
Article:
White, JBR, Silale, A, Feasey, M et al. (10 more authors) (2023) Outer membrane utilisomes mediate glycan uptake in gut Bacteroidetes. Nature, 618. pp. 583-589. ISSN 0028-0836

https://doi.org/10.1038/s41586-023-06146-w

Outer membrane utilisomes mediate glycan 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], Bert van den Berg[2#] and Neil A. Ranson[1#]

[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:
bert.van-den-berg@ncl.ac.uk; n.a.ranson@leeds.ac.uk

* Authors contributed equally

Summary

Bacteroidetes are abundant members of the human microbiota, utilising a myriad of diet- and host-derived glycans in the distal gut[1]. Glycan uptake across the bacterial outer membrane (OM) of these bacteria is mediated by SusCD protein complexes, comprising a membrane-embedded barrel and a lipoprotein lid, that is thought to open and close to facilitate substrate binding and transport. However, surface-exposed glycan binding proteins and glycoside hydrolases also play critical roles in the capture, processing and transport of large glycan chains. The interactions between these components in the OM are poorly understood, despite being crucial for nutrient acquisition by our colonic microbiota. Here we show that for both 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 cryogenic electron microscopy (cryoEM) structures in the absence and presence of substrate reveal concerted conformational changes that demonstrate the mechanism of substrate capture, and rationalise the role of each component in the utilisome.

Main Text:

The human large intestine is home to a diverse microbial community, the gut microbiome, which is essential for human health[2]. Complex dietary glycans which are inaccessible to the enzymes of the human digestive tract are the primary nutrient source for the microbiome, and the utilisation of these complex sugars is essential for their survival[1,3]. This utilisation is integral to mutualism between host and bacteria, leading, for example, to the generation of short-chain fatty acids that are associated with normal gastrointestinal physiology and systemic health benefits to the host[4,5].

The distal gut microbiome is dominated by two bacterial phyla: the Gram-positive Firmicutes, and the Gram-negative Bacteroidetes[6]. The OM of Gram-negative bacteria presents a formidable barrier to the uptake of large nutrients[7], and gut Bacteroidetes employ a common strategy for glycan utilisation, wherein the machinery for the uptake, processing, transport and metabolism of glycans is encoded in co-regulated gene clusters known as polysaccharide utilisation loci (PULs)[8]. Glycan transport across the OM is dependent on the SusCD core components of a PUL. SusC is an integral OM, TonB-dependent active transporter, whilst SusD is a surface-exposed OM lipoprotein[8,9]. SusCD complexes exist as $SusC_2D_2$ tetramers[10–12] (henceforth abbreviated 'SusCD'), creating a core transportation unit with a twin barrel structure, in with each barrel is associated with its own SusD that caps the extracellular face of the transporter. Recent work suggests that SusCD transporters utilise a 'pedal bin' mechanism, wherein SusD undergoes hinge-like movements that alternatively expose and occlude nutrient binding sites within the SusC barrel, facilitating substrate capture and transport[12].

Although most glycan breakdown occurs within the bacterial cell, initial binding and processing of long glycans occurs extracellularly. To achieve this, a PUL minimally encodes at least one SusCD pair, together with a surface-located, endo-acting glycanase (most commonly glycoside hydrolases; henceforth 'GH') and one (or more) surface glycan binding proteins (SGBPs)[8,9]. The model gut symbiont *B. thetaiotaomicron* (*B. theta*) has 88 predicted PULs, each likely dedicated to a specific glycan, but substrate specificity is known for only ~20 of them[13]. This extraordinary commitment to glycan utilisation (~20% of the *B. theta* genome) demonstrates the importance of complex glycans to bacterial survival in the distal gut.

One of the best-characterised PULs is that for levan, a plant- and bacterial-derived fructan polysaccharide of β2,6-linked fructose units with occasional β2,1 branches[14,15]. We previously characterised the binding and uptake of fructo-oligosaccharides (FOS) by the core SusCD

levan transporter (SusCD^{lev})[10,12]. However, important steps preceding FOS capture and transport remain unclear. Transport-competent FOS are generated from levan by a surface-exposed lipoprotein, a GH32-family levan endo-glycanase (GH^{lev})[14,16], while an SGBP (SGBP^{lev}) is thought to recruit levan at the cell surface[14,17]. A key unresolved question is whether and how these additional lipoproteins associate with their SusCD transporters to facilitate glycan processing and transport. In the archetypal starch PUL, which encodes a SusCD transporter, two SGBPs (SusE and SusF) and a surface amylase (SusG)[18,19], seminal studies suggested that SusCD and SGBP form a complex[19]. However, recent studies indicate that more dynamic and transient, substrate-induced complexes are formed [20], with the SGBPs acting as immobile starch binding centres around which SusCD transporters and the amylase assemble[21]. In contrast, the SusCD pair from an uncharacterised PUL in *B. theta* forms a stable complex with two lipoproteins of unknown function[10]. Thus, two broad models exist: (1) components transiently assemble in response to substrate, and (2) all components form a stable complex in the OM regardless of the presence of substrate.

Here we show, using quantitative proteomics and single-particle cryo-EM, that all four OM components of the levan PUL from *B. theta* (**ED Fig. 1a)** exist in a stable complex that we term a 'utilisome' in keeping with the "PUL" acronym. The four OM components of the *B. theta* dextran (an α1,6-linked glucose polymer with occasional α1,3 glucose branches) PUL (**ED Fig. 1b)** adopt a similar architecture, suggesting that utilisomes may be a generic feature of PULs in the Bacteroidetes. Upon addition of substrate, the levan utilisome undergoes large conformational changes, revealing the substrate-bound states of GH^{lev} and SGBP^{lev}, and demonstrating that the pedal-bin mechanism of nutrient capture operates in the presence of all OM-localised components of the PUL. Collectively, these data show that utilisomes constitute multi-component, macromolecular machines on the cell surface, the architecture of which is consistent with efficient capture, processing, and transport of complex glycans.

Utilisomes are stable without substrate

Previously, we used single particle cryoEM to capture the SusCD levan transporter (SusCD^{lev}) in several conformational states[12]. During structure determination, a minor subset (~10%) of the dataset was identified that contained additional density (**ED Fig. 1 c-e**). SDS-PAGE (and mass spectrometry) revealed co-purification of sub-stoichiometric amounts of GH^{lev} and SGBP^{lev} (**ED Fig. 1c**). Given this, we docked the crystal structure of *E. coli*-expressed, inactive GH^{lev} (PDB ID 7ZNR; **ED Table 1** and Methods) into the map, which fitted well into part of the unassigned density (**ED Fig. 1d,e**), leading us to assign the remainder to the SGBP^{lev}, for which no structure was available. The additional density was associated with only one of the two available SusC^{lev} barrels, and this seemed unlikely to represent an intact complex given

113 the dimeric nature of SusCD^{lev} . After moving the His$_6$-tag from SusD^{lev} to the C-terminus of
114 SGBP^{lev}, and using milder extraction detergent (methods), we were able to purify a complex
115 containing approximately stoichiometric amounts of all four components, for both the levan
116 complex and a related system for dextran utilisation (**Fig. 1a,b)**.
117
118 **The utilisome has four components**
119 Successful purification of two different detergent-solubilised four-component complexes
120 strongly suggests that the complexes are stable in the OM. To shed light on the relative
121 abundance of the components, we performed quantitative proteomics using intensity-based
122 absolute quantification (iBAQ)[22] on outer membrane fractions from *B. theta* grown on fructose
123 or dextran. While the additional lipoproteins of the putative levan four-component complex
124 (4CC) appear to be substoichiometric (dark bars, **Fig. 1c)**, SDS-PAGE and the monodisperse
125 peaks in size exclusion chromatography (**Fig. 1a,b**) suggest that the iBAQ analysis
126 underestimates the abundance of these additional lipoproteins in the OM. For the dextran
127 complex (obtained from *B. theta* grown on dextran), the four components are present in
128 roughly equimolar amounts based on the iBAQ analyses (dark bars, **Fig. 1d**).
129 As expected from transcription studies[14], the four OM components of the levan PUL are among
130 the most-highly expressed membrane proteins in *B. theta* when grown on minimal media with
131 fructose as the sole carbon source, and dextran PUL components were not detectable (**Fig.**
132 **1e)**. Conversely, in minimal media with dextran, the dextran PUL components are highly
133 expressed and levan PUL components are present at just ~1 % of the level found when cells
134 are grown on fructose. This allowed us to spike membrane proteome samples from cells grown
135 on fructose or dextran with detergent-purified dextran or levan complexes, respectively. For
136 both systems, the relative abundance of the four components in the spiked versus
137 endogenous membrane samples was very similar (light versus dark bars; **Fig. 1c,d**),
138 suggesting no excess of any component exists in the OM. Collectively these data support the
139 hypothesis that OM-localised PUL components are present as equimolar 4CCs, which we
140 term 'utilisomes' in keeping with the 'PUL' acronym.
141
142 **CryoEM shows an octameric levan utilisome**
143 The structure of the DDM-solubilised, substrate-free levan utilisome was assessed by single
144 particle cryo-EM, and 3D class averages showed clear evidence for additional components
145 (**ED Fig. 2a**), with ~52 % of the particles containing a SusCD^{lev} core transporter, as well as
146 two copies of both GH^{lev} and SGBP^{lev} (**Figure 2a-d**). We propose that this octameric complex
147 with two copies of each of the four components represents the complete levan utilisome. The
148 remainder of the dataset corresponded to utilisome sub-complexes comprising a SusCD^{lev}
149 pair, with a single copy of GH^{lev}/SGBP^{lev} (~28 %), a small population of 'naked' SusCD^{lev}

150 dimers (~12 %), and some 'junk' particles where SusC barrels are poorly defined (~8%) (**ED**
151 **Fig. 2a**). Based on the proteomics and biochemical data, we believe that the substoichiometric
152 classes are the result of partial denaturation of the complex occurring in the thin-film
153 environment that is present during grid preparation.
154 The structure of the apo levan utilisome was initially assessed using 3D classification. We
155 identified two regions of conformational heterogeneity: the C-terminal domain of SGBP^{lev}, and
156 in the position of the SusD^{lev} lids (**ED Fig. 3a-d;** Methods). All complexes within this dataset
157 have open lids, but they are open to differing degrees. Within a complex, the lids appear to
158 operate independently, such that differing lid positions break the two-fold symmetry of the
159 complex (**ED Fig. 3a-c**). Refinement of particle stacks corresponding to differing SusD^{lev} lid
160 positions gave reconstructions that were limited to ~6 Å. In contrast to our findings for the
161 LDAO-solubilised SusCD^{lev} core complex[10,12], all SusC^{lev} barrels contain density for the plug
162 domain that occludes the barrel. Thus, SusCD^{lev} transporters are open, plugged, and
163 competent to accept substrate. Despite the conformational heterogeneity in the position of the
164 SusD^{lev} lid, SusC^{lev}, GH^{lev}, and the N-terminal domain of SGBP^{lev} constitute a rigid unit with
165 two-fold (C2) symmetry. Application of a mask that excluded the density for the SusD^{lev} lids
166 followed by focused refinement allowed inclusion of all particles, regardless of SusD^{lev}
167 position, and resulted in a map with a global resolution of 3.5 Å (**Fig. 2e**), allowing us to refine
168 models for SusC^{lev} and GH^{lev}, and to build *de novo* the N-terminal domain of SGBP^{lev} (**Fig. 2f**).
169 The refined model for GH^{lev} within the utilisome is similar to the crystal structure of *E. coli*-
170 expressed GH^{lev} solved in isolation (PDB: 7ZNR)(Cα rmsd = 0.54 Å). GH^{lev} is mounted on the
171 lip of the SusC^{lev} barrel, with extracellular loops 1 and 9 of SusC^{lev} contributing to the binding
172 interface (**ED Fig. 4**). This GH^{lev} binding occurs at the opposite side of SusC^{lev} from the SusD^{lev}
173 binding site, in a position that does not obviously impede closure of the SusD^{lev} lid (**Figure 2a-**
174 **d**). The GH^{lev} active site is positioned close to the mouth of the SusCD^{lev} transporter (minimum
175 distance ~30 Å; see methods), consistent with concerted function with the transporter (**Figure**
176 **2d**). SGBP^{lev} is adjacent to GH^{lev} on the lip of SusC^{lev} (see **Figure 2b,d)**; **ED Fig. 4**). An
177 AlphaFold2 model[23] of SGBP^{lev} predicts two N-terminal Ig-like domains, and a C-terminal
178 levan binding domain (**ED Fig. 1a**)[24]. In this substrate-free utilisome structure, there is clear
179 density for only the two N-terminal, Ig-like domains with the second being weaker than the
180 first. Poorly-resolved density is present for the C-terminal, levan-binding domain which
181 extends away from the transporter, towards the extracellular space, with diffuse density visible
182 only at high threshold levels (**Fig. 2a,b**). The lip of SusC therefore serves as a platform for the
183 additional lipoproteins, perhaps explaining why SusC transporters are ~40 % larger than
184 classic, monomeric TonB dependent transporters[9]. Collectively, these data provide important
185 insight into how the utilisome architecture may contribute to efficient levan utilisation. The C-
186 terminal domain of SGBP^{lev} projects away from the cell, where it would facilitate capture of

