Insights into glycan import by a prominent gut symbiont

Declan A. Gray[1], Joshua B. R. White[2], Abraham O. Oluwole[3], Parthasarathi Rath[4], Amy J. Glenwright[1], Adam Mazur[4], Michael Zahn[4], Arnaud Baslé[1], Carl Morland[1], Sasha L. Evans[2], Alan Cartmell[5], Carol V. Robinson[3], Sebastian Hiller[4], Neil A. Ranson[2], David N. Bolam[1#] and Bert van den Berg[1#]

[1]Biosciences Institute, The Medical School, Newcastle University, Newcastle upon Tyne NE2 4HH, UK.

[2]Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, UK.

[3]Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford OX1 3QZ, UK.

[4]Biozentrum, University of Basel, Basel, Switzerland

[5]Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK.

[#] To whom correspondence should be addressed: bert.van-den-berg@ncl.ac.uk

david.bolam@ncl.ac.uk

Abstract

In Bacteroidetes, one of the dominant phyla of the mammalian gut, active uptake of large nutrients across the outer membrane is mediated by SusCD protein complexes via a “pedal bin” transport mechanism. However, many features of SusCD function in glycan uptake remain unclear, including ligand binding, the role of the SusD lid and the size limit for substrate transport. Here we characterise the β2,6 fructo-oligosaccharide (FOS) importing SusCD from *Bacteroides thetaiotaomicron* (Bt1762-Bt1763) to shed light on SusCD function. Co-crystal structures reveal residues involved in glycan recognition and suggest that the large binding cavity can accommodate several substrate molecules, each up to ~2.5 kDa in size, a finding supported by native mass spectrometry and isothermal titration calorimetry. Mutational studies *in vivo* provide functional insights into the key structural features of the SusCD apparatus and cryo-EM of the intact dimeric SusCD complex reveals several distinct states of the transporter, directly visualising the dynamics of the pedal bin transport mechanism.

Introduction

The human large intestine is home to a complex microbial community, known as the gut microbiota, which plays a key role in host biology[1-3]. One such role is to mediate the breakdown of complex glycans which would otherwise be unavailable to the host because the human digestive tract lacks the necessary enzymes[4-7]. The products of this anaerobic glycan metabolism are short chain fatty acids which provide a range of both localised and systemic health benefits to the host[8,9]. In the colon, the microbiota is dominated by two bacterial phyla, the Gram-negative *Bacteroidetes* and Gram-positive *Firmicutes*[1,10]. The *Bacteroidetes* are glycan generalists and individual species often have the capacity to utilise a wide diversity of polysaccharides from plant, microbial and host sources[7,11,12]. The glycan-degrading apparatus in *Bacteroidetes* is encoded by co-regulated gene clusters known as polysaccharide utilisation loci (PULs), with each PUL encoding all of the proteins required for the acquisition and degradation of a specific glycan[7,13]. For example, the model gut symbiont *Bacteroides thetaiotaomicron* (*B. theta*) has 88 predicted PULs[14], although only a limited number have been characterised[12]. Most glycan degradation occurs intracellularly, and import of the substrate molecules across the OM is mediated by a class of PUL-encoded TonB-dependent transporters (TBDTs) known as SusCs (we propose to re-purpose the term "Sus" for saccharide uptake system rather than starch utilisation system)[7,12,13,15]. SusC proteins are unique amongst TBDTs in that they are tightly associated with a SusD substrate binding lipoprotein[15-17] (Fig. 1a). Recently we showed that SusCD complexes mediate substrate uptake via a “pedal bin” mechanism[15,17]. The SusC transporter forms the barrel of the bin,

while SusD sits on top of the barrel, opening and closing like a lid to facilitate substrate binding. Previous structures of loaded SusCD complexes revealed a bound ligand which was completely encapsulated by the closed pedal bin, indicating that SusCD transporters may have a size limit for transport[15,17]. An investigation on the prototypical SusCD system revealed that a mutant strain lacking the surface endo-amylase preferentially utilised malto-oligosaccharides with a degree of polymerisation (DP) of ~5-16[18], suggesting this is the preferred size range imported by the Bt3701-02 SusCD transporter. Direct evidence for this notion is lacking, however, and this issue has not been explored for any other Sus systems[12]. Furthermore, many other key features of SusCD function remain unclear, including the role of the SusD lid and other conserved structural elements, as well as the identity of the glycan recognition elements of both SusC and SusD.

One example of a typical PUL is the *B. theta* levan utilisation locus, spanning Bt1754-Bt1765[19]. Levan is a fructan polysaccharide comprised of β2,6-linked fructose units, with occasional β2,1 fructose decorations and is produced mainly by bacteria as an exopolysaccharide, but also by some cereal plants such as wheat[20,21]. For the levan utilisation system, the cell-surface components are Bt1760 (a GH32 endo-levanase), Bt1761 (a surface glycan binding protein; SGBP), Bt1762 (SusD) and Bt1763 (SusC)[15,19,22] (Fig. 1a). The cell surface levanase Bt1760 and glycan binding proteins (Bt1761 and Bt1762) have been shown to be specific for levan, with no activity against, or binding to, β2,1-linked inulin-type fructans that are common to many plants[19,23] .

In this study we characterise binding and uptake of β2,6 fructo-oligosaccharides (FOS) by the Bt1762-63 SusCD transporter. Co-crystal structures of the closed complex with bound glycans reveal the residues involved in substrate recognition and suggest that the large binding cavity can contain several substrate molecules, each up to ~2.5 kDa in size, a finding supported by native mass spectrometry and isothermal titration calorimetry. Mutational studies *in vivo* provide insights into the key structural features of the SusCD apparatus and cryo-EM of the intact $SusC_2D_2$ complex reveals several distinct states of the transporter, *i.e*. open-open, open-closed and closed-closed. These structures directly visualise the dynamics of the pedal bin mechanism and suggest that the individual SusCD complexes function independently of each other. Taken together, these results provide important insights into the mechanism of glycan import by SusCD complexes of dominant gut bacteria.

Results

A FOS co-crystal structure reveals SusCD residues involved in ligand binding

In the *B. theta* levan PUL, the periplasmic enzymes are GH32 exo-acting fructosidases that release fructose from the imported β2,6 FOS (Fig. 1a)[19,22]. The fructose is recognised by the periplasmic domain of the PUL sensor-regulator BT1754, which upregulates expression of the PUL[19]. Thus, by growing *B. theta* on fructose, large amounts of the Bt1762-63 transporter can be obtained for structural work[15].

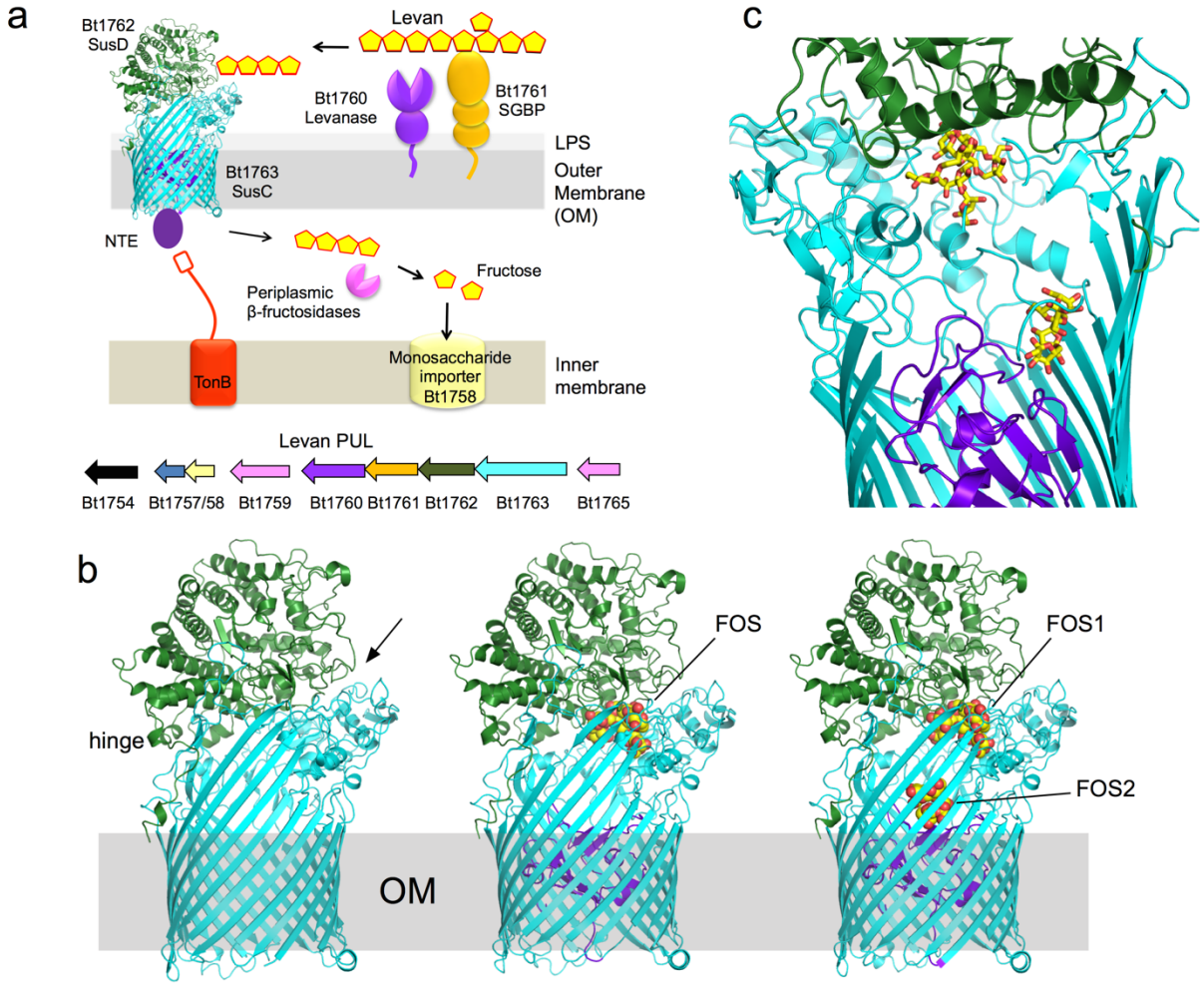

Figure 1 Structures of Bt1762-63 in the presence of β2,6 FOS. a, Schematic overview of the *B. theta* levan Sus apparatus (top), encoded by the PUL spanning Bt1754-65 (bottom). NTE is the N-terminal extension of Bt1763. **b**, Cartoon representations of Bt1762-63 without ligand (left panel; PDB ID 6Z8I), and Bt1762-63 with longer (middle; PDB ID 6Z9A) and shorter FOS (right; PDB ID 9ZAZ). Bt1762 (SusD) is coloured green and Bt1763 (SusC) cyan. The plug domain of Bt1763 is coloured purple. The bound FOS molecules are shown as space-filling models, with oxygens coloured red and carbons yellow. The N-terminal hinge at the back of BT1762, allowing opening of the transporter at the front (arrow) is labelled. **c**, Close-up of the two FOS binding sites inside the BT1762-63 binding cavity. FOS molecules are shown as stick models. X-ray structural figures were made with Pymol (https://pymol.org/2/).

A previous crystal structure of Bt1762-63 obtained in the absence of levan substrate revealed a dimeric ($SusC_2D_2$) closed state in which the SusC TonB-dependent transporter (Bt1763)

lacked the plug domain as a result of proteolytic cleavage[15]. We have now obtained a structure without substrate using a preparation that did not suffer from proteolysis (Fig. 1b, left panel; Supplementary Table 1). Interestingly, while this structure is very similar to that reported earlier, density for the plug domain is weak but clearly present. While this suggests that the plug has been ejected from the barrel in the majority of transporter molecules in the crystal, the relatively poor fit of the density with the native plug domain suggests increased dynamics of the *in situ* plug domain in the subset of transporters that contain a plug (Extended Data Fig. 1).

