to *susC* and *susD* in the genomes of human gut Bacteroidetes is likely due, at least in part, to gene duplication and diversification events (*8*). Thus, the variation in both amino acid sequence and substrate specificity of these factors likely represents a case of divergent evolution in which, over time, these PUL have diversified and evolved toward different glycan specificities. BT1043 is one of the most divergent SusD homologues, sharing only 6.7% sequence identity with SusD over all Cα atoms (based upon the structural alignment of the TPR domain), nearly that expected for two random sequences (~5.6%), whereas most structural homologues share at least 8−9% sequence identity (*40*). This implies that SusD and BT1043 (and BT3984) diverged very early in the evolution of the *Bacteroides* PUL.

To identify the repertoire of glycans that BT1043 can recognize, we examined the protein using the glycan array screening facility at the Consortium for Functional Glycomics Protein-Glycan Interaction Core (H) facility at Emory University. BT1043 was screened both by covalent labeling of the protein with a fluorescent tag and indirectly using fluorescently labeled antibodies to the N-terminal His tag of BT1043. Neither attempt identified a potential cognate glycan. However, this was not altogether surprising in light of previous analysis of SusD ligand binding (*17*). Maltooligosaccharide binding to SusD is dominated by the three-dimensional conformation of maltooligosaccharides and glycan backbone interactions rather than specific interactions with the composite glucose residues. Isothermal titration calorimetry confirmed that a circularized starch analogue such as α-, β-, or γ-cyclodextrin bound with nearly 20-fold greater affinity than similarly sized linear sugars even though the crystal structures suggested a similar mode of binding with little or no change in the number of hydrogen bonding or hydrophobic stacking interactions. Furthermore, the ϕ and ψ angles of maltotriose or maltoheptaose bound to SusD closely approached those of double-helical amylose, suggesting that SusD recognizes the native conformation of the polysaccharide. Therefore, it is possible that glycan recognition within the SusD family of proteins is driven by the native structure of the polysaccharide that is imposed by the linkages between composite sugars and not simply the stereochemistry of individual sugars. If this is the case, ligand binding to BT1043 or BT3984 will be dependent not only on the monosaccharide composition but also on the context of these sugars in terms of their linkages to each other and the overall size and shape of the glycan. Isothermal titration calorimetry experiments suggest that *N*-acetyllactosamine binds to BT1043 weakly with a K_d of ~12 mM. Consistent with this observation, the *BT1042−45* locus is upregulated ~4-fold in response to *N*-acetyllactosamine (*13*). The structure of BT1043 complexed with LacNAc demonstrates that the GlcNAc moiety is anchored by hydrogen bonding interactions at its reducing end and by aromatic stacking with the phenolic side chain of Y281. While it is likely that the cognate ligand of BT1043 contains GlcNAc, it is difficult to predict the preceding glycan linkages and monosaccharides that comprise the full-length oligosaccharide since mucin *O*-glycans can be quite variable in both length and composition. In addition, it is difficult to make predictions about glycan binding since BT1043 is markedly different from other CBMs or lectin structures.

It is quite possible that some structural rearrangement occurs in the glycan-binding pocket upon binding of a longer oligosaccharide to BT1043. In the structure of SusD with maltoheptaose, two small loops (residues 70−77 and 292−296) undergo rearrangement, allowing the ligand to bind, whereas maltotriose, bound in a position identical to those of three of the glucose residues of maltoheptaose, does not require such a structural perturbation for binding (*17*). Even though the binding pockets of BT1043 and BT3984 do not appear to have a similarly positioned flexible loop, it is possible that some rearrangement occurs to accommodate the larger, cognate glycan.