187 levan chains from the environment. The flexibility of the SGBP^{lev} (**ED Fig. 3d**) may facilitate
188 processing of those levan chains at the adjacent GH^{lev}, and any cleavage products would be
189 released close to their binding site at the SusCD^{lev} interface, promoting efficient loading of the
190 transporter.
191
192 **Utilisomes also exist for dextran**
193 We next studied the dextran utilisation system of *B. theta*, which, like the levan PUL, encodes
194 4 OM components (**ED Fig. 1b)**. Initial biochemical characterisation indicated that a utilisome
195 existed (**Fig. 1a,b**), but cryoEM of the dextran complex revealed it to be much more
196 aggregation prone than the levan utilisome (Methods). Despite this, initial 3D classification
197 revealed that the SusCD^{dex} complex is intact, (i.e. a $SusC_2D_2$ tetramer), and that the SusD^{dex}
198 lids are open. The positions of GH^{dex} and SGBP^{dex} were initially assigned by rigid-body docking
199 of AlphaFold2-predicted models into additional density decorating the lip of SusC^{dex}. However,
200 unlike in the levan system, no classes corresponding to an octameric dextran utilisome
201 existed, which we attribute to a higher propensity of the dextran utilisomes to
202 disassemble/aggregate, potentially at the air-water interface during cryoEM grid preparation.
203 We did observe heptameric complexes, containing two copies of the SGBP^{dex} and a single
204 GH^{dex} (**ED Fig. 5a-d**), together with hexamers and pentamers encompassing all possible
205 complements of GH^{dex} and SGBP^{dex} (**ED Fig. 5e**). Strikingly, although a dextran utilisome is
206 observed, the details of its realisation in 3D are different to the levan utilisome (**ED Figs. 4**
207 **and 5f**). For levan, GH^{lev} is nearest to the hinge of SusD^{lev}, whereas for dextran, the SGBP^{dex}
208 is nearest to SusD^{dex}. Despite this variation, for two distinct *Bacteroides* PULs these data show
209 that all OM components required for utilisation of a glycan are assembled into a utilisome.
210
211 **Substrate drives conformational changes**
212 To probe substrate capture and processing by the levan utilisome, ~0.5 mM levan FOS with
213 a degree of polymerisation of ~8-12 (DP8-12) was added to the detergent-purified utilisome
214 complex, before performing cryoEM. The resulting 3D structure reveals that FOS binding
215 drives major conformational change in the utilisome (**Fig. 3**). All SusD^{lev} lids are closed, tightly
216 capping the extracellular face of the SusC^{lev} barrels (**Fig. 3a**), relative to the open position in
217 the apo utilisomes (**Fig. 3b**). 3D classification reveals that the major source of heterogeneity
218 present in the data was the presence of sub-complexes missing one or more accessory
219 lipoproteins (**ED Fig. 2b**), so to probe substrate binding within the closed transporter complex,
220 we first performed targeted refinement of the SusCD^{lev} core, using particle subtraction to
221 remove the signal for GH^{lev} and SGBP^{lev} from the images. This allowed inclusion of all particles
222 in the refinement, regardless of their lipoprotein complement, and yielded a reconstruction of
223 SusCD^{lev} at 2.9 Å resolution that contains unambiguous density for FOS at two sites within the

224 SusC^{lev} barrels (**Fig. 3c**). FOS1 is found at the interface between SusC^{lev} and SusD^{lev}, and
225 consists of six β2,6-linked fructose units with a β2,1 decoration on fructose-4 (numbered from
226 the non-reducing end). This levan chain adopts a compact, twisted topology similar to that
227 observed in the SusCD crystal structure (6ZAZ)[12]. A second binding site (FOS2) most likely
228 consists of four β2,6-linked fructose units, and is bound at the bottom of the SusC^{lev} cavity
229 where it contacts the top of the plug domain.
230
231 Given the utilisome's biological role, substrate loading of the transporter should be signalled
232 across the OM, presumably via disordering of an N-terminal region of SusC (and all TonB-
233 dependent transporters) called the 'TonB box', at the periplasmic face of the complex[12]. An
234 interesting difference between the apo and substrate-bound structures is in the position of one
235 of the hinge points between SusC^{lev} and SusD^{lev}. In the apo structure, this region (SusC^{lev} loop
236 8) is mostly unresolved, likely as an artefact of masking during image processing (Methods),
237 whereas in the substrate-bound complex SusC^{lev} loop 8 contacts SusD^{lev} and FOS1 via F649
238 (**Fig. 3c,d**). However, the part of loop 8 proximal to the SusC barrel is resolved in both the apo
239 and substrate-bound structures (**Fig. 3d**), with subtle changes in the side chains of its amino
240 acid residues along the edge of the barrel and plug towards the periplasm. In the substrate-
241 bound structure, W685, which sits at the base of loop 8, shifts inwards, nudging F583 up and
242 S193 down (**Fig. 3e**). This in turn results in a downward shift of a plug loop containing S193.
243 In the apo structure, Y191, which is part of the same plug loop, forms a triple aromatic stack
244 with Y89 and F558 that links the wall of the barrel (F558), the TonB box (Y89) and the plug
245 domain (Y191) (**Fig. 3f**). When the plug loop with Y191 shifts towards the periplasm, this
246 stacking interaction is broken, resulting in release of the N-terminus and an ~35 Å shift of R93
247 (**Fig. 3g**), which in the substrate-bound complex is the first visible residue. This would plausibly
248 make the TonB box (D82-G88) available for interaction with TonB to disrupt the plug and
249 generate a channel for substrate diffusion into the periplasmic space (Supplementary
250 Discussion). The aromatic lock in SusC^{lev} is reminiscent of *E. coli* BtuB (the TBDT for vitamin
251 B12), where an 'ionic lock'[25] is broken upon vitamin B12 binding[26] to signal extracellular
252 substrate binding to the periplasmic face of the transporter.
253
254 **Long substrate tethers levan SGBP**
255 While concerted behaviour of the levan utilisome components is implied by their arrangement,
256 the structure of the complete levan binding protein remained unresolved. We reasoned that in
257 a utilisome containing a catalytically inactive GH^{lev} (with the active site D42A mutation), and
258 longer FOS chains (~DP 15-25), the binding of a single levan chain by both GH^{lev} and SGBP^{lev}
259 might reduce conformational variability. We therefore collected such a "long-FOS" cryoEM
260 dataset, and observed in 3D classification that a novel conformation was present, in which the

261 C-terminal domain of SGBP^{lev} adopted a 'docked' state proximal to both SusD^{lev} and GH^{lev} (**ED**
262 **Fig. 2c.**). Via focused classification without particle alignment (Methods), and using a mask
263 including only the docked SGBP^{lev} position, we obtained a class (29k particles) containing
264 well-resolved density for the entire SGBP^{lev} (**ED Fig. 2d-f**). The unmasked versions of those
265 29k particles were then used to generate a final reconstruction of the complete utilisome at
266 3.0 Å with well-resolved density for a complete SGBP^{lev} (**Fig. 3a-c**) (see Methods). SGBP^{lev} is
267 organised into three domains, as predicted by Alphafold. The lipidated N-terminal domain and
268 the central domain each have an Ig-like fold. The C-terminal levan-binding domain is globular,
269 with a central β-sheet decorated with α-helices. A distance-matrix alignment (DALI) analysis
270 of the C-terminal domain showed significant similarity to a ThuA-like protein (PDB ID 1T0B)
271 with a putative role in disaccharide binding and a homoserine O-succinyl transferase (PDB ID
272 7CBE) (Methods).

273

274 CryoEM density, which we attribute to a length of levan chain, is found at the two locations
275 previously described in the short FOS structure (FOS1 and FOS2). Interestingly, the substrate
276 densities within the SusCD^{lev} binding cavity differ from those observed previously for short
277 FOS **(ED Fig. 6)**, suggesting that FOS binding by the SusCD^{lev} core is promiscuous. In
278 addition, density is present at two additional sites that are unique to this long FOS structure.
279 Within the inactivated GH^{lev} active site, there is density for ~five β2,6-linked fructose units, and
280 we define this site as FOS3. A secondary or 'tethering' site of FOS binding is also observed
281 (FOS4) with density for FOS bridging between SGBP^{lev} and GH^{lev} (**Fig. 4f,g**). No obvious
282 density links the FOS3 and FOS4 binding sites. The long, bridging levan molecule binds to
283 GH^{lev} across the top of the GH^{lev} β-propeller domain, before proceeding towards its C-terminal
284 β-sandwich domain. The density is compatible with ~12 β2,6-linked fructose units in an
285 extended conformation, with a putative β2,1 decoration on Frc-7 (**Fig. 4g.**). Levan binding at
286 this tethering site is also seen in GH^{lev} in the absence of bridging interactions with SGBP^{lev}
287 and in crystal structures of the inactive levanase solved in the presence of FOS of DP ~7-8
288 (**ED Fig. 6)**. Indeed, secondary binding sites are relatively common in endo-acting GH
289 enzymes, where they may enhance substrate processivity[27]. No ligand was observed at the
290 secondary site in a GH^{lev} structure complexed to a 4 unit FOS (PDB ID 6R3U[28]), suggesting
291 the affinity for short FOS is low.

292

293 Whilst the resolution is in the region of the bridging levan is insufficient for a detailed
294 description of binding such as hydrogen bonds (3.5-4.0 Å), several tryptophan residues are
295 clearly involved, including W297 and W359 from SGBP^{lev}, and W217 from GH^{lev}, which appear
296 to cradle the levan chain via stacking interactions. Isothermal titration calorimetry (ITC) on

297 recombinant SGBP^{lev} confirmed the importance of W297 and W359 of the SGBP^{lev} for levan
298 binding (**Fig. 4h**), and suggested that SGBP^{lev} has a single binding site, and a minimum
299 binding unit of ~8 fructose units (**ED Fig. 7a**). The importance of specific stacking interactions
300 by tryptophan residues for glycan binding platforms is consistent with previous data[29,30] and is
301 supported by their conservation in levan binding proteins (**ED Fig. 8**). Indeed, ITC indicates
302 that SGBP^{lev} has little affinity for inulin, a fructose oligosaccharide similar to levan but with
303 different main chain linkages (β2,1 instead of β2,6) (**ED Fig. 7b**), confirming the specificity of
304 binding. We next attempted to dissect the affinities of the FOS3 and FOS4 binding sites, both
305 of which include numerous aromatic residues (**ED Fig. 7c,d**). Single substitution of aromatic
306 residues (to alanine) near the GH^{lev} active site (Y70A and W318A) decreased affinity for levan
307 25-30 fold, while equivalent substitutions at the tethering site FOS4 had a 6-fold decrease in
308 affinity (W217A) or no effect (F243A and Y437A). These results suggest the active site (FOS3)
309 is responsible for most of the affinity of GH^{lev} for levan. However, numerous weak interactions
310 within the extensive secondary site (FOS4) including polar residues, may result in a
311 substantial overall affinity and specificity.

312

313 The utilisome structure suggests why SusCD transporters exist as dimers. The SGBP^{lev} is
314 clearly inherently flexible, and in the dataset with inactive GH^{lev} and long FOS, a novel
315 conformation was observed in which an untethered SGBP^{lev} from one SusC^{lev} contacts the
316 tethered SGBP^{lev} associated with the other SusC^{lev} (**ED Fig. 3e**). This conformation could
317 result in both SGBP^{lev} interacting with the same stretch of a long levan chain (as would be
318 present *in vivo)*, increasing substrate avidity and helping to retain and position it near the
319 utilisome for processing and transport.

320

321 **Discussion**

322 Our utilisome cryo-EM structures provide new insight into glycan acquisition by gut
323 Bacteroidetes. As Bacteroidetes are found in a diverse range of terrestrial and marine
324 niches[31], and have all been shown to encode SusCD homologues[32], the data may also expand
325 our understanding of glycan utilisation outside the animal gut. We show that for two PULs
326 dedicated to the breakdown of simple glycans, all of the gene products that localise to the OM
327 form a stable utilisome complex. The way in which these utilisomes are realised in 3D are
328 subtly different (**ED Fig. 5f**), but they have closely-related architectures that rationalise
329 function (**Fig. 5**). In the absence of substrate (1), the glycan binding domain of an SGBP is
330 mobile, increasing its efficiency as a glycan grappling device. Likewise, the SusD lids, which
331 can open and close without clashing with the other components, are mobile but open. (2)
332 Glycan binding by the SGBP is followed by docking to the proximal GH, and the glycan binds

333 to both the tethering and active sites of the enzyme. Substrate cleavage by GH then generates
334 shorter oligosaccharides close to the mouth of the SusCD transporter. (3) Binding of the
335 oligosaccharide to SusCD promotes lid closure and signalling to the plug, breaks the aromatic
336 lock and exposes the Ton box to the periplasm. This is followed by TonB-dependent transport
337 events (Supplementary Discussion), and resetting of the transporter to the open state (4)[10,12].
338
339 The long levan chain between $SGBP^{lev}$ and GH^{lev} is much more extended than at other sites,
340 and may therefore represent a 'strained' state in which a bound levan chain is pulled taut.
341 However, we note that this conformation represents the least mobile and therefore most
342 resolvable state, and has thus been 'selected for' in cryoEM image processing. Furthermore,
343 although the GH^{lev} tethering site (FOS4) is ~25 Å away from its catalytic site (FOS3), we
344 speculate that the SGBP:GH interaction increases the levan concentration local to the active
345 site, potentially enhancing the efficiency of substrate cleavage. Flexibility of the levan chain
346 away from the interaction sites may preclude resolution of contiguous density in our structure,
347 or the FOS used here may be too short to bridge both sites.
348
349 Whilst our data argue for stable utilisome assemblies in the OM, recent studies on the starch
350 PUL propose a more dynamic model, with two SGBPs forming immobile starch binding
351 centres, around which a GH and transporter components transiently assemble[21]. Proteomics
352 data for the starch utilisation system reveal similar amounts of SusC and SusD at the OM
353 whilst SusEFG are much less abundant, suggesting the presence of complexes with differing
354 stoichiometries[21]. Intriguingly, co-immunoprecipitation with SusD antibodies captures twice as
355 much SusD as SusC, suggesting that even the core SusCD transporter may not form a stable
356 complex in the starch system. This contrasts with our work, where separate SusC or SusD
357 components have not been observed. Moreover, while it is not clear how similar the outer
358 membranes of *E. coli* and *B. theta* are, recent work has shown that *E. coli* OMP and LPS
359 mobility is limited [33,34], raising questions about the efficiency of a complex that dynamically
360 (dis)assembles in a crowded OM. It is not clear why the system for starch appears to operate
361 differently to those for levan and dextran. The dynamic model for the starch PUL is based on
362 live-cell fluorescence studies with C-terminal fusions of OM PUL components with relatively
363 large fluorescent protein tags, which may destabilise the utilisomes but not markedly affect
364 cell growth *in vitro*. On the other hand, the starch PUL is different as it encodes two SGBPs
365 (SusE and SusF), *i.e*. there are five OM PUL components and not four as in the systems
366 studied here. An AlphaFold2 structure prediction of the starch SusC (Bt3702) shows that it
367 has no extra interaction surface for the association of additional lipoprotein components
368 relative to the levan and dextran SusCs, making it unlikely that SusE, SusF and SusG can all

369 bind the same SusCD transporter and suggesting the starch utilisome may be more dynamic,
370 and/or perhaps forms distinct, specialised utilisomes.
371
372 Indeed, the diversity of known PULs is considerable. Determining whether four-component
373 utilisomes, as presented here, are a generic feature of PULs will require the study of more
374 systems from gut *Bacteroides spp.* (both *B. theta* and others), as well as from more distant
375 Bacteroidetes occupying different ecological niches such as soil and marine environments.
376 The levan and dextran utilisation systems described here target relatively simple glycans, and
377 both the number of PULs and their compositional complexity appears to correlate with the
378 complexity of their substrates, *i.e*. the number of encoded SGBPs, carbohydrate active
379 enzymes, and even SusCD pairs increases in-line with substrate complexity[35–37]. Whether the
380 components of these complex PULs can undergo mix-and-match style assembly, resulting in
381 the formation of several utilisome complexes with unique lipoprotein complements is unknown,
382 but represents an important target of future study.
383
384 **Main text References**

385 1. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the
386 human gut microbiota. *Nature Reviews Microbiology 2012 10:5* **10**, 323–335 (2012).
387 2. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nature*
388 *reviews. Microbiology* **19**, 55–71 (2021).
389 3. Hamaker, B. R. & Tuncil, Y. E. A Perspective on the Complexity of Dietary Fiber
390 Structures and Their Potential Effect on the Gut Microbiota. *Journal of Molecular*
391 *Biology* **426**, 3838–3850 (2014).
392 4. Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota
393 and their impact on human metabolism. *Gut microbes* **7**, 189–200 (2016).
394 5. Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. From Dietary Fiber to Host
395 Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell* **165**, 1332–1345
396 (2016).