To provide further insight into glycan recognition and transport by SusCD complexes, we next used the same protein preparation to determine a co-crystal structure using data to 3.1 Å resolution with β2,6-linked FOS. The FOS were generated by partial digestion of levan by Bt1760 endo-levanase, followed by size exclusion chromatography (SEC) and analysis by thin-layer chromatography (TLC) and mass spectrometry (MS; Methods). In this structure, containing FOS with a wide range of sizes (~DP15-25), the plug domain is present with normal occupancy, suggesting that it is more stable in the presence of substrate (Fig. 1b, middle panel and Extended Data Fig. 1). Like the oligopeptide ligands in the RagAB and Bt2261-64 structures[15,17], the FOS is bound at the top of a large, solvent-excluded cavity formed by the Bt1762-63 complex. Density for seven β2,6-linked fructose units can unambiguously be assigned in the structure and this was designated as the primary binding site (Fig. 2a). The bound oligosaccharide is compact and has a twisted, somewhat helical conformation. The ligand makes numerous polar contacts with side chains of residues in both Bt1762 and Bt1763 (Fig. 2b). For Bt1762 (SusD) these residues are D41, N43, D67, R368 and Y395, and for Bt1763 (SusC) T380, D383, D406 and N901. In addition, prominent stacking interactions are present between the ring of fructose 2 (Frc 2) and W85 of Bt1762. Interestingly, a β2,1 decoration is present in the bound ligand at Frc 4, and the branch point interacts with the extensive non-polar surface provided by the vicinal disulphide between Cys298 and Cys299 of Bt1762 (Fig. 2b).

We also determined a co-crystal structure of Bt1762-63 with shorter β2,6 FOS (~DP6-12) using data to 2.69 Å resolution (Supplementary Table 1). Strikingly, besides having very similar density for glycan at the principal binding site (FOS1; Fig 1b, right panel), this structure contains a second ligand molecule at the bottom of the SusC cavity, contacting the plug (FOS2; Fig 1b, right panel, and Fig. 1c). Density for four fructose units can be seen in this secondary binding site, and polar contacts are made with Bt1763 residues Q372, K454, E481 and W483 in the barrel wall, and with H169 and E170 in the plug domain (Fig. 2c). The fit to

the density is better for a 3-mer with a β2,1 decoration compared to a β2,6-linked 4-mer, suggesting the transporter may have some specificity for FOS with a β2,1 decoration, or alternatively, that *Erwinia* levan contains extensive β2,1 decorations such that most of the levanase products are branched. The relatively small size of the co-crystallised FOS, combined with the relative orientation and the large distance between FOS1 and FOS2 (> 20 Å; Fig. 1c), makes it highly plausible that there are two ligand molecules in the Bt1762-63 cavity. The co-crystal structure with the longer FOS also shows some density at the secondary site, but it is of insufficient quality to allow model building, perhaps due to the lower resolution.

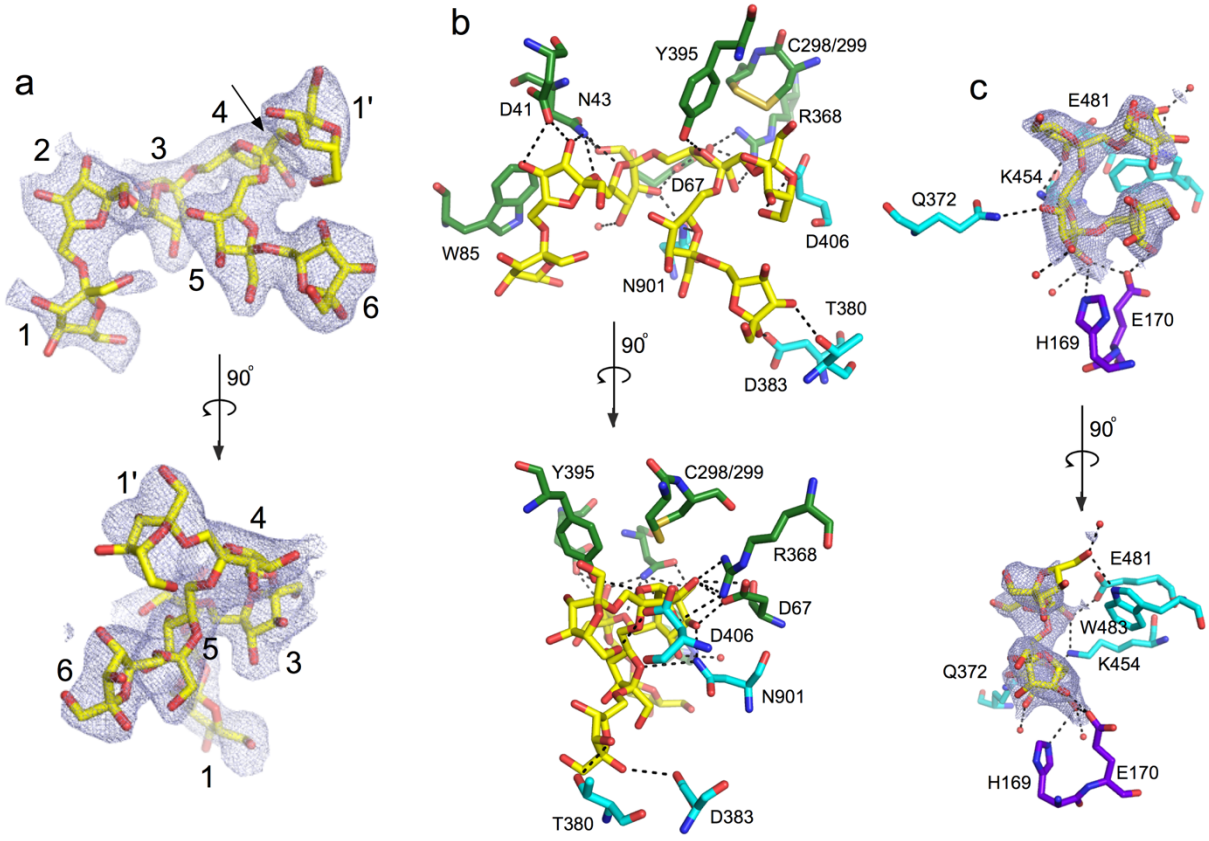

Figure 2 FOS densities and interactions with the transporter. a, 2Fo-Fc density (contoured at 1.0 σ, carve = 2) of FOS1 bound in the primary binding site rotated by 90 degrees. The glycan branch point is indicated with an arrow. **b**, Close-ups of the principal FOS binding site with residues making polar interactions with the substrate shown as stick models and coloured as in **b**. Polar interactions are shown as dashed lines, and water molecules as red spheres. For both Bt1762 and Bt1763, residue numbering is for the mature protein without the signal sequence, *i.e.* starting at C19 for Bt1762 and G26 for Bt1763. **c**, Close-ups of the FOS2 binding site with polar interactions indicated. Plug residues are coloured purple.

Single-particle cryo-EM reveals different conformational states of Bt1762-63

The very similar crystal structures of Bt1762-63 in the absence and presence of FOS appear at odds with the dynamics proposed in the pedal bin transport model[15], but are likely explained by the fact that crystallisation selects for compact, stable states. To visualise potential different conformational states, we used single-particle cryo-EM on detergent-solubilised apo-Bt1762-

63. Following initial 2D classification, it is clear that Bt1762-63 is dimeric in solution, demonstrating that the dimers in the crystal structures[15] are not artefacts. Strikingly, several distinct conformations of the $SusC_2D_2$ complex are present in a single dataset, and after further classification steps, three dominant states were identified. Following 3D reconstruction, electron density maps were obtained at resolutions allowing rigid-body placement and refinement of individual protomers. This yielded structures corresponding to the three possible combinations of open and closed dimeric transporters: closed-closed (CC; 4.2 Å), open-closed (OC; 4.7 Å) and open-open (OO; 3.9 Å) (Fig. 3 and Supplementary Table 2). With the exception of a few Bt1763 loops (L6 and L7; Fig. 3c) and the Bt1762 segment following the lipid anchor that serves as a pivot around which Bt1762 moves upon lid opening, the conformational changes between the three states correspond to rigid body movements of Bt1762. In contrast to the recently characterised RagAB peptide transporter from *Porphyromonas* gingivalis[17], the Bt1762-63 cryo-EM dataset shows evidence of intermediate states of lid opening (Extended Data Fig. 2), suggesting the presence of a number of minima in the energy landscape for opening and closing of the transporter. Interestingly, the presence of the plug domain in the interior of the Bt1763 barrel in solution is strictly correlated with the open *vs*. closed state of the transporter; only open states contain a plug domain, regardless of the state of the $SusC_2D_2$ dimer (Extended Data Fig. 3). Given that there are no direct contacts between Bt1762 and the Bt1763 plug domain, this suggests that the transition from the open to the closed state generates conformational changes in the barrel of Bt1763, leading to displacement of the plug domain from the barrel *in vitro*. Superposition of the crystal structures of plug-less and complete transporters indeed show non-uniform displacements of ~2-2.5 Å for C_α atoms of barrel wall residues, which is substantially more than the coordinate errors of the structures (~0.5 Å), showing that the barrels have subtly different shapes (Extended Data Fig. 4).

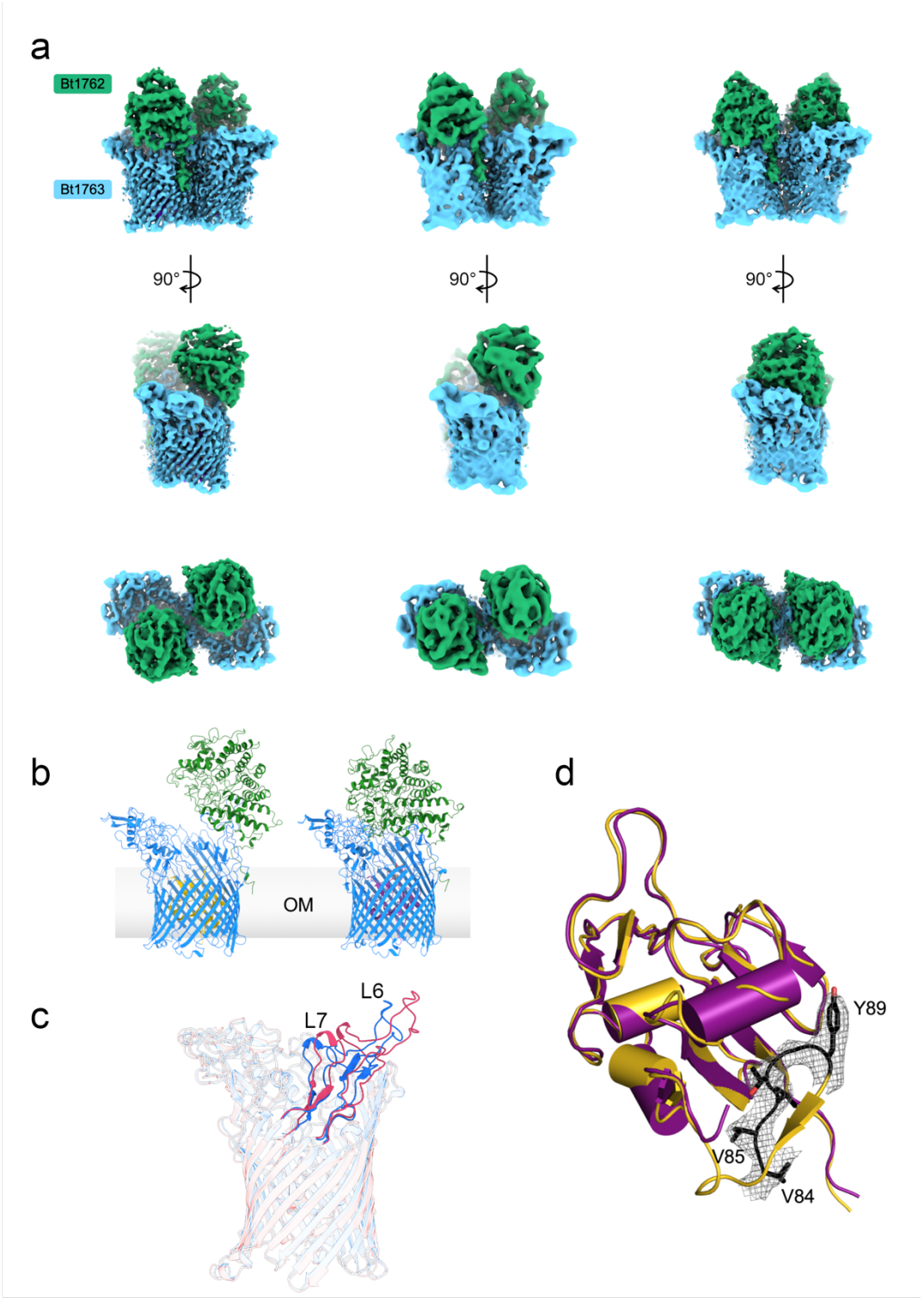

Figure 3 Dynamics of pedal bin opening revealed by single particle cryo-EM. a, Electron density filtered by local-resolution for Bt1762-63 viewed from the OM plane for the OO state (left panels), OC state (middle panels) and CC state (right panels). The bottom panels show views from the extracellular environment. **b**, Cartoon models of the individual open (left) and substrate-bound, closed (right) SusCD transporters structures derived by cryo-EM and X-ray crystallography, respectively. For clarity, transporters are shown as monomers. **c,** Superposition of open (red) and closed states (blue) of Bt1763, with both hinge loops L6 and L7 highlighted. **d**, Overlay of plug domains from open (gold) and closed (purple), substrate-bound forms of the transporter. The Ton box (^{82}DEVVVTG^{88}) is visible only in the open state