The lack of interactions between BT1043 and the galactose of LacNAc suggests that either galactose is not the preceding monosaccharide in the cognate ligand or the glycan linkage is different. It is possible that a β1,3 linkage off of GlcNAc, as opposed to the β1,4 linkage present in LacNAc, may more effectively span the glycan binding site and facilitate hydrogen bonding interactions with R101, Y104, and N64 in addition to a potential hydrophobic stacking interaction with Y281 (Figure 2B). β1,3-linked glycans are quite common in mucin *O*-glycan core structures and therefore are a likely possibility in the cognate glycan for BT1043 (*34, 37*). If the linkage were β1,3, or perhaps any other linkage besides

β1,4, then GalNAc may be a reasonable substitute for GlcNAc as the reducing sugar. In the complex of BT1043 with GlcNAc, the bridging oxygen of the β1,4 linkage is within 3.3 Å of Y281. However, it is not in an ideal orientation for hydrogen bonding and, therefore, may or may not play a role in dictating glycan specificity.

The hypothesis that BT1043 binds a mucin glycan containing a GlcNAc moiety is supported by the function of the other genes found in its PUL. In the majority of Sus-like PUL, one or more glycolytic enzymes contain N-terminal lipidation sites similar to that found in SusD and SusG, making it likely that they are also presented on the outer membrane. These enzymes are believed to begin degradation of the target polysaccharide for import into the cell via the associated TonB-dependent porin much as SusG cleaves large starch molecules into maltooligosaccharides. In the *BT1042*−*45* operon, a component of a larger PUL, *BT1044* encodes a predicted CAZy (carbohydrate-active enzymes) glycoside hydrolase family 18 enzyme endo-β-*N*-acetylglucosaminidase that could cleave adjacent to *N*-acetyllactosamine extensions, liberating reducing end *N*-acetylglucosamine-containing oligosaccharides from glycosylated mucin proteins (*13, 41*). Likewise, *BT1045* encodes a predicted concanavalin A-like lectin/glucanase, which can sometimes be found as part of multienzyme mucinase complexes (*42*). Both BT1044 and BT1045 contained predicted secretion and lipidation signal sequences for tethering to the outer membrane, suggesting that they may assist in processing mucin oligosaccharides prior to import via the SusC homologue, BT1042. Similarly, the *BT3983*−*88* PUL also encodes a predicted endo-β-*N*-acetylglucosaminidase (BT3987).

The mode of binding of LacNAc to BT1043 is quite different from that observed by other LacNAc and/or *O*-glycan binding carbohydrate binding modules (CBMs). The crystal structures of the family 32 CBM from *Clostridium perfringens N*-acetyl-β-hexosaminidase with bound galactose, LacNAc, and the type II blood group H-trisaccharide reveal that a series of hydrogen bonds with the nonreducing end galactose moiety dominate the protein−carbohydrate interface (*43*). Indeed, recognition of the nonreducing end of galactose was also observed in the crystal structures of the CBM 32 domain from the *Micromonospora viridifaciens* sialidase (*44*), and in the CBM 32 domain of the *C. perfringens* NanJ sialidase (*45*) with bound galactose. The likely function of these CBMs is to bind the nonreducing terminal end of mucin glycans, thus allowing attachment to glycans covering the surface of the mucin protein, while BT1043 seems to recognize the reducing end *N*-acetylglucosamine moiety of mucin *O*-glycans, suggesting that BT1043 binds free mucin glycans, thereby participating more in foraging rather than tissue adhesion.

Mucin glycan foraging by *B. thetaiotaomicron* is a significant contributing factor to this species adaptation to, and persistence in, the distal gut (*13*). Whole genome transcriptional profiling of *B. thetaiotaomicron* in the intestines of gnotobiotic mice revealed at least 12 Sus-like PUL that are upregulated 83−488-fold under conditions in which the bacteria are deprived of dietary polysaccharides and thus turn to host glycans as an alternative carbon source (*13*). Similarly, these and other PUL are also upregulated to varying degrees during growth on glycan fractions that are enriched for mucin *O*-glycans. In a *B. thetaiotaomicron* mutant with weakened ability to utilize several of its *O*-glycan targeting PUL, including those harboring BT1043 and BT3984, both persistence in the mouse intestines and mother-to-pup transmission are diminished, demonstrating the importance of this nutrient niche in vivo (*13*). The bacterium likely devotes multiple PUL to mucin *O*-glycan utilization because of the complexity and structural diversity of glycans in this class. Likewise, the abundance of mucin glycan PUL, with 12 systems highly upregulated in vivo, supports the necessity of this nutrient source to this species' survival in the competitive gut environment.

