Outer membrane utilisomes mediate oligosaccharide uptake in gut Bacteroidetes

Joshua B. R. White[1]*, Augustinas Silale[2]*, Matthew Feasey[2], Tiaan Heunis[2], Yiling Zhu[2], Hong Zheng[2], Akshada Gajbhiye[2], Susan Firbank[2], Arnaud Baslé[2], Matthias Trost[2], David N. Bolam[2], Neil A. Ranson[1] and Bert van den Berg[2]

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

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

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

* Authors contributed equally

Abstract

Bacteroidetes are abundant members of the human microbiota, with species occupying the distal gut capable of utilising a myriad of diet- and host-derived glycans. Transport of glycans across the outer membrane (OM) of these bacteria is mediated by SusCD protein complexes, comprising a membrane-embedded barrel and a lipoprotein lid, that are thought to operate via a 'pedal-bin' mechanism in which the lids open and close to facilitate substrate binding. However, additional cell surface-exposed lipoproteins, namely surface glycan binding proteins and glycoside hydrolases, play critical roles in the capture and processing of large glycan chains into transport-competent substrates. Despite constituting a crucial mechanism of nutrient acquisition by our colonic microbiota, the interactions between these components in the OM are poorly understood. Here we show that for the levan and dextran utilisation systems of *Bacteroides thetaiotaomicron,* the additional OM components assemble on the core SusCD transporter, forming stable glycan utilising machines which we term 'utilisomes'. Single particle electron cryogenic electron microscopy (cryo-EM) structures in the absence and presence of substrate reveal concerted conformational changes that rationalise the role of each component for efficient nutrient capture, as well as providing a direct demonstration of the pedal bin mechanism of substrate capture in the intact utilisome.

55 **Introduction**

56 The human large intestine is home to a diverse microbial community, known collectively as
57 the gut microbiome, which plays an essential role in human health[1]. Within the large
58 intestine, complex dietary glycans which are inaccessible to the enzymes of the human
59 digestive tract are the primary nutrient source, and the utilisation of these complex sugars by
60 gut microorganisms is essential for their survival[2,3]. Crucially, this nutrient supply is also
61 integral to mutualism between host and bacteria, via the generation of short-chain fatty acids
62 that are accessible to the host. The availability of such metabolites is associated with normal
63 gastrointestinal physiology and systemic health benefits[4,5]. The distal gut microbiome is
64 dominated by two bacterial phyla: the Gram-negative Bacteroidetes and the Gram-positive
65 Firmicutes[6]. The Bacteroidetes employ a common strategy for glycan utilisation. The
66 machinery required for the uptake, processing and metabolism of specific glycans is encoded
67 in co-regulated gene clusters known as polysaccharide utilisation loci (PULs)[7].

68

69 The OM of Gram-negative bacteria presents a formidable barrier to the uptake of large
70 nutrients[8]. Translocation of complex glycans across the Bacteroidetes OM is dependent on
71 the SusCD core components of a PUL, comprising a TonB-dependent active transporter (SusC)
72 and an associated lipoprotein (SusD)[7,9]. Previous structural studies have shown that these
73 SusCD complexes exist as $SusC_2D_2$ tetramers[10–12], creating a core transportation unit with a
74 twin barrel structure, where each barrel is associated with its own SusD component that caps
75 the extracellular face of the transporter barrel. Recent work has shown that these SusCD-like
76 transporters utilise a 'pedal bin' mechanism of nutrient uptake, where SusD is able to undergo
77 large, hinge-like movements that alternatively expose or occlude a binding site for the
78 nutrient within the lumen of the SusC barrel, facilitating substrate capture[12].

79

80 Although most glycan breakdown occurs within the bacterial cell, the initial binding and
81 processing of long glycan chains occurs extracellularly. To achieve this, an authentic PUL
82 minimally encodes, in addition to at least one SusCD pair, a surface-located, endo-acting
83 glycanase (most commonly glycoside hydrolases; GH) and one or more OM surface glycan
84 binding proteins (SGBPs)[7,9]. The model gut symbiont *B. thetaiotaomicron* (*B. theta*) has 88
85 predicted PULs comprising almost 20% of its genome, but the substrate specificity is known
86 for only ~20 of these[13]. The large number of PULs, each likely dedicated to the acquisition of

87 a specific glycan, demonstrates the importance of complex glycans to bacterial survival in the
88 distal gut.
89
90 One of the best-characterised PULs is that for the utilisation of levan, a plant and bacterial-
91 derived fructan polysaccharide, which consists of β2,6-linked fructose units with occasional
92 β2,1 branches[14,15]. We previously characterised the binding and uptake of β2,6 fructo-
93 oligosaccharides (FOS) by the core SusCD levan transporter[10,12]. However, the important steps
94 preceding levan capture and transport by SusCD remain unclear. Generation of transport-
95 competent FOS from levan is achieved by the OM lipoprotein Bt1760 (a GH32-family levan
96 endo-glycanase)[14,16], while a second OM lipoprotein, Bt1761 (the surface glycan binding
97 protein; SGBP) has a putative role in recruitment of levan to the cell surface[17]. These two
98 lipoproteins will henceforth be indicated as the levanase and the levan SGBP respectively, and
99 collectively as the "additional lipoproteins".
100
101 A key unresolved question in the field is whether and how the additional lipoproteins
102 associate with their cognate SusCD transporters to drive oligosaccharide uptake. Some
103 information is available for the archetypal starch utilisation system, which encodes two SGBPs
104 (SusE and SusF) in addition to the SusCD transporter and a surface amylase (SusG)[18,19].
105 Seminal studies offered evidence in support of SusCD-SGBP complex formation[19]. Since then,
106 a more dynamic picture of the starch utilisation system, involving all 5 OM components, has
107 emerged. Those studies indicate transient, substrate-induced complex formation[20],and
108 describe a role for the SGBPs SusE and SusF as immobile starch binding centres around which
109 the SusCD transporter and the SusG amylase can assemble[21]. In contrast, the Bt2263^{SusC}-
110 Bt2264^{SusD} complex from an uncharacterised PUL in *B. theta* form a stable complex with two
111 lipoproteins of unknown function (Bt2261 and Bt2262)[10].Thus, broadly speaking, two
112 opposing models exist: (1) The dynamic assembly model where free lipoproteins and the SusC
113 transporter transiently assemble, possibly in response to the presence of substrate, and (2)
114 The stable complex model where the additional lipoprotein components assemble on the
115 core SusCD transporter.
116
117 Here we demonstrate, using proteomics and single particle cryo-EM, that all four OM
118 components of the levan PUL from *B. theta* exist in a stable, octameric complex that we term

119 a ‘utilisome’ in keeping with the "PUL" acronym. We also show that the four OM components
120 of the *B. theta* dextran (a bacterial α-glucan) PUL adopt a similar architecture, suggesting that
121 utilisomes are a generic feature of glycan utilisation in gut *Bacteroides*. Cryo-EM structures of
122 both levan and dextran utilisomes in the absence of their cognate substrates suggest that the
123 auxiliary components function in concert with the core SusCD transporters. Upon addition of
124 FOS, the levan utilisome undergoes large, concerted conformational changes, revealing the
125 substrate-bound states of the Bt1760 levanase and the Bt1761 SGBP, and demonstrating that
126 the pedal-bin mechanism of nutrient import operates in the presence of all OM components
127 of the PUL. Collectively, we show that utilisomes constitute multi-component,
128 macromolecular machines on the cell surface, the architecture of which is consistent with
129 efficient capture and processing of polysaccharides and import of the generated
130 oligosaccharides.
131

132 ## Results

133 **Cryo-EM data of the SusCD levan transporter indicate substrate-independent association of**
134 **auxiliary lipoprotein components**

135 Previously, we used single particle cryo-EM to capture the $SusC_2D_2$ levan transporter
136 produced by fructose-grown *B. theta* in several conformational states (fructose induces
137 expression of the levan PUL)[12]. During 3D classification of these data, compositional
138 heterogeneity of the complex was also identified, with one class (~10% of the data) possessing
139 density in addition to that expected for the SusCD components (**Fig. 1**). SDS-PAGE and
140 downstream mass spectrometry revealed co-purification of sub-stoichiometric amounts of
141 the levanase (Bt1760) and the levan SGBP (Bt1761). To confirm whether the additional
142 density corresponded to one or both of these components, we docked the crystal structure
143 of *E. coli*-expressed inactive Bt1760 (PDB ID 7ZNR; Supplementary Table 1 and Methods) into
144 the map. The levanase was found to occupy only part of the unassigned density, leading us to
145 tentatively assign the remaining density to Bt1761, for which there was no available structure.
146

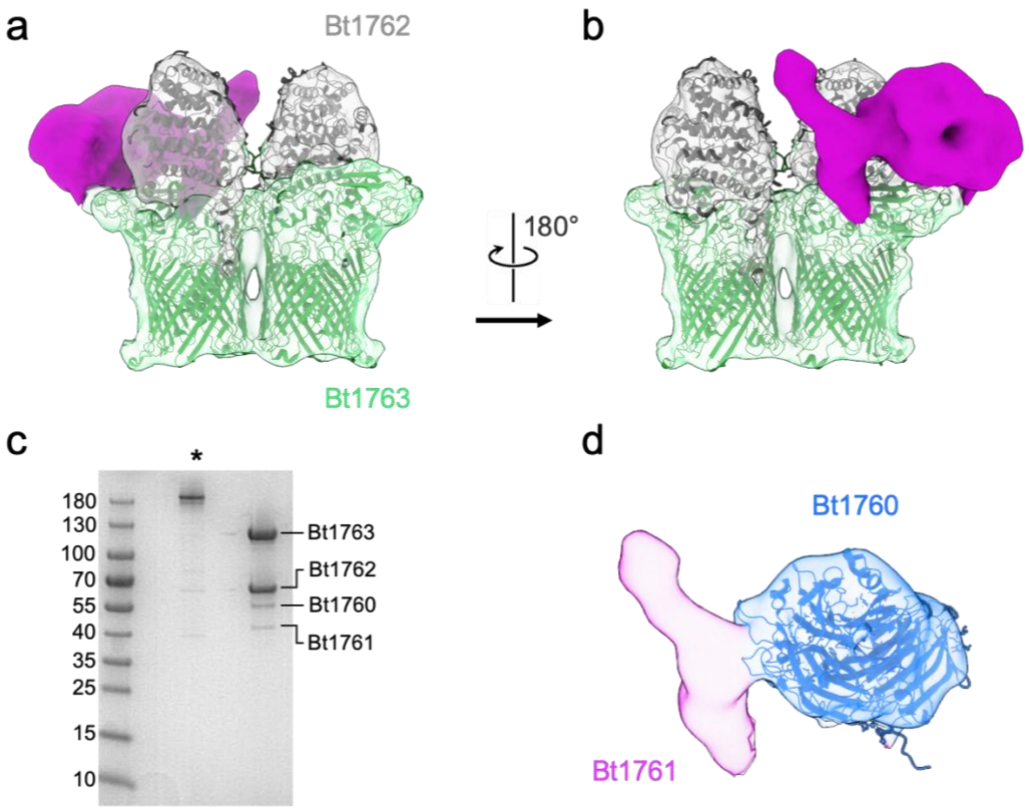


Figure 1: Additional density identified in 3D classification of the levan $SusC_2D_2$ transport complex. a, A class average obtained during 3D classification of the levan $SusC_2D_2$ transport complex. The SusC and SusD components (green and grey respectively) are docked into the corresponding density. A large region of density remains unassigned (magenta). **b**, as for **a** rotated 180 degrees. **c**, SDS-PAGE of the sample that was applied to cryo-EM grids before (asterisk) and after boiling. The boiled sample shows two clear bands in addition to those for Bt1762 (SusD) and Bt1763 (SusC), which were identified as Bt1760 levanase and Bt1761 LBP by mass spectrometry. **d**, Isolated view of the previously unassigned density with the crystal structure of Bt1760 (blue cartoon) fit as a rigid body. The structure fits only part of the unassigned density, now coloured blue. The remaining density was therefore attributed to Bt1761 and remains magenta.

The additional density we ascribed to the additional lipoproteins was associated with just one of the two available SusC barrels, and this seemed unlikely to represent a 'complete' complex given the dimeric nature of the $SusC_2D_2$ core complex. We therefore adapted the purification protocol to use milder detergents (a mixture of decyl-maltoside, DM and dodecyl-maltoside, DDM) rather than the harsher lauryldimethylamine-oxide (LDAO), to extract the complex from membranes. The remainder of the purification was carried out in DDM rather than in LDAO. In addition, we moved the His_6-tag from the C-terminus of SusD to the C-terminus of the levan SGBP Bt1761 (Methods). A qualitative assessment of newly purified sample by SDS-PAGE indicated the presence of all four components in approximately stoichiometric amounts (Bt1760-63), within a complex that ran as a single, monodisperse peak in analytical size

158 exclusion chromatography (**Fig. 2a,c**). Similar results were obtained for a dextran-grown *B.*
159 *theta* strain tagged on the SusD protein from the dextran PUL (Bt3089; **Fig. 2b,c**). In this case,
160 the purified four-component complex (4CC) consisted of Bt3087 (GH66 dextranase), Bt3088
161 (putative dextran SGBP), Bt3089 (SusD) and Bt3090 (SusC). Structures of the levan and
162 dextran PULs are shown in **ED Figure 1** alongside known and predicted structural information
163 for the OM components.

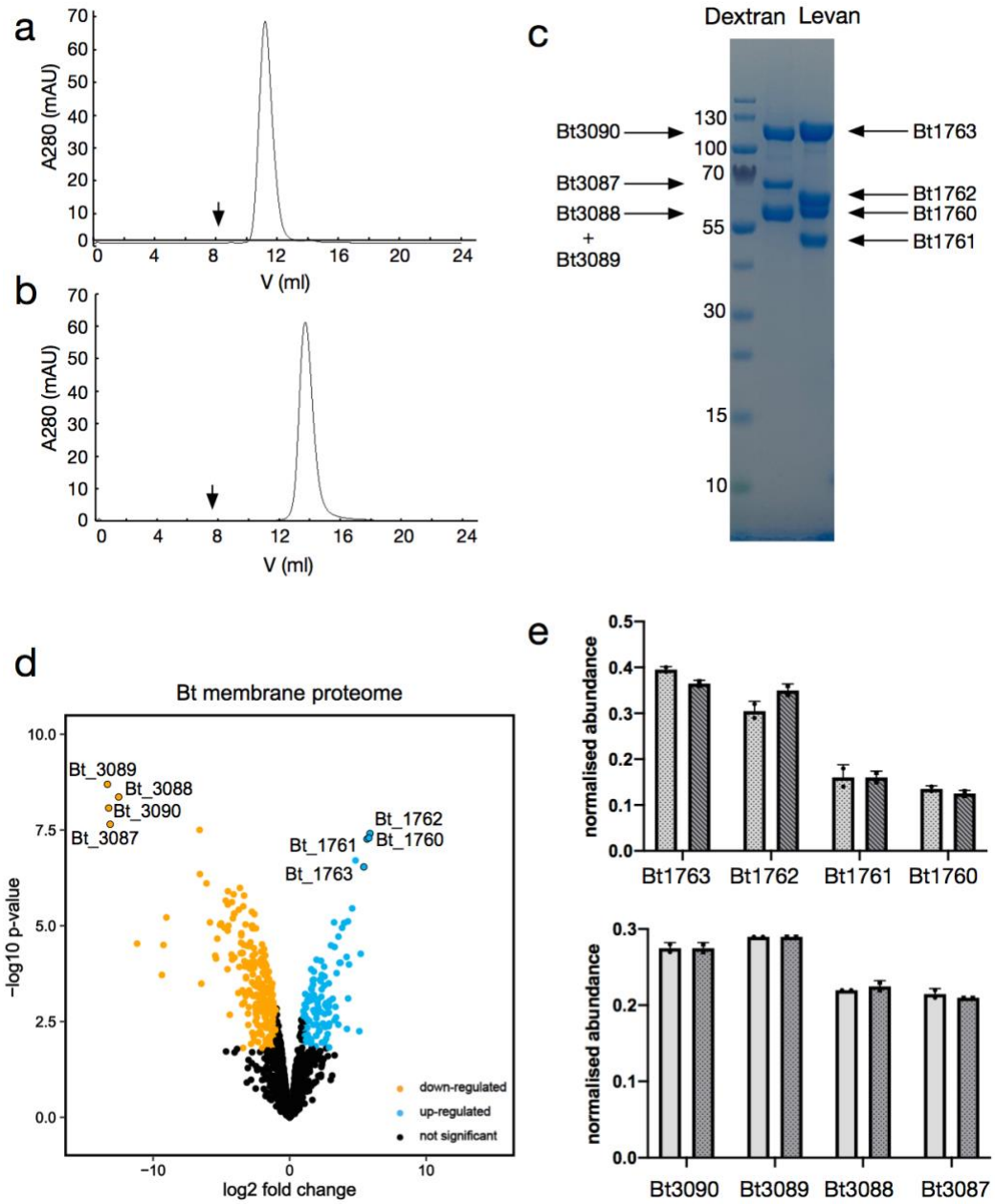


Figure 2: Purification of stable 4-component complexes from the OM of *B. thetaiotaomicron*. a,b Analytical SEC elution profiles for Bt1760-63 (levan 4CC; **a**) and Bt3087-90 (dextran 4CC; **b**) analysed on Superdex-200 and Superose-6, respectively. The void volume of the columns are indicated by the arrows. **c**, SDS-PAGE of the complexes purified in (**a,b**) with the identities of the bands indicated. Bt3088 and 3089 of the dextran 4CC are almost the same molecular weight and appear as a single band on the gel. **d**, Volcano plot of the *B. thetaiotaomicron* outer membrane proteome from fructose- vs. dextran-grown cells. **e**, Normalised OM

abundance of the four components of the levan (top panel) and the dextran (bottom panel) systems obtained from fructose- and dextran-grown cells, respectively (light bars). The dark bars show the normalised abundance of detergent-purified 4CCs spiked into the proteome samples of dextran- or fructose-grown cells (top and bottom panels, respectively).

164 **Proteomics data provide support for stable four-component complexes at the OM**

165 Successful purification of two different detergent-solubilised four-component complexes that
166 elute as monodisperse species from size exclusion chromatography strongly suggests that the
167 complexes are stable in the OM. To support this notion and to shed light on the relative
168 abundance of the components, we performed semi-quantitative proteomics using intensity-
169 based absolute quantification (iBAQ)[22] on the outer membrane fractions from *B. theta* grown
170 on either fructose or dextran. The four components of each complex are among the most-
171 highly expressed membrane proteins in *B. theta* when grown on minimal media with either
172 fructose or dextran as the sole carbon source (**Fig. 2d**). While the proteins of the dextran 4CC
173 are present in roughly equimolar amounts based on the iBAQ analyses, the additional
174 lipoproteins of the levan 4CC appear to be substoichiometric (**Fig. 2e**). However, the SDS-
175 PAGE gel and the monodisperse gel filtration peaks of detergent-purified complex suggest
176 that the iBAQ analysis, being semi-quantitative, may underestimate the abundance of the
177 additional lipoproteins of the levan 4CC.

178

179 As expected from transcription studies, the levan PUL is not expressed when *B. theta* is grown
180 on dextran, with the reverse true for expression of the dextran PUL (**Fig. 2d**). Thus, in dextran-
181 grown cells, components of the levan PUL are present at just ~1 % of the level found when
182 cells are grown on fructose, and in fructose-grown cells, dextran PUL components were not
183 detectable. This provided the opportunity to spike membrane proteome samples from cells
184 grown on fructose or dextran with detergent-purified dextran or levan 4CCs, respectively. For
185 both the levan and the dextran systems, the relative abundance of the four components in
186 the spiked versus endogenous membrane samples was very similar (**Fig. 2e**), suggesting that
187 there is no excess of any component in the OM. Collectively these data support the hypothesis
188 that the additional lipoproteins form stable, equimolar, 4CCs with their cognate SusCD
189 transporters that are dedicated to the generation and import of oligosaccharides, and which
190 we propose to name 'utilisomes' in keeping with the 'PUL' acronym.

191

192 **Single-particle cryo-EM reveals an octameric, four-component complex for levan utilisation**

193 The structure of the DDM-solubilised levan utilisome was assessed by single particle cryo-EM
194 in the absence of substrate. 2D class averages were markedly distinct from those of the core
195 complex, with unambiguous evidence for additional components in the most populated
196 classes (see **ED Figure 2**). Initial classification in 3D revealed extensive compositional
197 heterogeneity in the dataset. Given the biochemical evidence for approximately equimolar
198 complexes, we speculate that this compositional heterogeneity most likely arises from
199 denaturation and dissociation during cryo-EM grid preparation. However, >50 % of the
200 particles represented a complex containing a full complement of components, that is, the
201 $SusC_2D_2$ core, as well as two copies of both the levanase (Bt1760) and the levan SGBP (Bt1761)
202 (see **Figure 3**). We propose that this four-component, octameric complex (~570 kDa)
203 represents the complete OM levan utilisome. The dataset also contained classes
204 corresponding to partial complexes comprising a single copy of Bt1760 and Bt1761 in addition
205 to the $SusC_2D_2$ core (~40 % of particles), as well as a small population of 'naked' $SusC_2D_2$
206 complexes (~12 %).