and is colored black, with sidechains displayed as sticks. The equivalent region in the closed structure is not visible and is therefore assumed to be disordered and likely protrudes from the barrel, leaving the Ton box accessible to TonB. The visible density for the N-terminus starts at residue 96 of the "closed" plug and at residue 84 for the "open" plug. Cryo-EM figures were made with ChimeraX[52].

The established model of TonB-dependent transport[24] assumes that extracellular substrate binding to a site that includes residues from the plug domain induces a conformational change that is transmitted through the plug domain of the TBDT. This results in disordering of its N-terminal Ton box and increased accessibility of the Ton box for interaction with the C-terminal domain (CTD) of TonB. In the open, apo state of the SusCD transporter as observed via cryo-EM, the Ton box of Bt1763 (^{82}DEVVVTG^{88}) is visible, tucked away inside the barrel, and interacts with the body of the plug (Fig. 3d). By contrast, the density in the crystal structure of the FOS-bound BT1763 starts only at L97, indicating that the Ton box is disordered and would likely be exposed to the periplasmic space. Thus, our structures for Bt1762-63 are consistent with the proposed behaviour of the Ton box upon ligand binding by the transporter.

Structure-function studies of Bt1762-63

SusD mutants: The observed interactions of FOS1 with Bt1762 SusD in the co-crystal structures are in excellent agreement with previous ITC studies of levan binding to recombinant Bt1762 variants with single substitutions of aromatic residues and cysteines (Fig. 2b)[15]. The mutations W85A and C298A abolished *in vitro* levan binding by Bt1762, whereas the affinity of Y395A was ~ 6-fold weaker than that of wild-type (WT). Single substitutions of five other aromatic and cysteine residues remote from the ligand binding site did not affect levan binding[15]. To assess how single substitutions that abolish levan binding to Bt1762 *in vitro* affect the ability of *B. theta* to utilise levan *in vivo*, a W85A mutant strain was constructed and growth was analysed using 0.5 % *Erwinia* levan as the sole carbon source. Surprisingly, there was no growth defect observed for the W85A strain (Fig. 4a), which, combined with its WT levels of expression (Figs. 4d,e), indicates this variant is fully functional and that care is required in drawing conclusions from *in vitro* binding studies, especially of the SusD protein in isolation. Clearly, the context of the intact transporter ensures that the effects of SusD point mutations may be much less dramatic *in vivo*. When 7 residues in Bt1762 involved in direct interaction with FOS1 (Fig. 2b; green sticks; D41, N43, D67, W85, C298, R368 and Y395) were changed to alanine at the same time, the resulting strain ($SusD_{BR}$) grew only after a prolonged lag phase of approximately 8 h, which is comparable to a strain lacking Bt1762 (ΔSusD; Fig. 4a). However, the OM levels of the $SusD_{BR}$ mutant are much lower than that for WT and W85A strains (Figs. 4d,e), providing an explanation for the observed growth defects. Given these results for Bt1762 mutants, we opted not to investigate the effect of Bt1763

binding residue mutations but instead focused on other potentially important functional regions of the transporter.

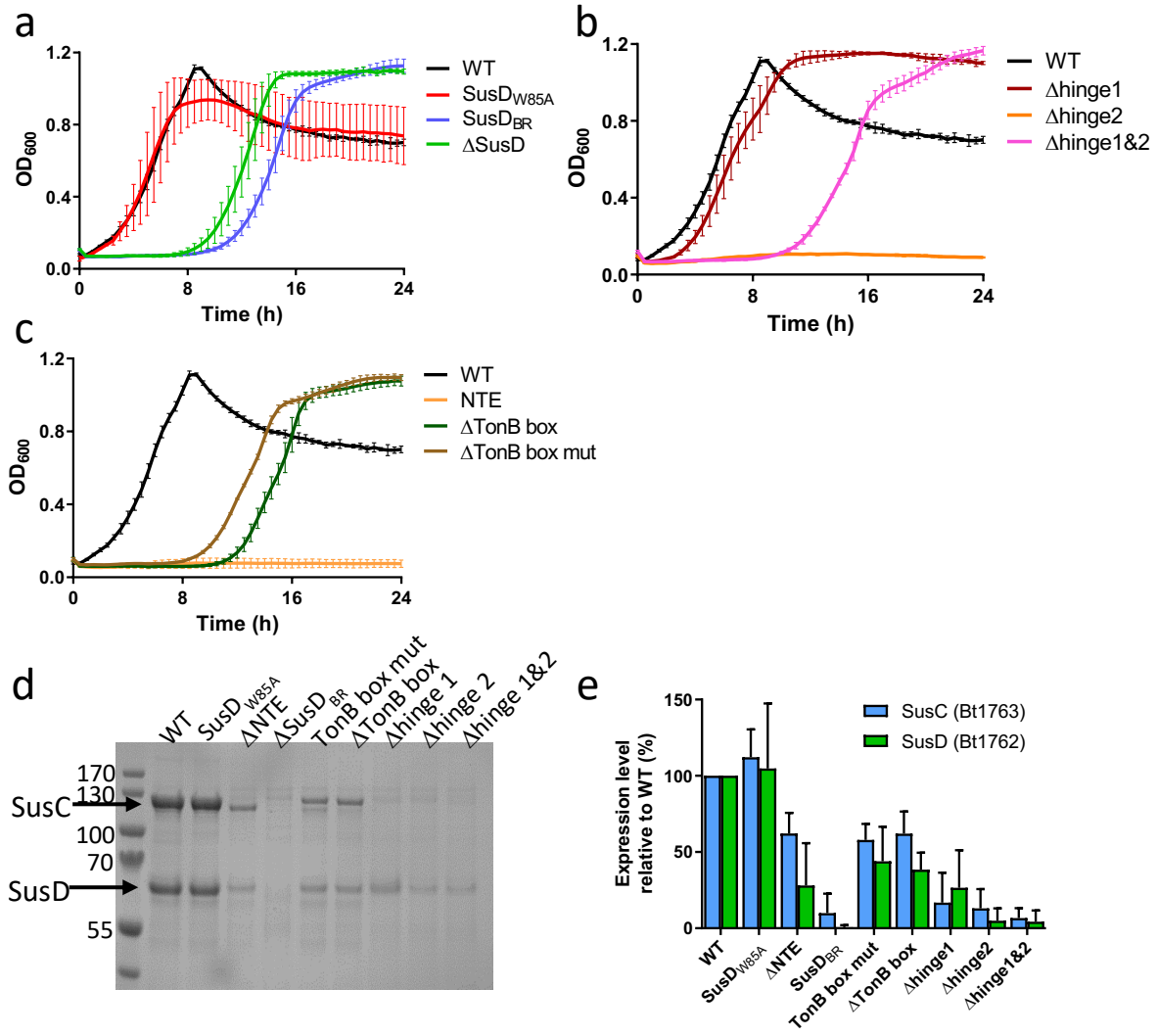

Figure 4 Structure-function studies of the Bt1762-63 SusCD levan importer. a-c, Growth curves on levan of Bt1762-SusD mutant strains (**a**), Bt1763-SusC hinge loop mutant strains (**b**), and Bt1763 ΔNTE and TonB box mutant strains (**c**). WT = wild type strain (with His6 tag on C-terminus of Bt1762). Growths were performed in triplicate on at least two separate occasions. **d**, Representative SDS-PAGE gel (n = 3) of IMAC-purified Bt1762-63 complexes from OM fractions of WT (Bt1762-his strain) and mutant strains grown on MM-fructose. The far left lane is MW marker with the sizes shown in kDa. **e**, Relative expression levels quantified from band intensities on SDS-PAGE gels using ImageJ (n = 3; means ± S.D are shown). Band intensities of WT Bt1762 and Bt1763 were set to 100 %.

Hinge region mutants: Loop L7 of Bt1763, identified as a 'hinge' loop in previous simulations of Bt2263-64[16] and hereafter named hinge1, appears likely to be important for Bt1762 lid opening (Fig. 1b and Fig. 3c). In addition, we identified from the cryo-EM structures a second loop in Bt1763, L6 (Fig. 3c), as potentially important for lid opening (designated hinge2). Hinge 1 and 2 change conformation substantially during lid opening and are responsible for the

majority of interactions between Bt1762 and Bt1763 in the open state (Fig. 3c). The deletion of hinge1 (Methods) caused little to no growth defect (Fig. 4b), which is surprising given that Bt1762-63$_{\Delta hinge1}$ expression is barely detectable (Figs. 4d,e). By contrast, the ∆hinge2 strain showed a complete lack of growth during the 24 h monitoring period, but expression of this mutant was also very low (Figs. 4b,d,e). Surprisingly, a strain in which both hinges were deleted (∆hinge1&2), grew similarly to the ∆SusD strain, *i.e*. after a ~8 hr lag phase (Fig. 3b).

TonB box and N-terminal extension mutants: SusC-like proteins are predicted to be TonB-dependent transporters (TBDTs), but direct evidence for this is lacking. We therefore examined the importance of the putative TonB box located at the N-terminus of Bt1763. In addition, SusC proteins have an N-terminal extension (NTE) domain of ~9 kD (Pfam 13715, Carboxypeptidase D regulatory-like domain) that is absent in other types of TBDTs and precedes the TonB box, but has an unknown function (Fig. 1a)[32]. We constructed two mutant *B. theta* strains in which the TonB box was either deleted (residues ^{81}VDEVVV^{86}; ∆TonB box) or mutated to all-Ala (TonB box mut; Methods), and a strain in which the complete NTE was deleted (residues 3-76; ∆NTE), but with the Ton box intact. Growth of the mutants on levan as the sole carbon source revealed that both TonB box mutant strains had similar growth defects to that of ∆SusD, and were expressed at reasonable levels (~50% of WT; Figs. 4c-e), demonstrating that Bt1763 is a bona-fide TBDT. Remarkably, the ∆NTE strain shows no detectable growth, despite expression levels of ~25-50% that of WT, indicating an important role for this domain in transporter function.