397 6. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human
398 microbiome. *Nature 2012 486:7402* **486**, 207–214 (2012).
399 7. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited.
400 *Microbiology and molecular biology reviews : MMBR* **67**, 593–656 (2003).
401 8. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex Glycan Catabolism
402 by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. *Journal of*
403 *Biological Chemistry* **284**, 24673–24677 (2009).
404 9. Bolam, D. N. & van den Berg, B. TonB-dependent transport by the gut microbiota: novel
405 aspects of an old problem. *Current Opinion in Structural Biology* vol. 51 35–43 Preprint
406 at https://doi.org/10.1016/j.sbi.2018.03.001 (2018).
407 10. Glenwright, A. J. *et al.* Structural basis for nutrient acquisition by dominant members of
408 the human gut microbiota. *Nature* **541**, 407–411 (2017).
409 11. Madej, M. *et al.* Structural and functional insights into oligopeptide acquisition by the
410 RagAB transporter from Porphyromonas gingivalis. *Nature Microbiology 2020 5:8* **5**,
411 1016–1025 (2020).
412 12. Gray, D. A. *et al.* Insights into SusCD-mediated glycan import by a prominent gut
413 symbiont. *Nature Communications* **12**, 1–14 (2021).
414 13. Terrapon, N. *et al.* PULDB: the expanded database of Polysaccharide Utilization Loci.
415 *Nucleic Acids Research* **46**, D677–D683 (2018).
416 14. Sonnenburg, E. D. *et al.* Specificity of polysaccharide use in intestinal bacteroides species
417 determines diet-induced microbiota alterations. *Cell* **141**, 1241–1252 (2010).
418 15. Öner, E. T., Hernández, L. & Combie, J. Review of Levan polysaccharide: From a century
419 of past experiences to future prospects. *Biotechnology Advances* vol. 34 827–844
420 Preprint at https://doi.org/10.1016/j.biotechadv.2016.05.002 (2016).

421 16. Mardo, K. *et al.* A Highly Active Endo-Levanase BT1760 of a Dominant Mammalian Gut
422 Commensal Bacteroides thetaiotaomicron Cleaves Not Only Various Bacterial Levans,
423 but Also Levan of Timothy Grass. *PLOS ONE* **12**, e0169989 (2017).
424 17. Bolam, D. N. & Sonnenburg, J. L. Mechanistic insight into polysaccharide use within the
425 intestinal microbiota. *Gut microbes* **2**, (2011).
426 18. Reeves, A. R., Wang, G. R. & Salyers, A. A. Characterization of four outer membrane
427 proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. *Journal*
428 *of bacteriology* **179**, 643–649 (1997).
429 19. Cho, K. H. & Salyers, A. A. Biochemical analysis of interactions between outer membrane
430 proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *Journal of*
431 *bacteriology* **183**, 7224–7230 (2001).
432 20. Karunatilaka, K. S., Cameron, E. A., Martens, E. C., Koropatkin, N. M. & Biteen, J. S.
433 Superresolution imaging captures carbohydrate utilization dynamics in human gut
434 symbionts. *mBio* **5**, (2014).
435 21. Tuson, H. H., Foley, M. H., Koropatkin, N. M. & Biteen, J. S. The Starch Utilization System
436 Assembles around Stationary Starch-Binding Proteins. *Biophysical Journal* **115**, 242–250
437 (2018).
438 22. Schwanhüusser, B. *et al.* Global quantification of mammalian gene expression control.
439 *Nature 2011 473:7347* **473**, 337–342 (2011).
440 23. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*
441 *2021 596:7873* **596**, 583–589 (2021).
442 24. Tamura, K. & Brumer, H. Glycan utilization systems in the human gut microbiota: a gold
443 mine for structural discoveries. *Current Opinion in Structural Biology* **68**, 26–40 (2020).

25. Nilaweera, T. D., Nyenhuis, D. A. & Cafiso, D. S. Structural intermediates observed only in intact Escherichia coli indicate a mechanism for TonB-dependent transport. *eLife* **10**, (2021).
26. Zmyslowski, A. M., Baxa, M. C., Gagnon, I. A. & Sosnick, T. R. HDX-MS performed on BtuB in E. coli outer membranes delineates the luminal domain's allostery and unfolding upon B12 and TonB binding. *Proceedings of the National Academy of Sciences of the United States of America* **119**, (2022).
27. Cuyvers, S., Dornez, E., Delcour, J.A. & Courtin, C.M. Occurrence and functional significance of secondary carbohydrate binding sites in glycoside hydrolases. *Critical Reviews in Biotechnology* **32**, 93-107 (2011).
28. Ernits, K., Eek, P., Lukk, T., Visnapuu, T. & Alamäe, T. First crystal structure of an endo-levanase – the BT1760 from a human gut commensal Bacteroides thetaiotaomicron. *Scientific Reports 2019 9:1* **9**, 1–13 (2019).
29. Tamura, K. *et al.* Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont Bacteroides ovatus. *Cellular and molecular life sciences : CMLS* **76**, 4319–4340 (2019).
30. Tamura, K., Dejean, G., Van Petegem, F. & Brumer, H. Distinct protein architectures mediate species-specific beta-glucan binding and metabolism in the human gut microbiota. *The Journal of biological chemistry* **296**, (2021).
31. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide utilization loci: Fueling microbial communities. *Journal of Bacteriology* **199**, (2017).
32. McKee, L.S., La Rosa, S.L., Westereng, B., Eijsink, V.G., Pope, P.B. & Larsbrink, J. Polysaccharide degradation by the Bacteroidetes: Mechanisms and nomenclature. *Environmental Microbiology Reports* **13**, 559-581 (2021)

468 33. Rassam, P. *et al.* Supramolecular assemblies underpin turnover of outer membrane
469 proteins in bacteria. *Nature 2015 523:7560* **523**, 333–336 (2015).
470 34. Benn, G. *et al.* Phase separation in the outer membrane of Escherichia coli. *Proceedings*
471 *of the National Academy of Sciences of the United States of America* **118**, (2021).
472 35. Briliūtė, J. *et al.* Complex N-glycan breakdown by gut Bacteroides involves an extensive
473 enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nat Microbiol* **4**,
474 1571–1581 (2019).
475 36. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by
476 two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).
477 37. Ndeh, D. *et al.* Complex pectin metabolism by gut bacteria reveals novel catalytic
478 functions. *Nature* **544**, 65–70 (2017).
479 38. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch Catabolism by a
480 Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices.
481 *Structure* **16**, 1105–1115 (2008).
482
483 **Figure Legends:**
484
485 **Figure 1. Stable 4-component complexes in the OM of *B. thetaiotaomicron*. (a)** SDS-
486 PAGE of the purified levan and dextran utilisome complexes. (**b**) Analytical SEC elution
487 profiles for the levan (blue) and dextran utilisomes (orange), analysed on Superdex-200 and
488 Superose-6, respectively (void volumes indicated by arrows). OM abundance of the four
489 components of (**c**) the levan (blue) and (**d**) the dextran (orange) systems obtained from
490 fructose- and dextran-grown cells, respectively (dark coloured bars). The light-coloured bars
491 show the normalised abundance of detergent-purified levan or dextran 4CCs spiked into the
492 proteome samples of dextran- or fructose-grown cells. The dots show the individual values of
493 two independent replicates. (**e**) Volcano plot of the *B. thetaiotaomicron* OM proteome from
494 fructose versus dextran grown cells. Three biological replicates of each condition were used
495 for proteomics analysis. Statistical analysis was performed using a two-sided moderated t-test
496 in limma[43] to identify differentially regulated proteins (fold-change > 2, adjusted P-value <
497 0.05). The Benjamini-Hochberg correction for multiple hypothesis testing was implemented.
498

499 **Figure 2. The organisation of the apo levan utilisome revealed by cryo-EM.** (**a**) A side
500 view (in the plane of the OM), and (**b**) top view (perpendicular to the OM plane from outside
501 the cell) of a ~6 Å 3D refinement of one conformational state of the apo utilisome obtained by
502 3D classification, shown at low threshold. $SusC^{lev}$ is green, $SusD^{lev}$ is grey, GH^{lev} is blue,
503 $SGBP^{lev}$ is magenta, and the weak density for the detergent micelle is transparent grey. (**c**)
504 and (**d**) show identically positioned and coloured views at high threshold. The position of the
505 GH^{lev} active site is shown by a yellow arrow in (**d**). (**e**) C2 symmetrized reconstruction at 3.5
506 Å from a focused refinement, with data masked to exclude the variable SusD subunit positions.
507 Note that all particles in this dataset possess SusD components, and their absence here is a
508 result of focused refinement. (**f**) *De novo* atomic model built into the density shown in (**e)**.
509
510 **Figure 3. Conformational change in the utilisome upon FOS binding.** (**a**) CryoEM
511 structure of the levan utilisome with short FOS (DP8-12) at 2.9 Å resolution. (**b**) Equivalent
512 view of the apo utilisome at 6Å resolution, showing that the $SusD^{lev}$ lids are open in the
513 absence of substrate and closed when it is bound. (**c**) Details of the FOS1 and FOS2 binding
514 sites with $SusC^{lev}$ side chains highlighted (apo in grey, FOS-bound in green) and FOS coloured
515 yellow. F649 at the tip of loop 8 is disordered in apo, but becomes ordered in the FOS-bound
516 utilisome. (**d**) Overview of the $SusC^{lev}$ structure, showing loop 8, and coloured as in (**c**). The
517 apo structure is overlaid in grey. N-terminal residues visible only in the absence of substrate
518 are coloured red. (**e**) Zoomed-in view of residues at the base of loop 8. (**f**) The 'aromatic lock'
519 in apo $SusC^{lev}$, with Y89 in the segment that immediately precedes the TonB box (D82-G88)
520 sandwiched between the wall of the channel (F588) and the plug (Y191). (**g**) Substrate binding
521 disrupts the aromatic lock, and the sequence preceding the TonB box is released into the
522 periplasm. Green arrows indicate movements in the FOS-bound structure relative to apo.
523
524 **Figure 4. Cryo-EM structure of the holo utilisome with long FOS.** Reconstruction of the
525 levan utilisome complex after focused classification viewed (**a**) perpendicular to the plane of
526 the membrane (from outside the cell), (**b**) rotated 90° for a side view from the OM plane, and
527 (**c**) rotated a further 90° to give an 'end' view. Subunits are coloured as indicated. One $SGBP^{lev}$
528 is now fully resolved (magenta), docked to the adjacent GH^{lev} subunit. (**d**) The 'active site' of
529 the inactive GH^{lev} (D42A) with density for bound FOS at site 3 indicated in yellow (within the
530 orange box). (**e**) Zoomed inset of the boxed region in (**d**), but now as a cartoon. The modelled
531 levan chain is yellow, with aromatic side chains shown. (**f**). Zoomed view of the FOS density
532 in the 4th (FOS4) binding site that bridges the $SGBP^{lev}$ and GH^{lev} subunits. (**g**) Proximal
533 aromatic residues, especially tryptophans are shown (CPK coloured) and coloured according
534 to the subunit to which they belong. The black arrow indicates the position of the putative β2,1
535 decoration on Frc-7. Maps displayed here have been filtered using LAFTER[64] and segmented
536 in UCSF Chimera. (**h**) Isothermal titration calorimetry experiments with 1 mM FOS DP~15
537 titrated into 50 μM recombinant wild type $SGBP^{lev}$, or the W297A/W359A double-mutant
538 (WAWA). Removing both tryptophan residues abolishes FOS binding.
539
540 **Figure 5. The transport pathway of utilisomes dedicated to the processing and import**
541 **of simple glycans.** Description of states (1-4) is provided in the main text. N.B. only one half
542 of the dimeric utilisome is shown for clarity. A red asterisk marks the active site of the glycoside
543 hydrolase (GH). OM=outer membrane.
544

545 **Methods**

546 **Construction of *B. theta* strains.** *B. theta* strains were made as described previously [12].
547 Briefly, the DNA sequence containing the desired alterations was constructed using the
548 sewing PCR method and ligated into the pExchange-tdk vector[38]. The vectors carrying the
549 altered DNA sequences were introduced into the *B. theta tdk*$^-$ strain via conjugation from *E.*
550 *coli* S17 λ *pir*. Chromosomal alterations were made by allelic exchange, followed by selection
551 for loss of the pExchange-tdk vector backbone. Mutations were confirmed by PCR
552 amplification of the region of interest and Sanger sequencing.

553

554 **Expression and purification of utilisomes from *B. theta.*** *B. theta* strains were grown at
555 37°C in a Don Whitley Scientific A32 anaerobic workstation. Brain-heart infusion (BHI) cultures
556 supplemented with 2 ug/ml hemin were inoculated with stabs from glycerol stocks of the
557 appropriate *B. theta* strain and incubated overnight. Defined minimal medium[12] was
558 supplemented with 2 ug/ml hemin and either 0.5% fructose (levan system) or 0.5% dextran
559 3.5 kDa (dextran system) and inoculated with the overnight BHI cultures (1:1000 dilution). The
560 cultures were harvested after 20 h by centrifugation and the pellets were stored at -20°C. 4
561 litres of culture were grown for SusC^{lev}-His (with wild type GH^{lev}) and SGBP^{lev}-His/GH^{lev}-D42A
562 (inactive GH^{lev}) strains for cryo-EM.