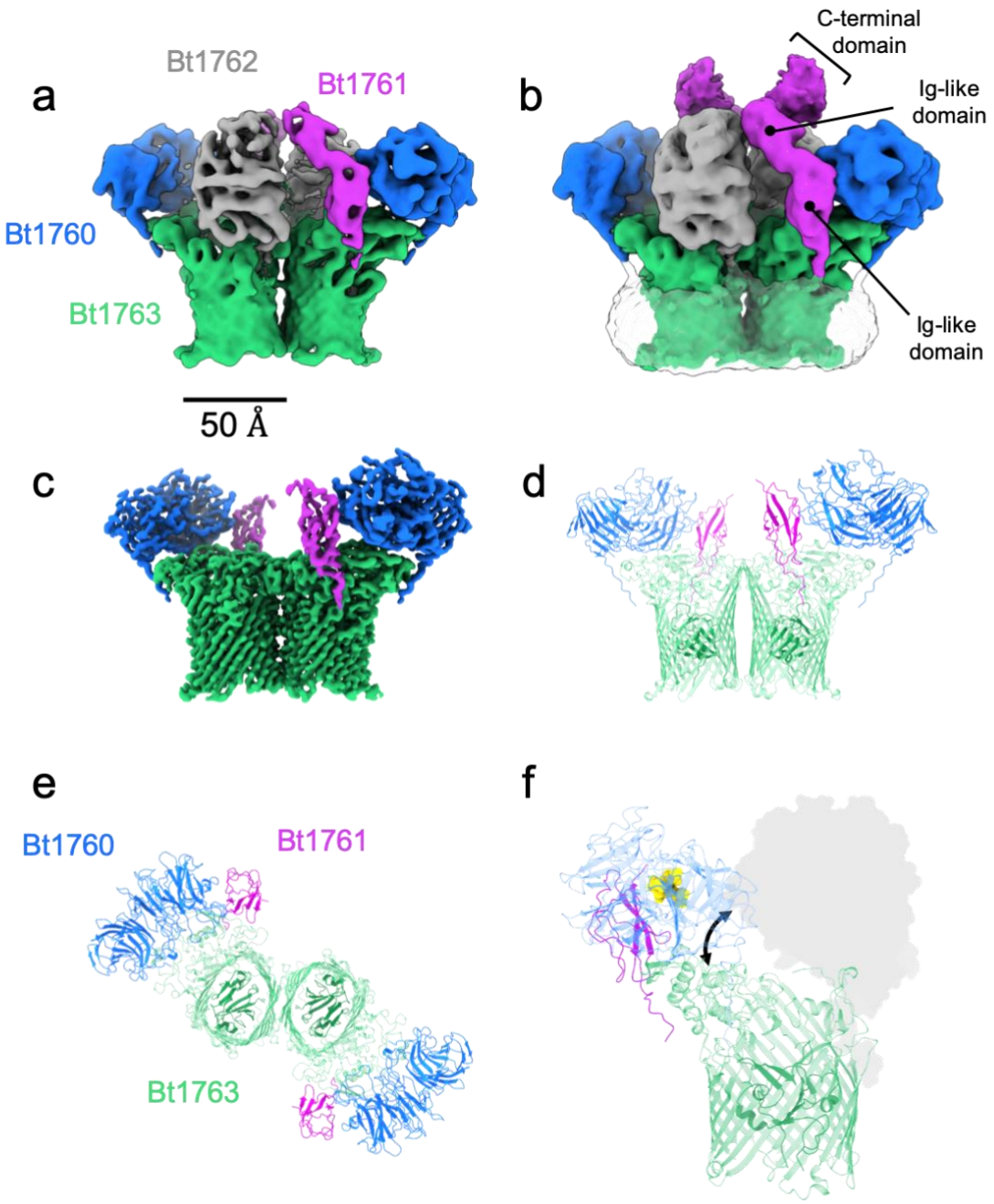


Fig 3 The organisation of the levan utilisome revealed by cryo-EM. a-b, Output of 3D refinement for one conformational state of the utilisome obtained by 3D classification. The map is shown at low (**a**) and high (**b**) threshold levels and is viewed in the plane of the membrane. The SusC and SusD components are green and grey respectively. The levan binding protein Bt1761 is shown in magenta and the endo-levanase Bt1760 is coloured blue. Density for the disordered micelle in the high threshold view is transparent grey. **c**, C2 symmetrical reconstruction at 3.5 Å from a focused refinement excluding the variable SusD subunits, viewed in the plane of the membrane. **d**, Atomic model built from the density shown in **c**. **e**, the same model but viewed top down from the extracellular side. **f**, One half of the complex is shown in the context of the SusD lid (grey shadow). Using the crystal structure (PDB ID 7ZNR), FOS DP5 (yellow) has been artificially displayed at the active site of the endo-levanase, highlighting proximity to the open 'mouth' of the SusCD transporter (marked by the double headed arrow). Note that all particles possess SusD components, and their absence here is a result of focused refinement.

207 **The arrangement of the levanase and levan SGBP suggests concerted function**

208 The levanase is mounted on the lip of the SusC barrel, at the opposite edge from the location
209 of the hinge contacts between SusC and SusD (**Figure 3**). Extracellular loops 1 and 9 contribute
210 to the binding interface which buries a surface area of ~820 Å. This arrangement means that,
211 crucially, the opening and closing of the SusD lid is not hindered by the presence of the
212 levanase. Furthermore, the levanase is oriented in a way that is consistent with concerted
213 function with the core transporter, with its active site facing, and close to, the mouth of the
214 SusCD transporter such that the minimum distance between FOS bound to the respective
215 subunits is ~30 Å (estimated by alignment of crystal structures of the levanase bound to short
216 FOS (7ZNR) and Bt1763 containing bound FOS (6ZAZ) with the corresponding components in
217 this structure) (**Figure 3**). The levan SGBP is positioned adjacent to the levanase. An
218 AlphaFold2-predicted model[23] of the levan SGBP has two N-terminal Ig-like domains, followed
219 by a globular C-terminal domain which we postulate is responsible for levan binding,
220 consistent with the domain organisation of other characterised SGBPs (**ED Figure 1**)[24]. In the
221 substrate-free utilisome structure, only the most N-terminal Ig-like domain is unambiguously
222 resolved at high resolution. Interactions between the levan SGBP and SusC again occur at
223 extracellular loop 1. Weaker, poorly-resolved density is present for the second, Ig-like
224 domain, which extends away from the transporter, towards the extracellular space. Finally,
225 the C-terminal domain of the levan SGBP is much more mobile, with weak, diffuse density
226 only visible at high threshold levels (**Figure 3**). The lip of SusC therefore serves as a platform
227 for the association of the auxiliary components, perhaps explaining why SusC-type TonB-
228 dependent transporters are ~40 % larger than their classical monomeric counterparts[9]. The
229 structure, alongside the observation that most of the levan SGBP subunits appear to be
230 flexible, provides important insight into how the utilisome architecture contributes to
231 efficient levan utilisation. The putative levan binding C-terminal domain of the levan SGBP is
232 projected away from the cell surface, facilitating initial capture of levan chains. The observed
233 flexibility of this region would then feasibly permit positioning of these chains for processing
234 by the adjacent levanase units. Finally, any cleavage products would be released close to their
235 binding site at the SusCD interface, promoting efficient loading of the transporter.

236

237 **Conformational heterogeneity of the apo utilisome complex**

238 We next performed further 3D classification of particles corresponding to the intact utilisome.
239 The major area of conformational heterogeneity within these data is in the position of the
240 SusD lids (**ED Figure 3**). All complexes within this dataset are open, but they are open to
241 differing degrees, suggesting a rugged energy landscape for the lid-open state with several
242 local minima. The lids within a complex appear to operate independently of one another such
243 that differing lid positions on either side of the complex break its C2 symmetry. The lid
244 conformations vary from a partially open state that is comparable to the predominant lid
245 position observed in the study of the core apo $SusC_2D_2$ complex[12], to a wide-open state also
246 reported previously (**ED Figure 3**). Importantly, no closed lids are observed and, in contrast to
247 our findings for the LDAO-solubilised $SusC_2D_2$ core complex[10,12], all SusC barrels contain
248 unambiguous density for the SusC plug domain that occludes the barrel. Thus, transporters
249 with open lids are plugged and competent to accept substrate, consistent with the proposed
250 'pedal-bin' hypothesis, Mobility of the C-terminal region of the levan SGBP is also evident in
251 the output of 3D classification (**ED Figure 3**).
252
253 Whilst the SusD lids have several conformations, SusC, the levanase and the N-terminal
254 domain of the levan SGBP constituted a rigid unit with C2 symmetry. To provide more detailed
255 structural information for these regions, a mask that excluded density for the SusD lids was
256 applied in a focused refinement, generating a reconstruction at 3.5 Å after sharpening (**Figure**
257 **3**). At this resolution, it was possible to dock and refine models for SusC and the levanase, and
258 to build the N-terminal domain of the levan SGBP. The refined model of Bt1760 within the
259 utilisome is similar to the crystal structure of *E. coli*-expressed Bt1760 solved in isolation (PDB:
260 7ZNR)(Cα rmsd = 0.54 Å).
261
262 **A stable, four-component utilisome is not unique to the levan utilisation system**
263 To investigate whether the levan utilisome architecture is a general feature of OM PUL
264 components, we studied a second example; the dextran utilisation system of *B. theta*. Dextran
265 is an α1,6 linked glucose polymer with occasional α1,4 glucose branches and, like the levan
266 utilisation system, the dextran PUL encodes 4 OM components: the SusC transporter and
267 SusD lid (Bt3090 and Bt3089), a GH66 endo-dextranase (Bt3087), and a putative dextran SBGP
268 (Bt3088) (**ED Figure 1**).
269

270 Cryo-EM of the dextran system revealed it to be more aggregation prone than the levan
271 system, requiring altered grid preparation conditions for data collection (Methods). Despite
272 this, initial 3D classification shows that the $SusC_2D_2$ complex is intact with SusD subunits
273 occupying an open position. As observed for the levan system, there is considerable
274 compositional heterogeneity in the data. Densities for the dextranase and dextran SGBP were
275 first assigned by rigid-body docking of AlphaFold2-predicted models (**ED Figure 4).** Both
276 lipoproteins were observed to assemble on the core $SusC_2D_2$ transporter although, unlike the
277 levan system, there were no classes corresponding to an octameric dextran utilisome, which
278 we attribute to a higher propensity of components of the dextran system to
279 disassemble/aggregate during grid preparation. The highest order complex observed was
280 heptameric, containing two copies of the dextran SGBP and a single dextranase. Hexameric
281 and pentameric complexes comprised of all possible complements of additional lipoproteins
282 (**ED Figure 5**) were also observed. Interestingly, the relative positions of the auxiliary
283 components on the core complex differ between the levan and dextran utilisomes (**ED Figure**
284 **6**). For the former, the endo-levanase is nearest to the hinge of SusD whereas for the latter
285 the glycan binding protein is nearest. Regardless, the arrangement of components in the
286 dextran utilisome is compatible with movement of the SusD lid. Together, the data suggest
287 that OM utilisomes are a general feature of *Bacteroides* PULs.
288
289 **Addition of levan causes concerted conformational changes in the utilisome**
290 To probe the mechanism of substrate capture and processing by the levan utilisome, ~0.5 mM
291 levan FOS with a degree of polymerisation (DP) of ~8-12 (DP8-12) was added to the detergent-
292 purified utilisome complex. The resulting 3D structure shows that a major, concerted
293 conformational change has occurred in the complex in the presence of FOS, with all SusD lids
294 closed and tightly capping the lumen on the extracellular side of the SusC barrels (**Figure 4**).
295 3D classification of these data suggested that all major heterogeneity was due to the
296 presence/absence of auxiliary lipoproteins (**ED Figure 7**). To confirm the presence or absence
297 of substrate within the closed transporter complex, we first performed targeted refinement
298 of the $SusC_2D_2$ core. To achieve this, we used particle subtraction to remove the signal for the
299 levanase and levan SGBP from the images, and were thus able to include all particles in our
300 refinement, regardless of their lipoprotein complement. A final reconstruction of the $SusC_2D_2$
301 core was obtained at 2.9 Å, and contains unambiguous density for FOS substrate at two

302 distinct sites within the SusC barrel. These densities are comparable to those observed in the
303 crystal structure of the substrate-bound $SusC_2D_2$ complex (6ZAZ) (**Figure 4**). The
304 oligosaccharide bound at the interface of SusC and SusD consists of six β2,6-linked fructose
305 units with a β2,1 decoration on Frc-4 (numbered from the non-reducing end). The FOS adopts
306 a compact, twisted topology similar to that observed for the FOS in the $SusC_2D_2$ crystal
307 structure (6ZAZ) and FOS DP5 in the levanase crystal structure. The second FOS molecule
308 occupies a site at the bottom of the SusC binding cavity, where it is in contact with the plug.
309 Here the density most likely corresponds to four β2,6-linked fructose units.
310

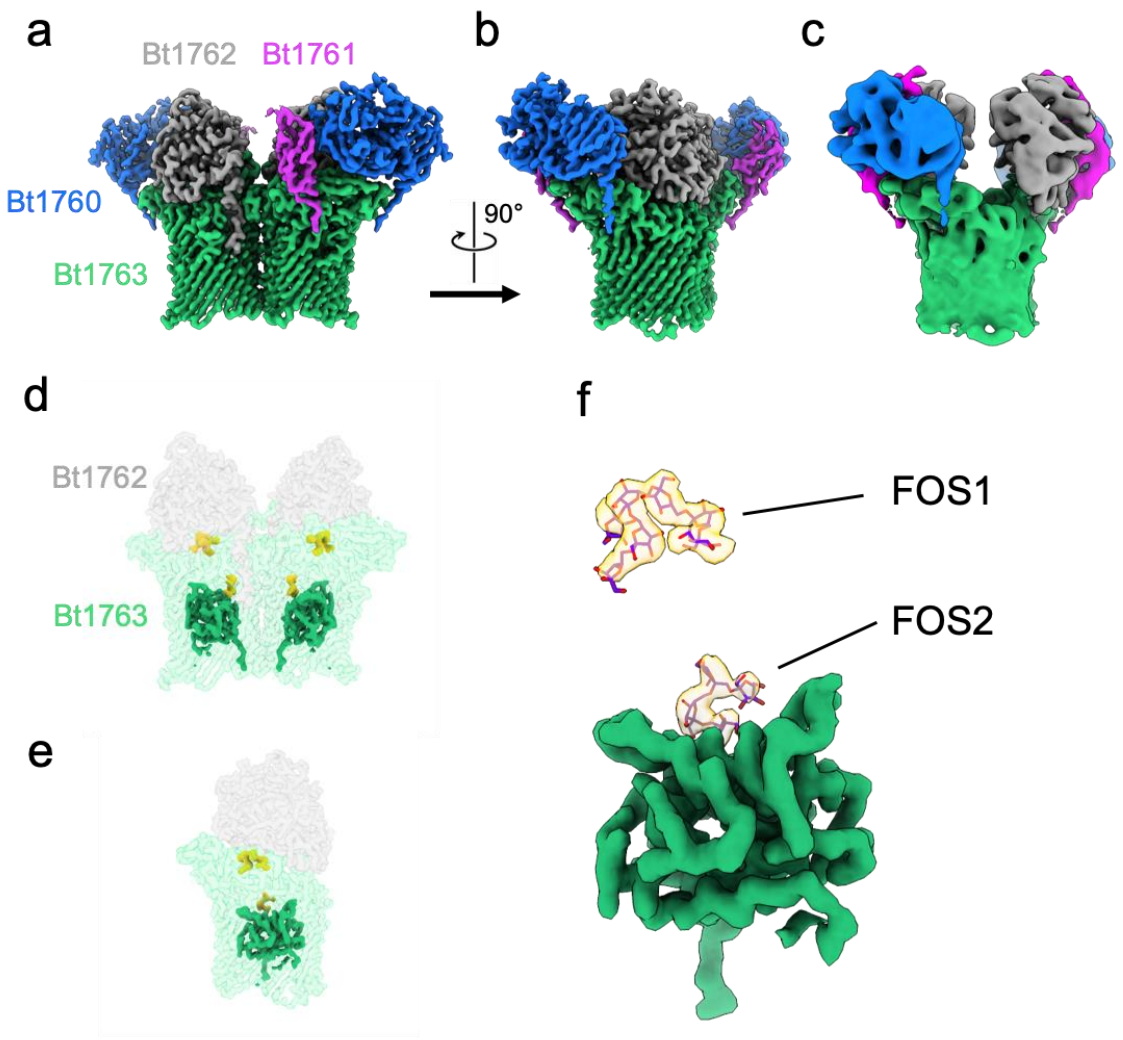


Fig 4 Cryo-EM shows a concerted conformational change of the levan utilisome upon addition of short FOS. **a-b**, Reconstruction at 3.2 Å of the closed substrate-bound transporter viewed in the plane of the membrane **(a)** and rotated 90° (**b**). As before, SusC is coloured green, SusD grey, levan SGBP magenta and levanase blue. **c**, Wide open state of the dimeric transporter seen in the absence of substrate in same orientation as in b, shown for comparison. **d**, Reconstruction of the $SusC_2D_2$ core complex generated from a focused C2 refinement after particle subtraction of additional lipoproteins. Densities for bound FOS (yellow) and the plug

domain (solid green) of SusC are highlighted. **e**, Isolated view of a single SusCD subunit from (**d**) shown from the dimerisation interface. **f,** Isolated view of plug and substrate density in the same orientation as (**e**). The FOS modelled are similar to those in the previously determined crystal structure of $SusC_2D_2$ (PDB: 6ZAZ). Maps for the substrate-bound structure are LAFTER filtered (see **Methods**).

311

312

313 Given the biological role of the utilisome, it is expected that substrate loading of the
314 transporter should be signalled across the OM. This would presumably occur via disordering
315 of the SusC Ton box region, at the periplasmic face of the complex[12]. Indeed, the Ton box and
316 the region preceding it are not visible in the substrate-bound structure (**ED Figure 8**). This is
317 in stark contrast to the apo state of the transporter in which the region preceding the Ton box
318 is resolved, and adopts a sharp turn towards the plug domain, constraining the Ton box within
319 the barrel. It should be noted that the N-terminal extension (NTE) domain of SusC, that we
320 have previously shown to be essential for function[12], is not resolved in either the apo or
321 substrate-bound states of the complex.

322

323 The positions of Bt1763 substrate binding residues are almost identical in the apo and FOS-
324 bound cryo-EM structures (**ED Figure 9**). Therefore, the conformational changes upon
325 substrate binding that lead to exposure of the Ton box are not obvious. One notable
326 difference between the apo and substrate-bound structures is the position of Hinge 2 (or L8)
327 of SusC[12]. Hinge 2 forms extensive contacts with SusD and interacts with FOS at the top of the
328 SusC cavity via F649 (**ED Figure 9a**). In our apo structure, most of Hinge 2 is not visible, likely
329 because it is masked out together with the flexible SusD lid during data processing (Methods).
330 However, the part of Hinge 2 proximal to the SusC barrel is resolved in both structures and
331 shows differences in the side chain orientations which propagated along the edge of the
332 barrel and plug towards the periplasm (**ED Figure 9**). In the substrate-bound structure, W685,
333 which sits at the base of Hinge 2, shifts down towards the periplasm, nudging F583 up and
334 S193 down. This results in a downward shift of the plug loop connecting α3 and β6, to which
335 S193 belongs. In the apo structure, Y191, which is also part of the α3-β6 plug loop, forms a
336 triple aromatic stack together with Y89 and F588. When the α3-β6 plug loop shifts towards
337 the periplasm, the stacking interaction between Y191 and Y89 is broken, and Y191 clashes
338 with the C-terminus of the Ton box resulting in its release into the periplasm. Recent studies
339 of *E. coli* BtuB (vitamin B TBDT) highlighted the presence of a salt bridge between the
340 transporter barrel and plug residues, which has been termed an 'ionic lock'[25], that is broken

341 upon vitamin B12 binding[26]. Similarly, we propose to name the Bt1763 Y89-Y191 stacking
342 interaction an 'aromatic lock.' Our cryo-EM structures reveal how conformational changes
343 caused by substrate-induced SusD lid closure on the extracellular side of the outer membrane
344 propagate through the TonB-dependent transporter, breaking the aromatic lock on the
345 periplasmic side and exposing the Ton box.
346
347 As expected, no substrate density was observed within the active site of the wild type
348 levanase. The addition of short FOS substrate had no influence on the conformations of the
349 auxiliary lipoproteins, and the flexibility of the levan SGBP prevented further structural
350 analysis. Notably, in both the apo and substrate-bound datasets, the levanase and levan SGBP
351 were invariably associated with one another, i.e. there is no evidence of pentameric or
352 heptameric complexes where one SusCD unit is associated with a single lipoprotein as is the
353 case for the dextran system.
354
355 **Longer substrate chains tether the levan SGBP**
356 While the cooperative behaviour of the levan utilisome components is implied by their
357 arrangement, the structure of the complete levan binding protein remained unresolved. We
358 reasoned that by using a catalytically inactive levanase (with the active site D42A mutation)
359 and longer FOS chains (DP ~15-25), cooperative binding of a single levan chain by the levan
360 SGBP and the levanase might restrict the mobility of the levan SGBP. We therefore collected
361 a new cryo-EM dataset for the utilisome with inactivated levanase and long-chain FOS, and
362 observed that in 3D classifications, a novel conformation of the levan SGBP was present. In
363 several classes, the levan SGBP adopted a 'docked' state, in which its C-terminal domain was
364 held proximal to both SusD and the levanase (**ED Figure 10**). However, a consensus
365 refinement of those classes containing evidence of the docked state still produced a map in
366 which the C-terminal domain of Bt1761 was poorly resolved (**ED Figure 10**), indicating that
367 local conformational heterogeneity remained. We therefore attempted to collate a more
368 homogeneous subset of particles from within the dataset by using an approach involving
369 focused classification without particle alignment (Methods), using a mask that only included
370 the docked levan SGBP position. This classification yielded a class with high-resolution
371 information for the C-terminal domain of the levan SGBP (**ED Figure 10**). Particles belonging
372 to this class were used to generate a final, sharpened reconstruction of the complete

373 utilisome at 3.0 Å with well-resolved density for the complete levan SGBP (**Figure 5**), which
374 we built into the cryo-EM map, starting from a partial model obtained by X-ray crystallography
375 (Methods).