Solution NMR structure shows an Ig-like fold for the NTE

The striking, dominant-negative effect of the ∆NTE mutant on growth made it important to determine the structure of the isolated NTE, since it is invisible in the X-ray structures and has very weak density in the cryo-EM structures. Crystallisation trials proved unsuccessful, and we therefore solved the structure by high-resolution NMR spectroscopy, using uniformly ^{15}N,^{13}C-labelled protein produced in *E. coli* (Extended Data Fig. 5). A 2D [^{15}N,^{1}H]-HSQC spectrum of the NTE showed good chemical shift dispersion and the expected number of resonances (Fig. 5a). Complete sequence-specific assignments for the backbone resonances were then obtained using a combination of 2D and 3D experiments (Supplementary Tables 3 and 4, Extended Data Fig. 5). Backbone chemical shifts were compared to the random coil values, resulting in the identification of 7 β-strands. A total of 964 distance constraints were obtained from 3D NOESY spectra, and served as input for structure calculations in CYANA. The final ensemble of 20 lowest energy structures had an r.m.s.d. of 0.52 Å for the backbone heavy atoms (Fig. 5b, Extended Data Fig. 5 and Supplementary Tables 3 and 4). The NTE

structure shows a well-defined core of an Ig-like fold with a 7-stranded barrel (Fig. 5c, left panel). The N-terminus (including the His-tag) and the C-terminus, corresponding to the Ton box, are flexibly unstructured, as evidenced from their random coil chemical shifts and the absence of long-range NOEs (Fig. 5b and Fig. 5c, left panel).

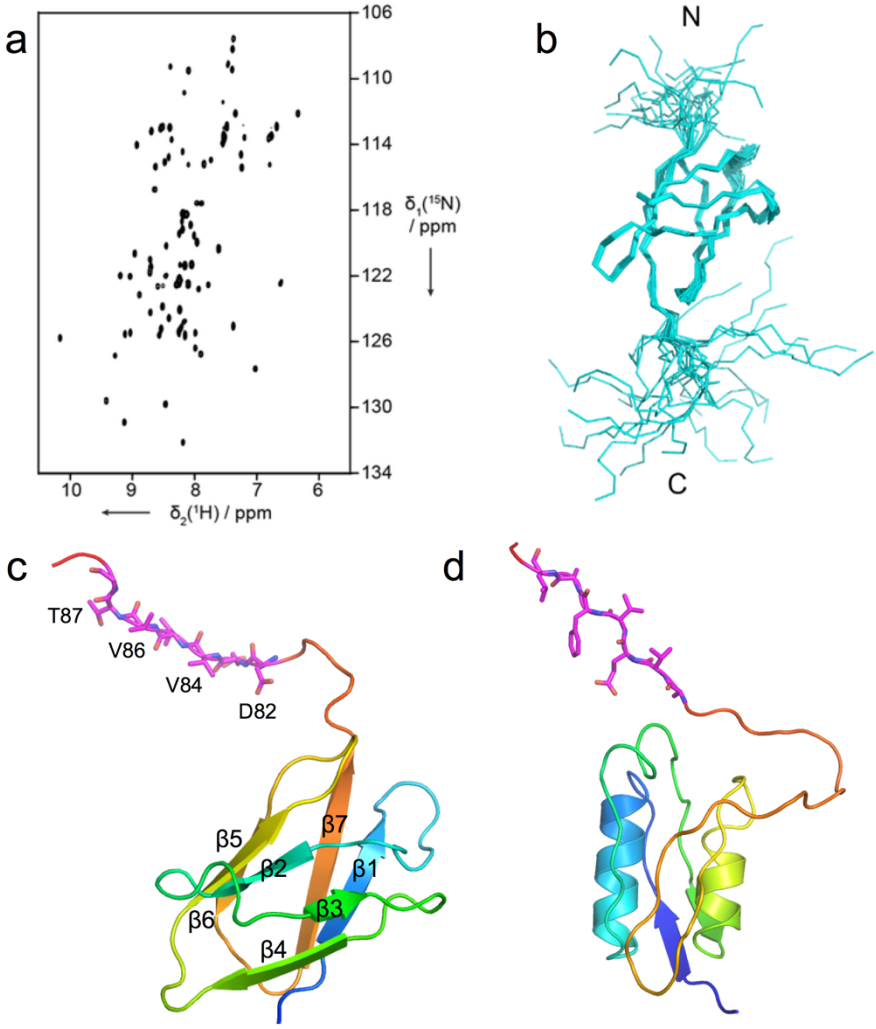

Figure 5 Solution NMR structure of the BT1763 NTE shows a 7-stranded barrel. a, 2D [^{15}N,^{1}H]-HSQC spectrum of the NTE. **b**, Ensemble of the 20 lowest energy structures of the NTE presented in stick view for the main chain. The N- and C-termini are labeled. **c**, Lowest-energy NMR structure from the ensemble of (**b**) in ribbon representation, coloured in rainbow mode (N-terminus; blue). **d**, STN domain of *Pseudomonas aeruginosa* FoxA (PDB ID 6I97) in the same coloring mode, for comparison. The Ton boxes of each domain (^{82}DEVVVTG^{88} in BT1763; based on the alignment with FoxA) is coloured magenta.

A DALI analysis[25] returned a number of structures with significant similarity, the highest of which had a Z-score of 10.2 (Transthyretin-like domain of Carboxypeptidase D, PDB ID 5aq0; Extended Data Fig. 6). However, none of the DALI hits provided any insights into a potential function of the NTE. In Proteobacteria, structures of TBDTs with a different N-terminal domain, designated STN (Pfam 07660), have been reported. This STN domain has a similar size as the NTE and has an established role in signalling, where it interacts directly with an anti-sigma

factor in the inner membrane to stimulate TBDT expression in response to the presence of their cognate substrates[24]. Notably, SusCs from ECF-sigma/anti-sigma controlled PULs also contain STN domains in addition to an NTE domain, with the NTE preceding the STN in all cases[26]. STN domains in SusCs have also been shown to be involved in anti-sigma signalling, indicating that the NTE domain in SusCs has a different role compared to the STN[26]. Interestingly, the structure of the FoxA STN in complex with the CTD of TonB shows that the STN is also composed of a small barrel with seven elements, some of which are helical instead of strands (Fig. 5c)[27]. This similarity suggests that, like the STN, the NTE might interact with a protein in the periplasmic space. In both domains, the Ton box is separated from the domain body and will thus be accessible to binding by the C-terminal domain of TonB. One possibility for a role for the NTE could be to provide interaction specificity for the multiple TonB orthologs present in the *B. theta* genome.

Investigation of the size range of FOS import by Bt1762-Bt1763

The co-crystal structures of Bt1762-63 with FOS, as well as the previously published structures of Bt2261-64[15] and RagAB[17], show that ligands are encapsulated within a large, solvent-inaccessible cavity formed by the closure of the SusD lid, strongly suggesting there is a size limit for substrate import. This hypothesis is further supported by previous studies showing that surface endo-acting enzymes are required for growth on large polysaccharides in several Sus[28,29]. In addition, import of monosaccharides and even small oligosaccharides does not appear to require a functional Sus apparatus, suggesting there may be also be a minimum size limit for transport by SusCD complexes[30]. To shed light on this issue we used isothermal titration calorimetry (ITC) to investigate the size range of FOS recognised by Bt1762-63. The β2,6 FOS was produced by digestion of *Erwinia* levan with Bt1760 endo-levanase, followed by separation via size exclusion chromatography (Methods). For estimation of the DP of each oligosaccharide we used both thin-layer chromatography (TLC) and mass spectrometry (MS; Extended Data Fig. 7). While this was effective for FOS with a DP of 4 to ~12, FOS larger than ~DP14 do not move from the origin in TLC, and the MS analysis of FOS larger than ~DP12 proved difficult because of poor ionisation efficiency of longer oligosaccharides. It is however clear that the higher DP FOS fractions contain a wide range of sizes.

ITC revealed that Bt1762-63 binds to the majority of FOS examined, with only the tetra-saccharide displaying no affinity for the transporter (Fig. 6, Extended Data Fig. 8 and Supplementary Table 5). For the larger FOS, affinity increased from DP5 to 6 (K_d ~30 and 17 μM, respectively) and plateaued at DP8 (tube 174, T174 K_d ~1 μM), with Bt1762-63 binding

to all FOS between DP8 and at least DP13-14 (T115) with similar affinity. These data are in broad agreement with the co-crystal structures, which show well-defined density for 7 fructose units in the primary binding site, suggesting that these provide the bulk of the binding interactions. Surprisingly, no binding was detected for the FOS in SEC fractions T114 and T113, despite these fractions having similar MS profiles to T115 with a broad range of oligosaccharides present (Fig. 6 and Extended Data Fig. 7). The average MW of the FOS in tube T114 (M_n >2666) is larger than that of T115 (M_n >2193; Extended Data Fig. 7), and it may be that this increase in average size is enough to preclude binding to the transporter. Furthermore, based on the co-crystal structure we can see that at least some, and perhaps all, of the bound *Erwinia* levan-derived FOS has a β2,1 decoration, which may influence binding to Bt1762-63. However, it was not possible to identify β2,1 decorations in the TLC or MS analysis. Thus, T115 FOS might contain significantly more branched species than T114 and this could explain a higher affinity for the T115 fraction. Taken together, however, these data indicate there is both an upper and lower size limit for FOS binding to the Bt1762-63 transporter *in vitro*, with the lower limit being DP5 and the upper limit ~DP15. In addition to wild-type Bt1762-63, we also measured binding of ~DP9 FOS to the $Bt1762_{(W85A)}$-63 variant (Fig. 6). Surprisingly, no binding is observed for the mutant, even though the $Bt1762_{W85A}$-63 strain grows as well as wild type on levan (Fig. 3a), suggesting that FOS binding by Bt1762 is not essential for Bt1762-63 function *in vivo*.