563

564 The cell pellets were thawed, supplemented with DNase I and homogenised in Tris-buffered
565 saline (TBS, 20 mM Tris-HCl pH 8.0, 300 mM NaCl). The cells were lysed with a single pass
566 at 22 kpsi through a cell disruptor (0.75 kW; Constant Systems). Membranes were isolated by
567 ultracentrifugation for 45 min at 42,000 rpm (45 Ti rotor, Beckman), 4°C. The membranes
568 were solubilised at 4°C for 1 h in TBS with 1% DDM while stirring. Insoluble material was
569 pelleted by ultracentrifugation for 30 min at 42,000 rpm (45 Ti rotor) at 4°C. The supernatants
570 were supplemented with 20 mM imidazole and loaded onto an 8 ml chelating sepharose
571 column charged with Ni^{2+}, by gravity flow at room temperature. The column was washed with
572 20 column volumes TBS with 30 mM imidazole and 0.15% DDM. The bound proteins were
573 eluted with 3 column volumes TBS with 250 mM imidazole and 0.15% DDM. The eluates were
574 concentrated in an Amicon Ultra filtration device with a 100 kDa cut-off membrane. The
575 samples were then loaded on a HiLoad 16/600 Superdex 200 pg column (Cytiva) in 10 mM
576 HEPES-NaOH pH 7.5, 100 mM NaCl, 0.03% DDM. Fractions containing pure protein were
577 pooled, concentrated, flash-frozen in liquid nitrogen and stored at -80°C.

578

579 **Sample preparation for mass spectrometry.** *B. theta* cells were grown in minimal medium
580 with 0.5% fructose or 0.5% dextran 3.5 kDa, harvested by centrifugation and lysed as above.
581 Total membrane pellets were extracted twice with 0.5% sarkosyl in 20 mM Hepes pH 7.5 (20

582 mins at room temperature) to remove inner membrane components. Each extraction was
583 followed by centrifugation for 30 min at 42,000 rpm (45 Ti rotor), and the pellet was retained.
584 OM samples were suspended in 5% sodium dodecyl sulfate (SDS) in 50 mM
585 triethylammonium bicarbonate (TEAB) pH 7.5 and sonicated using an ultrasonic homogenizer
586 (Hielscher) for 1 minute. Samples were centrifuged at 10,000 xg for 10 minutes to pellet debris.
587 Proteins (40 µg) were subsequently reduced by incubation with 20 mM tris(2-
588 carboxyethyl)phosphine for 15 minutes at 47 °C, and alkylated with 20 mM iodoacetamide for
589 15 minutes at room temperature in the dark. Proteomic sample preparation was performed
590 using the suspension trapping (S-Trap) sample preparation method[39,40], with minor
591 modifications as recommended by the supplier (ProtiFi™, Huntington NY). Briefly, 2.5 µl of
592 12% phosphoric acid was added to each sample, followed by the addition of 165 µl S-Trap
593 binding buffer (90% methanol in 100 mM TEAB pH 7.1). The acidified samples were added,
594 separately, to S-Trap micro-spin columns and centrifuged at 4,000 xg for 1 minute until all the
595 solution has passed through the filter. Each S-Trap micro-spin column was washed with 150
596 µl S-trap binding buffer by centrifugation at 4,000 xg for 1 minute. This process was repeated
597 for a total of five washes. Twenty-five µl of 50 mM TEAB containing 4 µg trypsin was added
598 to each sample, followed by proteolytic digestion for 2 hours at 47 °C using a thermomixer
599 (Eppendorf). Peptides were eluted with 50 mM TEAB pH 8.0 and centrifugation at 3,000 xg
600 for 1 minute. Elution steps were repeated using 0.2% formic acid and 0.2% formic acid in 50%
601 acetonitrile, respectively. The three eluates from each sample were combined and dried using
602 a speed-vac before storage at -80°C.
603
604 **Mass spectrometry.** Peptides were dissolved in 2% acetonitrile containing 0.1% formic acid,
605 and each sample was independently analysed on an Orbitrap Q Exactive HF mass
606 spectrometer (Thermo Fisher Scientific), connected to an UltiMate 3000 RSLCnano System
607 (Thermo Fisher Scientific). Peptides were injected on a PepMap 100 C18 LC trap column
608 (300 µm ID × 5 mm, 5 µm, 100 Å) followed by separation on an EASY-Spray nanoLC C18
609 column (75 µm ID × 50 cm, 2 µm, 100 Å) at a flow rate of 250 nl/min. Solvent A was water
610 containing 0.1% formic acid, and solvent B was 80% acetonitrile containing 0.1% formic acid.
611 The gradient used for analysis was as follows: solvent B was maintained at 2% B for 5 min,
612 followed by an increase from 2 to 30% B in 110 min, 30% to 42% B in 10 min, 42-90% B in
613 0.5 min, maintained at 90% B for 4 min, followed by a decrease to 2% in 0.5 min, and
614 equilibration at 2% for 20 min. The Orbitrap Q Exactive HF was operated in positive-ion data-
615 dependent mode. The precursor ion scan (full scan) was performed in the Orbitrap (OT) in the
616 range of 350-1,500 m/z with a resolution of 60,000 at 200 m/z, an automatic gain control (AGC)
617 target of 3×10^6, and an ion injection time of 50 ms. MS/MS spectra were acquired in the OT
618 using the Top 20 precursors fragmented by high-energy collisional dissociation (HCD)

619 fragmentation. Precursors were isolated using the quadrupole using a 1.6 m/z isolation width.
620 An HCD collision energy of 25% was used, the AGC target was set to 2×10^5 and an ion
621 injection time of 50 ms was allowed. Dynamic exclusion of ions was implemented using a 45 s
622 exclusion duration. An electrospray voltage of 1.8 kV and capillary temperature of 280°C, with
623 no sheath and auxiliary gas flow, was used.

624 **Mass spectrometry data analysis**. All spectra were analysed using MaxQuant 1.6.14.0 [41],
625 and searched against the *Bacteroides thetaiotaomicron* Uniprot proteome database
626 (UP000001414) downloaded on 22 September 2020. Peak list generation was performed
627 within MaxQuant and searches were performed using default parameters and the built-in
628 Andromeda search engine [42]. The enzyme specificity was set to consider fully tryptic peptides,
629 and two missed cleavages were allowed. Oxidation of methionine and N-terminal acetylation
630 were set as variable modifications. Carbamidomethylation of cysteine was set as a fixed
631 modification. A protein and peptide false discovery rate (FDR) of less than 1% was employed
632 in MaxQuant. Proteins that contained similar peptides and that could not be differentiated
633 based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Reverse
634 hits, contaminants, and proteins only identified by site modifications were removed before
635 downstream analysis. Ranking of protein abundance was performed using iBAQ intensity
636 values[22] obtained from MaxQuant. Label-free quantification was performed using LFQ
637 intensities obtained from MaxQuant. LFQ intensities were log2 transformed and filtered to
638 contain at least three valid values in one of the groups. Missing values were imputed using
639 the minProb function in the imputeLCMD package (https://cran.r-
640 project.org/web/packages/imputeLCMD) in R version 4.1.1. Statistical analysis was performed
641 using limma[43] and the Benjamini-Hochberg correction for multiple hypothesis testing was
642 implemented.

643 **Construction of plasmids for protein expression in *E. coli***. The nucleotide sequences
644 encoding GH^{lev} (Bt1760; residues 2-503) and $SGBP^{lev}$ (Bt1761; residues 2-438) were amplified
645 from *B. theta* genomic DNA, excluding the signal sequence and the lipid anchor cysteine. In
646 all cases, the protein numbering starts with the first residue of the mature sequence,
647 corresponding to C21 in GH^{lev} and C24 in $SGBP^{lev}$ precursor amino acid sequences. The PCR
648 product encoding GH^{lev} was digested with NcoI and XhoI and ligated into pET28b, resulting in
649 fusion of the coding sequence to a C-terminal LEHHHHHH tag. The PCR product encoding
650 $SGBP^{lev}$ was digested with NdeI and XhoI and ligated into pET28a, resulting in fusion of the
651 $SGBP^{lev}$ coding sequence to an N-terminal MGSSHHHHHHSSGLVPRGSHM tag. TOP10
652 cells were transformed with the ligation mixtures and plated on LB agar plates with kanamycin.
653 After overnight incubation at 37°C, clones were screened for successful ligation by colony

654 PCR. GH^{lev} and $SGBP^{lev}$ variants were generated using the Q5 site directed mutagenesis kit
655 (NEB). All constructs were verified by Sanger sequencing.
656
657 **Expression and purification of proteins from *E. coli***. GH^{lev}, $SGBP^{lev}$ and their variants were
658 overexpressed in *E. coli*. Electrocompetent *E. coli* BL21(DE3) cells were transformed with the
659 appropriate plasmid, plated on LB kanamycin plates and incubated at 37°C overnight. Starter
660 LB kanamycin cultures were inoculated by scraping the transformants and incubated at 37°C,
661 180 rpm for 2 hours. 13 ml of the starter culture were used to inoculate each litre of LB
662 kanamycin. The cultures were incubated at 37°C, 180 rpm until OD600 0.5-0.6 and induced
663 with 0.2 mM IPTG. The temperature was then lowered to 20°C and the cultures were
664 incubated for a further 19-21 h. The cells were harvested by centrifugation and the pellets
665 were stored at -20°C.
666
667 The pellets were thawed, supplemented with DNase I and homogenised in TBS buffer. The
668 cells were lysed with a single pass at 25 kpsi through a cell disruptor (Constant Systems). The
669 lysates were supplemented with 1 mM PMSF. Unbroken cells were pelleted by centrifugation
670 for 30 min at 30,000g, 4°C. The supernatants were loaded on a 5 ml Ni^{2+}-charged chelating
671 sepharose column by gravity flow at room temperature. The column was washed with 40
672 column volumes TBS buffer containing 30 mM imidazole, and the bound proteins were eluted
673 with 5 column volumes TBS buffer containing 250 mM imidazole. The eluates were
674 concentrated in an Amicon Ultra filtration device (30 kDa cut-off) by centrifugation. The
675 samples were then loaded in batches on a HiLoad 16/600 Superdex 200 pg column (Cytiva)
676 in 10 mM HEPES-NaOH pH 7.5, 100 mM NaCl. Elution fractions were collected and analysed
677 by SDS-PAGE for purity. Fractions containing the proteins of interest were pooled,
678 concentrated, flash-frozen in liquid nitrogen, and stored at -80°C.
679
680 **Fructooligosaccharide production**. FOS used for cryoEM, crystallography and ITC were
681 generated by partial digestion of *Erwinia herbicola* levan (Sigma) by GH^{lev} (Bt1760) as
682 described previously[12].
683
684 **Isothermal titration calorimetry**. Protein samples were thawed, centrifuged to remove any
685 aggregates, and diluted to 25 or 50 mM in 10 mM HEPES-NaOH pH 7.5, 100 mM NaCl. Levan
686 from *E. herbicola* or defined-length FOS were dissolved in the same buffer to 8 mg/ml and 1
687 mM, respectively. ITC was performed using a Microcal PEAQ-ITC instrument (Malvern
688 Panalytical). Levan or FOS was injected into the sample cell containing protein or buffer. The
689 titrations were performed at 25°C. The sample cell was stirred at 750 rpm. After an initial delay
690 of 60 s, an injection of 0.4 ml was done (which was discarded from data analysis) followed by

691 18 injections of 2 ml. The spacing between injections was 150 s. Ligand to buffer control
692 titrations were subtracted from all experiments. The experiments were repeated at least twice.
693 Data were fitted to a single binding site model using the Microcal PEAQ-ITC Analysis software
694 v1.40. It was impossible to determine the precise molar concentration of the levan titrant due
695 to heterogeneity in chain length. Therefore, the molarity of available binding sites was
696 estimated during data fitting. For GH^{lev}, the only secondary binding site substitution that had
697 an effect on the affinity was W217A. We assumed that all the affinity observed for this variant
698 could be attributed to binding to the active site alone. Therefore, the stoichiometry was fixed
699 to 1 and the ligand concentration was floated during data fitting. The estimated molar
700 concentration of 0.8% levan was 829 μM. N.B. the "molarity" in this case corresponds to the
701 number of accessible binding sites for the enzyme on the polymeric levan ligand, rather than
702 number of molecules, per volume. This titrant concentration was fixed for all other data fits for
703 GH^{lev}. Similarly, by fixing n to 1, the levan titrant concentration for $SGBP^{lev}$ was determined to
704 be 1.48 mM, suggesting ~2x the number of binding sites on levan available to the SGBP as
705 to the enzymes active site. Notably, the affinity of $SGBP^{lev}$ for levan determined this way was
706 similar to that determined using defined-length FOS with known molarity (**ED figure 9**).
707
708 **Protein crystallisation**. GH^{lev} (Bt1760_SeMet) was crystalized in the presence of 200 mM
709 potassium/sodium tartrate, 100 mM sodium citrate pH 5.6, 1.4 M ammonium sulphate and 500
710 mM fructose. GH^{lev}_D42N crystal forms were crystalized in 1.5 M lithium sulphate and 200
711 mM ammonium sulphate, 100 mM MES pH 6.5 and 30% PEG 5000 MME respectively.
712 $SGBP^{lev}$ (both native and SeMet protein) was crystallised using 1.8-2.2 M $(NH_4)_2SO_4$, 0.1 M
713 MES pH 6.5. The protein concentrations were in the range of 10 mg/ml. The drops, composed
714 of 0.1 µl or 0.2 µl of protein solution plus 0.1 µl of reservoir solution, were set up using a
715 Mosquito crystallization robot (SPT Labtech). The vapor diffusion sitting drop method was
716 used and the plates were incubated at 20 °C. If required, crystal hits were optimised via
717 hanging drop vapour diffusion with larger volume drops (typically 1-1.5 µl). GH^{lev}_SeMet
718 samples did not require additional cryoprotection. GH^{lev}_D42N samples were cryoprotected
719 with paratone-N and with addition of 20% PEG 400 to the reservoir respectively. $SGBP^{lev}$
720 samples were cryoprotected by adding 4-fold excess of 3.5 M $(NH_4)_2SO_4$ to the crystal drops.
721
722 **Data collection, structure solution, model building, refinement and validation**. Diffraction
723 data were collected at the synchrotron beamlines I02, I03 and I04 of Diamond Light Source
724 (Didcot, UK) at a temperature of 100 K. The data set for GH^{lev} SeMet was integrated with
725 DIALS[44] via XIA2[45] and scaled with Aimless[46]. The space group was confirmed with
726 Pointless[47]. The phase problem was solved by experimental phasing with Crank2[48]. Mutant
727 GH^{lev}_D42N data sets were integrated by XDS[49] and processed as above. After phase transfer