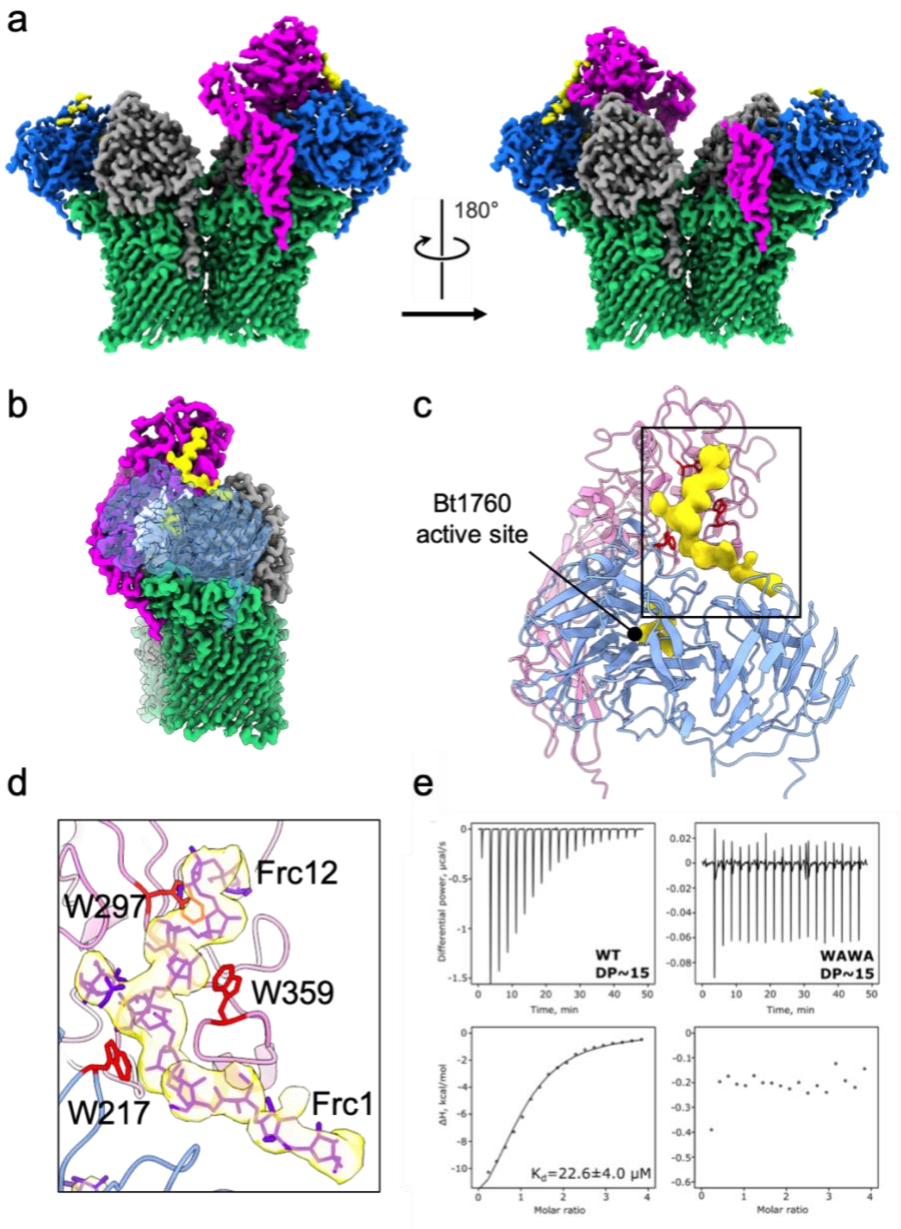


Fig 5 Cryo-EM structure of the holo utilisome. a, Reconstruction of the levan utilisome complex after focused classification viewed in the plane of the membrane (left) and rotated 180° (right). One levan SGBP is now fully resolved (magenta). **b**, side-on view of the complex with transparent density for Bt1760, highlighting the position of the cooperatively bound levan chain (yellow). **c**, Structures of the levan SGBP and inactive endo-levanase with additional density attributed to the cooperatively bound FOS shown in yellow (boxed). **d**, Zoomed inset of the boxed region in **c**. The modelled levan chain is shown as purple sticks. Tryptophan residues proposed to contribute to binding are shown in red and from top to bottom correspond to W297 and W359 from Bt1761 levan SGBP and W217 from Bt1760 levanase. Note that the FOS that bridges the levan SGBP and the levanase occupies a secondary binding site on the latter, distinct from the active site. Maps displayed here have been filtered using LAFTER[27]. **e**, Isothermal titration calorimetry experiments with 1 mM FOS DP~15 titrated into 50 μM recombinant wild type levan SGBP, or the W297A/W359A double-mutant (WAWA). Removing both tryptophan residues completely abolished binding.

376

377 In agreement with the AlphaFold2 structure prediction, the levan SGBP is organised into three
378 domains, with the lipidated N-terminal and central domain each possessing an Ig-like fold.
379 The C-terminal levan-binding domain (residues 205-321) is globular, with a central β-sheet
380 decorated with α-helices. A distance-matrix alignment (DALI) analysis of the C-terminal
381 domain returned a number of structures with significant similarity. A ThuA-like protein (PDB
382 ID 1T0B) with a putative role in the conversion of disaccharides to their 3-keto derivatives and
383 a homoserine O-succinyl transferase (PDB ID 7CBE), have Z-scores of ~15 and ~14
384 respectively.

385

386 Additional density forms a bridge between the levanase and levan SGBP. We attribute this
387 density to a length of levan chain, bound by the C-terminal domain of the levan SGBP and the
388 inactive levanase. The density is compatible with a stretch of 12 β2,6-linked fructose units,
389 with a putative β2,1 decoration observed on Frc-7 (numbered from non-reducing end) (**Figure**
390 **5**). Interestingly, the bound levan chain adopts a relatively extended conformation, markedly
391 different from the compact twisted conformations observed for FOS bound to the levanase
392 and within the SusCD units. Whilst the resolution of the density is insufficient for a detailed
393 description of binding interactions, several tryptophan residues are clearly implicated in
394 binding. These include W297 and W359 from the levan SGBP, and W217 from the levanase,
395 which appear to cradle the FOS chain via stacking interactions. Isothermal titration
396 calorimetry (ITC) on recombinant protein produced in *E. coli* confirmed the importance of
397 W297 and W359 for levan binding by the levan SGBP (**Figure 5e, ED Figure 11**). In addition,
398 the ITC data suggest that the binding site observed in the cryo-EM structure is the only levan
399 binding site on the levan SGBP. The minimum FOS size that the levan SGBP can bind
400 reasonably strongly is an 8-mer (**ED figure 11**).

401

402 Interaction of the long, bridging, levan chain with the levanase does not occur at the levanase
403 active site. Instead, the bridging levan binds across the top of the β-propeller domain of the
404 enzyme and the levan density then proceeds towards the C-terminal β-sandwich domain (**Fig.**
405 **5, ED Figures 12, 13**). Evidence of levan binding at this secondary 'tethering' site on the
406 levanase is also seen in the absence of bridging interactions with the levan SGBP and in crystal

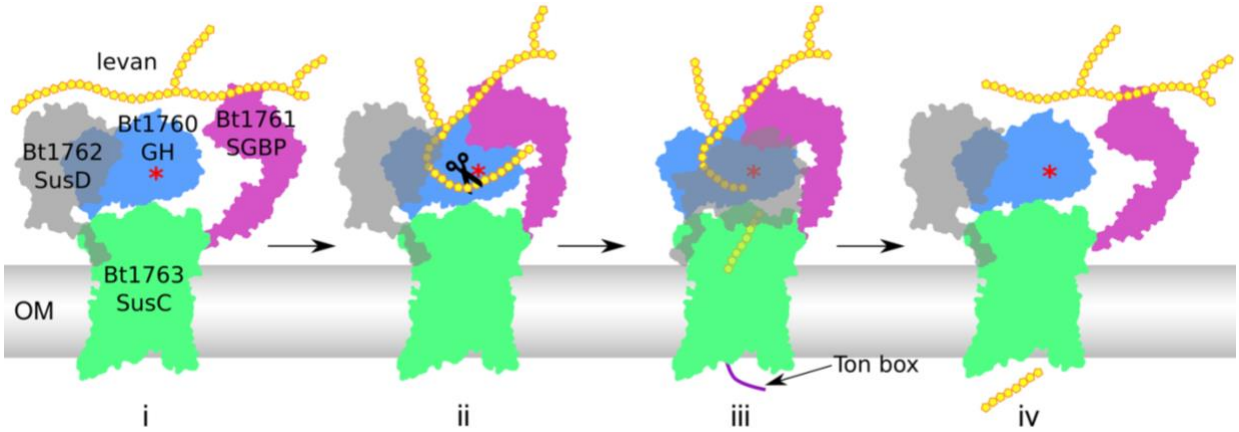
407 structures of the inactive levanase (D42N) solved in the presence of FOS of DP ~7-8 (**ED Figure**
408 **13)**. Indeed, the presence of a secondary binding site (SBS) is relatively common in endo-
409 acting GHs and is proposed to be involved in enhancing enzyme-substrate targeting and
410 processivity. Notably, no ligand was observed at the secondary site in a levanase structure
411 complexed to levantetraose (PDB ID 6R3U; [28]), suggesting the affinity of this site for very short
412 FOS is low. The mutated active site of the enzyme is occupied by a distinct FOS molecule, with
413 density for five β2,6 linked fructose units that adopt the same twisted conformation as
414 levantetraose in 6R3U (ED figure 13). No obvious levan density is observed between the
415 tethering site and active site.
416
417 We attempted to dissect the affinities of the levanase active site and the SBS for levan. The
418 protein-FOS interactions are extensive in both binding sites and include numerous aromatic
419 residues that are often found at carbohydrate interaction interfaces (**ED Figure 14**). We tested
420 the effects of substituting these aromatic residues to alanine on levan binding (**ED Figure 15**).
421 Substitutions near the active site (Y70A and W318A) decreased the affinity for levan 25-30
422 fold. Single substitutions in the SBS either had no effect (F243A and Y437A) or reduced the
423 affinity 6-fold (W217A). These results suggest the residues near the active site are responsible
424 for most of the affinity of the levanase for levan, and the SBS provides a modest contribution.
425 However, it is possible that the numerous weak interactions of the SBS with levan, including
426 polar residues, result in a substantial overall affinity.
427
428 Why SusCD transporters exist as dimers is one of the key unanswered questions in the field.
429 While the functional relevance of dimerisation is not clear when thinking about the core
430 $SusC_2D_2$ complex in isolation, considering the question in the context of the utilisome provides
431 some important insights. During 3D classification of both the active and inactive substrate-
432 bound utilisome datasets, the inherent flexibility of the levan SGBP was evident. However, in
433 the inactive utilisome dataset with long FOS, a novel conformation was observed in which an
434 untethered levan SGBP appears to reach across and contact a tethered levan SGBP associated
435 with the other SusC subunit (**ED Figure 16**). The fact this 'reach over' conformation was
436 observed exclusively in the inactive utilisome dataset with the longer FOS perhaps indicates
437 that this conformation is the result of both levan SGBPs interacting with the same stretch of
438 the long levan chain that would likely be present *in vivo*. Such cooperative binding of long

439 glycan chains by the levan SGBPs would result in higher avidity for the substrate and would
440 bring/keep more substrate in the vicinity of the utilisome, offering a potential explanation as
441 to how dimerisation contributes to function.
442
443 **FOS binding by the SusCD core is promiscuous**
444 Interestingly, in the presence of longer FOS and an inactive levanase, the substrate density
445 within the SusCD binding cavity differs from that observed previously for short FOS. At the
446 FOS1 site, density for the putative β2,1 decoration on Frc4 is missing (**ED Figure 17)**.
447 Conversely, contiguous density can be seen beyond the previously defined reducing terminus
448 of FOS1, with a novel β2,1 decoration on Frc5. The substrate bound at the FOS2 site follows
449 a similar trend with the β2,1 linked fructose side chain modelled previously being much
450 weaker with longer FOS, and additional density attributed to another β2,6 linked monomer
451 extending the chain in the direction of the FOS1 site. Strikingly, at higher threshold levels,
452 diffuse density is visible that connects both FOS binding sites, indicating that longer FOS
453 (~DP15) can occupy both sites simultaneously. The density between the FOS1 and FOS2 sites
454 is relatively weak and indicative of multiple conformations, consistent with the absence of
455 any contacts from SusC in this bridging segment. These data confirm that the transporter has
456 considerable substrate promiscuity and that, as suggested previously, relatively long FOS (~15
457 DP) can be accommodated[12].
458

459 ## Discussion

460 Our cryo-EM structures of the levan utilisome provide unprecedented insight into glycan
461 acquisition by *Bacteroides spp*. Since all sequenced Bacteroidetes to date contain SusCD
462 homologues, our findings likely apply across multiple systems, potentially including SusCD
463 systems involved in the transport of non-glycan macromolecules[10]. As Bacteroidetes are
464 found in a diverse range of terrestrial and marine niches[29], the data also expand our
465 understanding of glycan utilisation outside the animal gut. We demonstrate that for two
466 distinct PULs, all of the gene products that localise to the bacterial OM form a stable utilisome
467 complex with a defined architecture. The position of each component within this complex is
468 consistent with their respective functions, encapsulated in our general model for utilisome
469 function (**Fig. 6**). In the absence of substrate (i), the SGBP is in an extended conformation with

470 a mobile glycan-binding domain, increasing its efficiency as a glycan grappling device.
471 Likewise, the SusD lids, which can open and close without clashing with the other
472 components, are mobile and occupy predominantly open states. (ii) Glycan binding by the
473 SGBP is followed by docking to the proximal glycanase and binding by the SBS and active site
474 of the enzyme. Substrate cleavage by the endo-acting enzyme generates shorter
475 oligosaccharides in close vicinity to the mouth of the SusCD transporter. (iii) Concerted
476 binding of the oligosaccharide to SusD and SusC promotes the formation of the closed
477 conformation of the transporter, leading to conformational changes along the SusC barrel
478 and plug, breaking of the aromatic lock and exposure of the Ton box to the periplasm. This is
479 followed by TonB-dependent transport events (not shown), and resetting of the transporter
480 to the open state (iv)[10,12].
481



482
483 **Figure 6. The transport cycle of glycan utilisomes.** Description of states i-iv is provided in the main text. N.B.
484 only one half of the dimeric utilisome is shown for clarity. The red asterisk marks the active site of the glycoside
485 hydrolase (GH). OM, outer membrane.
486

487 The bridge formed by long levan chains (DP ~15-25) between the levan binding domain of the
488 SGBP and the levanase raises several interesting points. As outlined above, its extended
489 conformation is unique, perhaps arising because the chain is held under tension. The
490 utilisome structure captured here may therefore represent a state that is 'strained', with a
491 bound levan chain that is pulled taut. However, it is important to note that this conformation
492 has been selected for during image processing owing to the fact that it represents the least
493 mobile and therefore most resolvable state. Second, even though the SBS of the levanase is
494 ~25 Å away from its catalytic site, we speculate that the levan SGBP-levanase interaction
495 serves to increase the local levan concentration around the active site, optimising the
496 efficiency of substrate cleavage. Flexibility of the levan chain away from the interaction sites

497 may preclude resolution of contiguous density in our structure or alternatively, the FOS
498 present in our sample are not long enough to bridge both sites.
499
500 Whilst our data argue in favour of stable utilisome assemblies in the OM, recent studies on
501 the archetypal starch PUL propose a much more dynamic model, with immobile starch
502 binding centres comprised of the SGBPs SusE and SusF, around which transporter
503 components and the endo-amylase (SusG) can transiently assemble[21]. Proteomics data for
504 the starch utilisation system reveal similar amounts of SusC and SusD at the OM whilst SusEFG
505 are much less abundant, providing support for the presence of complexes with differing
506 stoichiometries[21]. Intriguingly, co-immunoprecipitation with SusD antibodies captures twice
507 as much SusD as SusC, suggesting that even the core SusCD transporter may not form a stable
508 complex in the starch system. This clearly contrasts with our work, where separate SusC or
509 SusD components have never been observed during purification, regardless of whether the
510 purification tag was located on SusC or SusD. Moreover, while it is not clear how much the *E.*
511 *coli* OM resembles that of *B. theta*, recent work has shown that *E. coli* OMP and LPS mobility
512 is limited [30,31], raising questions regarding the efficiency of dynamically (dis)assembling
513 transport complexes in a crowded OM.
514
515 It is not clear why the starch system appears to operate differently to the levan and dextran
516 systems presented here. One consideration is that these models are based on live-cell
517 fluorescence studies with C-terminal fusions of OM PUL components with relatively large
518 fluorescent protein tags. Considering the compact utilisome architecture (**ED Figure 4**), it is
519 possible that the fluorescent tags cause utilisome destabilisation or interfere with assembly,
520 resulting in dynamic behaviour that might not affect cell growth *in vitro*. On the other hand,
521 the starch PUL is different as it encodes two SGBPs (SusE and SusF), *i.e*. there are five OM PUL
522 components and not four as in the systems studied here. An AlphaFold2 structure prediction
523 of the starch SusC (Bt3702) shows that it has no extra interaction surface for the association
524 of additional lipoprotein components relative to the levan and dextran SusC units (**ED Figure**
525 **6**), making it unlikely that SusE, SusF and SusG can all bind the same SusCD transporter and
526 suggesting the starch utilisome may be more dynamic.
527

528 A key remaining question is how utilisomes are assembled. Surface lipoproteins, including the
529 SusD-like components, glycoside hydrolases and SGBPs, must be flipped from the inner to the
530 outer leaflet of the OM for presentation on the cell surface. The size and complexity of these
531 subunits suggests that dedicated cellular machinery is required, although candidates have yet
532 to be identified. It is conceivable that this process is coordinated with the folding and insertion
533 of SusC components into the OM, in which case the β-barrel assembly machinery (BAM)
534 would be closely associated with the putative lipoprotein flippase.
535

536 **Author contributions**

537 JBRW performed cryo-EM and determined cryo-EM structures, supervised by NR. AS purified
538 proteins, determined structures and carried out ITC. MF purified proteins. YL prepared OM
539 samples for proteomics. TH and AG performed proteomics, supervised by MT. HZ purified
540 Bt1760, supervised by DNB. SF collected X-ray crystallography data for Bt1760. AB solved
541 Bt1760 crystal structures and managed the Newcastle Structural Biology Laboratory. BvdB
542 generated *B. theta* strains, purified proteins, and crystallised Bt1761. BvdB coordinated the
543 project. JBRW, AS, DNB, NAR and BvdB wrote the manuscript.
544

545 **Data availability**

546 The data supporting the findings of this study are available from the corresponding authors
547 upon reasonable request. Cryo-EM reconstructions and corresponding coordinates have been
548 deposited in the Electron Microscopy Data Bank and the Protein Data Bank respectively:
549 Substrate free levan utilisome (EMD-15288, PDB ID 8A9Y), levan utilisome with FOS DP 8-12
550 (EMD-15289, PDB ID 8AA0), $SusC_2D_2$ core from levan utilisome with FOS DP 8-12 (EMD-15290,
551 PDB ID 8AA1), Inactive levan utilisome with FOS DP 15-25 (EMD-15291, PDB ID 8AA2), $SusC_2D_2$
552 core from inactive levan utilisome with FOS DP 15-25 (EMD-1592, PDB ID 8AA3), dextran
553 utilisome consensus refinement (EMD-15293, PDB ID 8AA4). Raw cryo-EM movies will be
554 deposited in the Electron Microscopy Public Image Archive. Coordinates and structure factors
555 from X-ray crystallography experiments that support the findings of this study have been
556 deposited in the Protein Data Bank under the accession codes 7ZNR and 7ZNS. The mass
557 spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via
558 the PRIDE partner repository with the dataset identifier PXD034863.

559 ## Acknowledgements

560 J.B.R.W. was supported by a Wellcome Trust 4-year Ph.D. studentship (215064/Z/18/Z).
561 B.v.d.B is funded by a Wellcome Trust Investigator award (214222/Z/18/Z), supporting AS and
562 YL. MF is supported by a Newcastle University studentship. We acknowledge the Diamond
563 Light Source (Didcot, UK) for beam time (proposals mx306, mx1221, mx13587 and mx18598),
564 and thank the staff of beamlines I02, I03, I04 and I24 for support. All cryoEM was performed
565 at the Astbury Biostructure Laboratory, which was funded by the University of Leeds and
566 Wellcome (108466/Z/15/Z, 221524/Z/20/Z). We thank R. Thompson, E. Hesketh and D.
567 Maskell for EM support. Protein ID mass spectrometry was performed at the University of
568 Leeds Mass Spectrometry Facility. We thank J. Ault and R. George for performing this analysis.
569 For the purpose of open access, the authors have applied a CC BY public copyright licence to
570 any Author Accepted Manuscript version arising from this submission.

571

572

573 ## Methods

574 **Construction of *B. theta* strains.** *B. theta* strains were made as described previously [12]. Briefly,
575 the DNA sequence containing the desired alterations was constructed using the sewing PCR
576 method and ligated into the pExchange-tdk vector[32]. The vectors carrying the altered DNA
577 sequences were introduced into the *B. theta tdk*$^-$ strain via conjugation from *E. coli* S17 λ *pir*.
578 Chromosomal alterations were made by allelic exchange, followed by selection for loss of the
579 pExchange-tdk vector backbone. Mutations were confirmed by PCR amplification of the
580 region of interest and Sanger sequencing.