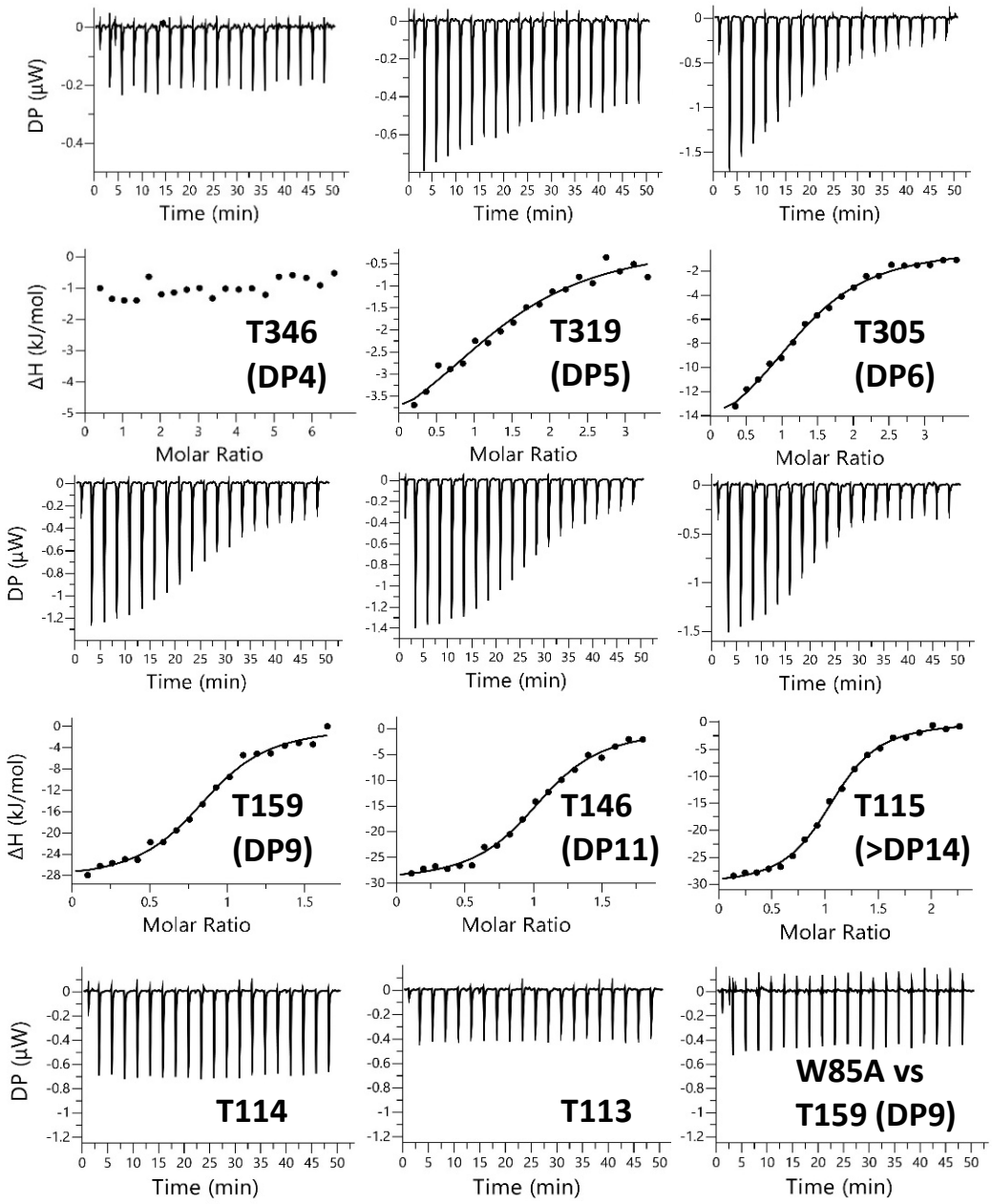

Figure 6 FOS binding to Bt1762-63 analysed by ITC. β2,6 FOS fractions from SEC of partially digested *Erwinia* levan were titrated into pure Bt1762-63 SusCD (25 µM or 50 µM) in 100 mM Hepes, pH 7.5 containing 0.05 % LDAO. The identity of the fraction used is indicated as is the approximate DP of the main FOS species present in each fraction as determined by MS. The upper parts of each titration show the raw heats of injection and the lower parts the integrated heats fit to a one set of sites model if binding was observed.

We used native mass spectrometry to shed further light on the ligand binding property of the Bt1762-63 complex. For the ligand-free protein, the mass spectrum revealed charge states corresponding to the intact $SusC_2D_2$ dimeric complex and low intensity peaks assigned to SusCD monomer (Extended Data Fig. 5a). Ligand binding appeared to induce considerable destabilization of the $SusC_2D_2$ complex as reflected in the higher intensity of monomer peaks in the presence of FOS molecules. Specifically, binding of FOS fractions characterised by

DP>14 (T115) and DP>16 (T114) yielded low intensity adducts to the intact $SusC_2D_2$, consistent with substrate binding for both samples. While this appears surprising, it should be noted that higher ratios of FOS : protein were used in MS compared to ITC (~6-fold *vs* ~2-fold respectively), providing a possible explanation for the lack of observed binding in ITC for T114. The complexity of the binding pattern in the spectrum is consistent with polydispersity of the T114 and T115 fractions (Extended Data Fig. 9). More useful insights were obtained with the T159 sample, which consists mainly of FOS with 8-10 fructose units (Fig. 7 and Extended Data Fig. 7b). These medium-chain oligosaccharides bind preferentially to the intact $SusC_2D_2$ dimer rather than to the SusCD monomer such that no ligand-free dimer was evident in the spectrum, potentially suggesting some kind of cooperativity for ligand binding in the dimer. Interestingly, the relative proportions of protein-bound FOS mirrored their abundance in the T159 sample (Fig. 7b), supporting the similar affinities of FOS with 8-10 fructose units for Bt1762-63 as measured by ITC (Fig. 6 and Supplementary Table 5). At the higher FOS concentrations, binding of more than one FOS molecule per SusCD transporter was observed (Fig. 7b), confirming the observation from our co-crystal structure that more than one ligand molecule can be present in the binding cavity, at least for the relatively small FOS.

Finally, we wanted to confirm the upper FOS size limit *in vivo* by using testing growth of a strain lacking the surface endo-levanase BT1760 against FOS of different sizes as the sole carbon source. The Δ1760 strain was previously reported to lack the ability to grow on levan[19], which would provide another indication that the Bt1762-63 complex cannot import high molecular weight substrates. Surprisingly, however, the growth rate of the Δ1760 strain on levan from several different sources was similar or only slightly slower than that of the wild type strain (Extended Data Fig. 10). PCR of the Δ1760 cells taken from stationary phase of the cultures confirmed the deletion of the BT1760 gene from the cells, indicating the phenotype was not due to contamination with wild-type strain (Extended Data Fig. 10). These data suggest that all the levans tested contained enough low DP FOS to allow growth without needing digestion by the surface endo-levanase. It was therefore not possible to determine an upper FOS size limit of the Bt1762-63 importer *in vivo*.

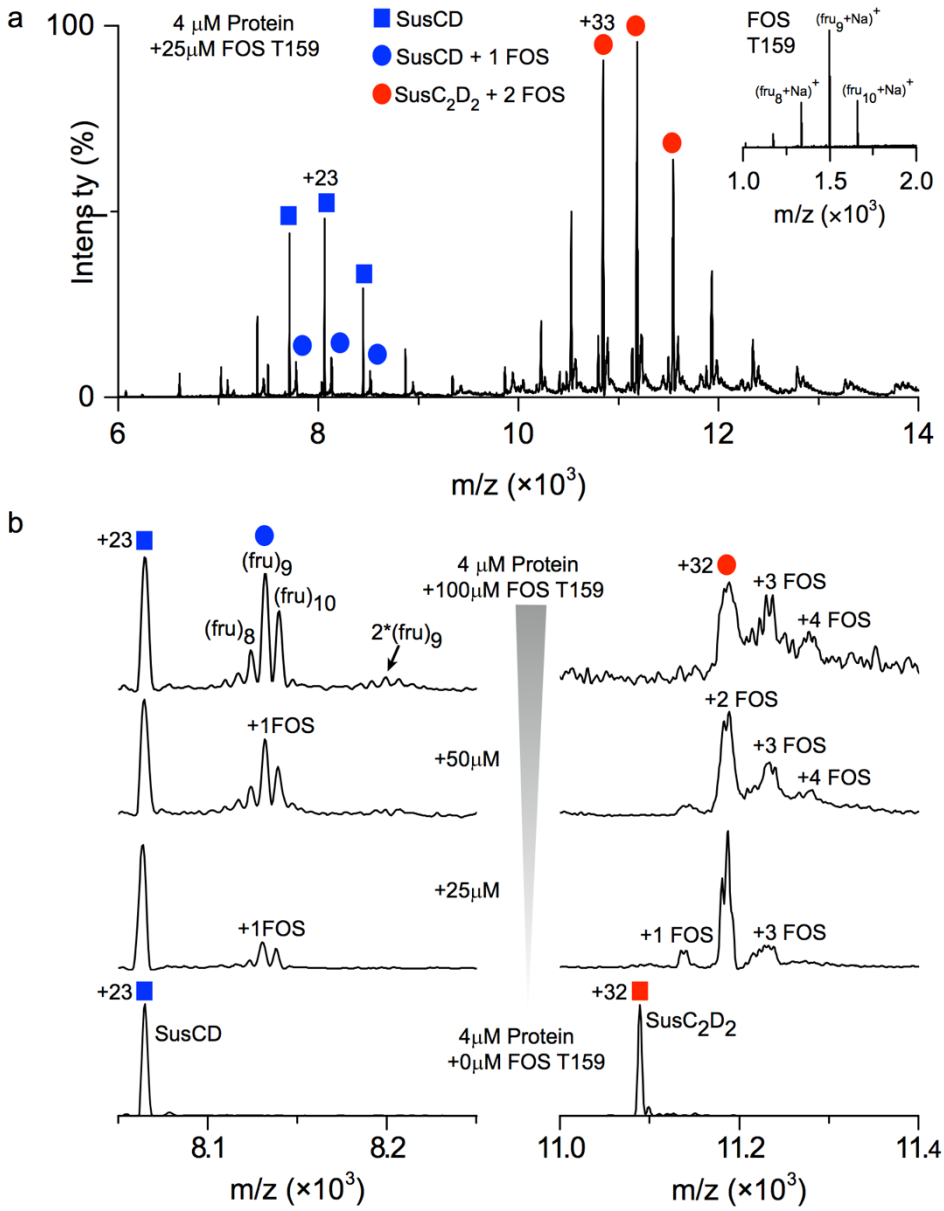

Figure 7 BT1762-63 binding to medium-chain fructo-oligosaccharide (FOS). a Native mass spectrum of BT1762-63 bound to FOS T159 fraction (~DP9) revealed the presence of both bound and unbound SusCD monomers, but only FOS-bound $SusC_2D_2$ dimers. The insert shows the spectrum of the FOS T159 fraction. **b,** Spectrum of BT1762-63 incubated with different concentrations of FOS T159. Shown are individual charge states of the monomer (+23) and of the dimer (+32). Up to two and four FOS molecules (mainly 8-10 fructose units) were found to bind to SusCD and $SusC_2D_2$ respectively. Masses of all species observed are listed in Supplementary Table 6.

Discussion

The difficulty in ionising large oligosaccharides in the mass-spectrometer[31] complicates reliable experimental determination of an upper substrate size limit of SusCD-like systems. Even if large oligosaccharides could be detected via mass spectrometry, the inability to separate high DP FOS (>~DP12) by SEC precludes estimation of detection efficiency, and consequently quantitation. Despite this, some important insights can be obtained via simple structural considerations. The ligand binding cavity of Bt1762-63 has a volume of ~10,000 Å^3 as determined via CASTp[32], which is similar to that of the peptide importer RagAB[17]. While no information is available for FOS, data for sucrose[33] allow an estimation of ~500 Å^3 per fully hydrated molecule, with 5 waters per sucrose. For one bound water per sucrose, the molar volume is lowered by ~20%, leading to a value of 400 Å^3 per sucrose. Thus, depending on FOS hydration inside the Bt1762-63 cavity and assuming ideal packing, there would be space for 20-25 sucrose molecules, *i.e*. 40-50 sugar monomers. Given that it is unlikely that the twisted structures of levan FOS could pack very efficiently, we estimate that the maximum number of fructose units occupying the Bt1762-63 cavity at any one time could be no more than ~30-35, putting an upper limit on total bound FOS of ~5 kDa. Considering our ITC and native MS data, it is likely that this total mass would comprise several individual molecules, rather than one large molecule. As there are unlikely to be large structural differences among SusCD-like systems, we suggest ~5 kDa as a general total size limit for these transporters, which is consistent with recent data for the archetypal Sus[18].

Our structures that show FOS in the principal binding site at the Bt1762-63 interface raise an important question: how is ligand occupancy relayed to the plug domain, and how does this lead to increased accessibility of the Ton box? This key issue is likely unique to SusCD systems, in particular those SusCs without the long plug loop present in *e.g*. Bt2264[15] that is able to contact ligand in the principal binding site and relay binding site occupancy directly to the plug domain (Extended Data Fig. 11). In Bt1763, the smallest distance between the visible part of the substrate in the principal binding site and the plug is 15 Å, and so an optimal-sized FOS (in terms of binding affinity) of ~DP8-12 could not contact the plug directly. The presence of a second substrate molecule at the bottom of the binding cavity (Fig. 1b) might be a way to overcome this problem, implying a mechanism in which the binding cavity "bin" is filled first via two or more substrate binding-release cycles to provide plug contacts, that collectively increase accessibility of the Ton box and binding to the CTD of TonB.