728 from experimental phasing the automated model building program task CCP4build on
729 CCP4cloud[50] delivered models with Rfactors below 30 %. The models were refined with
730 Refmac[51] and manual model building with COOT[52]. The final models were validated with
731 COOT and MolProbity. Data collection and refinement statistics are presented in **ED Table 1**.
732 Other software used were from CCP4 suite[53]. Data collected for SeMet $SGBP^{lev}$ allowed
733 solving the phase problem and partial model building via single anomalous dispersion (Se-
734 SAD) using Phenix AUTOSOL[54]. Iterative rounds of manual building within COOT and model
735 refinement in Phenix resulted in a partial model with R_{free} ~35%, which was used as the input
736 for model completion in the cryo-EM maps of the levan utilisome with long FOS. The segments
737 missing from the $SGBP^{lev}$ X-ray model could not be modelled using the complete cryo-EM
738 structure due to the poor quality of the X-ray electron density maps in the missing regions.
739
740 **Levan utilisome cryo-EM sample preparation, data collection and image processing**. A
741 sample of the purified apo levan utilisome complex solubilised in DDM-containing buffer (10
742 mM HEPES, pH 7.5, 100 mM NaCl, 0.03% DDM) was prepared at 3 mg/ml. Lacy carbon 300-
743 mesh copper grids (Agar Scientific) were glow-discharged in air (10 mA, 30s, Cressington
744 208). A sample volume of 3.5 mL was applied to the grid. Blotting and plunge freezing into
745 liquid nitrogen-cooled liquid ethane were carried out using an FEI Vitrobot Mark IV (Thermo
746 Fisher Scientific) with chamber conditions set at a temperature of 4 °C and 100% relative
747 humidity. The grid was blotted for 6 s with a blot force of 6. Micrograph movies were collected
748 on a Titan Krios Microscope (Thermo Fisher Scientific) operating at 300 kV with a Falcon III
749 direct electron detector operating in counting mode. All cryoEM data in this study was collected
750 using EPU v1.20.3.10 (Thermo Fisher Scientific). Data acquisition parameters can be found
751 in **ED Table 2**. Density maps coloured according to local resolution, together with angular
752 distribution plots, are included for each EM map described in this study (Supplementary Figure
753 4).
754
755 An initial dataset comprising 2057 micrograph movies was collected and image processing
756 was carried out in Relion3.1[55]. Drift correction and dose-weighting was carried out using
757 MOTIONCOR2[56]. CTF estimation of motion corrected micrographs was performed using
758 Gctf[57]. Template-based particle picking within Relion was hindered by the large amount of
759 carbon present in many micrographs. The micrograph stack was therefore manually culled to
760 remove micrographs containing >50% carbon, leaving 1093 micrographs for further
761 processing. Final particle picking was performed using the crYOLO general model[58], yielding
762 96,639 particles. This particle stack was subjected to several rounds of 2D classification, after
763 which 89,305 particles remained. A 3D starting model was generated *de novo* from the data
764 using the stochastic gradient descent algorithm within RELION. 3D classification was used to

765 isolate particles corresponding to the complete octameric utilisome complex (45,594). These
766 particles were subjected to further rounds of classification in 3D to assess the conformational
767 heterogeneity of the complex. Classification revealed considerable heterogeneity in the
768 position of the SusD lids, with positions described as 'wide open' (W) and 'narrow open' (N)
769 identified in all possible combinations. WW, WN and NN states contained 16,155, 22,452 and
770 6,987 particles respectively. To increase particle numbers and improve the results of
771 downstream processing a second dataset of 3142 movies was collected. These were
772 processed in the same way as described for the initial dataset, with particles picked using
773 crYOLO. Classification in 3D yielded 146,512 particles that corresponded to the complete
774 octameric utilisome complex. To improve the resolution for the more static, C2 symmetric
775 portions of the utilisome ($SusC^{lev}$, GH^{lev} and N-terminal region of the $SGBP^{lev}$) particles stacks
776 corresponding to the complete octameric complex from both datasets (192,106 total) were
777 combined and subject to focused refinement with C2 symmetry. The mask applied in focused
778 refinement excluded the SusD subunits. Particles were subjected to iterative rounds of CTF-
779 refinement and Bayesian polishing (run separately for each dataset) until no further
780 improvement in resolution was seen. Post-processing was performed using a soft, extended
781 mask and yielded a global sharpened reconstruction at 3.5 Å, as estimated by the gold
782 standard Fourier shell correlation using the 0.143 criterion.
783
784 **Active levan utilisome in the presence of FOS (DP8-12) cryo-EM sample preparation,**
785 **data collection and image processing**. A sample of the levan utilisome containing an active
786 GH^{lev} solubilised in a DDM-containing buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.03%
787 DDM) was prepared at 3 mg/ml and incubated with 0.5 mM levan FOS with a degree of
788 polymerisation of ~8-12 for at least one hour before grid preparation. Quantifoil carbon grids
789 (R1.2/1.3, 300 mesh) were glow discharged (30 mA, 60 s, Quorum GloQube) in the presence
790 of amylamine vapour. A sample volume of 3.5 mL was applied to the grid. Blotting and plunge-
791 freezing into liquid nitrogen-cooled liquid ethane were carried out using an FEI Vitrobot Mark
792 IV (Thermo Fisher Scientific) with chamber conditions set at a temperature of 4 °C and 100%
793 relative humidity. The grid was blotted for 6 s with a blot force of 6. Micrograph movies were
794 collected on a Titan Krios Microscope (Thermo Fisher Scientific) operating at 300 kV with a
795 Falcon III direct electron detector operating in counting mode. Data acquisition parameters
796 can be found in **ED Table 2**.
797
798 A dataset comprising 974 micrograph movies was collected and image processing was carried
799 out in RELION 3.1 [55]. Drift correction and dose-weighting was done using MOTIONCOR2 [56].
800 CTF estimation of motion corrected micrographs was performed using Gctf [57]. Particle picking
801 was performed using the general model of crYOLO and yielded 72,373 particles [58]. Unwanted

802 particles/contamination were removed from the particle stack through two rounds of 2D
803 classification, after which 63,789 particles remained. Classification in 3D was used to address
804 compositional heterogeneity. Good classes containing unambiguous $SusC_2D_2$ density
805 represented the complete octameric utilisome, a hexameric assembly which lacked one GH^{lev}
806 and one $SGBP^{lev}$ subunit and a naked $SusC_2D_2$ complex. Contributing particles numbers were
807 17,045, 31,789 and 7390, respectively. SusD lids invariantly occupied a closed position and
808 conformational heterogeneity was limited to the position of the levan binding protein.
809
810 Complete utilisome particles were subjected to iterative rounds of CTF-refinement and
811 Bayesian polishing until no further improvement in resolution was seen. Post-processing was
812 performed using a soft, extended mask and yielded a global sharpened reconstruction at 3.2
813 Å, as estimated by the gold standard Fourier shell correlation using the 0.143 criterion.
814
815 To extract higher resolution information for the $SusCD^{lev}$ core complex, particle subtraction
816 was performed to remove signal for additional lipoprotein components from all experimental
817 images contributing to good classes (as defined above). A soft, expanded mask
818 encompassing only the $SusCD^{lev}$ core was generated using the volume eraser tool within
819 Chimera [59] before using the resulting carved volume as in input for mask creation in RELION.
820 Subtracted particles were used in a focused refinement with the same mask applied while
821 enforcing C2 symmetry. Iterative rounds of CTF-refinement and Bayesian polishing were
822 employed until no further improvement in resolution was observed. Post-processing resulted
823 in a final sharpened reconstruction at 2.9 Å.
824
825 **Inactive levan utilisome in the presence of FOS (DP15-25) cryo-EM sample preparation,**
826 **data collection and image processing**. A sample of the levan utilisome containing
827 inactivated GH^{lev} (D42A), solubilised in a DDM-containing buffer (10 mM HEPES, pH 7.5, 100
828 mM NaCl, 0.03% DDM), was prepared at 3 mg/ml and incubated with ~0.5 mM levan FOS
829 with a degree of polymerisation 15-25 for at least one hour before grid preparation. Grid type,
830 preparation, microscope and detector were the same as for the active levan utilisome
831 described above. Data acquisition parameters can be found in **ED Table 2**.
832
833 A dataset comprising 1388 micrograph movies was collected and image processing was
834 carried out in RELION 3.1 [55]. Drift correction and dose-weighting were carried out using
835 MOTIONCOR2 [56]. Particle picking was performed using the general model of crYOLO and
836 yielded 157,953 particles[58]. Unwanted particles/contamination were removed from the particle
837 stack through two rounds of 2D classification, after which 146,056 particles remained.
838 Classification in 3D was used to address compositional heterogeneity. Good classes

839 representing the complete octameric utilisome and the hexameric assembly lacking one
840 SGBP^{lev} and one GH^{lev} subunit were observed and contained 78,469 and 42,488 particles,
841 respectively. Conformational heterogeneity in the position of the levan binding protein was
842 observed with some classes possessing a conformation where this subunit was held close to
843 GH^{lev}. Particles contributing to all classes with evidence of this docked conformation of the
844 levan binding protein were pooled (98,755) and a consensus refinement was carried out. CTF
845 refinement and Bayesian polishing were performed iteratively until no further improvement in
846 resolution was observed. The resulting reconstruction possessed weaker density for the C-
847 terminal domain of the levan binding protein than for the N-terminal portions. To improve this,
848 a focused classification approach without alignment was used.
849
850 A mask encompassing only the docked position of the SGBP^{lev} with some surrounding density
851 was created using the volume eraser tool in Chimera followed by mask creation in RELION.
852 A focused 3D classification job without alignment was run using the aforementioned mask and
853 the output from the aforementioned refinement as a reference model. The reference model
854 was low-pass filtered to 3.5 Å, just above the resolution of 3.3 Å reported for the consensus
855 refinement, thus allowing classification on high resolution features. Several T values ranging
856 from 20 to 70 were empirically tested and a T value of 40 was found to give the best results.
857 A single class, containing 27,310 particles, was identified that possessed well resolved density
858 for the C-terminal domain of SGBP^{lev}. A particle star file containing information for particles
859 contributing to this class was created manually via command line arguments. From this, two
860 new star files were generated that contained random half sets of the selected data. Using
861 relion_reconstruct, these star files were used to generate two independent half maps that
862 corresponded to the unmasked structure. Post-processing using these generated half maps
863 yielded a sharpened reconstruction of 3.0 Å, as estimated by gold standard Fourier Shell
864 correlations using the 0.143 criterion. The density for the C-terminal domain of the SGBP^{lev}
865 was improved, and density corresponding to levan chain that links SGBP^{lev} to the GH^{lev} was
866 also visible.
867
868 To obtain the highest quality density for FOS molecules occupying the SusCD^{lev} binding cavity,
869 all particles were considered regardless of lipoprotein complement or conformation. A particle
870 subtraction and focused refinement strategy targeting the SusCD^{lev} core of the complex was
871 used as described for the active levan utilisome. Post-processing of the model arising from
872 this final C2 symmetrised refinement resulted in a sharpened reconstruction at 2.7 Å.
873
874 **Dextran utilisome cryo-EM sample preparation, data collection and image processing**.
875 A sample of the dextran utilisome complex (Bt3087-Bt3090) solubilised in a DDM-containing

876 buffer (10 mM HEPES, 100 mM NaCl, pH 7.5, 0.03 % DDM) was prepared at 0.05 mg/ml.
877 Lacy carbon 300-mesh copper grids coated with a <3 nm continuous carbon film (Agar
878 Scientific) were glow-discharged in air (10 mA, 30 s). A sample volume of 3.5 mL was applied
879 to the grid. Blotting and plunge-freezing were performed 10 seconds after loading the sample
880 onto the grid using an FEI Vitrobot Mark IV (Thermo Fisher Scientific) with chamber conditions
881 set at a temperature of 4 °C and 100% relative humidity. The grid was blotted for 6 s with a
882 blot force of 0. Micrograph movies were collected on a Titan Krios Microscope (Thermo Fisher
883 Scientific) operating at 300 kV with a Falcon IV direct electron detector operating in counting
884 mode. Data acquisition parameters can be found in **ED Table 2.**
885
886 A dataset comprising 6,331 micrographs was collected and image processing was carried out
887 in RELION 3.1 [55]. Drift correction and dose-weighting were performed using RELION's own
888 implementation of MOTIONCOR2. CTF estimation of motion corrected micrographs was done
889 using CTFFIND4 [60]. Particle picking was done using the crYOLO general model which
890 identified 820,184 particles in the micrographs. Junk particles and contaminants were
891 removed through several rounds of 2D classification, after which 477,707 particles remained.
892 An initial model was generated *de novo* from the data. Extensive 3D classification was used
893 to address the considerable compositional heterogeneity that was present in the data (see **ED**
894 **Fig. 6**). Each unique composition was refined and sharpened independently. A consensus
895 refinement was carried out, with iterative rounds of CTF-refinement and Bayesian polishing
896 until no further improvement in resolution was seen. A final, sharpened consensus
897 reconstruction was obtained at 3.1 Å.
898
899 **Model building into electron microscopy maps**. Buccaneer (part of CCP-EM v1.5.0)[61,62]
900 was used to build the initial protein models into the post-processed consensus inactive
901 levanase map, resulting in almost complete protein models. An AlphaFold2 [23] prediction of the
902 dextran SusC was generated and used as an initial model. Manual modelling and real space
903 refinement of protein and FOS chains were performed iteratively using COOT[52] and Phenix[54],
904 respectively. The completed protein and FOS models were placed into other maps and real
905 space refined. The acyl-cysteine was designated as the first residue of each lipoprotein. Model
906 refinement statistics are presented in **ED Table 2**.
907
908 **Density analysis and figure making**. Investigation and comparison of EM density maps was
909 performed using Chimera[59] and COOT[52]. Figures of maps and models were generated using
910 Chimera and ChimeraX[63]. To aid interpretability of EM density in generated figures, some
911 maps were filtered using LAFTER[64]. Maps processed in this way are clearly indicated in the

912 corresponding figure legend. The masks supplied in filtering were the same masks used for
913 post-processing within RELION.
914
915 **Additional References**

916 39. Zougman, A., Selby, P. J. & Banks, R. E. Suspension trapping (STrap) sample preparation
917 method for bottom-up proteomics analysis. *PROTEOMICS* **14**, 1006–1000 (2014).
918 40. HaileMariam, M. *et al.* S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun
919 Proteomics. *Journal of Proteome Research* **17**, 2917–2924 (2018).
920 41. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized
921 p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature*
922 *Biotechnology 2008 26:12* **26**, 1367–1372 (2008).
923 42. Cox, J. *et al.* Andromeda: A peptide search engine integrated into the MaxQuant
924 environment. *Journal of Proteome Research* **10**, 1794–1805 (2011).
925 43. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing
926 and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
927 44. Winter, G. *et al.* DIALS: Implementation and evaluation of a new integration package.
928 *Acta Crystallographica Section D: Structural Biology* **74**, 85–97 (2018).
929 45. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction.
930 *Journal of Applied Crystallography* **43**, 186–190 (2010).
931 46. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta*
932 *Crystallographica Section D: Biological Crystallography* **69**, 1204–1214 (2013).
933 47. Evans, P. Scaling and assessment of data quality. in *Acta Crystallographica Section D:*
934 *Biological Crystallography* vol. 62 72–82 (International Union of Crystallography, 2006).