581

582 **Expression and purification of utilisomes from *B. theta.*** *B. theta* strains were grown at 37°C
583 in a Don Whitley Scientific A32 anaerobic workstation. Brain-heart infusion (BHI) cultures
584 supplemented with 2 ug/ml hemin were inoculated with stabs from glycerol stocks of the
585 appropriate *B. theta* strain and incubated overnight. Defined minimal medium[12] was
586 supplemented with 2 ug/ml hemin and either 0.5% fructose (levan system) or 0.5% dextran
587 (dextran system) and inoculated with the overnight BHI cultures (1:1000 dilution). The
588 cultures were harvested after 20 h by centrifugation and the pellets were stored at -20°C. 4

589 litres of culture were grown for Bt1762-His (with wild type levanase) and Bt1761-His/Bt1760-
590 D42A (inactive levanase) strains for cryo-EM.
591
592 The cell pellets were thawed, supplemented with DNase I and homogenised in Tris-buffered
593 saline (TBS, 20 mM Tris-HCl pH 8.0, 300 mM NaCl). The cells were lysed with a single pass at
594 22 kpsi through a cell disruptor (0.75 kW; Constant Systems). Membranes were isolated by
595 ultracentrifugation for 45 min at 42,000 rpm (45 Ti rotor, Beckman), 4°C. The membranes
596 were solubilised at 4°C for 1 h in TBS with 1% DDM while stirring. Insoluble material was
597 pelleted by ultracentrifugation for 30 min at 42,000 rpm (45 Ti rotor), 4°C. The supernatants
598 were supplemented with 20 mM imidazole and loaded onto an 8 ml chelating sepharose
599 column charged with Ni^{2+}, by gravity flow at room temperature. The column was washed with
600 20 column volumes TBS with 30 mM imidazole and 0.15% DDM. The bound proteins were
601 eluted with 3 column volumes TBS with 250 mM imidazole and 0.15% DDM. The eluates were
602 concentrated in an Amicon Ultra filtration device with a 100 kDa cut-off membrane. The
603 samples were then loaded on a HiLoad 16/600 Superdex 200 pg column (Cytiva) in 10 mM
604 HEPES-NaOH pH 7.5, 100 mM NaCl, 0.03% DDM. Fractions containing pure protein were
605 pooled, concentrated, flash-frozen in liquid nitrogen and stored at -80°C.
606
607 **Sample preparation for mass spectrometry.** *Bacteroides thetaiotaomicron* membrane
608 samples were suspended in 5% sodium dodecyl sulfate (SDS) in 50 mM triethylammonium
609 bicarbonate (TEAB) pH 7.5 and sonicated using an ultrasonic homogenizer (Hielscher) for 1
610 minute. Samples were centrifuged at 10,000 xg for 10 minutes to pellet debris. Proteins (40
611 µg) were subsequently reduced by incubation with 20 mM tris(2-carboxyethyl)phosphine for
612 15 minutes at 47 °C, and alkylated with 20 mM iodoacetamide for 15 minutes at room
613 temperature in the dark. Proteomic sample preparation was performed using the suspension
614 trapping (S-Trap) sample preparation method[33,34], with minor modifications as recommended
615 by the supplier (ProtiFi™, Huntington NY). Briefly, 2.5 µl of 12% phosphoric acid was added to
616 each sample, followed by the addition of 165 µl S-Trap binding buffer (90% methanol in 100
617 mM TEAB pH 7.1). The acidified samples were added, separately, to S-Trap micro-spin
618 columns and centrifuged at 4,000 xg for 1 minute until all the solution has passed through the
619 filter. Each S-Trap micro-spin column was washed with 150 µl S-trap binding buffer by
620 centrifugation at 4,000 xg for 1 minute. This process was repeated for a total of five washes.

621 Twenty-five µl of 50 mM TEAB containing 4 µg trypsin was added to each sample, followed
622 by proteolytic digestion for 2 hours at 47 °C using a thermomixer (Eppendorf). Peptides were
623 eluted with 50 mM TEAB pH 8.0 and centrifugation at 3,000 xg for 1 minute. Elution steps
624 were repeated using 0.2% formic acid and 0.2% formic acid in 50% acetonitrile, respectively.
625 The three eluates from each sample were combined and dried using a speed-vac before
626 storage at -80°C.
627
628 **Mass spectrometry.** Peptides were dissolved in 2% acetonitrile containing 0.1% formic acid,
629 and each sample was independently analysed on an Orbitrap Q Exactive HF mass
630 spectrometer (Thermo Fisher Scientific), connected to an UltiMate 3000 RSLCnano System
631 (Thermo Fisher Scientific). Peptides were injected on a PepMap 100 C18 LC trap column
632 (300 µm ID × 5 mm, 5 µm, 100 Å) followed by separation on an EASY-Spray nanoLC C18
633 column (75 µm ID × 50 cm, 2 µm, 100 Å) at a flow rate of 250 nl/min. Solvent A was water
634 containing 0.1% formic acid, and solvent B was 80% acetonitrile containing 0.1% formic acid.
635 The gradient used for analysis was as follows: solvent B was maintained at 2% B for 5 min,
636 followed by an increase from 2 to 30% B in 110 min, 30% to 42% B in 10 min, 42-90% B in
637 0.5 min, maintained at 90% B for 4 min, followed by a decrease to 2% in 0.5 min, and
638 equilibration at 2% for 20 min. The Orbitrap Q Exactive HF was operated in positive-ion data-
639 dependent mode. The precursor ion scan (full scan) was performed in the Orbitrap (OT) in the
640 range of 350-1,500 m/z with a resolution of 60,000 at 200 m/z, an automatic gain control
641 (AGC) target of 3×10^6, and an ion injection time of 50 ms. MS/MS spectra were acquired in
642 the OT using the Top 20 precursors fragmented by high-energy collisional dissociation (HCD)
643 fragmentation. Precursors were isolated using the quadrupole using a 1.6 m/z isolation width.
644 An HCD collision energy of 25% was used, the AGC target was set to 2×10^5 and an ion
645 injection time of 50 ms was allowed. Dynamic exclusion of ions was implemented using a 45 s
646 exclusion duration. An electrospray voltage of 1.8 kV and capillary temperature of 280°C, with
647 no sheath and auxiliary gas flow, was used.
648
649 **Mass spectrometry data analysis**. All spectra were analysed using MaxQuant 1.6.14.0 [35], and
650 searched against the *Bacteroides thetaiotaomicron* Uniprot proteome database
651 (UP000001414) downloaded on 22 September 2020. Peak list generation was performed
652 within MaxQuant and searches were performed using default parameters and the built-in

653 Andromeda search engine [36]. The enzyme specificity was set to consider fully tryptic peptides,
654 and two missed cleavages were allowed. Oxidation of methionine and N-terminal acetylation
655 were set as variable modifications. Carbamidomethylation of cysteine was set as a fixed
656 modification. A protein and peptide false discovery rate (FDR) of less than 1% was employed
657 in MaxQuant. Proteins that contained similar peptides and that could not be differentiated
658 on the basis of MS/MS analysis alone were grouped to satisfy the principles of parsimony.
659 Reverse hits, contaminants, and proteins only identified by site modifications were removed
660 before downstream analysis. Ranking of protein abundance was performed using iBAQ
661 intensity values [22] obtained from MaxQuant. Proteomics data analysis was performed using
662 the R package Limma [37]. Paired Student's t-tests were adjusted using a Benjamini-Hochberg
663 and a FDR threshold of p-value <0.05 was applied.
664
665
666 **Construction of plasmids for protein expression in *E. coli***. The nucleotide sequences
667 encoding Bt1760 (residues 2-503) and Bt1761 (residues 2-438) were amplified from B. theta
668 genomic DNA, excluding the signal sequence and the lipid anchor cysteine. In all cases, the
669 protein numbering starts with the first residue of the mature sequence, corresponding to C21
670 in Bt1760 and C24 in Bt1761 precursor amino acid sequences. The PCR product encoding
671 Bt1760 was digested with NcoI and XhoI and ligated into pET28b, resulting in fusion of the
672 coding sequence to a C-terminal LEHHHHHH tag. The PCR product encoding Bt1761 was
673 digested with NdeI and XhoI and ligated into pET28a, resulting in fusion of the Bt1761 coding
674 sequence to an N-terminal MGSSHHHHHHSSGLVPRGSHM tag. TOP10 cells were transformed
675 with the ligation mixtures and plated on LB agar plates with kanamycin. After overnight
676 incubation at 37°C, clones were screened for successful ligation by colony PCR. Bt1760 and
677 Bt1761 variants were generated using the Q5 site directed mutagenesis kit (NEB). All
678 constructs were verified by Sanger sequencing.
679
680 **Expression and purification of proteins from *E. coli***. Bt1760, Bt1761 and their variants were
681 overexpressed in *E. coli*. Electrocompetent *E. coli* BL21(DE3) cells were transformed with the
682 appropriate plasmid, plated on LB kanamycin plates and incubated at 37°C overnight. Starter
683 LB kanamycin cultures were inoculated by scraping the transformants and incubated at 37°C,
684 180 rpm for 2 hours. 13 ml of the starter culture were used to inoculate each litre of LB

685 kanamycin. The cultures were incubated at 37°C, 180 rpm until OD600 0.5-0.6 and induced
686 with 0.2 mM IPTG. The temperature was then lowered to 20°C and the cultures were
687 incubated for a further 19-21 h. The cells were harvested by centrifugation and the pellets
688 were stored at -20°C.
689
690 The pellets were thawed, supplemented with DNase I and homogenised in TBS buffer. The
691 cells were lysed with a single pass at 25 kpsi through a cell disruptor (Constant Systems). The
692 lysates were supplemented with 1 mM PMSF. Unbroken cells were pelleted by centrifugation
693 for 30 min at 30,000g, 4°C. The supernatants were loaded on a 5 ml Ni^{2+}-charged chelating
694 sepharose column by gravity flow at room temperature. The column was washed with 40
695 column volumes TBS buffer containing 30 mM imidazole, and the bound proteins were eluted
696 with 5 column volumes TBS buffer containing 250 mM imidazole. The eluates were
697 concentrated in an Amicon Ultra filtration device (30 kDa cut-off) by centrifugation. The
698 samples were then loaded in batches on a HiLoad 16/600 Superdex 200 pg column (Cytiva) in
699 10 mM HEPES-NaOH pH 7.5, 100 mM NaCl. Elution fractions were collected and analysed by
700 SDS-PAGE for purity. Fractions containing the proteins of interest were pooled, concentrated,
701 flash-frozen in liquid nitrogen, and stored at -80°C.
702
703 **Fructooligosaccharide production**. FOS used for cryo-EM, crystallography and ITC were
704 generated by partial digestion of *Erwinia herbicola* levan (Sigma) by Bt1760 as described
705 previously[12].
706
707 **Isothermal titration calorimetry**. Protein samples were thawed, centrifuged to remove any
708 aggregates, and diluted to 25 or 50 μM in 10 mM HEPES-NaOH pH 7.5, 100 mM NaCl. Levan
709 from *E. herbicola* or defined-length FOS were dissolved in the same buffer to 8 mg/ml and 1
710 mM, respectively. ITC was performed using a Microcal PEAQ-ITC instrument (Malvern
711 Panalytical). Levan or FOS was injected into the sample cell containing protein or buffer. The
712 titrations were performed at 25°C. The sample cell was stirred at 750 rpm. After an initial
713 delay of 60 s, an injection of 0.4 μl was done (which was discarded from data analysis)
714 followed by 18 injections of 2 μl. The spacing between injections was 150 s. Ligand to buffer
715 control titrations were subtracted from all experiments. The experiments were repeated at

716 least twice. Data were fitted to a single binding site model using the Microcal PEAQ-ITC
717 Analysis software v1.40.
718
719 It was impossible to determine the molar concentration of the levan titrant due to its
720 heterogeneity in chain length. Therefore, the molarity of available binding sites was estimated
721 during data fitting. For Bt1760, the only secondary binding site substitution that had an effect
722 on the affinity was W217A. We assumed that all the affinity observed for this variant could
723 be attributed to binding to the active site alone. Therefore, the stoichiometry was fixed to 1
724 and the ligand concentration was floated during data fitting. The estimated molar
725 concentration of 0.8% levan was 829 μM. N.B. the "molarity" in this case corresponds to the
726 number of accessible binding sites for the enzyme on the polymeric levan ligand, rather than
727 number of molecules, per volume. This titrant concentration was fixed for all other data fits
728 for Bt1760. Similarly, by fixing n to 1, the levan titrant concentration for Bt1761 was
729 determined to be 1.48 mM, suggesting ~2x the number of binding sites on levan available to
730 the SGBP as to the enzymes active site. Notably, the affinity of Bt1761 for levan determined
731 this way was similar to that determined using defined-length FOS with known molarity (ED
732 figure 10).
733
734 **Protein crystallisation**. Bt1760_SeMet was crystalized in the presence of 200 mM
735 potassium/sodium tartrate, 100 mM sodium citrate pH 5.6, 1.4 M ammonium sulphate and
736 500 mM fructose. Bt1760_D42N crystal forms were crystalized in 1.5 M lithium sulphate and
737 200 mM ammonium sulphate, 100 mM MES pH 6.5 and 30% PEG 5000 MME respectively.
738 Bt1761 (both native and SeMet protein) was crystallised using 1.8-2.2 M $(NH_4)_2SO_4$, 0.1 M
739 MES pH 6.5. The protein concentrations were in the range of 10 mg/ml. The drops, composed
740 of 0.1 µl or 0.2 µl of protein solution plus 0.1 µl of reservoir solution, were set up using a
741 Mosquito crystallization robot (SPT Labtech). The vapor diffusion sitting drop method was
742 used and the plates were incubated at 20 °C. If required, crystal hits were optimised via
743 hanging drop vapour diffusion with larger volume drops (typically 1-1.5 µl). Bt1760_SeMet
744 samples did not require additional cryoprotection. Bt1760_D42N samples were
745 cryoprotected with paratone-N and with addition of 20% PEG 400 to the reservoir
746 respectively. Bt1761 samples were cryoprotected by adding 4-fold excess of 3.5 M $(NH_4)_2SO_4$
747 to the crystal drops.

748

749 **Data collection, structure solution, model building, refinement and validation**. Diffraction
750 data were collected at the synchrotron beamlines I02, I03 and I04 of Diamond Light Source
751 (Didcot, UK) at a temperature of 100 K. The data set for Bt1760 SeMet was integrated with
752 DIALS[38] via XIA2[39] and scaled with Aimless[40]. The space group was confirmed with Pointless[41].
753 The phase problem was solved by experimental phasing with Crank2[42]. Mutant BT1762_D42N
754 data sets were integrated by XDS[43] and processed as above. After phase transfer from
755 experimental phasing the automated model building program task CCP4build on CCP4cloud[44]
756 delivered models with Rfactors below 30 %. The models were refined with Refmac[45] and
757 manual model building with COOT[46]. The final models were validated with COOT and
758 MolProbity. Data collection and refinement statistics are presented in Supplementary Table
759 1. Other software used were from CCP4 suite[47]. Data collected for SeMet Bt1761 allowed
760 solving the phase problem and partial model building via single anomalous dispersion (Se-
761 SAD) using Phenix AUTOSOL[48]. Iterative rounds of manual building within COOT and model
762 refinement in Phenix resulted in a partial model with R_{free} ~35%, which was used as the input
763 for model completion in the cryo-EM maps of the levan utilisome with long FOS. The segments
764 missing from the Bt1761 X-ray model could not be modelled using the complete cryo-EM
765 structure due to the poor quality of the X-ray electron density maps in the missing regions.
766

767 **Levan utilisome cryo-EM sample preparation, data collection and image processing**. A
768 sample of the purified apo levan utilisome complex (Bt1760-Bt1763) solubilised in DDM-
769 containing buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.03% DDM) was prepared at 3
770 mg/ml. Lacy carbon 300-mesh copper grids (Agar Scientific) were glow-discharged in air (10
771 mA, 30s, Cressington 208). A sample volume of 3.5 mL was applied to the grid. Blotting and
772 plunge freezing into liquid nitrogen-cooled liquid ethane were carried out using an FEI
773 Vitrobot Mark IV (Thermo Fisher Scientific) with chamber conditions set at a temperature of
774 4 °C and 100% relative humidity. The grid was blotted for 6 s with a blot force of 6.
775 Micrograph movies were collected on a Titan Krios Microscope (Thermo Fisher Scientific)
776 operating at 300 kV with a Falcon III direct electron detector operating in counting mode.
777 Data acquisition parameters can be found in Supplementary Table 2.
778

779 An initial dataset comprising 2057 micrograph movies was collected and image processing
780 was carried out in Relion3.1[49]. Drift correction and dose-weighting was carried out using
781 MOTIONCOR2[50]. CTF estimation of motion corrected micrographs was performed using
782 Gctf[51]. Template-based particle picking within Relion was hindered by the large amount of
783 carbon present in many micrographs. The micrograph stack was therefore manually culled to
784 remove micrographs containing >50% carbon, leaving 1093 micrographs for further
785 processing. Final particle picking was performed using the crYOLO general model[52], yielding
786 96,639 particles. This particle stack was subjected to several rounds of 2D classification, after
787 which 89,305 particles remained. A 3D starting model was generated *de novo* from the data
788 using the stochastic gradient descent algorithm within RELION. 3D classification was used to
789 isolate particles corresponding to the complete octameric utilisome complex (45,594). These
790 particles were subjected to further rounds of classification in 3D to assess the conformational
791 heterogeneity of the complex. Classification revealed considerable heterogeneity in the
792 position of the SusD lids, with positions described as 'wide open' (W) and 'narrow open' (N)
793 identified in all possible combinations. WW, WN and NN states contained 16,155, 22,452 and
794 6,987 particles respectively. To increase particle numbers and improve the results of
795 downstream processing a second dataset of 3142 movies was collected. These were
796 processed in the same way as described for the initial dataset, with particles picked using
797 crYOLO. Classification in 3D yielded 146,512 particles that corresponded to the complete
798 octameric utilisome complex. To improve the resolution for the more static, C2 symmetric
799 portions of the utilisome (SusC, levanase and N-terminal region of the levan-binding protein
800 subunits) particles stacks corresponding to the complete octameric complex from both
801 datasets (192,106 total) were combined and subject to focused refinement with C2
802 symmetry. The mask applied in focused refinement excluded the SusD subunits. Particles
803 were subjected to iterative rounds of CTF-refinement and Bayesian polishing (run separately
804 for each dataset) until no further improvement in resolution was seen. Post-processing was
805 performed using a soft, extended mask and yielded a global sharpened reconstruction at 3.5
806 Å, as estimated by the gold standard Fourier shell correlation using the 0.143 criterion.
807
808 **Active levan utilisome in the presence of FOS (DP8-12) cryo-EM sample preparation, data**
809 **collection and image processing**. A sample of the levan utilisome containing an active
810 levanase solubilised in a DDM-containing buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.03%

811 DDM) was prepared at 3 mg/ml and incubated with 0.5 mM levan FOS with a degree of
812 polymerisation of ~8-12 for at least one hour before grid preparation. Quantifoil carbon grids
813 (R1.2/1.3, 300 mesh) were glow discharged (30 mA, 60 s, Quorum GloQube) in the presence
814 of amylamine vapour. A sample volume of 3.5 μL was applied to the grid. Blotting and plunge-
815 freezing into liquid nitrogen-cooled liquid ethane were carried out using an FEI Vitrobot Mark
816 IV (Thermo Fisher Scientific) with chamber conditions set at a temperature of 4 °C and 100%
817 relative humidity. The grid was blotted for 6 s with a blot force of 6. Micrograph movies were
818 collected on a Titan Krios Microscope (Thermo Fisher Scientific) operating at 300 kV with a
819 Falcon III direct electron detector operating in counting mode. Data acquisition parameters
820 can be found in Supplementary Table 3.
821
822 A dataset comprising 974 micrograph movies was collected and image processing was carried
823 out in RELION 3.1 [49]. Drift correction and dose-weighting was done using MOTIONCOR2 [50].
824 CTF estimation of motion corrected micrographs was performed using Gctf [51]. Particle picking
825 was performed using the general model of crYOLO and yielded 72,373 particles [52]. Unwanted
826 particles/contamination were removed from the particle stack through two rounds of 2D
827 classification, after which 63,789 particles remained. Classification in 3D was used to address
828 compositional heterogeneity. Good classes containing unambiguous $SusC_2D_2$ density
829 represented the complete octameric utilisome, a hexameric assembly which lacked one
830 levanase and one LBP subunit and a naked $SusC_2D_2$ complex. Contributing particles numbers
831 were 17,045, 31,789 and 7390, respectively. SusD lids invariantly occupied a closed position
832 and conformational heterogeneity was limited to the position of the levan binding protein.
833
834 Complete utilisome particles were subjected to iterative rounds of CTF-refinement and
835 Bayesian polishing until no further improvement in resolution was seen. Post-processing was
836 performed using a soft, extended mask and yielded a global sharpened reconstruction at 3.2
837 Å, as estimated by the gold standard Fourier shell correlation using the 0.143 criterion.
838
839 To extract higher resolution information for the $SusC_2D_2$ core complex, particle subtraction
840 was performed to remove signal for additional lipoprotein components from all experimental
841 images contributing to good classes (as defined above). A soft, expanded mask encompassing