The substrates for the transporter are generated by the combined action of the Bt1760 endo-levanase and the Bt1761 surface glycan binding protein (SGBP; Fig. 1a). It is not yet clear if, and how, these two proteins are spatially and temporally connected to the Bt1762-63

transporter. For the archetypal Sus, data suggest that all five OM components of the PUL (SusCDEFG) are arranged in one stable complex[16,34]. Other, more recent studies paint a much more dynamic picture, with the SGBPs (SusE and SusF) and the endo-amylase SusG transiently associating with the SusCD core complex[35,36]. Besides depending on the type of levan[23], the FOS sizes delivered to Bt1762-63 will depend critically on the binding kinetics of the Bt1760 levanase and on the proximity of Bt1761: a close association between the two would most likely favour production of uniformly-sized FOS of relatively small size which, as we have shown, are preferred substrates. Likewise, a close association between the enzyme and Bt1762-63 will enhance capture of the generated FOS by SusD and subsequent delivery to SusC. With regards to this last step, it is interesting to note that, in contrast to *in vitro* conditions, the substrate binding function by Bt1762 is not necessary *in vivo* (Fig. 4 and Fig. 6), *i.e*. when all other OM components are present (Bt1760/61/63), and similar data have recently been obtained for other Sus[18,37]. These observations would suggest that the SGBP can assume the substrate-delivery role of SusD and argues in favour of an intimate association of all OM components of a PUL. This then raises the question of why the SusD lid has evolved at all. The fact that the *presence* of SusD is important *in vivo* (Fig. 4) suggests that its function as a lid that can open and close is vital.

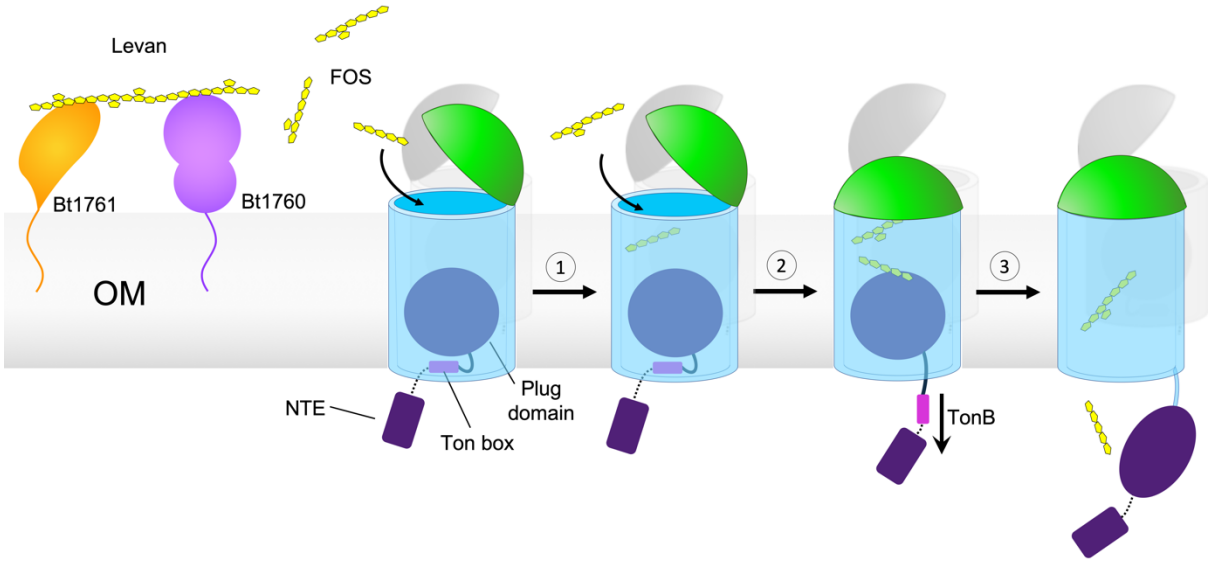

Figure 8 Model demonstrating the proposed mechanism of substrate capture and translocation by Bt1762-63. Levan polysaccharide is initially bound and hydrolysed at the cell surface by the SGBP Bt1761 and GH32 endo-levanase Bt1760, respectively, although the precise role of the SGBP is unclear. A lid-open state of the transporter permits binding of transport-competent FOS. **1.** Contributions from both SusC and SusD to FOS binding elicits closure of the lid. **2.** Multiple cycles of lid opening and closing occur until the SusC 'bin' is fully loaded with substrate, forming the transport-ready state of the complex. Substrate loading is communicated across the outer membrane by direct contact with the Bt1763 plug domain, inducing perturbation of the Ton box region on the periplasmic side of the plug, rendering it accessible to TonB. **3.** TonB-mediated disruption/extraction of the plug permits substrate translocation. Details of the 'reset' mechanism are unclear but re-insertion of the plug is likely a prerequisite to restoring the open state of the transporter.

Our structural data provide important clues about the function of SusD and about glycan import in general (Fig. 8). The basis for these clues is the unprecedented observation that closed, but empty transporters lack the entire plug domain. This spontaneous expulsion of the plug is likely to be non-physiological and caused by a loss of lateral membrane pressure due to detergent solubilisation. Nevertheless, it does suggest that Bt1762 lid closure causes conformational changes within the Bt1763 barrel that decrease the "affinity" of the plug for the barrel. This may facilitate the removal of the entire plug domain from the barrel by TonB action, as opposed to local unfolding and formation of a relatively narrow channel as has been proposed for non-Sus TBDTs[38-40]. To prevent plug removal in the absence of substrate resulting in futile transport cycles, we postulate that only the direct contact of substrate with the plug (as observed in the co-crystal structure with FOS2 and the in BT2263/64-peptide complex[15]) leads to increased accessibility of the TonB box and interaction with TonB (Fig. 8). In our model, the impermeability of the OM, which otherwise would be compromised due to the formation of a very large channel of ~20-25 Å diameter, would be preserved by the seal provided by the closed SusD lid. Upon reinsertion of the plug, the transporter would revert back to its open state (Fig. 8). The most important function of SusD proteins during glycan import may therefore be to provide a seal to preserve the OM permeability barrier.

Author contributions

D.A.G made *B. theta* mutants, expressed and purified proteins, performed growth assays, generated FOS substrates and carried out ITC. J.B.R.W. and S.L.E. determined cryo-EM structures, supervised by N.A.R. A.O.O. carried out native mass spectrometry, supervised by C.V.R. P.R., A.M. and M.Z. determined the NTE structure, supervised by S.H. A.J.G. contributed to all early stages of the project. C.M. made *B. theta* mutants. A.B. collected X-ray crystallography data. A.C. helped analyse crystallography data. B.v.d.B. crystallised proteins and determined crystal structures. The manuscript was written by B.v.d.B, D.N.B and D.A.G, with input from J.B.R.W., N.R., C.V.R. and S.H.

Acknowledgements

D.A.G., D.N.B and B.v.d.B. were supported by a Biotechnology and Biological Sciences Research Council grant (BB/P003192/1). J.B.R.W. was supported by a Wellcome Trust four-year PhD studentship (215064/Z/18/Z). All electron microscopy was performed at the Astbury Biostructure Laboratory which was funded by the University of Leeds and the Wellcome Trust (108466/Z/15/Z). We thank R. Thompson, E. Hesketh and D. Maskell for electron microscopy support.). S.H. was supported by the Swiss National Science Foundation (grant 167125).

A.O.O. and C.V.R. are supported by a Medical Research Council grant (MR/N020413/1). We would further like to thank personnel of the Diamond Light Source beamlines i03, i04 and i04-1 for beam time (Block Allocation Group numbers mx-13587 and mx-18598) and assistance with data collection.

Ethics declarations

Competing interests

The authors declare no competing interests.

Materials and methods

Maintenance and growth of bacterial strains. Brain heart infusion (BHI) media and agar, supplemented with 1 µg/ml haematin, or where needed with gentamycin (0.2 µg/ml) and erythromycin (0.025 µg/ml), was used for routine selection and maintenance of *Bacteroides thetaiotaomicron* VPI-5482 (*B. theta*) wild type and mutant strains. Luria-Bertani (LB) media and agar, supplemented where needed with kanamycin (50 µg/ml) or ampicillin (100 µg/ml), was used for routine selection and maintenance of *E. coli* strains. Cultures for *B. theta* protein purification were grown in a Chemically Defined Minimal Media (MM), NH_4SO_4 1 mg/ml, Na_2CO_3 1 mg/ml, cysteine 0.5 mg/ml, KPO_4 100 mM, vitamin K 1 µg/ml, $FeSO_4$ 4 µg/ml, vitamin B12 5 ng/ml, mineral salts 50 µl/ml (NaCl 0.9 mg/ml, $CaCl_2$ 26.5 µg/ml, $MgCl_2$ 20 µg/ml, $MnCl_2$ 10 µg/ml and $CoCl_2$ 10 µg/ml) and hematin 1 µg/ml. These cultures were supplemented with D-fructose (0.5 % final) as the sole carbon source (MM-Frc). Frc activates the levan PUL, but does not require Bt1762-63 for import. All *B. theta* growths were carried out under anaerobic conditions at 37 °C (A35 Workstation, Don Whitely Scientific).

Gene cloning and construction of *B. theta* mutants. The desired genetic sequence containing mutations or deletions was amplified through two rounds of PCR, including ~1000 bp upstream and downstream of the deletion or mutation. The primers used to generate these PCR products are listed in Supplementary Table 7. Briefly, for the TonB deletion strain primers TB-fwd, TBD-rev, and TBD-fwd, TB-rev respectively were used to make 1000 bp fragments which underwent an additional PCR to construct the 2000 bp sequence. The same process was used to make mutants TonB mutation (TB-fwd, TBM-rev and TBM-fwd, TB-rev), ∆hinge1 (Hinge-fwd, H1D-rev and HD1-fwd, Hinge-rev), ∆hinge2 (Hinge-fwd, HD2-rev and HD2-fwd, hingerev), W85A (Trp1-fwd, Trp2-rev and Trp3-fwd, Trp4-rev), and ∆NTE (NTE1-fwd, NTE2-rev and NTE3-fwd, NTE4-rev).The final PCR product was ligated into pExchange-tdk vector[30].

578 *B. theta* genetic deletions and mutations were created by allelic exchange using the
579 pExchange-tdk vector. Briefly, the constructed pExchange-tdk plasmids, containing the
580 mutations/deletions plus ~1000 bp flanking up- and downstream, were transformed into S17
581 λ pir *E. coli* cells, in order to achieve conjugation with the *B. theta* recipient strain[30]. The
582 conjugation plates were scraped to generate a culture containing *B. theta* and *E. coli*. *B. theta*
583 cells undergoing a single recombination event were selected for by plating on BHI-hematin
584 agar plates containing gentamycin (200 µg/ml) and erythromycin (25 µg/ml), 8 – 12 colonies
585 were restreaked on fresh BHI-hematin-gent-ery plates. Single colonies were cultured in BHI-
586 hematin and pooled. To select for the second recombination event pooled cultures were plated
587 on BHI-hematin agar plates containing FUdR (200 µg/ml). FUdR resistant colonies, 8 – 12,
588 were restreaked on fresh BHI-hematin-FUdR. From these, single colonies were cultured in
589 BHI and genomic DNA was extracted and screened for the correct mutations using diagnostic
590 PCR and sequencing. DNA sequencing was performed by MWG-Eurofins.
591
592 **Expression and purification of native proteins.** The required strain of *B. theta* was
593 inoculated from -80 °C stocks into 5 ml brain heart infusion (BHI) media supplemented with
594 hematin, 1 µg/ml. These were cultured overnight and used to inoculate 500 ml Duran bottles
595 containing 500 ml minimal media (MM) supplemented with hematin and either fructose or
596 levan (0.5 %) as the sole carbon source and grown for 18-20 h. The following morning cells
597 were harvested by centrifugation at 11305 x g for 25 min, and the pellets resuspended in TSB
598 and stored at -20 °C.
599
600 One protease inhibitor tablet and DNase were added to the completely thawed cell samples
601 before being lysed using a cell disrupter (Constant Systems) at 23 kilopounds per square inch
602 (KPSI). The lysed cells were centrifuged for 1 h at 42,000 rpm (204,526 x g). The resultant
603 pellet, containing total membranes, was homogenised in 0.5 % Sodium Lauroyl Sarcosine
604 (sarkosyl), 20 mM Hepes. The sample was allowed to stir at room temperature for 20 min
605 before undergoing centrifugation for 30 min at 204,526 x g. The supernatant was discarded
606 and the pellet was homogenised into a Lauryldimethylamine-N-oxide (LDAO), 10 mM Hepes,
607 100 mM NaCl pH 7.5 buffer, and incubated at 4 °C stirring for 1 h. The sample underwent
608 centrifugation at 204,526 x g for 30 min and the supernatant, containing the outer membrane
609 proteins (OMP), was kept. The total protein concentration was determined using a BCA
610 Protein Assay Kit (Thermo Fisher Scientific).
611
612 Bt1762-63 was purified using an Immobilised Metal Affinity Chromatography (IMAC) Ni
613 column using the C-terminal Hisx6-tag on SusD (Bt1762)[15]. The OM fraction of MM-Frc grown
614 *B. theta* cells was passed through an IMAC Ni column equilibrated with 0.2 % LDAO TSB. The

column was washed with TSB 0.2 % LDAO 25 mM imidazole, and bound proteins were eluted with TSB 0.2 % LDAO 200 mM imidazole. IMAC samples were visualised by SDS-PAGE. Samples from IMAC were further purified by size exclusion chromatography (SEC). An ÄKTA pure system (GE healthcare) was used in conjunction with a HiLoad 16/60 Superdex 200pg 120 ml gel filtration column (GE Healthcare). The column was equilibrated with the required buffer containing the necessary detergent. The IMAC samples were concentrated to 3-5 ml and loaded onto the column. Proteins were eluted at 1.2 ml/min and 3 ml fractions were collected. Peak samples were also visualised by SDS-PAGE. Buffer exchange was used to transfer the protein into the final desired buffer.