REFERENCES

1. Savage, D. C. (1977) Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol. 31*, 107–133.
2. 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.
3. Palmer, C., Bik, E. M., Digiulio, D. B., Relman, D. A., and Brown, P. O. (2007) Development of the Human Infant Intestinal Microbiota. *PLoS Biol. 5*, e177.
4. Shroff, K. E., Meslin, K., and Cebra, J. J. (1995) Commensal enteric bacteria engender a self-limiting humoral mucosal immune response while permanently colonizing the gut. *Infect. Immun. 63*, 3904–3913.
5. Mazmanian, S. K. (2008) Capsular polysaccharides of symbiotic bacteria modulate immune responses during experimental colitis. *J. Pediatr. Gastroenterol. Nutr. 46* (Suppl. 1), E11–E12.
6. 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.
7. Salyers, A. A., Vercellotti, J. R., West, S. E., and Wilkins, T. D. (1977) Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. *Appl. Environ. Microbiol. 33*, 319–322.
8. Xu, J., Mahowald, M. A., Ley, R. E., Lozupone, C. A., Hamady, M., Martens, E. C., Henrissat, B., Coutinho, P. M., Minx, P., Latreille, P., Cordum, H., Van Brunt, A., Kim, K., Fulton, R. S., Fulton, L. A., Clifton, S. W., Wilson, R. K., Knight, R. D., and Gordon, J. I. (2007) Evolution of Symbiotic Bacteria in the Distal Human Intestine. *PLoS Biol. 5*, e156.
9. 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.
10. Tancula, E., Feldhaus, M. J., Bedzyk, L. A., and Salyers, A. A. (1992) Location and characterization of genes involved in binding of starch to the surface of *Bacteroides thetaiotaomicron*. *J. Bacteriol. 174*, 5609–5616.
11. Cheng, Q., Yu, M. C., Reeves, A. R., and Salyers, A. A. (1995) Identification and characterization of a *Bacteroides* gene, csuF, which encodes an outer membrane protein that is essential for growth on chondroitin sulfate. *J. Bacteriol. 177*, 3721–3727.
12. Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C. H., Westover, B. P., Weatherford, J., Buhler, J. D., and Gordon, J. I. (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science 307*, 1955–1959.
13. Martens, E. C., Chiang, H., and Gordon, J. I. (2008) Mucosal glycan foraging enhances persistent and transmission by a human guy symbiont. *Cell Host Microbe 4*, 447–457.
14. Bjursell, M. K., Martens, E. C., and Gordon, J. I. (2006) 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.
15. 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.
16. D'Elia, J. N., and Salyers, A. A. (1996) Effect of regulatory protein levels on utilization of starch by *Bacteroides thetaiotaomicron*. *J. Bacteriol. 178*, 7180–7186.