842 only the $SusC_2D_2$ core was generated using the volume eraser tool within Chimera [53] before
843 using the resulting carved volume as in input for mask creation in RELION. Subtracted particles
844 were used in a focused refinement with the same mask applied while enforcing C2 symmetry.
845 Iterative rounds of CTF-refinement and Bayesian polishing were employed until no further
846 improvement in resolution was observed. Post-processing resulted in a final sharpened
847 reconstruction at 2.9 Å.
848
849 **Inactive levan utilisome in the presence of FOS (DP15-25) cryo-EM sample preparation, data**
850 **collection and image processing**. A sample of the levan utilisome containing inactivated
851 levanase (D42A), solubilised in a DDM-containing buffer (10 mM HEPES, pH 7.5, 100 mM NaCl,
852 0.03% DDM), was prepared at 3 mg/ml and incubated with ~0.5 mM levan FOS with a degree
853 of polymerisation 15-25 for at least one hour before grid preparation. Grid type, preparation,
854 microscope and detector were the same as for the active levan utilisome described above.
855 Data acquisition parameters can be found in Supplementary Table 4.
856
857 A dataset comprising 1388 micrographs movies was collected and image processing was
858 carried out in RELION 3.1 [49]. Drift correction and dose-weighting were carried out using
859 MOTIONCOR2 [50]. Particle picking was performed using the general model of crYOLO and
860 yielded 157,953 particles[52]. Unwanted particles/contamination were removed from the
861 particle stack through two rounds of 2D classification, after which 146,056 particles remained.
862 Classification in 3D was used to address compositional heterogeneity. Good classes
863 representing the complete octameric utilisome and the hexameric assembly lacking one levan
864 binding protein and one levanase subunit were observed and contained 78,469 and 42,488
865 particles, respectively. Conformational heterogeneity in the position of the levan binding
866 protein was observed with some classes possessing a conformation where this subunit was
867 held close to the levanase. Particles contributing to all classes with evidence of this docked
868 conformation of the levan binding protein were pooled (98,755) and a consensus refinement
869 was carried out. CTF refinement and Bayesian polishing were performed iteratively until no
870 further improvement in resolution was observed. The resulting reconstruction possessed
871 weaker density for the C-terminal domain of the levan binding protein than for the N-terminal
872 portions. To improve this, a focused classification approach without alignment was used.
873

874 A mask encompassing only the docked position of the levan binding protein with some
875 surrounding density was created using the volume eraser tool in Chimera followed by mask
876 creation in RELION. A focused 3D classification job without alignment was run using the
877 aforementioned mask and the output from the aforementioned refinement as a reference
878 model. The reference model was low-pass filtered to 3.5 Å, just above the resolution of 3.3 Å
879 reported for the consensus refinement, thus allowing classification on high resolution
880 features. Several T values were empirically tested and a T value of 40 was found to give the
881 best results. A single class, containing 27,310 particles, was identified that possessed well
882 resolved density for the C-terminal domain of the levan binding protein. A particle star file
883 containing information for particles contributing to this class was created manually via
884 command line arguments. From this, two new star files were generated that contained
885 random half sets of the selected data. Using relion_reconstruct, these star files were used to
886 generate two independent half maps that corresponded to the unmasked structure. Post-
887 processing using these generated half maps yielded a sharpened reconstruction of 3.0 Å, as
888 estimated by gold standard Fourier Shell correlations using the 0.143 criterion. The density
889 for the C-terminal domain of the levan binding protein was improved, and density
890 corresponding to levan chain that links the levan binding protein to the levanase was also
891 visible.
892
893 To obtain the highest quality density for FOS molecules occupying the SusCD binding cavity,
894 all $SusC_2D_2$-containing particles were considered regardless of lipoprotein complement or
895 conformation. A particle subtraction and focused refinement strategy targeting the $SusC_2D_2$
896 core of the complex was used as described for the active levan utilisome. Post-processing of
897 the model arising from this final C2 symmetrised refinement resulted in a sharpened
898 reconstruction at 2.7 Å.
899
900 **Dextran utilisome cryo-EM sample preparation, data collection and image processing**. A
901 sample of the dextran utilisome complex (Bt3087-Bt3090) solubilised in a DDM-containing
902 buffer (10 mM HEPES, 100 mM NaCl, pH 7.5, 0.03 % DDM) was prepared at 0.05 mg/ml. Lacy
903 carbon 300-mesh copper grids coated with a <3 nm continuous carbon film (Agar Scientific)
904 were glow-discharged in air (10 mA, 30 s). A sample volume of 3.5 μL was applied to the grid.
905 Blotting and plunge-freezing were performed 10 seconds after loading the sample onto the

906 grid using an FEI Vitrobot Mark IV (Thermo Fisher Scientific) with chamber conditions set at a
907 temperature of 4 ° C and 100% relative humidity. The grid was blotted for 6 s with a blot
908 force of 0. Micrograph movies were collected on a Titan Krios Microscope (Thermo Fisher
909 Scientific) operating at 300 kV with a Falcon IV direct electron detector operating in counting
910 mode. Data acquisition parameters can be found in Supplementary Table 5.
911
912 A dataset comprising 6,331 micrographs was collected and image processing was carried out
913 in RELION 3.1 [49]. Drift correction and dose-weighting were performed using RELION's own
914 implementation of MOTIONCOR2. CTF estimation of motion corrected micrographs was done
915 using CTFFIND4 [54]. Particle picking was done using the crYOLO general model which identified
916 820,184 particles in the micrographs. Junk particles and contaminants were removed through
917 several rounds of 2D classification, after which 477,707 particles remained. An initial model
918 was generated *de novo* from the data. Extensive 3D classification was used to address the
919 considerable compositional heterogeneity that was present in the data (see ED Figure 5). Each
920 unique composition was refined and sharpened independently. A consensus refinement was
921 carried out, with iterative rounds of CTF-refinement and Bayesian polishing until no further
922 improvement in resolution was seen. A final, sharpened consensus reconstruction was
923 obtained at 3.1 Å.
924
925 **Model building into electron microscopy maps**. Buccaneer (part of CCP-EM v1.5.0)[55,56] was
926 used to build the initial protein models into the post-processed consensus inactive levanase
927 map, resulting in almost complete protein models. An AlphaFold 2[23] prediction of the dextran
928 SusC was generated and used as an initial model. Manual modelling and real space refinement
929 of protein and FOS chains were performed iteratively using COOT[46] and Phenix[48], respectively.
930 The completed protein and FOS models were placed into other maps and real space refined.
931 The acyl-cysteine was designated as the first residue of each lipoprotein. Model refinement
932 statistics are presented in Supplementary Tables 2-5.
933
934 **Density analysis and figure making**. Investigation and comparison of EM density maps was
935 performed using Chimera[53] and COOT[46]. Figures of maps and models were generated using
936 Chimera and ChimeraX[57]. To aid interpretability of EM density in generated figures, some

937 maps were filtered using LAFTER[27]. Maps processed in this way are clearly indicated in the
938 corresponding figure legend. The masks supplied in filtering were the same masks used for
939 post-processing within RELION.
940
941 **References**

942 1. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat.*
943 *Rev. Microbiol.* **19**, 55–71 (2021).
944 2. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes
945 the human gut microbiota. *Nat. Rev. Microbiol. 2012 105* **10**, 323–335 (2012).
946 3. Hamaker, B. R. & Tuncil, Y. E. A Perspective on the Complexity of Dietary Fiber
947 Structures and Their Potential Effect on the Gut Microbiota. *J. Mol. Biol.* **426**, 3838–
948 3850 (2014).
949 4. Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut
950 microbiota and their impact on human metabolism. *Gut Microbes* **7**, 189–200 (2016).
951 5. Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. From Dietary Fiber to
952 Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell* **165**, 1332–
953 1345 (2016).
954 6. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human
955 microbiome. *Nat. 2012 4867402* **486**, 207–214 (2012).
956 7. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex Glycan
957 Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. *J.*
958 *Biol. Chem.* **284**, 24673–24677 (2009).
959 8. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited.
960 *Microbiol. Mol. Biol. Rev.* **67**, 593–656 (2003).
961 9. Bolam, D. N. & van den Berg, B. TonB-dependent transport by the gut microbiota:
962 novel aspects of an old problem. *Current Opinion in Structural Biology* **51**, 35–43
963 (2018).
964 10. Glenwright, A. J. *et al.* Structural basis for nutrient acquisition by dominant members
965 of the human gut microbiota. *Nature* **541**, 407–411 (2017).
966 11. Madej, M. *et al.* Structural and functional insights into oligopeptide acquisition by the
967 RagAB transporter from Porphyromonas gingivalis. *Nat. Microbiol. 2020 58* **5**, 1016–
968 1025 (2020).

969 12. Gray, D. A. *et al.* Insights into SusCD-mediated glycan import by a prominent gut
970 symbiont. *Nat. Commun.* **12**, 1–14 (2021).
971 13. Terrapon, N. *et al.* PULDB: the expanded database of Polysaccharide Utilization Loci.
972 *Nucleic Acids Res.* **46**, D677–D683 (2018).
973 14. Sonnenburg, E. D. *et al.* Specificity of polysaccharide use in intestinal bacteroides
974 species determines diet-induced microbiota alterations. *Cell* **141**, 1241–1252 (2010).
975 15. Öner, E. T., Hernández, L. & Combie, J. Review of Levan polysaccharide: From a
976 century of past experiences to future prospects. *Biotechnology Advances* **34**, 827–844
977 (2016).
978 16. Mardo, K. *et al.* A Highly Active Endo-Levanase BT1760 of a Dominant Mammalian
979 Gut Commensal Bacteroides thetaiotaomicron Cleaves Not Only Various Bacterial
980 Levans, but Also Levan of Timothy Grass. *PLoS One* **12**, e0169989 (2017).
981 17. Bolam, D. N. & Sonnenburg, J. L. Mechanistic insight into polysaccharide use within
982 the intestinal microbiota. *Gut Microbes* **2**, (2011).
983 18. Reeves, A. R., Wang, G. R. & Salyers, A. A. Characterization of four outer membrane
984 proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. *J.*
985 *Bacteriol.* **179**, 643–649 (1997).
986 19. Cho, K. H. & Salyers, A. A. Biochemical analysis of interactions between outer
987 membrane proteins that contribute to starch utilization by Bacteroides
988 thetaiotaomicron. *J. Bacteriol.* **183**, 7224–7230 (2001).
989 20. Karunatilaka, K. S., Cameron, E. A., Martens, E. C., Koropatkin, N. M. & Biteen, J. S.
990 Superresolution imaging captures carbohydrate utilization dynamics in human gut
991 symbionts. *MBio* **5**, (2014).
992 21. Tuson, H. H., Foley, M. H., Koropatkin, N. M. & Biteen, J. S. The Starch Utilization
993 System Assembles around Stationary Starch-Binding Proteins. *Biophys. J.* **115**, 242–
994 250 (2018).
995 22. Schwanhüusser, B. *et al.* Global quantification of mammalian gene expression
996 control. *Nat. 2011 4737347* **473**, 337–342 (2011).
997 23. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nat.*
998 *2021 5967873* **596**, 583–589 (2021).
999 24. Tamura, K. & Brumer, H. Glycan utilization systems in the human gut microbiota: a
1000 gold mine for structural discoveries. *Curr. Opin. Struct. Biol.* **68**, 26–40 (2020).

1001 25. Nilaweera, T. D., Nyenhuis, D. A. & Cafiso, D. S. Structural intermediates observed
1002 only in intact Escherichia coli indicate a mechanism for TonB-dependent transport.
1003 *Elife* **10**, (2021).
1004 26. Zmyslowski, A. M., Baxa, M. C., Gagnon, I. A. & Sosnick, T. R. HDX-MS performed on
1005 BtuB in E. coli outer membranes delineates the luminal domain's allostery and
1006 unfolding upon B12 and TonB binding. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).
1007 27. Ramlaul, K., Palmer, C. M. & Aylett, C. H. S. A Local Agreement Filtering Algorithm for
1008 Transmission EM Reconstructions. *J. Struct. Biol.* **205**, 30–40 (2019).
1009 28. Ernits, K., Eek, P., Lukk, T., Visnapuu, T. & Alamäe, T. First crystal structure of an endo-
1010 levanase – the BT1760 from a human gut commensal Bacteroides thetaiotaomicron.
1011 *Sci. Reports 2019 91* **9**, 1–13 (2019).
1012 29. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide
1013 utilization loci: Fueling microbial communities. *J. Bacteriol.* **199**, (2017).
1014 30. Rassam, P. *et al.* Supramolecular assemblies underpin turnover of outer membrane
1015 proteins in bacteria. *Nat. 2015 5237560* **523**, 333–336 (2015).
1016 31. Benn, G. *et al.* Phase separation in the outer membrane of Escherichia coli. *Proc. Natl.*
1017 *Acad. Sci. U. S. A.* **118**, (2021).
1018 32. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch Catabolism by a
1019 Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices.
1020 *Structure* **16**, 1105–1115 (2008).
1021 33. Zougman, A., Selby, P. J. & Banks, R. E. Suspension trapping (STrap) sample
1022 preparation method for bottom-up proteomics analysis. *Proteomics* **14**, 1006–1000
1023 (2014).
1024 34. HaileMariam, M. *et al.* S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun
1025 Proteomics. *J. Proteome Res.* **17**, 2917–2924 (2018).
1026 35. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized
1027 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat.*
1028 *Biotechnol. 2008 2612* **26**, 1367–1372 (2008).
1029 36. Cox, J. *et al.* Andromeda: A peptide search engine integrated into the MaxQuant
1030 environment. *J. Proteome Res.* **10**, 1794–1805 (2011).

1031 37. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers
1032 differential expression analyses for RNA-sequencing and microarray studies. Nucleic
1033 Acids Res 43: e47
1034 38. Winter, G. *et al.* DIALS: Implementation and evaluation of a new integration package.
1035 *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 85–97 (2018).
1036 39. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction.
1037 *J. Appl. Crystallogr.* **43**, 186–190 (2010).
1038 40. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution?
1039 *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
1040 41. Evans, P. Scaling and assessment of data quality. in *Acta Crystallographica Section D:*
1041 *Biological Crystallography* **62**, 72–82 (International Union of Crystallography, 2006).
1042 42. Skubak, P. *et al.* A new MR-SAD algorithm for the automatic building of protein
1043 models from low-resolution X-ray data and a poor starting model. *urn:issn:2052-2525*
1044 **5**, 166–171 (2018).
1045 43. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125 (2010).
1046 44. Krissinel, E., Uski, V., Lebedev, A., Winn, M. & Ballard, C. Distributed computing for
1047 macromolecular crystallography. *Acta Crystallogr. Sect. D, Struct. Biol.* **74**, 143–151
1048 (2018).
1049 45. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and
1050 guidelines for its use. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2184–2195 (2004).
1051 46. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of
1052 Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
1053 47. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta*
1054 *Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235 (2011).
1055 48. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons
1056 and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D, Struct. Biol.*
1057 **75**, 861–877 (2019).
1058 49. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure
1059 determination in RELION-3. *Elife* **7**, (2018).
1060 50. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for
1061 improved cryo-electron microscopy. *Nat. Methods 2017 144* **14**, 331–332 (2017).
1062 51. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12

(2016).

52. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol. 2019 21* **2**, 1–13 (2019).

53. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–12 (2004).

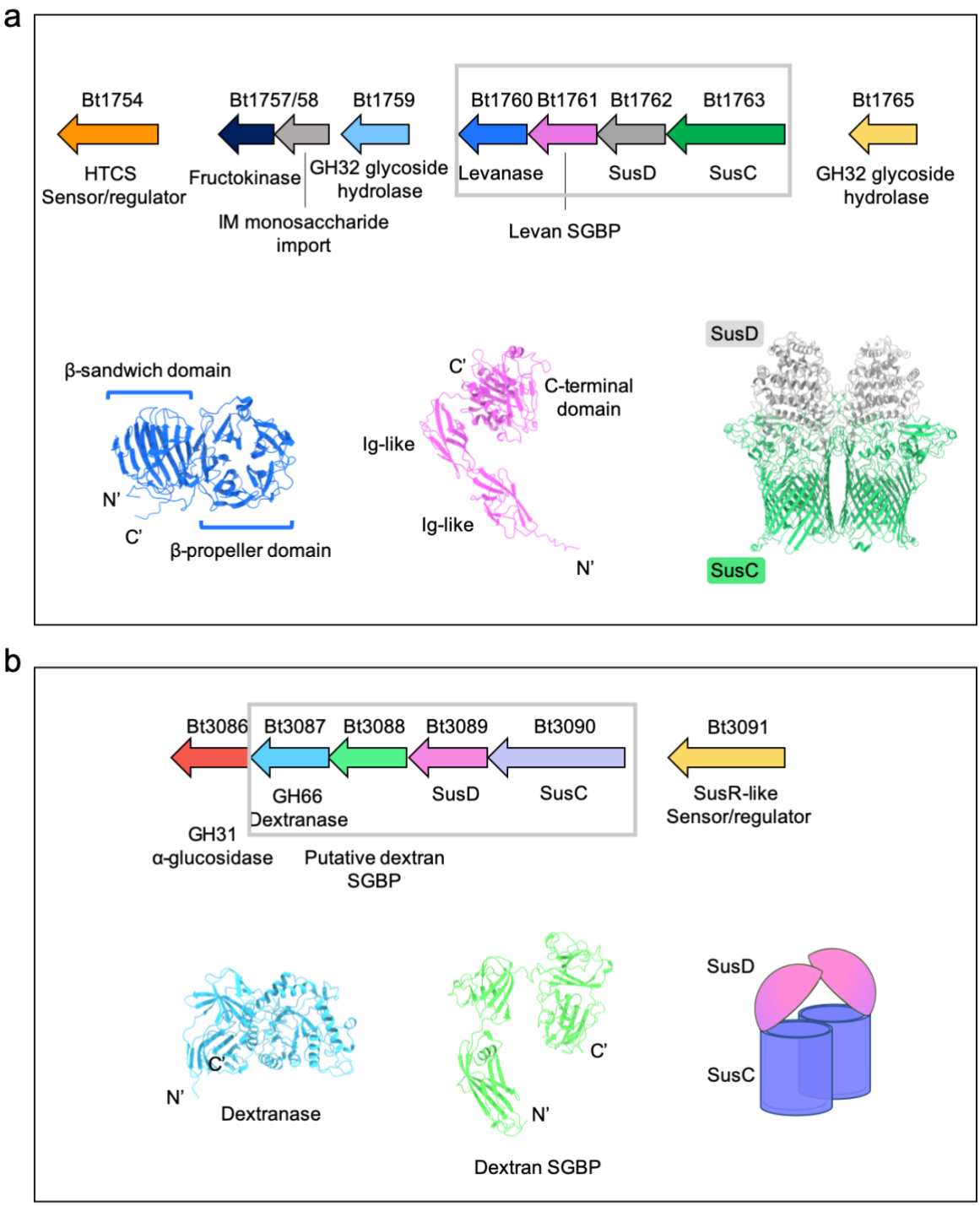
54. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

55. Hoh, S. W., Burnley, T. & Cowtan, K. Current approaches for automated model building into cryo-EM maps using Buccaneer with CCP-EM. *Acta Crystallogr. Sect. D, Struct. Biol.* **76**, 531–541 (2020).

56. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr. Sect. D, Struct. Biol.* **73**, 469–477 (2017).