Crystallisation and structure determination. For crystallisation, the final SEC purification step was carried out in 10 mM Hepes/100 mM NaCl/0.4% C_8E_4 pH 7.5. Fractions were pooled and concentrated to ~15 mg/ml, aliquoted and flash-frozen in liquid nitrogen.Sitting drop crystallisation trials were set up using a Mosquito crystallisation robot (TTP Labtech) with commercial screens (Molecular Dimensions MemGold 1 and 2 and Morpheus). Optimisation was performed with larger-volume hanging drops set up manually. Well-diffracting crystals for apo-BT1762-63 were obtained from MemGold 2, condition 18 (0.15 M sodium formate, 0.1 M Hepes pH 7-7.5, 16-20% % w/v PEG 3350), and were cryoprotected by the addition of ~20-25% PEG400 for ~5-10 s before being flash-frozen in liquid nitrogen. Diffraction data were collected at 100 K at the Diamond Light Source (Didcot, UK) on beamline i03. Data were processed via Xia2[41] or Dials[42]. The structure was solved by molecular replacement with Phaser[43], using data to 2.62 Å resolution. The previous structure of Bt1762-63 (PDB 5T3R)[15] was used as search model. The model was built iteratively by manually building in COOT[44], and was refined with Phenix[45] using TLS refinement with 1 group per chain. Structure validation was carried out with MolProbity[46]. The model has 94.0% of residues in the favoured regions of the Ramachandran plot, and 0.7% outliers. The structures of Bt1762-63 with bound FOS were obtained by incubation of purified complex with 2.5 mM of the appropriate FOS at room temperature, followed by setting up crystal plates as above. Models of the bound FOS were generated with JLigand within CCP4[47] and fit manually into the density in COOT. Refinement was carried out within Phenix[45] as above. The FOS-bound structures have 90.5/1.0% (PDB ID 6Z9A) and 94.1/0.6% (PDB ID 6ZAZ) of residues in the favourable/disallowed regions of the Ramachandran plot.

CryoEM sample preparation and data collection. A sample of purified BT1762-BT1763 solubilised in an LDAO-containing buffer (10 mM HEPES pH 7.5, 100 mM NaCl, 0.05% LDAO) was prepared at 0.02 mg/ml. Lacy carbon 300-mesh copper grids coated with a <3 nm continuous carbon film (Agar Scientific) were glow-discharged in air (10 mA, 30 seconds). A

sample volume of 3.5 μL was applied to each grid. Blotting and plunge freezing were carried out using a Vitrobot Mark IV (FEI) with the chamber conditions set at a temperature of 4 °C and 100% relative humidity. A blot force of 6 and a blot time of 6 s were used prior to vitrification in liquid nitrogen-cooled liquid ethane. Micrograph movies were collected on a Titan Krios microscope (Thermo Fisher) operating at 300 kV with a GIF energy filter (Gatan) and K2 summit direct electron detector (Gatan) operating in counting mode. Data acquisition parameters can be found in Supplementary Table 2.

Image processing. Following cryo-EM data collection, the RELION (v3.0) pipeline was used for image processing[48,49]. Drift correction was performed using MotionCor2[50] and contrast transfer functions were estimated using gCTF[51]. Approximately 3000 particles were manually picked, extracted and classified in 2D to provide templates for automated particle picking within RELION. Particles were extracted in 264 x 264 pixel boxes binned to 132 x 132 for initial rounds of 2D and 3D classification. The 3D starting model was generated *de novo* from the EM data by stochastic gradient descent in RELION. The particle stack was cleaned using both 2D and 3D classification and the remaining particles were un-binned. Further rounds of 3D classification revealed significant conformational heterogeneity in the data. Three predominant conformational states of the levan transport system were identified: open-open (OO), open-closed (OC) and closed-closed (CC). Additional classes, whilst less populated, revealed states with the BT1762 'lid' open to differing extents (Extended Data Fig. 2). Invariably, the closed position of BT1762 was associated with an absence of density for the plug domain of BT1763. Global 3D classification was unable to distinguish 'true' closed conformations from those where BT1762 occupied a marginally open state. As a result, a masked 3D classification approach was employed to achieve homogeneous particle stacks. The masked classification was performed without image alignment and, since the region of interest is relatively small, the regularization parameter, T, was set to 20. Intermediate results and further details are provided in Extended Data Fig. 12. Clean particle stacks for the three principle conformational states were subject to multiple rounds of CTF refinement and Bayesian polishing[49]. C2 symmetry was applied to both the OO and CC reconstructions. Post-processing was performed using soft masks and yielded reconstructions for the OO, OC, and CC states of 3.9, 4.7 and 4.2 Å respectively, as estimated by gold standard Fourier Shell correlations using the 0.143 criterion.

Model building into cryoEM maps. Comparing the maps to the crystal structure of BT1762-BT1763 revealed that their handedness was incorrect. Maps were therefore Z-flipped in UCSF Chimera[52,53]. The reconstruction of the OO state was of sufficient resolution for model building

and refinement. Bt1762 and Bt1763 subunits from the crystal structure were independently rigid-body fit to the local resolution filtered map and later subjected to several iterations of manual refinement in COOT[44] and 'real space refinement' in Phenix[45]. The asymmetric unit was symmetrised in Chimera after each iteration. Molprobity[46] was used for model validation. The reconstructions of the OC and CC states were of insufficient resolution to permit model building and refinement owing to low particle numbers and a poor distribution of viewing angles respectively. Instead, the crystal structure of Bt1762-Bt1763 was rigid-body fit to the CC state. The ligand was removed from the model and an inspection in COOT showed that no density extended past Lys213 in the direction of the N-terminus. All residues N-terminal of Lys213 were therefore removed from the model before rigid-body fitting. The open state from the OO EM structure and the closed state from the crystal structure (modified as described above) were rigid-body fit to their corresponding densities in the OC state. Rigid-body fitting was performed in Phenix.

Expression and purification of recombinant enzymes. Recombinant pET vectors (containing either BT1760 levanase[19] or Bacillus levansucrase genes) were transformed into *E. coli* Tuner cells (Novagen) with appropriate antibiotic selection. One-litre cultures in 2 l flasks were grown to mid-exponential phase at 37 °C in an orbital shaker at 150 rpm, then cooled to 16 °C and induced with 1 mM isopropyl β-d-thiogalactopyranoside (IPTG). These cultures were then incubated for a further 16 h at 16 °C and 150 rpm. Cells were harvested by centrifugation, lysed by sonication and the recombinant His-tagged protein purified from cell-free extracts using IMAC (Talon resin, Clontech) as previously described[19].

Expression and purification of the NTE

A pBAD24 vector carrying the Bt1763 NTE domain with an N-terminal His6 tag (Extended Data Fig. 5d), arabinose promoter and ampicillin resistance was transformed into *E. coli* BL21 (DE3) Lemo cells for expression. Cells were grown in M9 minimal media containing $^{15}NH_4Cl$ and ^{13}C-glucose as the source of nitrogen and carbon for protein labelling. Cells were grown at 37 °C for protein expression. At the optical density of 0.7, cells were induced with 0.1% L-arabinose for 3 h. Cells were then harvested by centrifugation at 5000 rpm for 15 min, and resuspended in buffer A (20 mM HEPES, 150 mM NaCl, pH 7.5) containing lysozyme, DNAse and protease inhibitor. Homogenized cells were lysed by three rounds passing through a high-pressure microfluidizer. The lysate was centrifuged at 12000 rpm for 30 min at 4 °C. The supernatant was loaded on a Ni-NTA His-trap column (GE healthcare), washed with 10 column volume of buffer A containing 30 mM imidazole. ^{15}N, ^{13}C-labelled NTE was then eluted with three column volume buffer A containing 200 mM imidazole. Eluted proteins were loaded on Superdex S75 gel filtration column pre-equilibrated with NMR buffer containing 20 mM

sodium phosphate, 150 mM NaCl and pH 7.5. Fractions corresponding to pure NTE were concentrated though a 5 kDa cut-off filter for NMR spectra acquisitions. The NMR sample contained 250 µl 1 mM *U*-^{15}N, ^{13}C-NTE in NMR buffer containing 5% $^{2}H_2O$.

NMR experiments

All NMR spectra were recorded at 20 °C on a Bruker Avance-700 spectrometer equipped with a cryogenic triple-resonance probe and using Topspin 3.6. The proton chemical shifts were referenced to water, and ^{15}N and ^{13}C were indirectly referenced. 2D [^{15}N,^{1}H]-HSQC, 2D [^{13}C,^{1}H]-HSQC, 3D HNCACB, 3D CBCA(CO)NH, 3D ^{15}N-resolved-[^{1}H,^{1}H]-NOESY and 3D ^{13}C-resolved-[^{1}H,^{1}H]-NOESY were acquired for backbone resonance assignment and structure determination of NTE. The NMR experiments and respective acquisition parameters are mentioned in Supplementary Tables 3 and 4.

Calculation of the three-dimensional solution structure of NTE

The side-chain chemical shifts were assigned automatically using FLYA algorithm with fixed backbone chemical shifts[54]. ^{1}H-^{15}N and ^{1}H-^{13}C NOE cross peaks were automatically assigned followed by manual corrections resulting in total 964 unambiguously assigned peaks. Backbone torsion angle constraints were derived from chemical shift values using the program TALOS[55]. Using these values, solution structure calculation was performed using the program CYANA[56].

Fructo-oligosaccharide (FOS) production and purification. Two grams of *Erwinia herbicola* levan (Sigma-Aldrich) were dissolved in 100 ml PBS (Oxoid) by autoclaving. The solution was cooled to room temperature before the addition of 100 nM Bt1760 GH32 endo-acting levanase was added incubated at 37 °C for 30 min to partially digest the levan to a mixture of different size β2,6 FOS. The enzyme was then heat inactivated by boiling for 20 min. The resultant sample was freeze dried using a Christ Alpha 1-2 Freeze Drier at -45 °C. The freeze dried FOS mixture was resuspended in 5 ml 50 mM acetic acid and loaded onto a column (two 2.5 x 80 cm Glass Econo-Columns, connected in series with a flow adaptor; Bio-Rad) packed with P2 Bio-gel size exclusion resin (Bio-Rad) and pre-equilibrated, with 50 mM acetic acid. The column was run at 0.25 ml/min using a peristaltic pump (LKB Bromma 2132 microperpex) and 2 ml fractions were collected continuously for 48 h using a Bio-Rad model 2110 fraction collector. Fractions were initially analysed by TLC and any that contained sugar were freeze-dried to remove acetic acid.