935 48. Skubak, P. *et al.* A new MR-SAD algorithm for the automatic building of protein models
936 from low-resolution X-ray data and a poor starting model. *urn:issn:2052-2525* **5**, 166–
937 171 (2018).
938 49. Kabsch, W. XDS. *Acta Crystallographica Section D: Biological Crystallography* **66**, 125
939 (2010).
940 50. Krissinel, E., Uski, V., Lebedev, A., Winn, M. & Ballard, C. Distributed computing for
941 macromolecular crystallography. *Acta crystallographica. Section D, Structural biology*
942 **74**, 143–151 (2018).
943 51. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and
944 guidelines for its use. *Acta crystallographica. Section D, Biological crystallography* **60**, 2184–
945 2195 (2004).
946 52. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot.
947 *Acta crystallographica. Section D, Biological crystallography* **66**, 486–501 (2010).
948 53. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta*
949 *Crystallographica Section D: Biological Crystallography* **67**, 235 (2011).
950 54. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and
951 electrons: recent developments in Phenix. *Acta crystallographica. Section D, Structural*
952 *biology* **75**, 861–877 (2019).
953 55. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure
954 determination in RELION-3. *eLife* **7**, (2018).
955 56. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for
956 improved cryo-electron microscopy. *Nature Methods 2017 14:4* **14**, 331–332 (2017).
957 57. Zhang, K. Gctf: Real-time CTF determination and correction. *Journal of structural biology*
958 **193**, 1–12 (2016).

959 58. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for
960 cryo-EM. *Communications Biology 2019 2:1* **2**, 1–13 (2019).
961 59. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and
962 analysis. *Journal of computational chemistry* **25**, 1605–12 (2004).
963 60. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from
964 electron micrographs. *Journal of Structural Biology* **192**, 216–221 (2015).
965 61. Hoh, S. W., Burnley, T. & Cowtan, K. Current approaches for automated model building
966 into cryo-EM maps using Buccaneer with CCP-EM. *Acta crystallographica. Section D,*
967 *Structural biology* **76**, 531–541 (2020).
968 62. Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software
969 suite. *Acta crystallographica. Section D, Structural biology* **73**, 469–477 (2017).
970 63. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and
971 analysis. *Protein science : a publication of the Protein Society* **27**, 14–25 (2018).
972 64. Ramlaul, K., Palmer, C. M. & Aylett, C. H. S. A Local Agreement Filtering Algorithm for
973 Transmission EM Reconstructions. *Journal of Structural Biology* **205**, 30–40 (2019).
974
975 **Acknowledgements**
976 JBRW was supported by a Wellcome Trust 4-year Ph.D. studentship (215064/Z/18/Z). BvdB
977 is funded by a Wellcome Trust Investigator award (214222/Z/18/Z), supporting AS and YL.
978 MF is supported by a Newcastle University studentship. We acknowledge the Diamond Light
979 Source (Didcot, UK) for beam time (proposals mx306, mx1221, mx13587 and mx18598), and
980 thank the staff of beamlines I02, I03, I04 and I24 for support. All cryoEM was performed at the
981 Astbury Biostructure Laboratory, which was funded by the University of Leeds and the
982 Wellcome Trust (108466/Z/15/Z, 221524/Z/20/Z). We thank R. Thompson, E. Hesketh and D.
983 Maskell for EM support. Protein ID mass spectrometry was performed at the University of
984 Leeds Mass Spectrometry Facility. We thank J. Ault and R. George for performing this
985 analysis.

987
988 **Author contributions**
989 JBRW performed cryo-EM and determined cryo-EM structures, supervised by NAR. AS
990 purified proteins, determined X-ray crystal structures and carried out ITC, supervised by BvdB.
991 MF purified proteins. YL prepared OM samples for proteomics. TH and AG performed
992 proteomics, supervised by MT. HZ purified Bt1760, supervised by DNB. SF collected X-ray
993 crystallography data for Bt1760. AB solved Bt1760 crystal structures and managed the
994 Newcastle Structural Biology Laboratory. BvdB generated *B. theta* strains, purified proteins,
995 and crystallised Bt1761. JBRW, AS, DNB, BvdB and NAR wrote the manuscript.
996
997 **Competing Interests Statement**
998 The authors declare no competing interests.
999
1000
1001
1002 **Data availability**
1003 The data supporting the findings of this study are available from the corresponding authors
1004 upon reasonable request. Cryo-EM reconstructions and corresponding coordinates have been
1005 deposited in the Electron Microscopy Data Bank and the Protein Data Bank respectively:
1006 Substrate free levan utilisome (EMD-15288, PDB ID 8A9Y), levan utilisome with FOS DP 8-
1007 12 (EMD-15289, PDB ID 8AA0), $SusC_2D_2$ core from the levan utilisome with FOS DP 8-12
1008 (EMD-15290, PDB ID 8AA1), inactive levan utilisome with FOS DP 15-25 (EMD-15291, PDB
1009 ID 8AA2), $SusC_2D_2$ core from inactive levan utilisome with FOS DP 15-25 (EMD-1592, PDB
1010 ID 8AA3), dextran utilisome consensus refinement (EMD-15293, PDB ID 8AA4). Raw cryo-
1011 EM movies will be deposited in the EMPIAR database. Coordinates and structure factors from
1012 X-ray crystallography experiments for GH^{lev} have been deposited in the Protein Data Bank
1013 under the accession codes 7ZNR and 7ZNS. The mass spectrometry proteomics data have
1014 been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with
1015 the dataset identifier PXD034863. Raw data from this study are available at the University of
1016 Leeds Data Repository: https://doi.org/10.5518/1329
1017
1018
1019
1020 **Extended Data Legends**
1021 **Extended Data Figure 1. The levan and dextran utilisomes. (a)** Organisation of the levan
1022 PUL showing relative gene positions within the PUL, with functions indicated. The four OM-
1023 associated PUL components ($SusC^{lev}$, $SusD^{lev}$, GH^{lev} and $SGBP^{lev}$) are highlighted by the grey

box. An X-ray structure of GH^{lev} (Bt1760; GH32 endo-levanase) is shown (blue; PDB-ID: 6R3R). The AlphaFold2-predicted model for $SGBP^{lev}$ (Bt1761) is shown (pink) oriented such that the N-terminus is at the bottom and the proposed (C-terminal) levan binding domain is at the top. Note that the N-termini of GH^{lev} and $SGBP^{lev}$ will be lipidated and associated with the outer leaflet of the OM. The cryo-EM structure of the dimeric $SusCD^{lev}$ complex in its open-open state is shown ($SusC^{lev}$ is green, $SusD^{lev}$ is 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 GH^{dex} (Bt3087; GH66 endo-dextranase), the putative $SGBP^{dex}$ (Bt3088) and the $SusCD^{dex}$ complex, are shown coloured as for the levan PUL in (**a**). **(c)** SDS-PAGE of the previously-studied sample of LDAO-purified $SusCD^{lev}$ [10] before (asterisk) and after boiling. The boiled sample shows two weak bands in addition to those for $SusC^{lev}$ and $SusD^{lev}$, which were subsequently identified as GH^{lev} and $SGBP^{lev}$ by mass spectrometry. (**d**) A class average obtained during 3D classification of the levan $SusC_2D_2$ core complex. The SusC and SusD components (green and grey respectively) are docked into the density. A large region of density remains unassigned (orange). (**e**) Isolated view of the previously unassigned density with the crystal structure of GH^{lev} (blue cartoon) fitted into the EM density (blue) as a rigid body. The remaining density was therefore attributed to $SGBP^{lev}$ and is coloured magenta.

Extended Data Figure 2. Classification of levan utilisome Data. (**a**) Output of the first round of 3D classification for apo utilisome data. Yellow, purple and pink classes represent the octameric complex *i.e*. the complete octameric utilisome. The green class shows the additional lipoproteins associated with just one SusC unit whilst the blue class shows that a small proportion of $SusC_2D_2$ core complex was present. (**b**) Output of 3D classification for the levan utilisome with an active levanase in the presence of FOS DP8-12. Classes (viewed in the plane of the membrane) containing particles of the complete octameric complex were observed (blue and green) as well as hexameric complexes containing a single copy of the $SGBP^{lev}$ and GH^{lev} (pink and yellow). A class containing $SusCD^{lev}$ in isolation is also present (purple). (**c**) Outputs of 3D classification for long FOS (DP15-25) showing that $SGBP^{lev}$ can adopt a 'docked' conformation proximal to both the SusD and levanase. (**d**) A consensus refinement of all classes containing at least one docked SGBP (yellow, pink, cyan and green in panel (**c**)). A mask was created around the region of interest (transparent yellow). (**e**) Outputs of focused classification on the masked region without alignment. A class displaying high resolution for the region of interest is marked with a red asterisk. Independent half maps were reconstructed using unmasked particles belonging to this class. (**f**) Sharpened reconstruction generated with the aforementioned half maps showing improved density for $SGBP^{lev}$.

1061

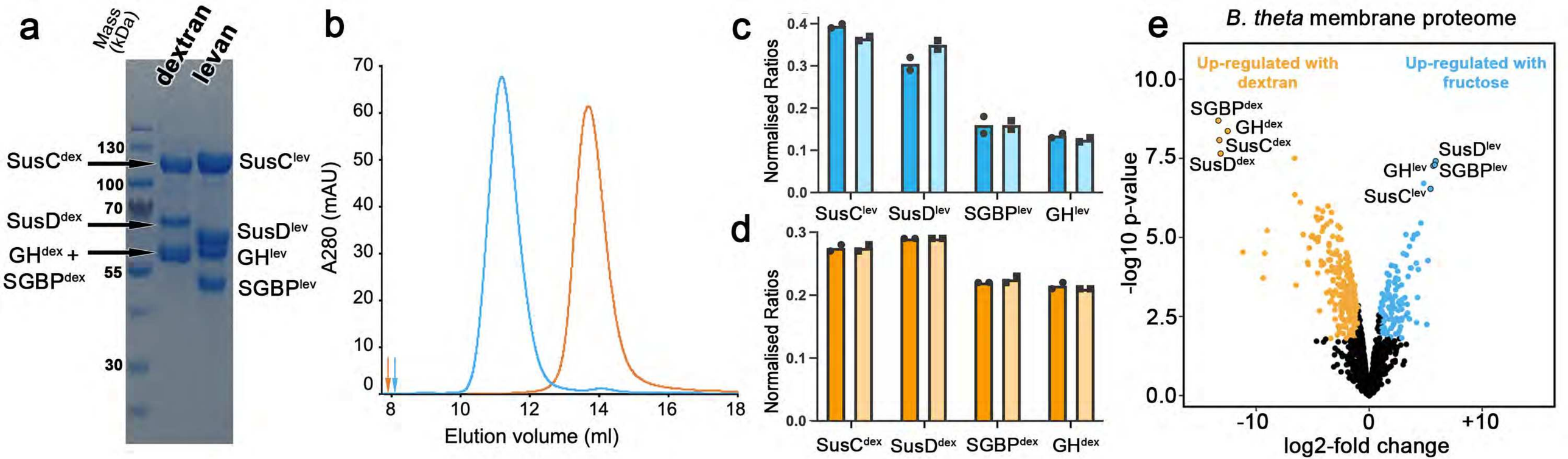
1062 **Extended Data Figure 3. Conformational variability in apo and substrate-bound levan**
1063 **utilisomes.** (**a**) 3D classification of apo levan utilisome viewed from outside the cell. $SusC^{lev}$
1064 (green), $SusD^{lev}$ (grey), $SGBP^{lev}$ (magenta), and GH^{lev} (blue). Classes are separated on their
1065 $SusD^{lev}$ lid positions. Wide-wide (WW), normal-wide (NW), and normal-normal (NN) open
1066 states (from left to right) (**b**) Overlay of the wide (SusD grey) and normal (SusD orange) open
1067 states of the complex. (**c**) Overlay of atomic models for the normal versus wide open state
1068 generated by a rigid-body fit of $SusD^{lev}$ into the cryoEM density. A monomer is shown for clarity
1069 and an asterisk marks the same $SusD^{lev}$ helix in both models. (**d**) A view of the utilisomes
1070 shown at high threshold in the plane of the membrane (left). Different conformations of the
1071 $SGBP^{lev}$ observed in 3D classification are overlaid to demonstrate the flexibility of this subunit
1072 (boxed region). The same view rotated 90° is shown (right). Disordered micelle density is
1073 shown as translucent grey. **(e)** Variability of the $SGBP^{lev}$ position in the substrate-bound
1074 utilisomes with short FOS (~DP8-12) and an active GH^{lev}, and long FOS (DP15-25) with an
1075 inactive GH^{lev}. A novel state is uniquely observed in the long FOS structure with one $SGBP^{lev}$
1076 (orange) reaching across and contacting the $SGBP^{lev}$ associated with the other SusC subunit
1077 that is present in a docked state. This conformation is consistent with both $SGBP^{lev}$ subunits
1078 in the utilisome interacting with the same chain of substrate.