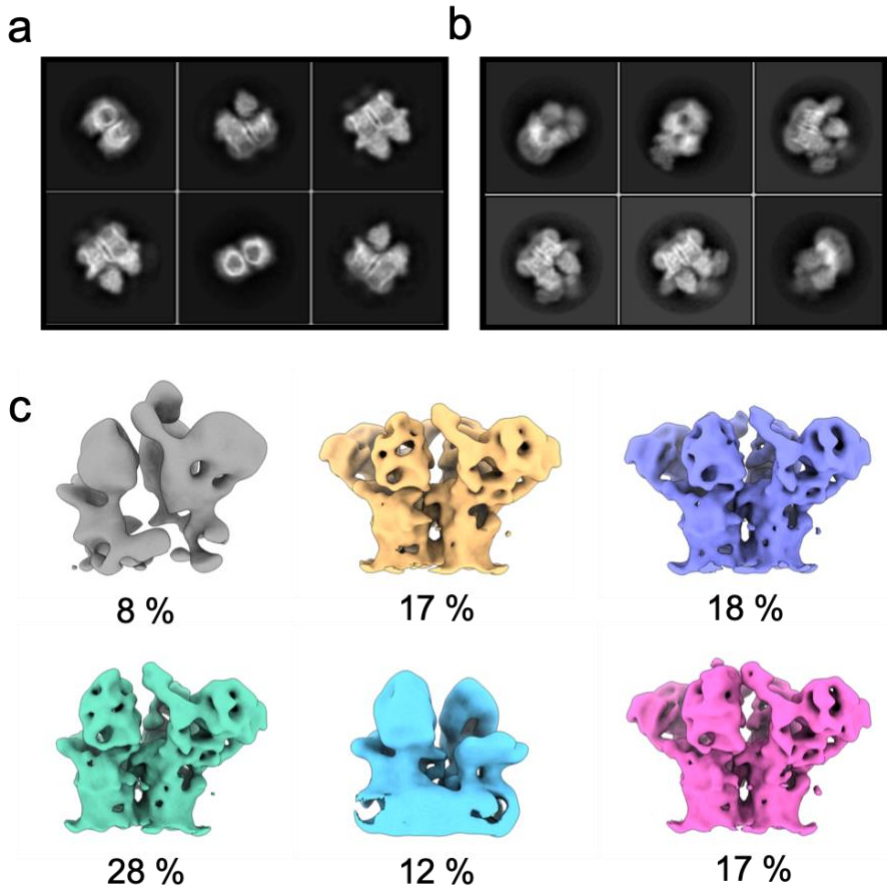
57. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

Extended Data

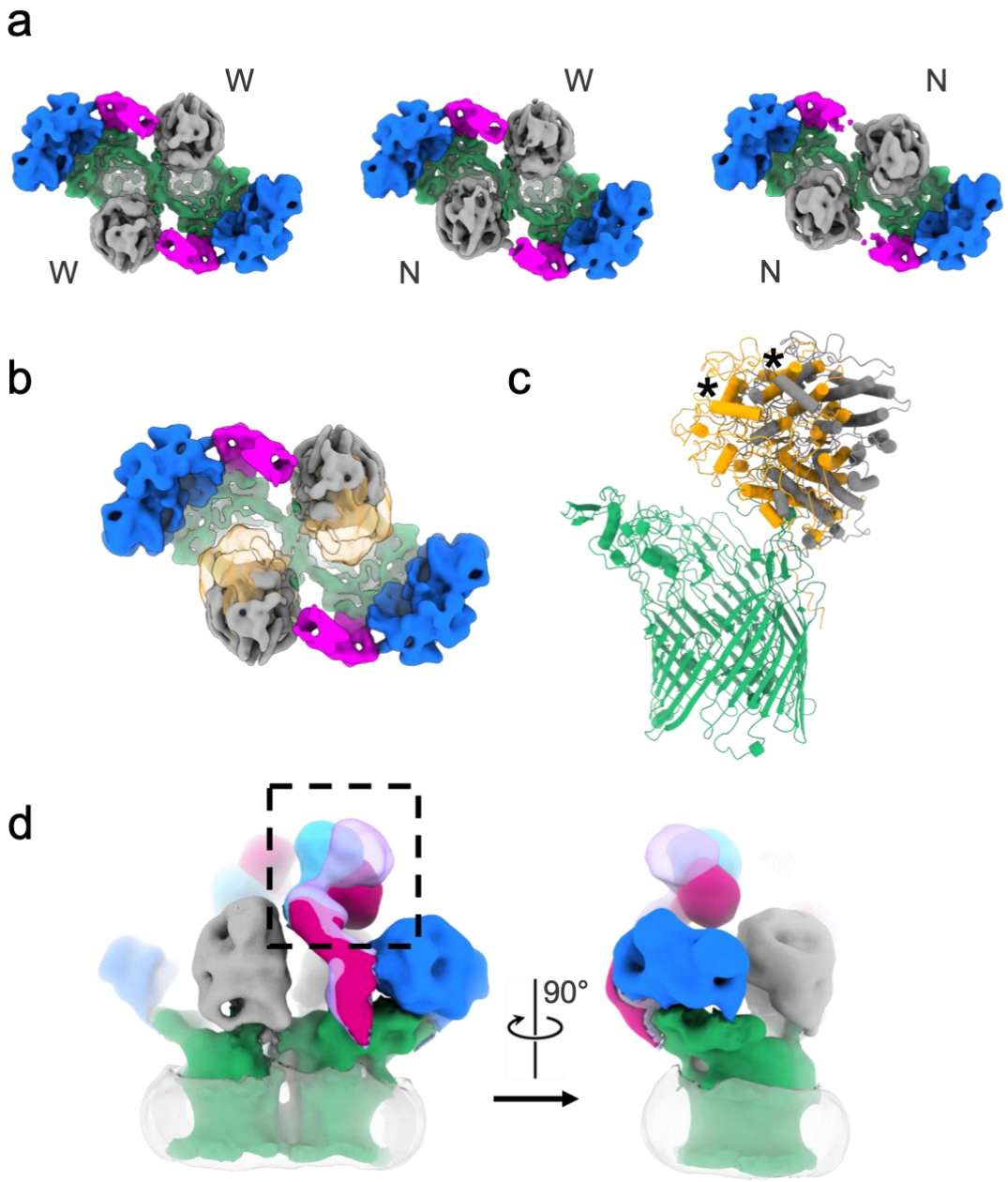


Extended Data Figure 1. Arrangement of levan and dextran PULs with known and predicted structural information for OM-associated components. a, Organisation of the levan PUL showing gene positions within the locus with functions labelled. OM-associated PUL components are within the grey box. An X-ray crystal structure of the levanase Bt1760 is shown (blue) with its β-sandwich and catalytic β-propeller domains clearly identifiable (PDB ID 6R3R). The AlphaFold2-predicted model for Bt1761 is shown (pink) oriented such that the N-terminus is at the bottom and the globular C-terminal domain hypothesised to bind levan is at the top. Note that the N-termini of the additional lipoproteins will be lipidated and associated with the OM. The cryo-EM structure of the dimeric SusCD complex in the open-open state is shown with SusC components in green

and SusD components in grey. **b**, Organisation of the dextran PUL showing gene positions within the locus with functions labelled. OM-associated PUL components are boxed in grey. AlphaFold2-predicted models for the dextranase (Bt3087) and putative dextran SGBP (Bt3088) are shown in blue and green, respectively.

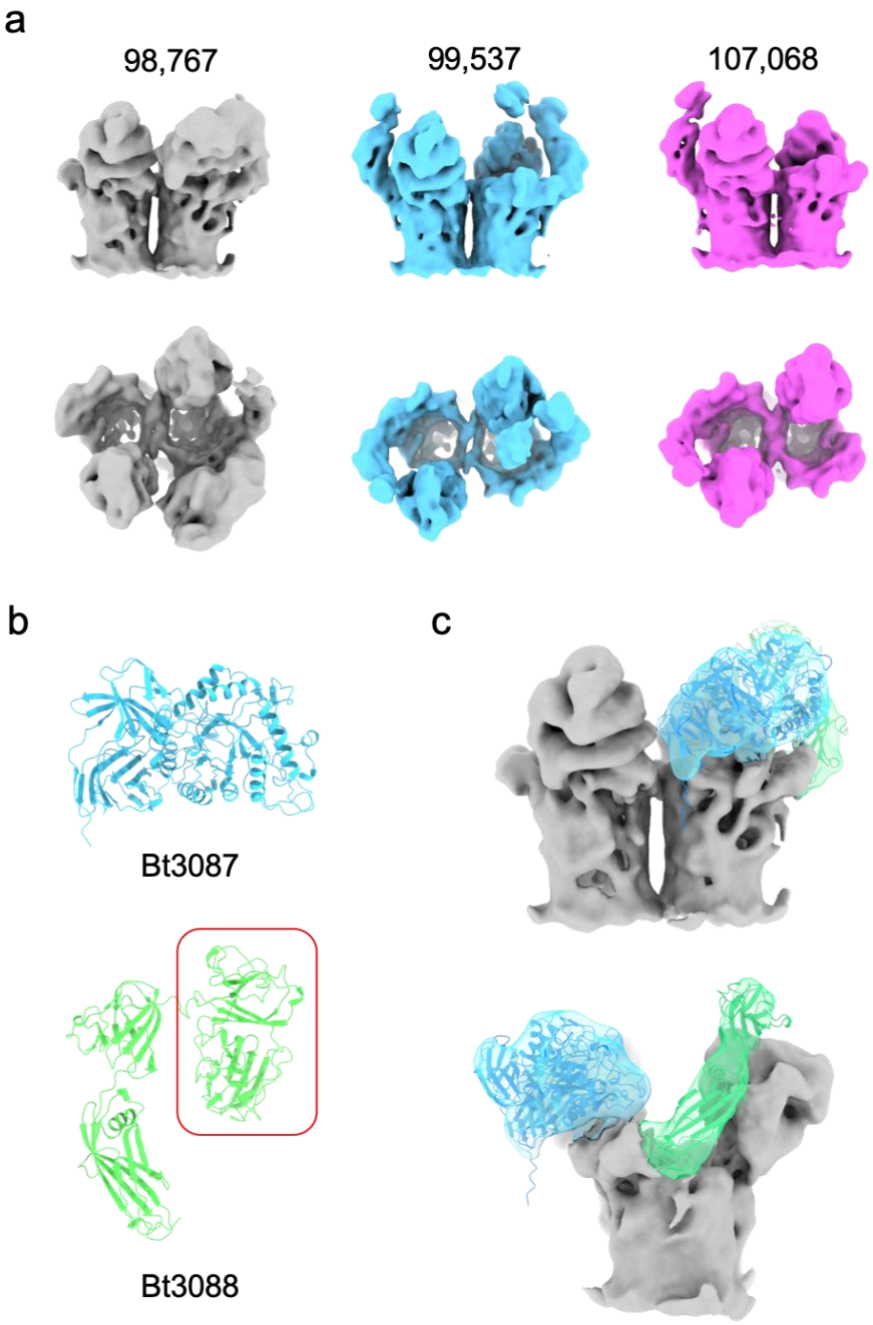


Extended Data Figure 2. Early cryo-EM processing outputs of the apo levan $SusC_2D_2$ core complex and apo levan utilisome. a, Example 2D class averages of the LDAO-solubilised levan core complex from our previous study. **b**, Example 2D class averages of the levan utilisome solubilised in DDM, and with the HIS-tag moved from SusD to the levan SGBP. Density in addition to that of the core complex (**a**) can be seen in all views. **c**, Output of the first round of 3D classification. Yellow, purple and pink classes represent the octameric complex i.e. the complete utilisome. The green class shows the additional lipoproteins associated with just one SusC unit whilst the blue class shows that a small proportion of $SusC_2D_2$ complexes were present in isolation.



Extended Data Figure 3. Conformational variability of the substrate-free levan utilisome. a, Refined outputs from 3D classification of the octameric utilisome viewed from the extracellular space. Bt1763 SusC (green), Bt1762 SusD (grey), Bt1761 levan SGBP (magenta), Bt1760 levanase (blue). Classes are separated on the basis of SusD lid position. Wide-wide, normal-wide, and normal-normal open states are presented from left to right with wide and normal positions labelled 'W' and 'N', respectively. **b**, Overlay of the wide-wide (grey) and normal-normal (orange) open states of the complex. **c**, Overlay of models for the normal open versus wide open state of the transporter generated by a rigid-body fit of SusD into the density of the respective maps. A monomer is shown for clarity and an asterisk marks the same SusD helix in both models. **d**, A view of the utilisome shown at high threshold levels in the plane of the membrane (left). Different conformations of the levan binding protein observed in 3D classification are overlaid to demonstrate the flexibility of this subunit (boxed region). The same view rotated 90° is shown (right). Disordered micelle density is shown as translucent grey.

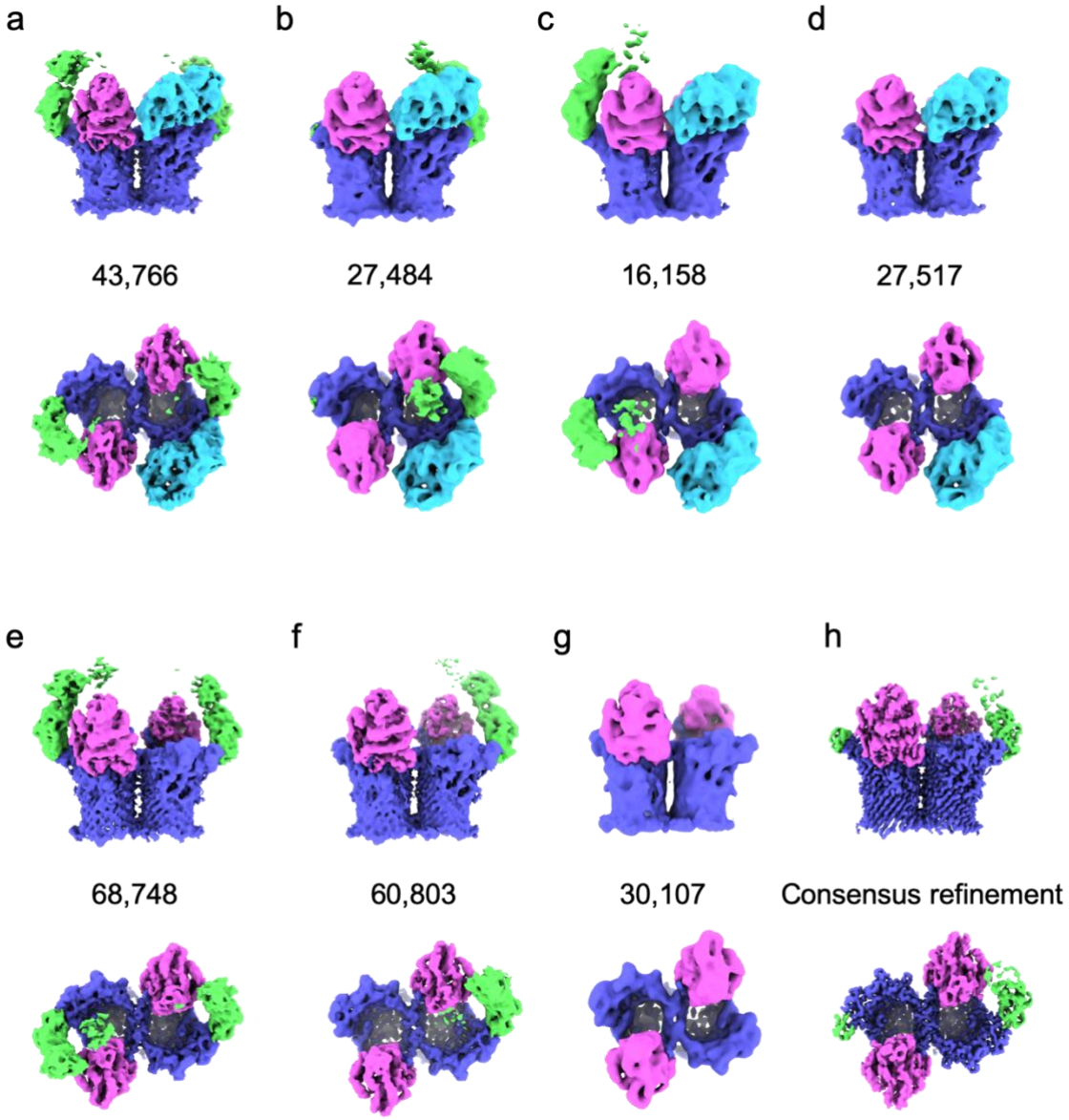
1120



Extended Data Figure 4. Initial outputs from 3D classification of the dextran utilisome and assignment of lipoprotein density. a, Good 3D classes with clear density for the core transporter viewed in the plane of the membrane (top) and from the extracellular space (bottom). Unassigned density was also present in all classes. Particle numbers contributing to each class are displayed. **b**, AlphaFold2 predictions of the dextranase (Bt3087; blue) and putative dextran SGBP (Bt3088; green) of the dextran utilisome. **c**, Grey class shown in (**a**) with additional densities assigned to the dextranase and putative SGBP on the basis of predicted structures. Note that the C-terminal domain of the SGBP (red box) is not visible in the EM map and the docked model for the SGBP has been truncated accordingly.

1121

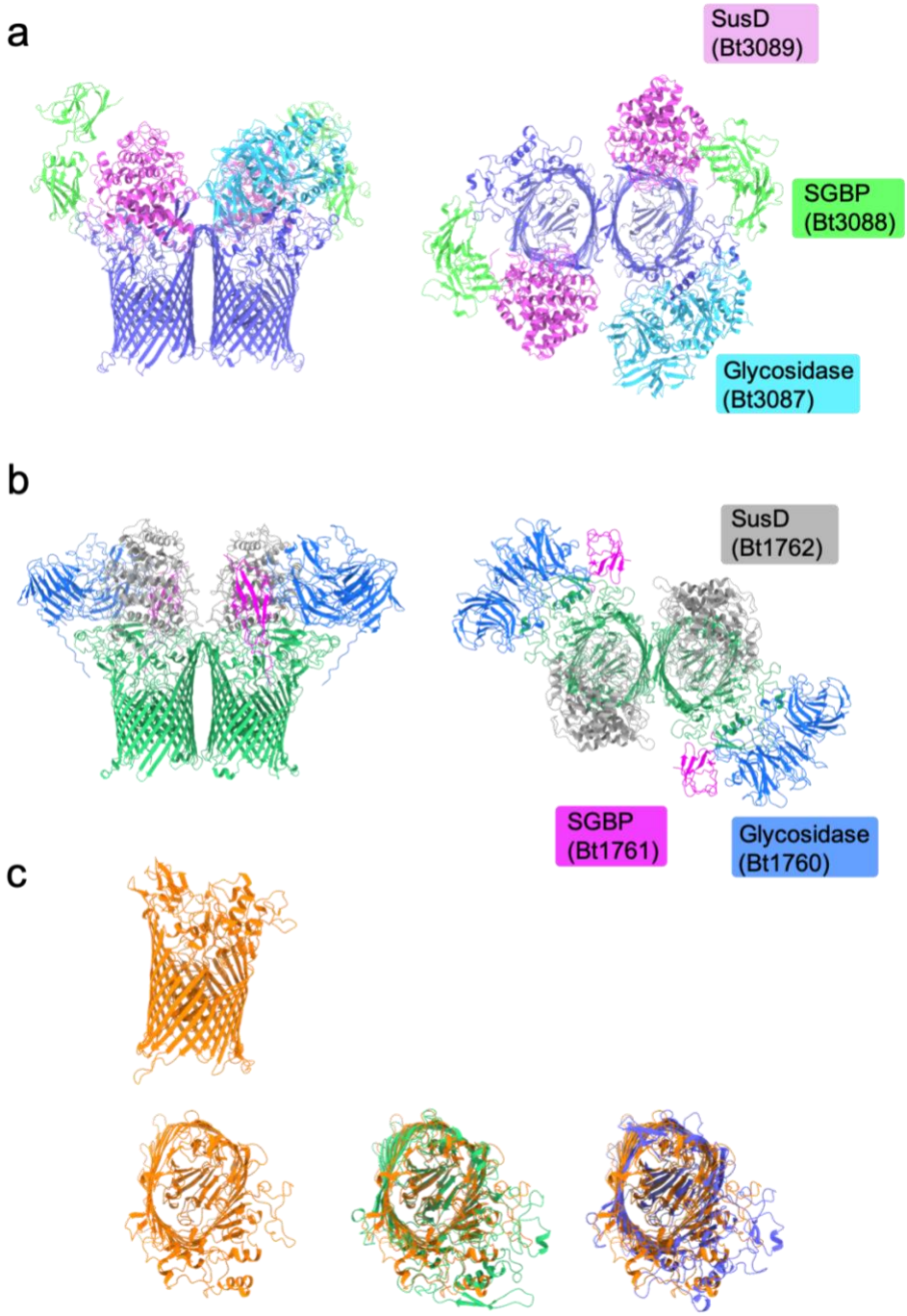
1122



Extended Data Figure 5. Compositional heterogeneity of the dextran utilisome observed by cryo-EM. a-g, Refined outputs of 3D classification viewed in the plain of the membrane (top panels) and from the extracellular space (bottom panels), where each map corresponds to a unique complement or arrangement of auxiliary components. SusC (Bt3090) is purple, SusD (Bt3089) is pink, the dextran SGBP (Bt3088) is green and the dextranase (Bt3087) is cyan. **h**, A consensus refinement at a global resolution of 3.1 Å.