17. 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.
18. D'Elia, J. N., and Salyers, A. A. (1996) Contribution of a neopullulanase, a pullulanase, and an α-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. *J. Bacteriol. 178*, 7173–7179.
19. 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.
20. 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.
21. 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.
22. 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.
23. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology* (Carter, C. W. J., and Sweet, R. M., Eds.) pp 307−326, Academic Press, New York.
24. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D55* (Part 4), 849–861.
25. Terwilliger, T. C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D56*, 965–972.
26. Jones, T. A., Zou, J.-Y., and Cowan, S. W. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A47*, 110–119.
27. Brunger, A. T., Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system (CNS): A new software system for macromolecular structure determination. *Acta Crystallogr. D54*, 905–921.
28. Navaza, J. (1994) AMoRe: An automated package for molecular replacement. *Acta Crystallogr. A50*, 157–163.
29. Collaborative Computational Project No. 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D50*, 760–763.
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the sterochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283–291.
31. D'Andrea, L. D., and Regan, L. (2003) TPR proteins: The versatile helix. *Trends Biochem. Sci. 28*, 655–662.
32. Lapouge, K., Smith, S. J., Walker, P. A., Gamblin, S. J., Smerdon, S. J., and Rittinger, K. (2000) Structure of the TPR domain of p67phox in complex with Rac·GTP. *Mol. Cell 6*, 899–907.
33. Taylor, P., Dornan, J., Carrello, A., Minchin, R. F., Ratajczak, T., and Walkinshaw, M. D. (2001) Two structures of cyclophilin 40: Folding and fidelity in the TPR domains. *Structure 9*, 431–438.
34. Karlsson, N. G., Nordman, H., Karlsson, H., Carlstedt, I., and Hansson, G. C. (1997) Glycosylation differences between pig gastric mucin populations: A comparative study of the neutral oligosaccharides using mass spectrometry. *Biochem. J. 326* (Part 3), 911–917.
35. Varki, A. (1999) *Essentials of glycobiology*, Cold Spring Harbor Laboratory Press, Plainview, NY.
36. Choi, J.-M., Hutson, A. M., Estes, M. K., and Prasad, B. V. V. (2008) Atomic resolution structural characterization of recognition of histo-blood group antigens by Norwalk virus. *Proc. Natl. Acad. Sci. U.S.A. 105*, 9175–9180.
37. Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J. (1999) *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Plainview, NY.
38. Koropatkin, N. M., Martins, E. C., Gordon, J. I., and Smith, T. J. (2008) Starch catabolism by a prominent human symbiont is directed by the recognition of amylose helices. *Structure 16*, 1105–1115.
39. Hashimoto, H. (2006) Recent structural studies of carbohydrate-binding modules. *Cell. Mol. Life Sci. 63*, 2954–2967.
40. Rost, B. (1997) Protein structures sustain evolutionary drift. *Folding Des. 2*, S19–S24.
41. Coutinho, P. B., and Henrissat, B. (1999) Carbohydrate-active enzymes: An intergrated database approach. In *Recent Advances in Carbohydrate Engineering* (Henrissat, B., and Svenson, B., Eds.) pp 3−12, The Royal Society of Chemistry, Cambridge, U.K.
42. Crennell, S., Garman, E., Laver, G., Vimr, E., and Taylor, G. (1994) Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain. *Structure 2*, 535–544.
43. Ficko-Blean, E., and Boraston, A. B. (2006) The interaction of a carbohydrate-binding module from a *Clostridium perfringens* N-acetyl-β-hexosaminidase with its carbohydrate receptor. *J. Biol. Chem. 281*, 37748–37757.
44. Newstead, S. L., Watson, J. N., Bennet, A. J., and Taylor, G. (2005) Galactose recognition by the carbohydrate-binding module of a bacterial sialidase. *Acta Crystallogr. D61*, 1483–1491.
45. Boraston, A. B., Ficko-Blean, E., and Healey, M. (2007) Carbohydrate recognition by a large sialidase toxin from *Clostridium perfringens*. *Biochemistry 46*, 11352–11360.
46. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem. 25*, 1605–1612.
47. Nicholls, A. (1993) GRASP: Graphical Representation and Analysis of Surface Properties, Columbia University, New York.
48. Nicholls, A., and Honig, B. (1991) A rapid finite difference algorithm, utilizing successive over relaxation to solve the Poisson Boltzmann equation. *J. Comput. Chem. 12*, 435–445.

BI801942A