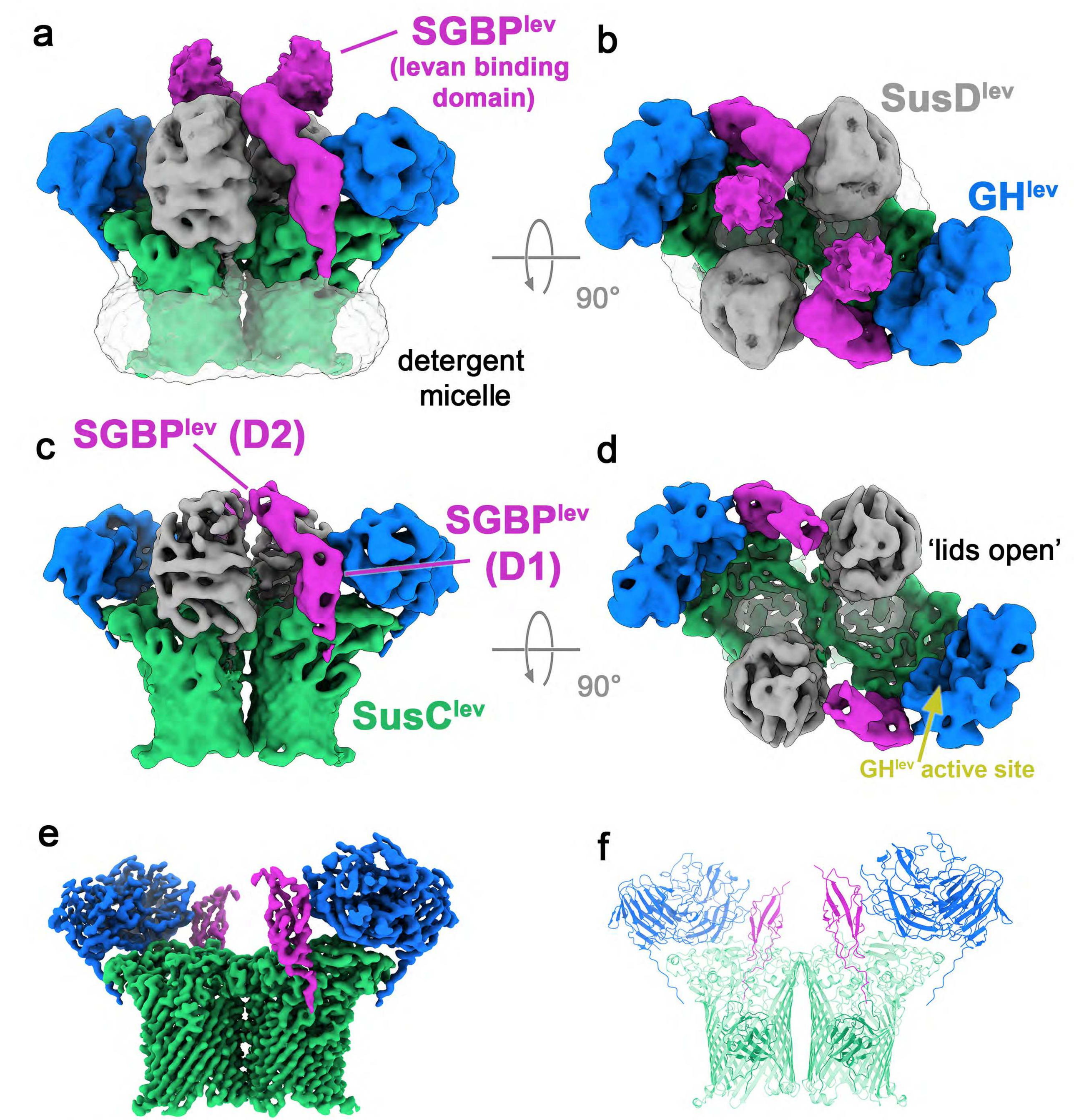
1079 **Extended Data Figure 4. SusC extracellular loops that contribute to lipoprotein interface**
1080 **interactions differ in the levan and dextran utilisomes. a,** Arrangement of GH^{lev} and
1081 $SGBP^{lev}$ on SusC in the levan utilisome. GH^{lev} makes contacts with extracellular loop 1 (gold)
1082 and extracellular loop 9 (red), while $SGBP^{lev}$ only makes contacts with extracellular loop 1. **b,**
1083 Arrangement of GH^{dex} and $SGBP^{dex}$ on SusC in the dextran utilisome. Here, extracellular loop
1084 1 of $SusC^{dex}$ is the primary site of interaction for GH^{dex}, while extracellular loop 9 comprises
1085 the interface with $SGBP^{dex}$. For clarity, one half of the utilisome is shown in each case, and
1086 SusD components are omitted. Note that the dextran utilisome model is a composite of cryo-
1087 EM structures ($SusC^{dex}$) and predicted models from AlphaFold2 (GH^{dex} and $SGBP^{dex}$).

1088 **Extended Data Figure 5. CryoEM structure and heterogeneity of the dextran utilisome**
1089 **observed by cryo-EM.** Side (**a**) and top (**b**) view of the heptameric dextran utilisome map.
1090 The identical side (**c**) and top (**d**) views of a composite atomic model for dextran utilisome is
1091 shown. CryoEM data permitted refinement of $SusC^{dex}$. AlphaFold2 structure predictions for
1092 $SusD^{dex}$ and GH^{dex} were docked into the cryoEM map for the heptameric complex. An
1093 AlphaFold2 structure prediction for part of $SGBP^{dex}$ was also fit to the cryoEM map.
1094 Unambiguous density was visible only for the first two domains of $SGBP^{dex}$, and the predicted
1095 model was truncated prior to the C-terminal domain. $SusC^{dex}$=purple, $SusD^{dex}$=pink,

1096 GH^{dex}=cyan and $SGBP^{dex}$=green. The refinement for the heptameric complex had a global
1097 resolution of 3.1 Å. (**e**) Refined outputs of 3D classification viewed where each map
1098 corresponds to a unique complement or arrangement of auxiliary components (as labelled).
1099 (**f**) Schematic of the architecture for two apo glycan utilisomes. The levan utilisome (left) is
1100 coloured as in the main text ($SusC^{lev}$=green, $SusD^{lev}$=gray, GH^{lev}=blue, and
1101 $SGBP^{lev}$=magenta). The equivalent schematic for the substrate-free dextran utilisome is on
1102 the right. Note the different arrangement of the GH and SGBP components relative to SusD
1103 in the levan and dextran systems.
1104
1105 **Extended Data Figure 6. FOS binding by $SusCD^{lev}$.** (**a**) Isolated FOS density obtained from
1106 the levan utilisome dataset with active GH^{lev} and short FOS (DP8-12)[12]. Density for substrate
1107 (yellow) is shown at high (left) and low (right) thresholds. (b) Isolated FOS density obtained
1108 from the utilisome structure with inactive GH^{lev} and long FOS (DP15-25). Levan density
1109 (orange) is shown at high (left) and low (right) thresholds. Arrows indicate missing fructose
1110 branches relative to (**a**). At the FOS1 site, density for the putative β2,1 decoration on Frc4 is
1111 missing. Conversely, contiguous density extends beyond the previously resolved density at
1112 FOS2, with a novel β2,1 decoration on Frc5. The substrate bound at the FOS2 site follows a
1113 similar trend with the previously modelled β2,1 linked fructose side chain being much weaker
1114 with longer FOS, while additional density attributed to another β2,6 linked monomer extends
1115 the chain towards the FOS1 site. At higher threshold levels, density connects the FOS1 and
1116 FOS2 binding sites, indicating that longer FOS (~DP15) can occupy both sites simultaneously.
1117 The connecting density is weak and indicative of multiple conformations, consistent with the
1118 absence of any contacts from $SusC^{lev}$ to this segment. These data confirm that the transporter
1119 has considerable substrate binding promiscuity and that, as suggested previously, relatively
1120 long FOS (~15 DP) can be accommodated[12]. FOS models shown are from the original X-ray
1121 crystal structure of the $SusCD^{lev}$ complex determined in the presence of short FOS (DP6-12)[12].
1122 (**c)** Cryo-EM structure of the inactive GH^{lev} with FOS bound (blue) superposed with the two
1123 crystal structures (7ZNR and 7ZNS; orange, grey). (**d, e)** Comparison of FOS bound in the
1124 FOS3 (the active site) and FOS4 (secondary) binding sites of GH^{lev}. The arrowheads point to
1125 breaks in the FOS chain in the crystal structures, possibly as a result of using a lower DP FOS
1126 for co-crystallization than for cryo-EM. Views in (**d)** and **(e)** are generated from a superposition.
1127
1128 **Extended Data Figure 7. ITC of glycan binding to recombinant $SGBP^{lev}$and GH^{lev}.** (a)
1129 Titration of 1 mM defined-length FOS into 50 μM wild type $SGBP^{lev}$, suggests that ~15 fructose
1130 units are required for full affinity, which is abolished by the WAWA (W297A/W359A) mutation.
1131 (b) ITC titrations of 8 mg/ml levan, inulin or dextran 500 into 50 μM $SGBP^{lev}$ shows its

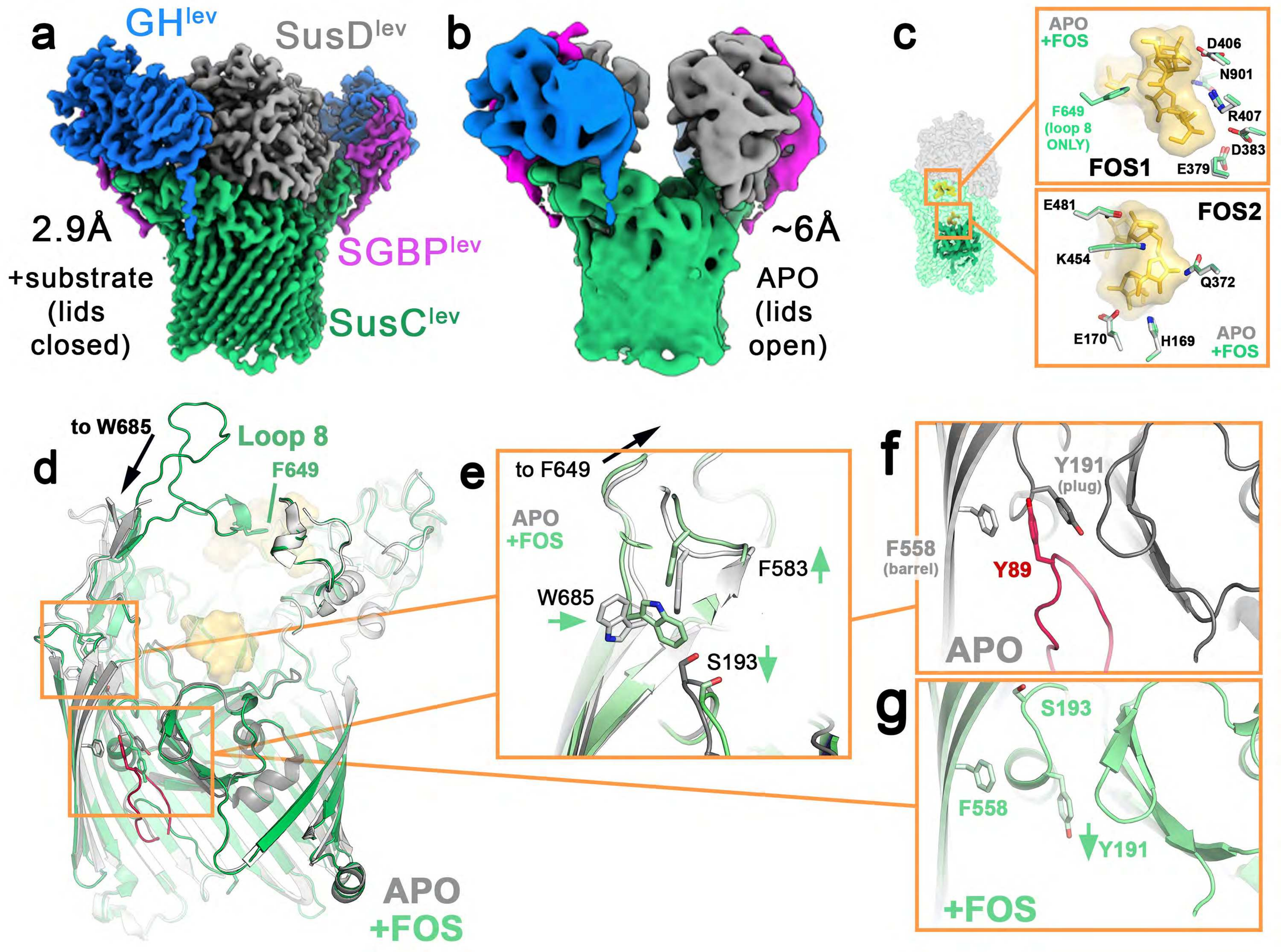
1132 specificity for levan. (c) ITC data from titrations of GH^{lev} variants (all indicated residues mutated
1133 to alanine in the inactive D42A GH^{lev} background). Levan (8 mg/ml) was titrated into 50 μM of
1134 indicated GH^{lev} variant. Data fitting assumptions are described in the methods. (d) Surface
1135 representation of the GH^{lev} model, with FOS shown as yellow sticks. Inset are zoomed views
1136 of the FOS3 (active site) and FOS4 (secondary) binding sites, in which atomic models in
1137 cartoon representation for FOS3 are shown with side chains for aromatic residues (Y70A,
1138 W318A). For the secondary binding site these residues are W217A, F243A, Y437A.
1139
1140 **Extended Data Figure 8. Conservation of the $SGBP^{lev}$ β2,6-FOS binding site. a,** Close up
1141 view of the $SGBP^{lev}$ FOS binding site. The aromatic and polar residues that likely interact with
1142 the FOS are shown as grey stick models. The cryo-EM structure of $SGBP^{lev}$ was aligned with
1143 selected homologue AlphaFold2-predicted models (**b**-**e**). **b,** *Bacteroides* sp. D2 $SGBP^{lev}$
1144 (UniProt E5CCB3). **c,** *Prevotella oralis* ATCC 33269 $SGBP^{lev}$ (E7RM14). **d,** *Flavobacterium*
1145 *commune* $SGBP^{lev}$ (A0A1D9P8I4). **e,** *F. cellulosilyticum* $SGBP^{lev}$ (A0A4R5CJN9). FOS-
1146 binding residues equivalent to those in **a** are shown as grey stick models (if present). The FOS
1147 chain from the $SGBP^{lev}$ cryo-EM model is shown in **b-e** for reference (orange and red). The
1148 identity indicated in each panel corresponds only to the C-terminal levan-binding domain
1149 sequence compared to $SGBP^{lev}$ from *B. theta*. Although we could not confidently identify which
1150 $SGBP^{lev}$ residues form hydrogen bonds with FOS from the cryo-EM maps, binding site
1151 conservation analysis indicates that N295, T350, Q352 and N384 of $SGBP^{lev}$ are likely
1152 involved in FOS binding. The amino acid sequence alignment of the models shown here can
1153 be found in Supplementary Figure 3.
1154
1155 **Extended Data Table 1. Crystallography data statistics and refinement details.**
1156
1157 **Extended Data Table 2. CryoEM**
1158 Data acquisition parameters and refinement statistics for the apo levan utilisome, the levan
1159 utilisome with short FOS, the levan utilisome with inactivated GH and long FOS, and the
1160 Dextran utilisome.
1161