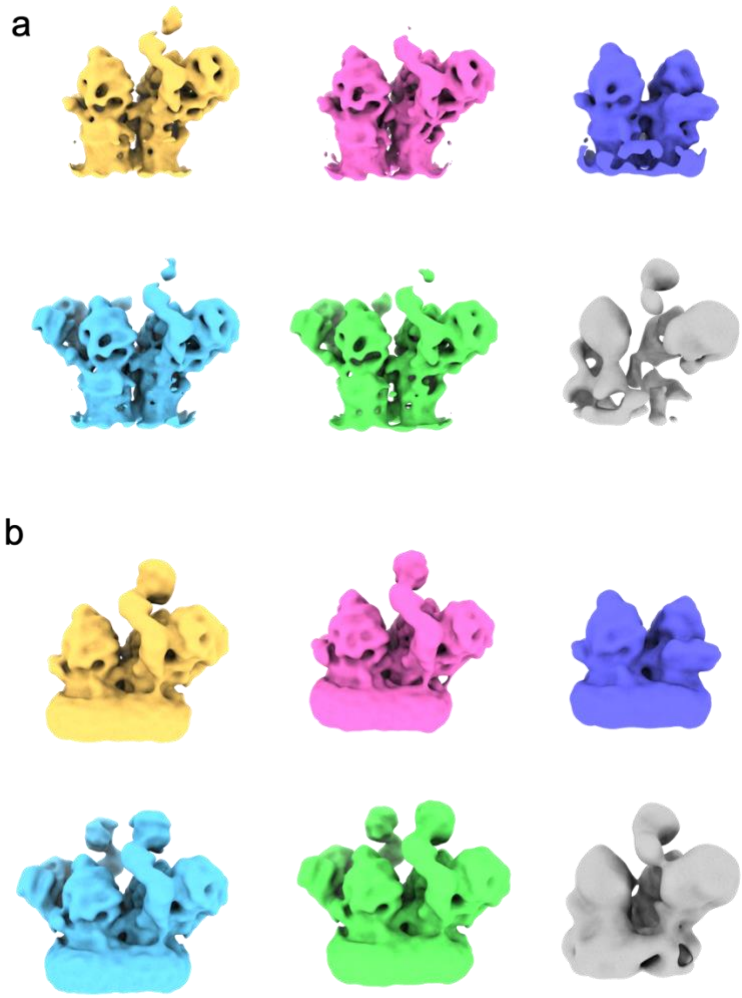
1123
1124

1125



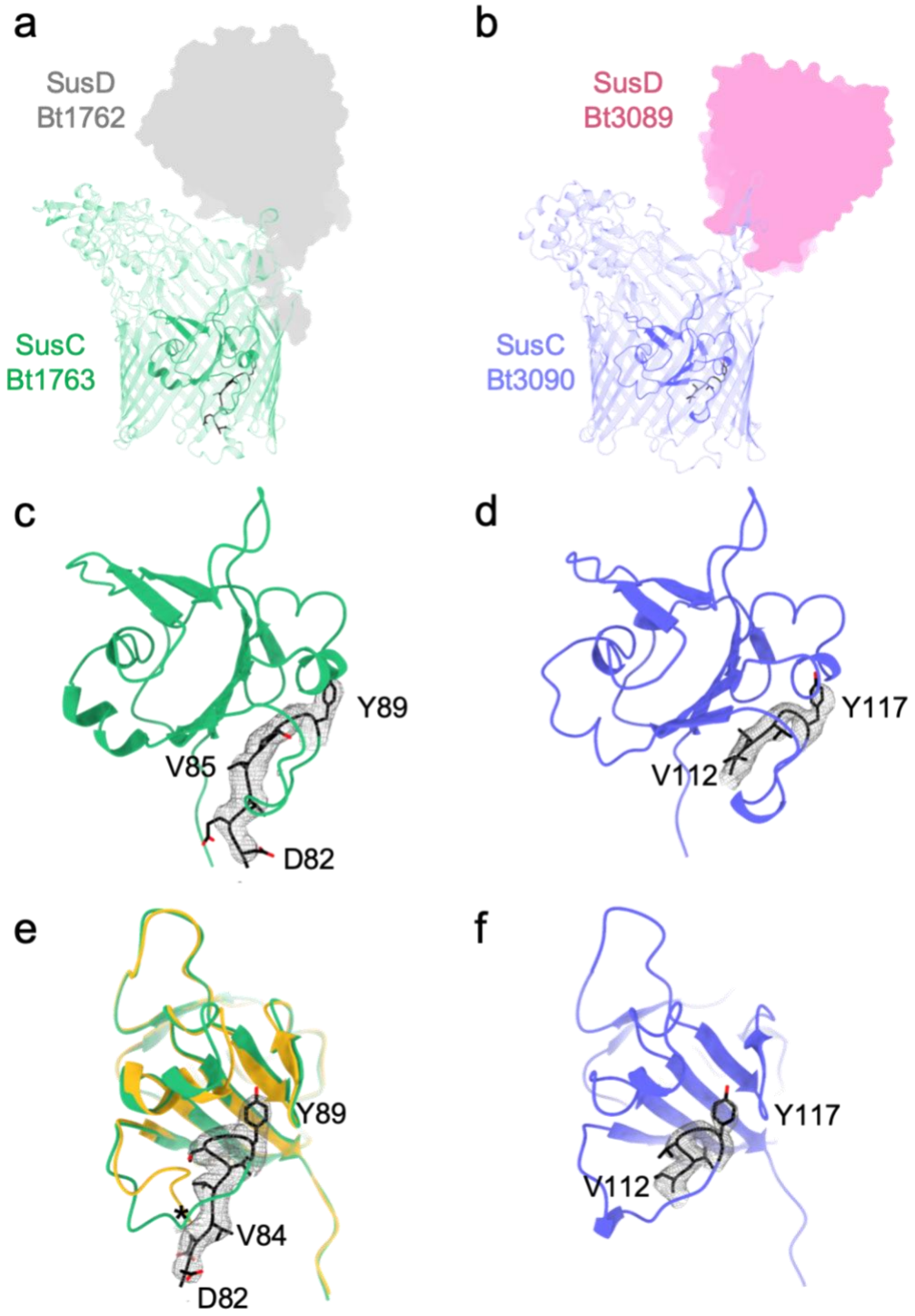
Extended Data Figure 6: Arrangement of auxiliary proteins in apo glycan utilisomes. a, Composite model demonstrating the arrangement of the dextran utilisome. Cryo-EM data permitted refinement of the SusC components. AlphaFold2 structure predictions for SusD (Bt3089) and the dextranase (Bt3087) were obtained and docked into the cryo-EM map for the heptameric complex. An AlphaFold2 structure prediction for part of the SGBP (Bt3088) was also rigid fit to the cryo-EM density. Unambiguous density is visible only for the first two domains of the SGBP. The predicted model was therefore truncated prior to the C-terminal domain. **b**, Structure of the substrate-free levan utilisome. SusD components (Bt1762) are positioned based on a rigid body fit to a low-resolution map in which unambiguous density for the lids was observed. Note the different arrangement of the glycosidase and SGBP components relative to SusD in the levan and dextran systems. **c**, AlphaFold2 structure prediction of Bt3702 SusC from the starch utilisation system (orange) viewed in the plane of the membrane (top) and from the extracellular space (bottom). Alignments of the predicted structure with experimentally determined structures for SusC components of the levan (green) and dextran (violet) systems are shown.

1126

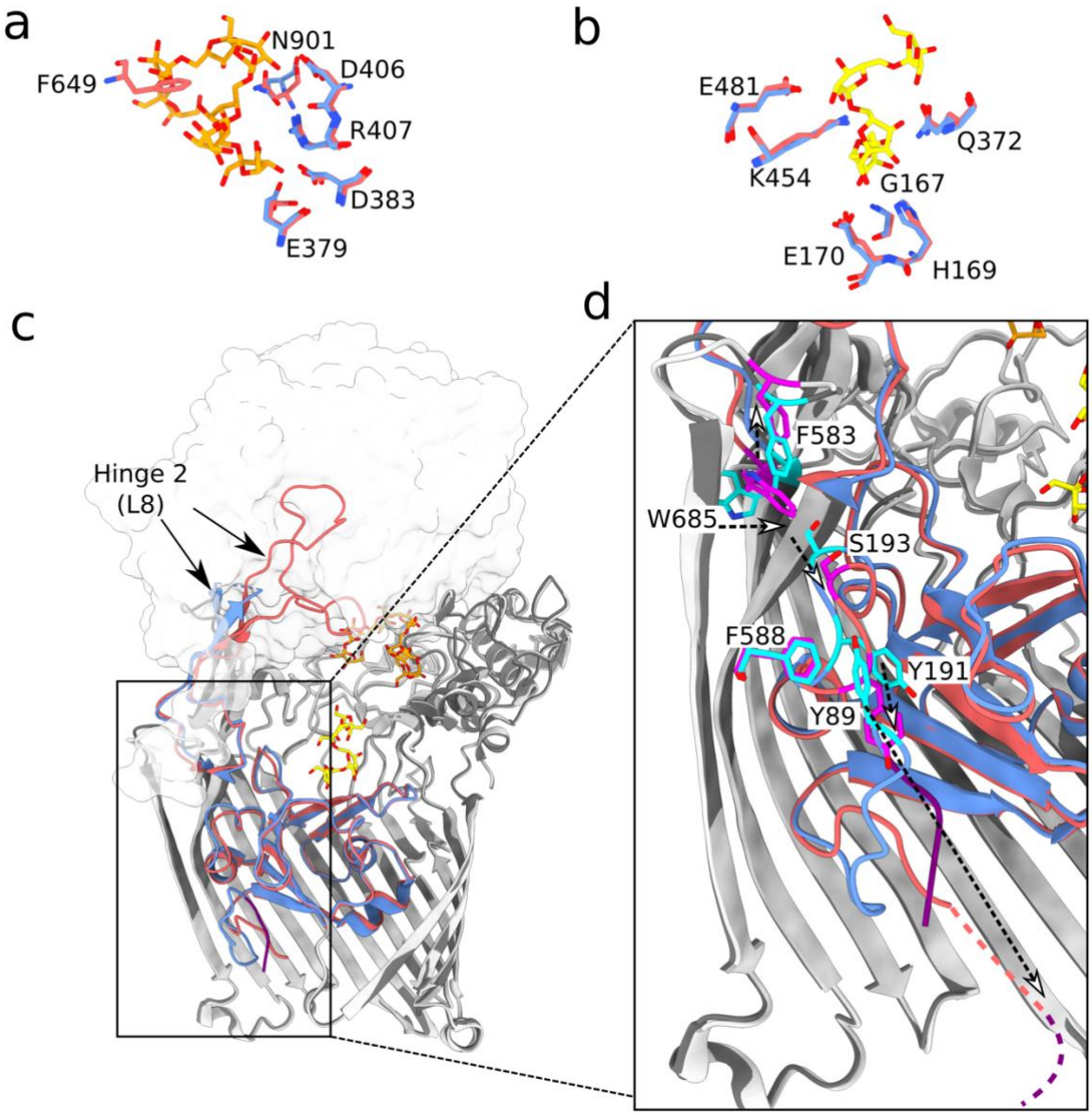


Extended Data Figure 7. Output of 3D classification for the levan utilisome with an active levanase in the presence of FOS DP8-12. a, Classes viewed in the plane of the membrane. Classes containing particles of the complete octameric complex were observed (blue and green) as well as hexameric complexes containing just a single copy of the LBP and glycosidase (pink and yellow). A class containing $SusC_2D_2$ in isolation is also present (purple). **b**, The same classes as in (**a**) viewed at an increased contour level.

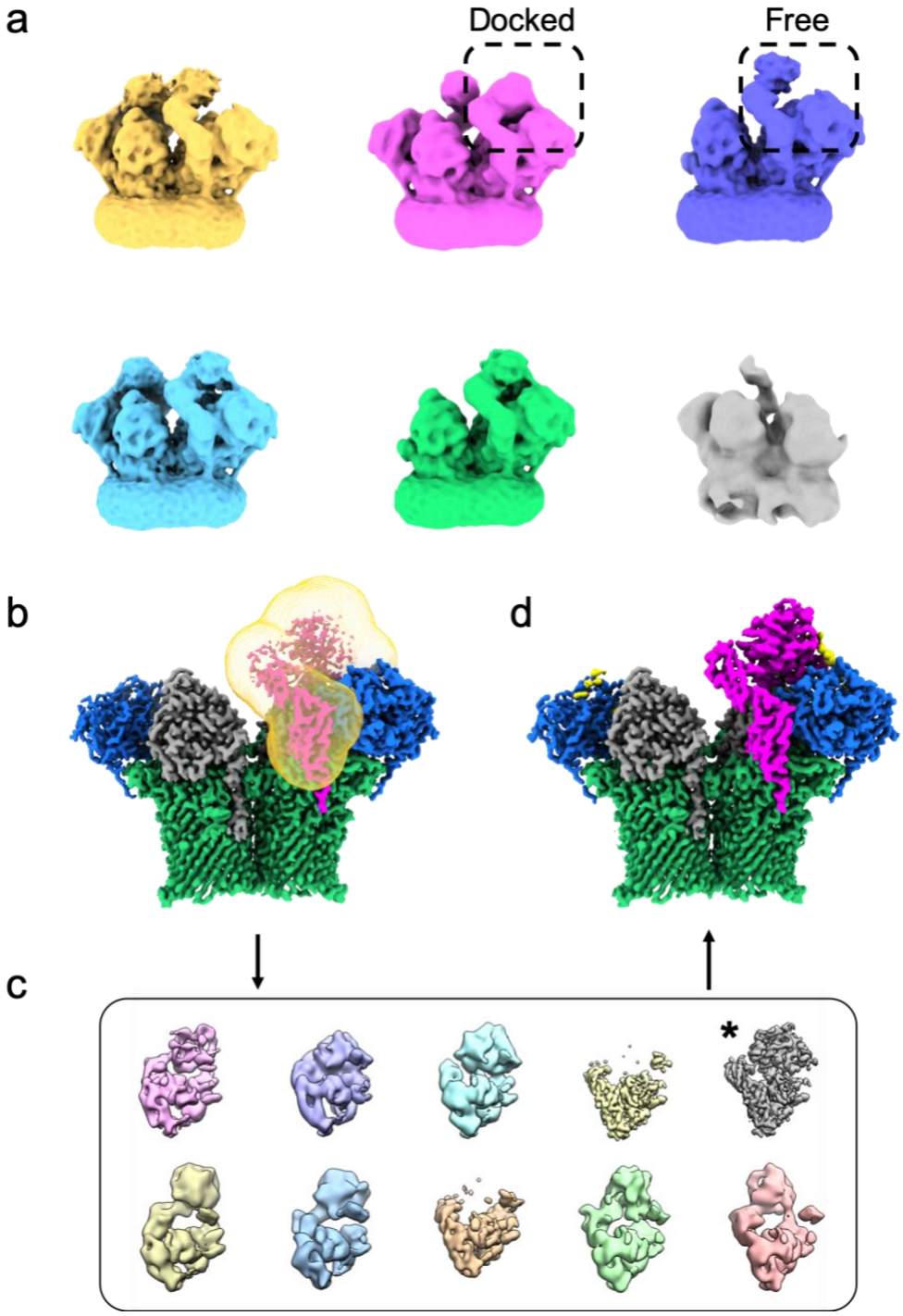
1127



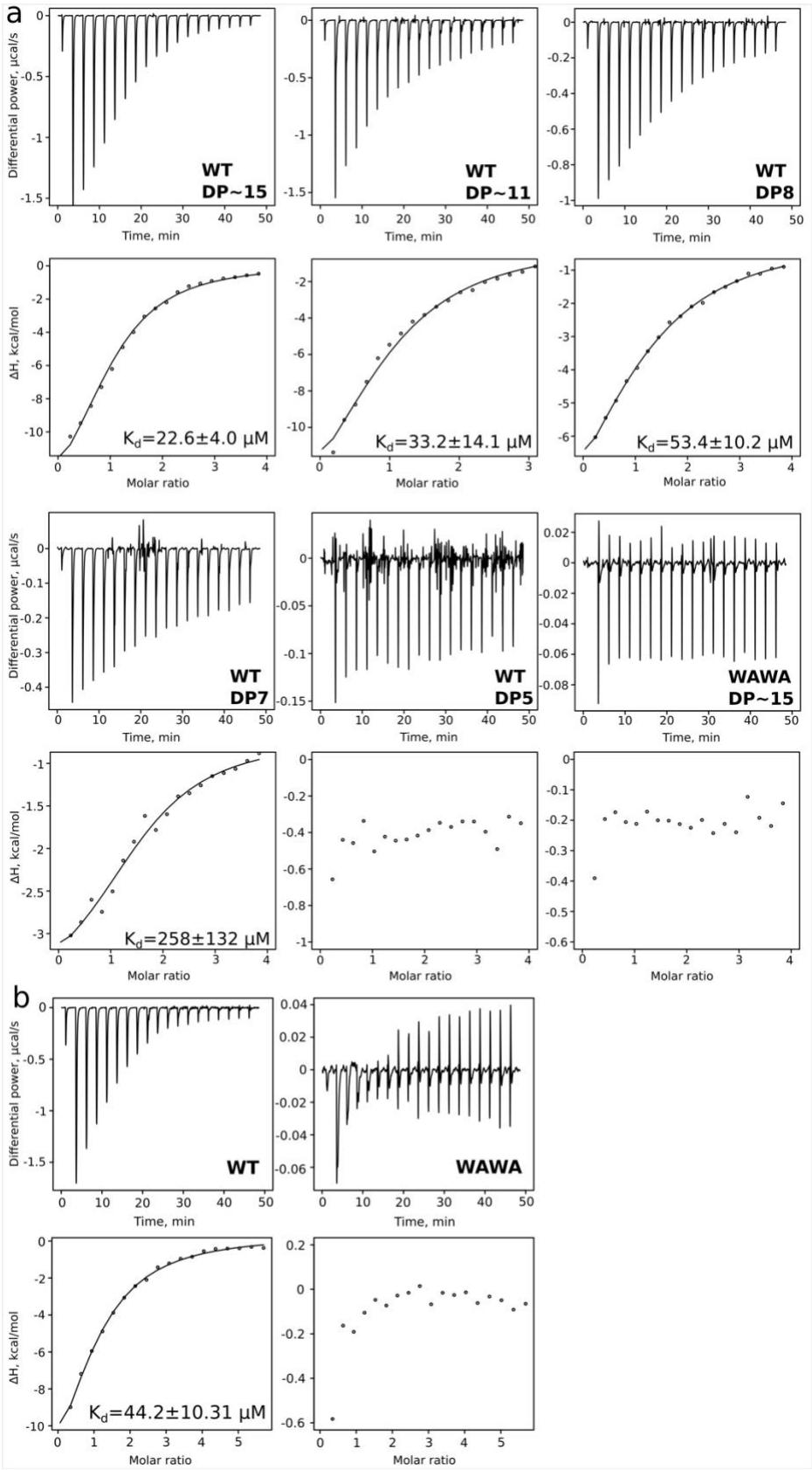
Extended Data Figure 8. Analysis of the TonB box region in cryo-EM structures of the levan and dextran transporters. a, Structure of SusC of the levan transporter (Bt1763) in the absence of substrate. **b**, Structure of SusC of the dextran transporter (Bt3090) in the absence of substrate. Ton box regions are shown in black. The corresponding SusD components are shown as shaded zones; both occupy an open position. Isolated plug domains for each transporter can be seen in **c** (levan) and **d** (dextran). Density is displayed for the Ton box regions (residues in black stick form). **e**, Superposition of plug domains for the apo (green) and substrate-bound (yellow) structures of the Bt1763; rotated 90° relative to (**c**). The visible N-terminus is D82 for the substrate-free transporter compared with R93 for the substrate-bound transporter (marked with an asterisk). **f**, A view of the Bt3090 plug from the same angle as (**e**).



Extended Data Figure 9. Subtle conformational changes upon substrate binding to Bt1763 SusC result in exposure of the Ton box. a and **b**, Overlay of apo (blue) and substrate-bound (salmon) Bt1763 FOS binding residues in the top (a) and bottom (b) binding sites. G167, H169 and E170 are part of the plug domain. **c**, Overlay of apo (dark grey) and FOS-bound (light grey) Bt1763 cryo-EM structures. Hinge 2 (L8[12])and the plug domain in the apo and substrate-bound structures are coloured in blue and salmon, respectively. β-strands 13-22 of the SusC barrel are hidden. The closed SusD lid is represented as a transparent surface. **d**, a ~90˚ rotated close-up view of the inset in **c**. Residues involved in transmitting substrate binding information from the extracellular side of the outer membrane to the periplasmic side are coloured in cyan and magenta in the apo and substrate-bound structures, respectively. The dashed white-headed arrows indicate residue movement from the apo to the substrate-bound state. The Ton box in the apo structure is shown in solid purple. The dashed line indicates the possible location of unresolved residues preceding R93 in the substrate-bound structure based on weak density observed in cryo-EM maps (Ton box in purple, residues 89-92 in salmon).

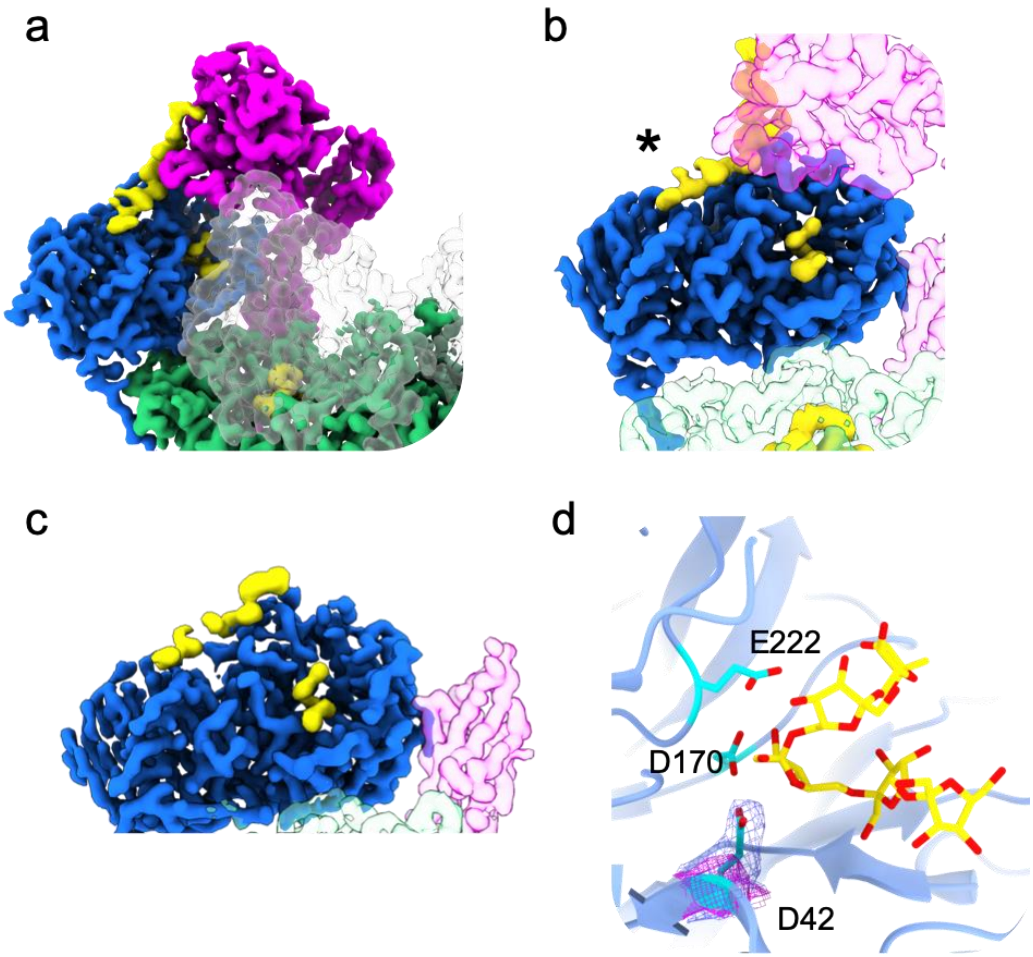


Extended Data Figure 10. Addition of FOS DP ~15-25 FOS to levan utilisomes containing an inactive levanase allows structure determination of the levan SGBP. a, Outputs of 3D classification showing that the levan SGBP can adopt a 'docked' conformation proximal to both the SusD and levanase. **b**, A consensus refinement of all classes containing at least one docked SGBP (yellow, pink, cyan and green). A mask was created around the region of interest (transparent yellow). **c**, Outputs of focused classification on the masked region without alignment. A class displaying high resolution for the region of interest is marked with an asterisk. Independent half maps were reconstructed using particles belonging to this class. **d**, Sharpened reconstruction generated with the aforementioned half maps showing improved density for the levan SGBP. Note that the handedness of maps displayed here is incorrect. This was addressed prior to model building.