Thin Layer Chromatography (TLC). TLC plates (Silica gel 60, Sigma-Aldrich) were cut to the required size, and samples were spotted 1 cm from the bottom of the plate. These were

dried using a hair dryer and placed into a tank containing 1 cm of running buffer (1-butanol, acetic acid and water at 2:1:1). Once the running buffer migrated to 1 cm from the top of the plate, the plate was then dried and run again. To visualise sugars the plates were completely dried and submerged in developer solution (sulphuric acid, ethanol and water at 3:70:20 with 1% orcinol) for 5–10 s Finally the plate was dried using a hair dryer and incubated at 65 °C in a drying oven until developed (usually ~1 h).

Isothermal Titration Calorimetry (ITC). SEC purified Bt1762-63-SusCD complex was dialysed overnight at 4 °C with 50kD cut off dialysis tubing into 100 mM Hepes, pH 7.5 containing 0.05 % LDAO. The dialysis buffer was used to resuspend SEC purified FOS fractions produced by partial digestion of *Erwinina* levan. ITC was performed using a MicroCal PEAQ-ITC machine (Malvern). Briefly, pure SusCD complex (25-50 µM; concentration determined by A_{280}) in a 200 µl reaction well was injected 20 times with 2 µl aliquots of FOS (0.25 – 4.5 mg/ml) at 25 °C. Integrated heats were fit to a one set of sites model using the Microcal PEAQ-ITC software (Malvern) to obtain K_d, ΔH and N. For most titrations, the molar concentration of ligand used for the fits was based on the DP of the major FOS species present as determined by the TLC and MS analysis. For fractions 118 to 115 the concentration of ligand was varied such that N = 1.

Native mass spectrometry. Proteins were buffer exchanged into 0.5% C8E4, 0.2 M ammonium acetate, pH6.9 using a micro biospin 6 column (Bio-Rad). The FOS sample was diluted to ~500 µM with 0.5 M ammonium acetate, pH6.9. Mass spectra were acquired on a Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) optimized for transmission and detection of high molecular weight protein complexes. About 3 µl of aliquot of the sample was transferred into gold-coated borosilicate capillary (Harvard Apparatus) prepared in-house and capillary and mounted on the nano ESI source. The instrument settings were 1.2 kV capillary voltage, S-lens RF 200%, argon UHV pressure 3.1×10^{-10} mbar, capillary temperature 100°C. Voltages of the ion transfer optics – injection flatapole, inter-flatapole lens, bent flatapole, and transfer multipole were set to 5 V, 3 V, 2 V, and 30 V respectively. The noise level was set at 3. Protein ions were activated with -120 V with the in-source trapping mode and a collisional activation voltage 300 V. Data were visualized and exported for processing using the Qual browser of Xcalibur 4.2 (Thermo Scientific).

Growth curves. Growth curves were performed in an Epoch microplate spectrometer (Biotek Instruments Ltd.) with 96-well Costar culture plate (Sigma-Aldrich) inside an anaerobic cabinet at 37 °C (Don Whitely Scientific, A35 workstation). Media was inoculated 1:10 with bacterial

cultures previously grown overnight in BHI. Final culture volumes of 200 µl were used, and each condition was performed in triplicate. Optical density at 600 nm was measured in each well at 30 minute intervals.

Sources of levan used. In most cases the levan used was from *Erwinia herbicola* (Sigma). For analysis of growth of the Δ1760 strain, bacterial levans from *Bacillus* sp. (Montana Polysaccharides) and *Zymomonas mobilis*[57] (a kind gift from Prof. Dr. Yekta Göksunger, Ege University, Izmir, Turkey) were also used as well as levan from Timothy grass (Megazyme) and *in vitro* synthesised levan. *In vitro* synthesised levan was made using *Bacillus subtilis* levansucrase incubated with 20 % sucrose in PBS for 24 h at 37 °C. Proteinase K (100 µg/ml) was then added to remove the protein, before being precipitated out using 0.5% TCA final. The levan was then extensively dialysed against water and freeze dried.

Data availability

The data supporting the findings of this study are available from the corresponding authors upon reasonable request. Coordinates and structure factors that support the findings of this study have been deposited in the Protein Data Bank with accession codes 6Z8I (Bt1762-63), 6Z9A (Bt1762-63 with longer FOS) and 6ZAZ (Bt1762-63 with shorter FOS). EM structure coordinates have been deposited in the Electron Microscopy Data Bank with accession codes xxx. The raw cryoEM movie mode micrographs for the primary dataset containing the CC, OC and OO structures will be deposited in the EMPIAR database.

References

1. Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-14 (2012).
2. Marchesi, J.R. et al. The gut microbiota and host health: a new clinical frontier. *Gut* **65**, 330-339 (2016).
3. Sommer, F. & Backhed, F. The gut microbiota--masters of host development and physiology. *Nat Rev Microbiol* **11**, 227-38 (2013).
4. Cantarel, B.L., Lombard, V. & Henrissat, B. Complex carbohydrate utilization by the healthy human microbiome. *PLoS One* **7**, e28742 (2012).
5. Hamaker, B.R. & Tuncil, Y.E. A Perspective on the Complexity of Dietary Fiber Structures and Their Potential Effect on the Gut Microbiota. *Journal of Molecular Biology* **426**, 3838-3850 (2014).
6. Koropatkin, N.M., Cameron, E.A. & Martens, E.C. How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology* **10**, 323-335 (2012).
7. Cockburn, D.W. & Koropatkin, N.M. Polysaccharide Degradation by the Intestinal Microbiota and Its Influence on Human Health and Disease. *J Mol Biol* **428**, 3230-3252 (2016).
8. Makki, K., Deehan, E.C., Walter, J. & Backhed, F. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe* **23**, 705-715 (2018).

9. Tan, J. et al. The role of short-chain fatty acids in health and disease. *Adv Immunol* **121**, 91-119 (2014).
10. Wexler, A.G. & Goodman, A.L. An insider's perspective: Bacteroides as a window into the microbiome. *Nature Microbiology* **2**(2017).
11. Martens, E.C. et al. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).
12. Ndeh, D. & Gilbert, H.J. Biochemistry of complex glycan depolymerisation by the human gut microbiota. *FEMS Microbiol Rev* **42**, 146-164 (2018).
13. Martens, E.C., Koropatkin, N.M., Smith, T.J. & Gordon, J.I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284**, 24673-7 (2009).
14. Martens, E.C., Chiang, H.C. & Gordon, J.I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447-57 (2008).
15. Glenwright, A.J. et al. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature* **541**, 407-411 (2017).
16. Shipman, J.A., Berleman, J.E. & Salyers, A.A. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *J Bacteriol* **182**, 5365-72 (2000).
17. Madej, M. et al. Structural and functional insights into oligopeptide acquisition by the RagAB transporter from Porphyromonas gingivalis. *Nature Microbiology* (2020).
18. Foley, M.H., Martens, E.C. & Koropatkin, N.M. SusE facilitates starch uptake independent of starch binding in B. thetaiotaomicron. *Mol Microbiol* **108**, 551-566 (2018).
19. Sonnenburg, E.D. et al. Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations. *Cell* **141**, 1241-U256 (2010).
20. Oner, E.T., Hernandez, L. & Combie, J. Review of Levan polysaccharide: From a century of past experiences to future prospects. *Biotechnol Adv* **34**, 827-844 (2016).
21. Jalan, N. et al. Studies on production of fructo-oligosaccharides (FOS) by gamma radiation processing of microbial levan. *Carbohydr Polym* **96**, 365-70 (2013).
22. Bolam, D.N. & Sonnenburg, J.L. Mechanistic insight into polysaccharide use within the intestinal microbiota. *Gut Microbes* **2**, 86-90 (2011).
23. Mardo, K. et al. A Highly Active Endo-Levanase BT1760 of a Dominant Mammalian Gut Commensal Bacteroides thetaiotaomicron Cleaves Not Only Various Bacterial Levans, but Also Levan of Timothy Grass. *PLoS One* **12**, e0169989 (2017).
24. Noinaj, N., Guillier, M., Barnard, T.J. & Buchanan, S.K. TonB-dependent transporters: regulation, structure, and function. *Annu Rev Microbiol* **64**, 43-60 (2010).
25. Holm, L. Benchmarking fold detection by DaliLite v.5. *Bioinformatics* **35**, 5326-5327 (2019).
26. Martens, E.C., Roth, R., Heuser, J.E. & Gordon, J.I. Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J Biol Chem* **284**, 18445-57 (2009).
27. Josts, I., Veith, K. & Tidow, H. Ternary structure of the outer membrane transporter FoxA with resolved signalling domain provides insights into TonB-mediated siderophore uptake. *Elife* **8**(2019).
28. Rogowski, A. et al. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* **6**, 7481 (2015).
29. Shipman, J.A., Cho, K.H., Siegel, H.A. & Salyers, A.A. Physiological Characterization of SusG, an Outer Membrane Protein Essential for Starch Utilization byBacteroides thetaiotaomicron. *Journal of Bacteriology* **181**, 7206-7211 (1999).
30. Koropatkin, N.M., Martens, E.C., Gordon, J.I. & Smith, T.J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105-1115 (2008).

31. Kailemia, M.J., Ruhaak, L.R., Lebrilla, C.B. & Amster, I.J. Oligosaccharide analysis by mass spectrometry: a review of recent developments. *Anal Chem* **86**, 196-212 (2014).
32. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res* **46**, W363-W367 (2018).
33. Gharsallaoui, A., Barbara, R., Génotelle, J. & Mathlouthi, M. Relationships between hydration number, water activity and density of aqueous sugar solutions. *Food Chemistry $V 106*, 1443-1453 (2008).
34. Cho, K.H. & Salyers, A.A. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **183**, 7224-30 (2001).
35. Karunatilaka, K.S., Cameron, E.A., Martens, E.C., Koropatkin, N.M. & Biteen, J.S. Superresolution imaging captures carbohydrate utilization dynamics in human gut symbionts. *mBio* **5**, e02172 (2014).
36. Tuson, H.H., Foley, M.H., Koropatkin, N.M. & Biteen, J.S. The Starch Utilization System Assembles around Stationary Starch-Binding Proteins. *Biophys J* **115**, 242-250 (2018).
37. Tauzin, A.S. et al. Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont. *mBio* **7**, e02134-15 (2016).
38. Fanucci, G.E., Cadieux, N., Kadner, R.J. & Cafiso, D.S. Competing ligands stabilize alternate conformations of the energy coupling motif of a TonB-dependent outer membrane transporter. *Proc Natl Acad Sci U S A* **100**, 11382-7 (2003).
39. Pawelek, P.D. et al. Structure of TonB in complex with FhuA, E. coli outer membrane receptor. *Science* **312**, 1399-402 (2006).
40. Shultis, D.D., Purdy, M.D., Banchs, C.N. & Wiener, M.C. Outer membrane active transport: structure of the BtuB:TonB complex. *Science* **312**, 1396-9 (2006).
41. Winter, G., Lobley, C.M. & Prince, S.M. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* **69**, 1260-73 (2013).
42. Winter, G. et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85-97 (2018).
43. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
45. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
46. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
47. Lebedev, A.A. et al. JLigand: a graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* **68**, 431-40 (2012).
48. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-30 (2012).
49. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).
50. Li, X., Zheng, S.Q., Egami, K., Agard, D.A. & Cheng, Y. Influence of electron dose rate on electron counting images recorded with the K2 camera. *J Struct Biol* **184**, 251-60 (2013).
51. Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12 (2016).
52. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
53. Goddard, T.D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018).
54. Schmidt, E. & Guntert, P. A new algorithm for reliable and general NMR resonance assignment. *J Am Chem Soc* **134**, 12817-29 (2012).

939 55. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein
940 backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-23 (2009).
941 56. Guntert, P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-78
942 (2004).
943 57. Silbir, S., Dagbagli, S., Yegin, S., Baysal, T. & Goksungur, Y. Levan production by Zymomonas
944 mobilis in batch and continuous fermentation systems. *Carbohydrate Polymers* **99**, 454-461
945 (2014).

946