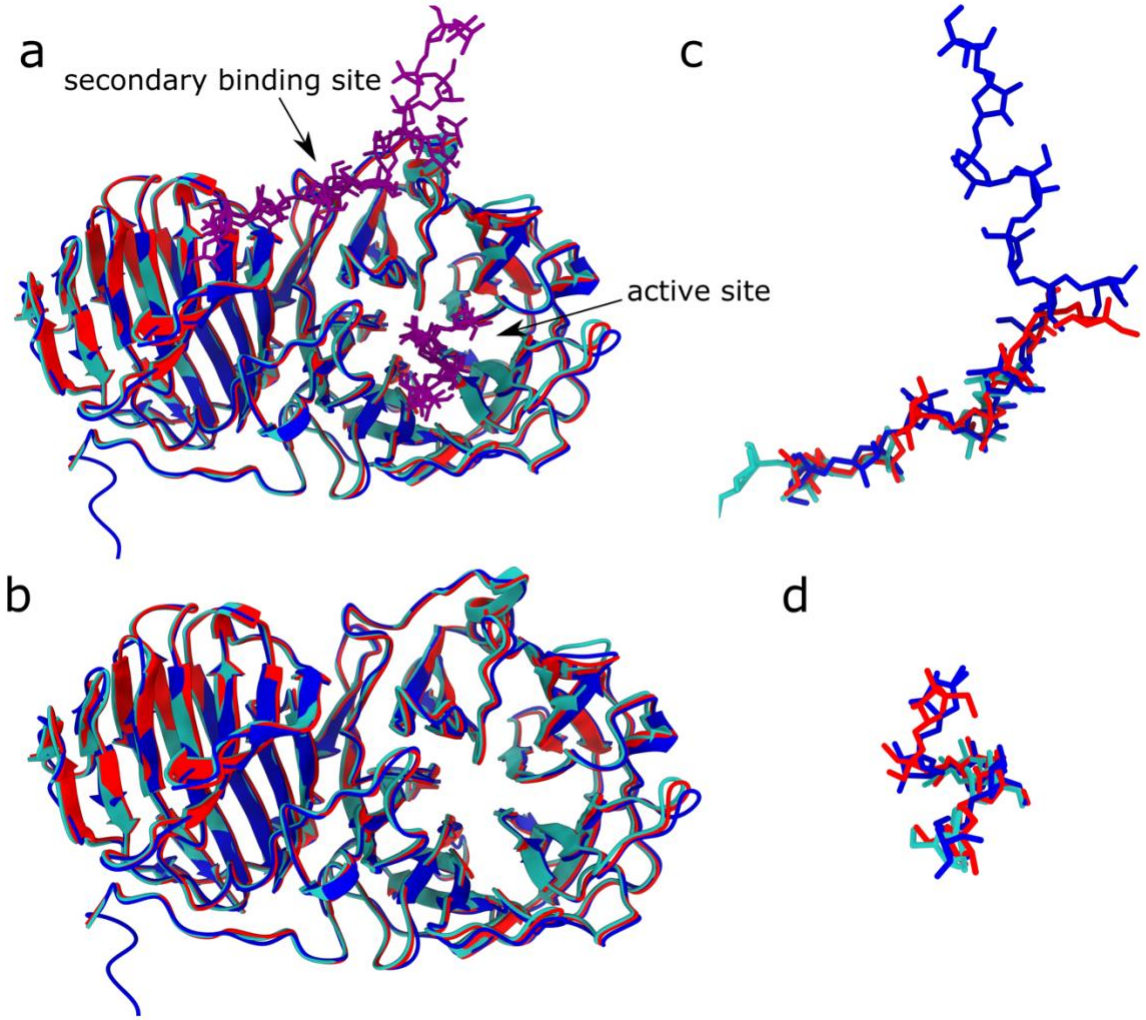


Extended Data Figure 11. ITC with recombinant Bt1761 levan SGBP demonstrates importance of Trp297 and Trp359 for levan binding. a, Titration of 1 mM defined-length FOS into 50 µM wild type or W297A/W359A double-mutant (WAWA) Bt1761. The traces with FOS DP~15 are the same ones shown in Figure 5e, and are repeated here for comparison with FOS of different sizes. The minimal FOS size for reasonably strong binding is

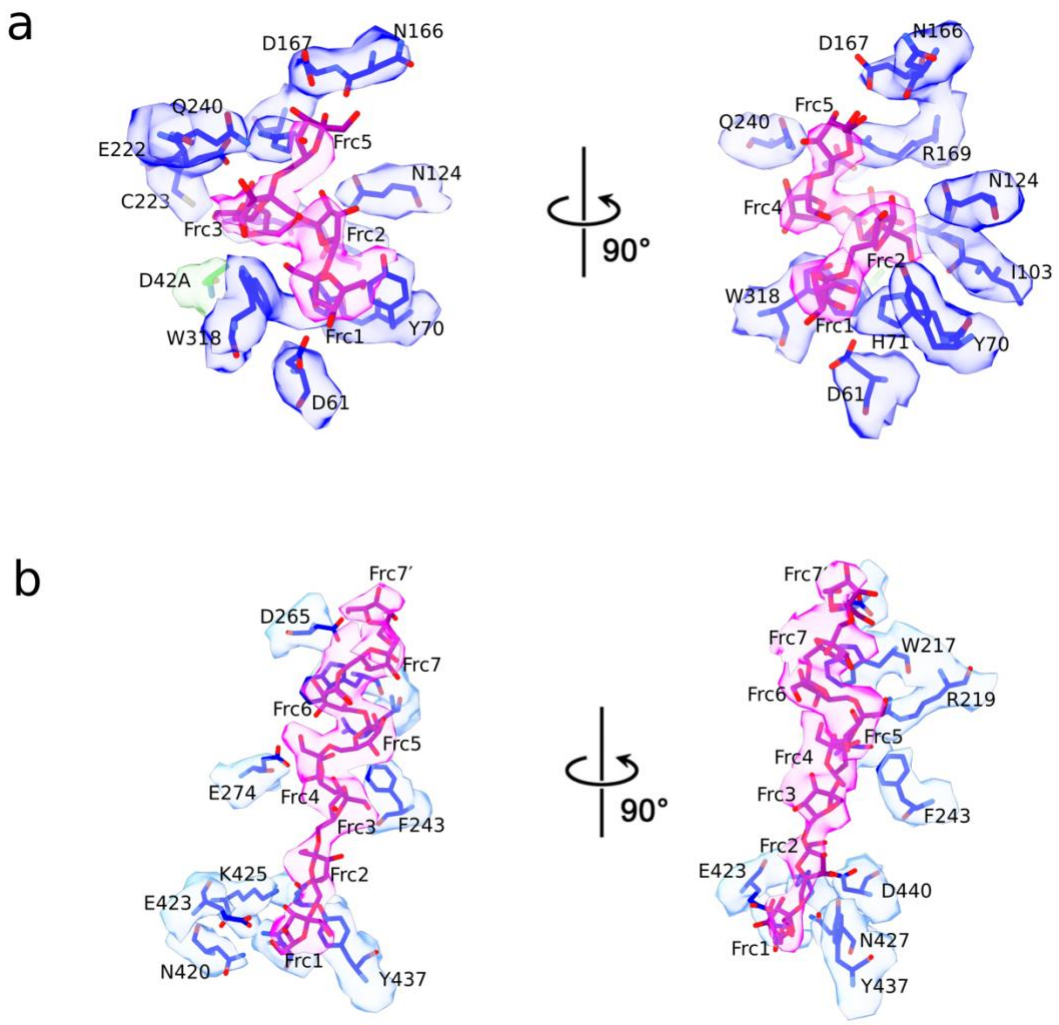
1146 DP 8. **b,** Titration of 8 mg/ml levan into 25 μM Bt1761. The affinity estimates, where possible, were obtained by
1147 fitting the data to a single binding site model (Methods).
1148



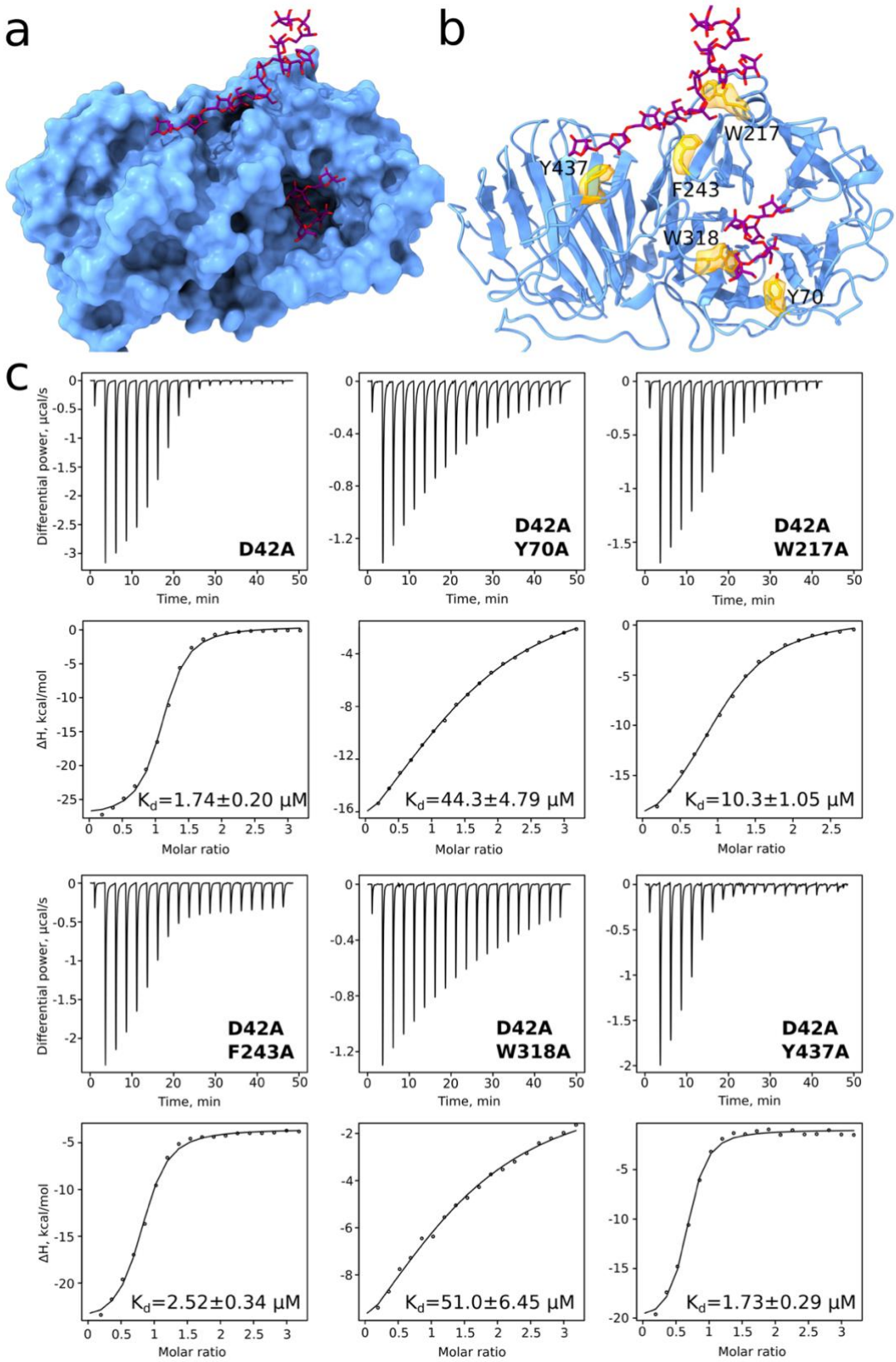
Extended Data Figure 12. FOS binds to two distinct sites on the levanase. a, Zoomed view of one half of the levan utilisome highlighting the tethered position of the levan SGBP relative to the levanase, the SusD subunit (translucent density) and the top of the SusC barrel (green). **b**, Face on view of the levanase showing the distinct densities for levan (yellow) bound at the secondary site (marked with an asterisk) and at the active site. No density linking these two sites was observed. In both (**a**) and (**b**) FOS bound at the upper site in SusC is observed. The minimum distance between FOS at the active site of the levanase and in the binding cavity of SusC is ~30 Å. **c**, View of the levanase associated with the alternate SusC unit where no tethering of the SGBP is observed but density for levan is still evident, bound across the top of the β-propeller and β-sandwich domains of the levanase. **d**, Zoomed view showing levan FOS modelled into the active site of the inactive levanase. Residues comprising the catalytic triad[28] are shown as stick models. The inactive levanase possesses a D42A mutation which is clearly resolved when overlaying density from the active (blue) and inactivated (pink) levanase reconstructions.



Extended Data Figure 13. FOS-bound levanase structures determined by cryo-EM and X-ray crystallography are very similar. a, b, Cartoon representation of structures aligned using ChimeraX Matchmaker: the cryo-EM structure from the inactive levanase complex (blue), and two crystal structures of the levanase alone (red, cyan; 7ZNR and 7ZNS), with the FOS from all structures coloured in purple (not shown in **b**). **c** and **d**, Alignment of FOS in the secondary binding site and the active site, respectively. Colours are the same as the protein chains in **a** and **b.**

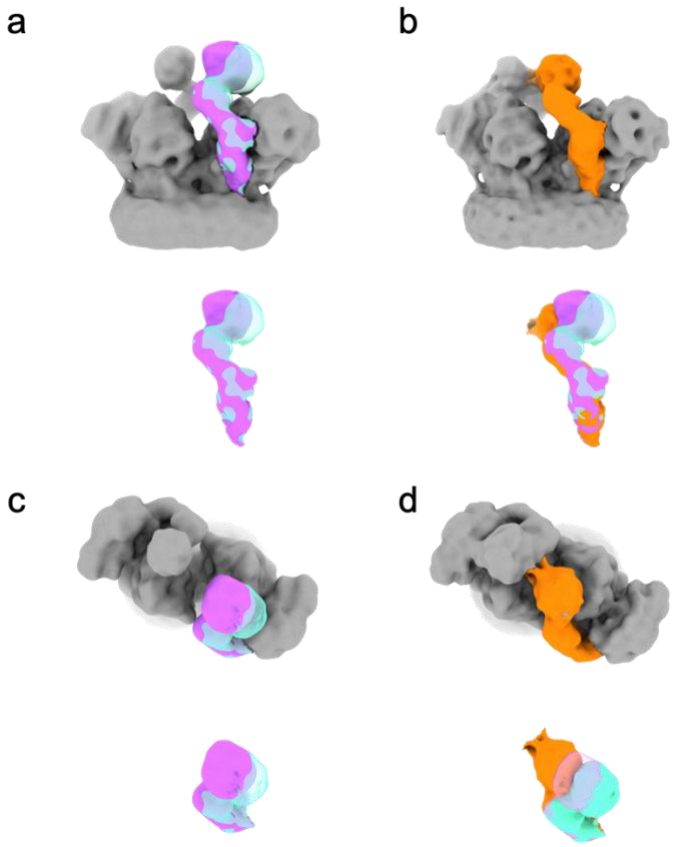


Extended Data Figure 14. The cryo-EM structure shows that the interaction network between FOS and the levanase is extensive. Residues likely interacting with FOS in the active site of Bt1760 levanase, **a**, and in the secondary binding site, **b,** shown in blue. The FOS is displayed in magenta. The active site mutation, D42A, is shown in green. Frc7′ denotes the β2,1 decoration on Frc7. The EM density from the inactive levanase reconstruction is shown around each residue.



Extended Data Figure 15. Aromatic residues near the levanase active site, but not the secondary binding site, are responsible for much of the affinity for levan. a, Surface representation of the levanase model, with FOS shown as purple sticks. **b**, model and EM density for the residues near the active site (Y70A, W318A) and in the secondary binding site (W217A, F243A, Y437A) that were substituted to alanines are shown in gold. All substitutions were done in the inactive levanase background (D42A). **c**, ITC data from titrations of 8 mg/ml levan into 50 µM of indicated levanase variant. Data fitting assumptions are described in the methods.

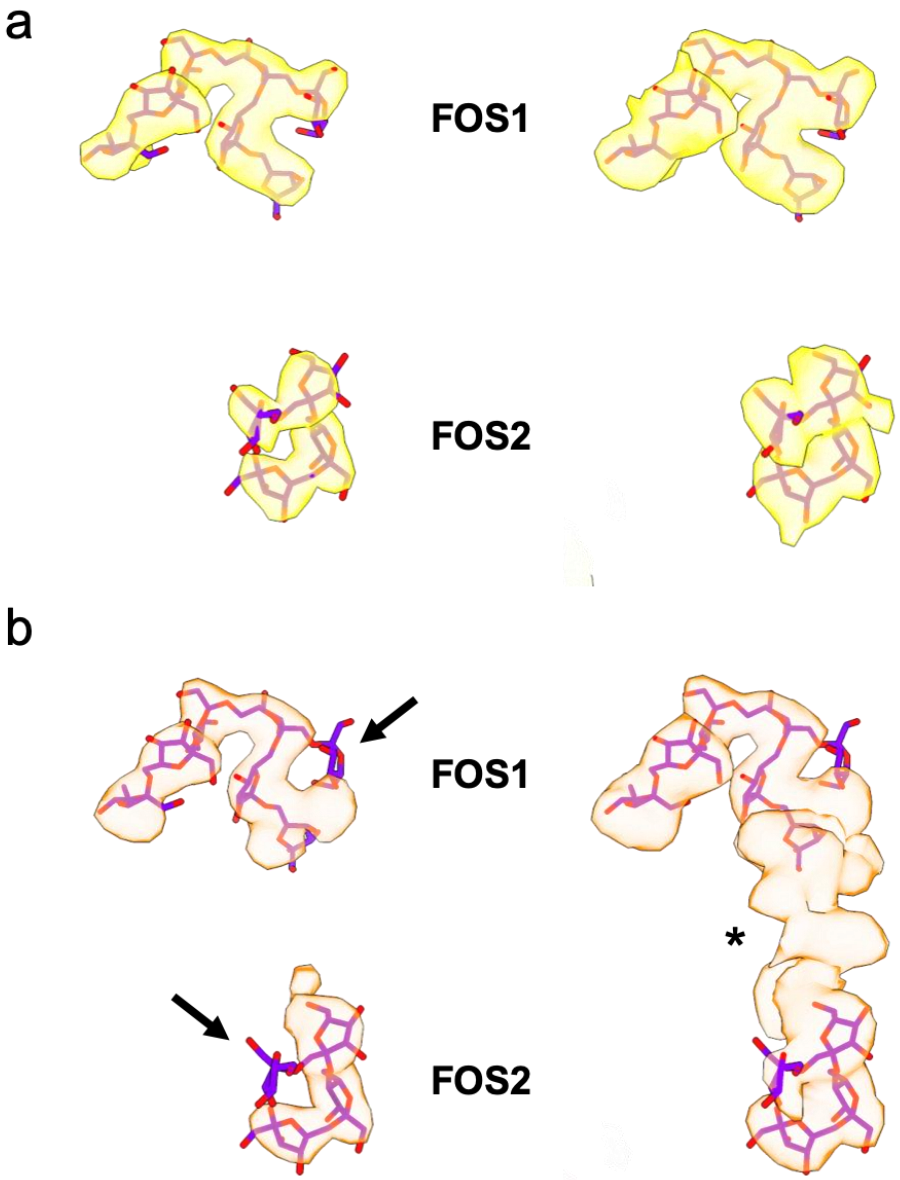
1172
1173
1174



Extended Data Figure 16: Evidence for cooperative substrate binding by both SGBPs of the levan utilisome. **a**, Variability of the levan SGBP observed for the substrate-bound utilisome with active levanase and short FOS (~DP8-12). **b**, A novel state observed in 3D classification of the substrate-bound utilisome with inactive levanase and longer FOS (~DP15-25). Here, one SGBP of the utilisome (orange) appears to reach across and contact the SGBP associated with the other SusC subunit that is present in a docked state. Given that this state was exclusively observed in the presence of long FOS and inactive levanase, we propose that this conformation is the result of both SGBPs in the utilisome interacting with the same stretch of substrate. **c** and **d** show the same maps as **a** and **b** rotated 90° such that they are viewed from the extracellular space.

1175

1176



Extended Data Figure 17. Differences in substrate density for short and long FOS bound in the cavity of SusC. a, Isolated FOS density obtained from the levan utilisome dataset with active Bt1760 levanase and short FOS (DP ~8-12). Density for substrate (yellow) is shown at high (left) and low (right) thresholds. **b**, Isolated FOS density obtained from the dataset with inactive Bt1760 and long FOS (DP ~15-25). Density for substrate (orange) is shown at high (left) and low (right) thresholds. Arrows indicate missing fructose branches relative to (**a**). Note the presence of unmodeled substrate density bridging the binding sites visible at high thresholds (asterisk). FOS models shown are from the original X-ray crystal structure of the Bt1763-62 $SusC_2D_2$ complex determined in the presence of short FOS (DP6-12)[12].

1177
1178