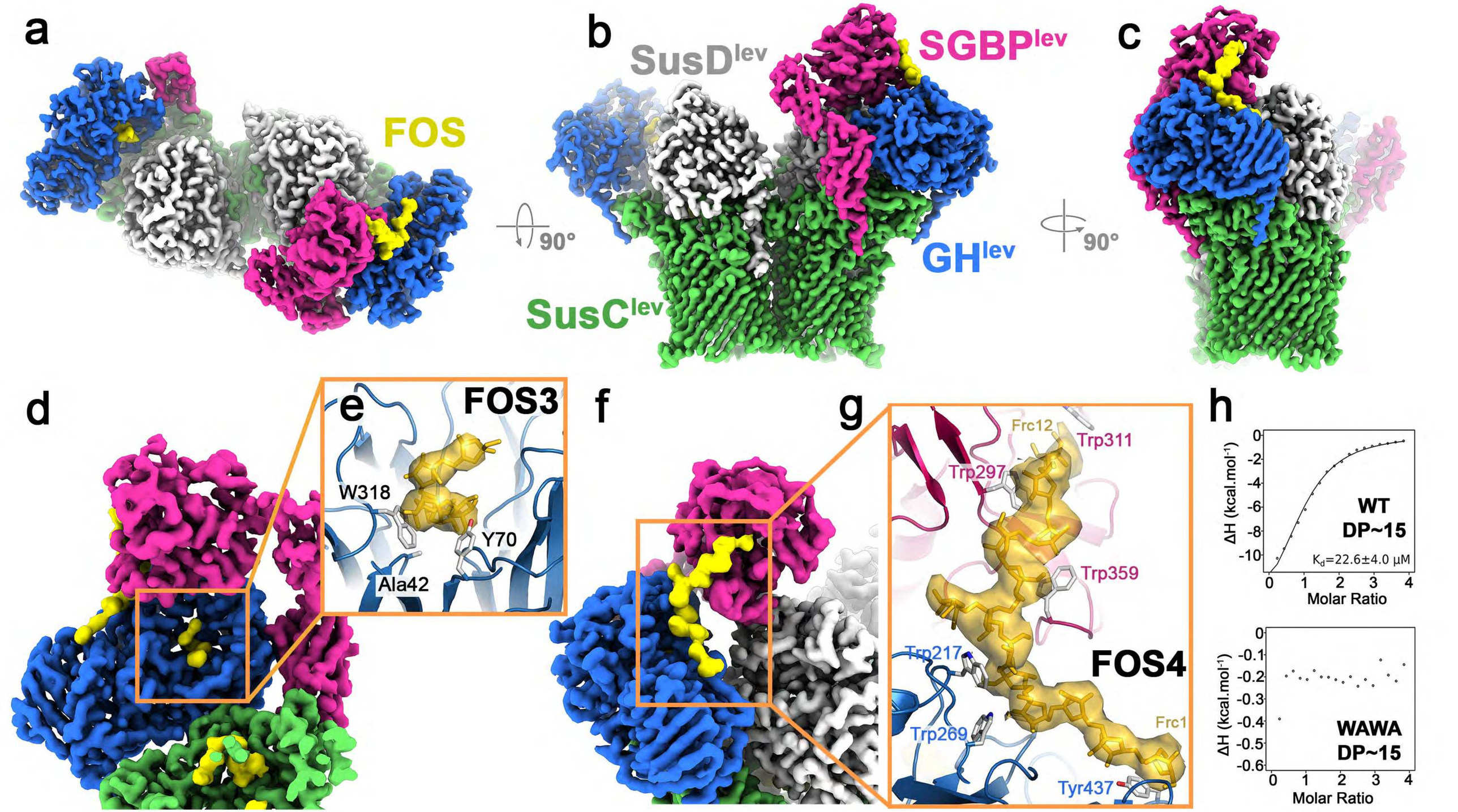

a
Mass (kDa)
dextran
levan
130
100
70
55
30
SusC^dex
SusD^dex
GH^dex +
SGBP^dex
SusC^lev
SusD^lev
GH^lev
SGBP^lev
b
A280 (mAU)
Elution volume (ml)
c
Normalised Ratios
SusC^lev
SusD^lev
SGBP^lev
GH^lev
d
Normalised Ratios
SusC^dex
SusD^dex
SGBP^dex
GH^dex
e
B. theta membrane proteome
Up-regulated with dextran
Up-regulated with fructose
SGBP^dex
GH^dex
SusC^dex
SusD^dex
SusD^lev
GH^lev
SGBP^lev
SusC^lev
-log10 p-value
log2-fold change

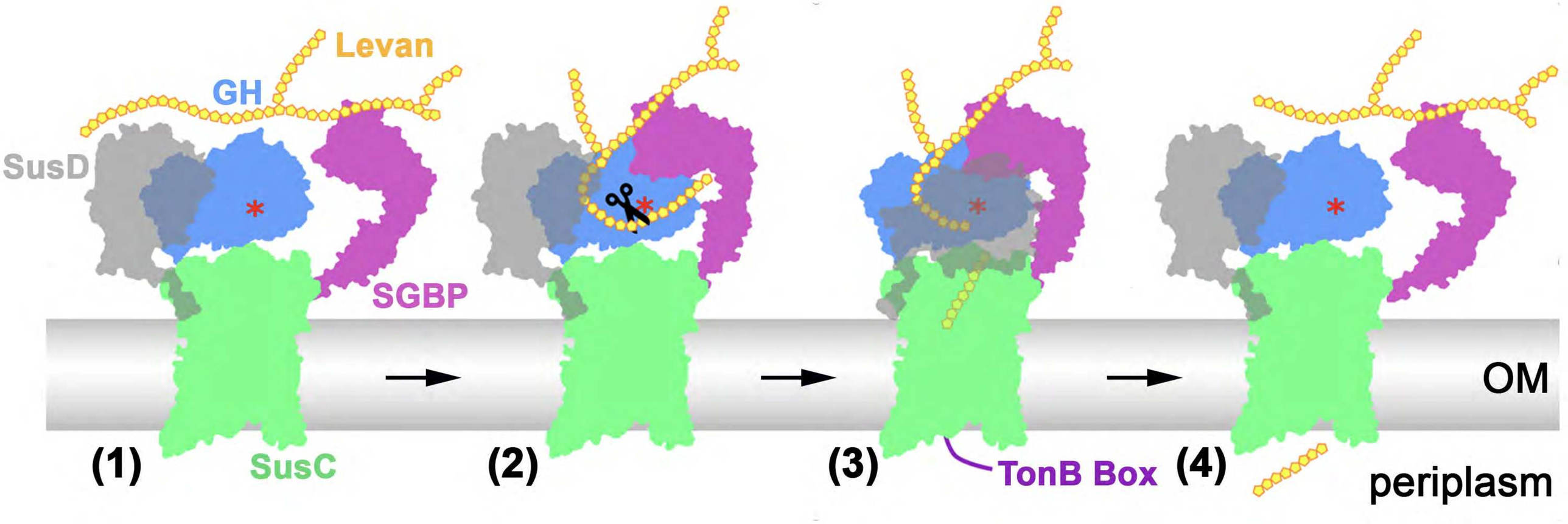


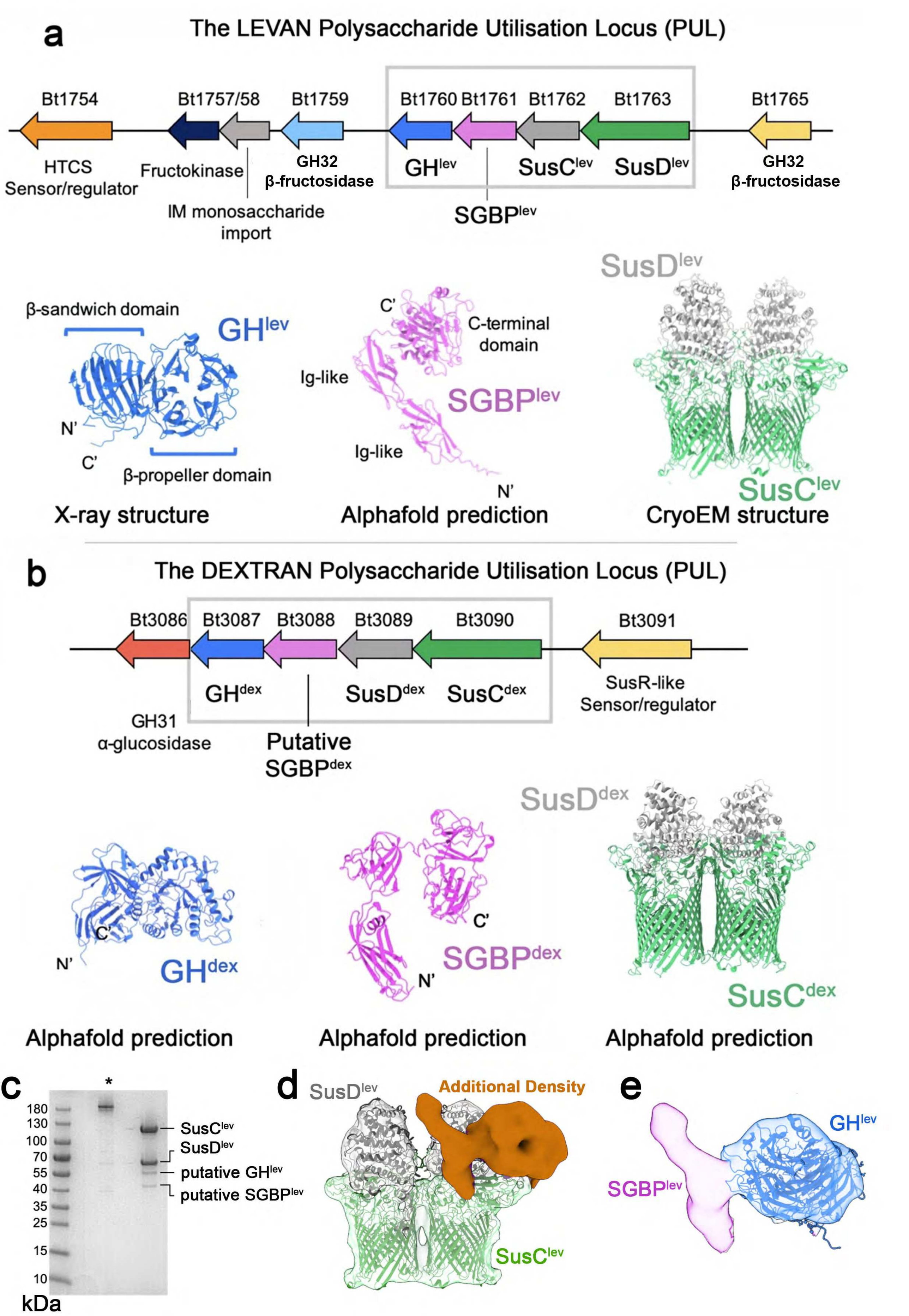
Focused refinement without SusD at 3.5Å resolution

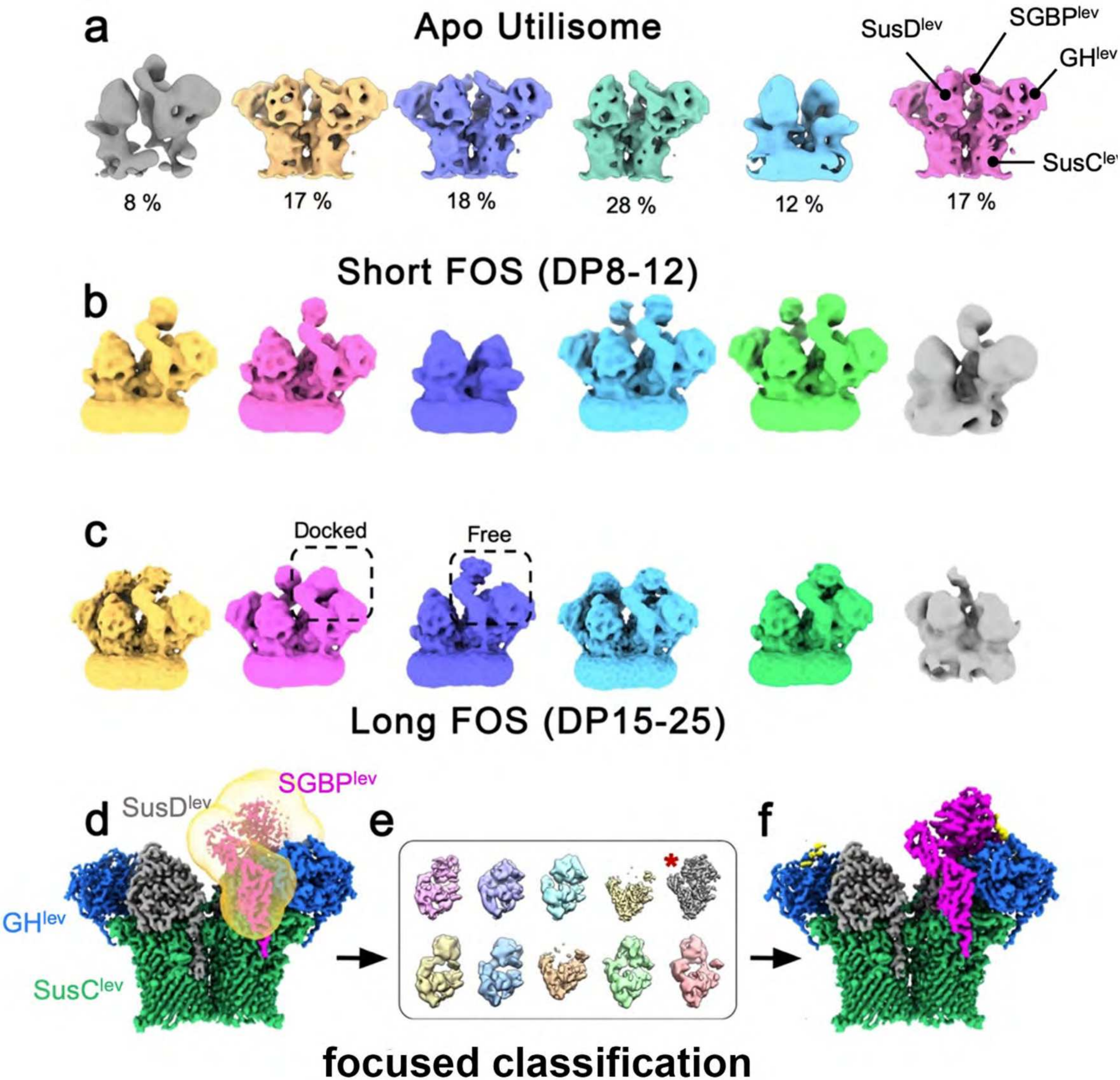
De novo atomic model

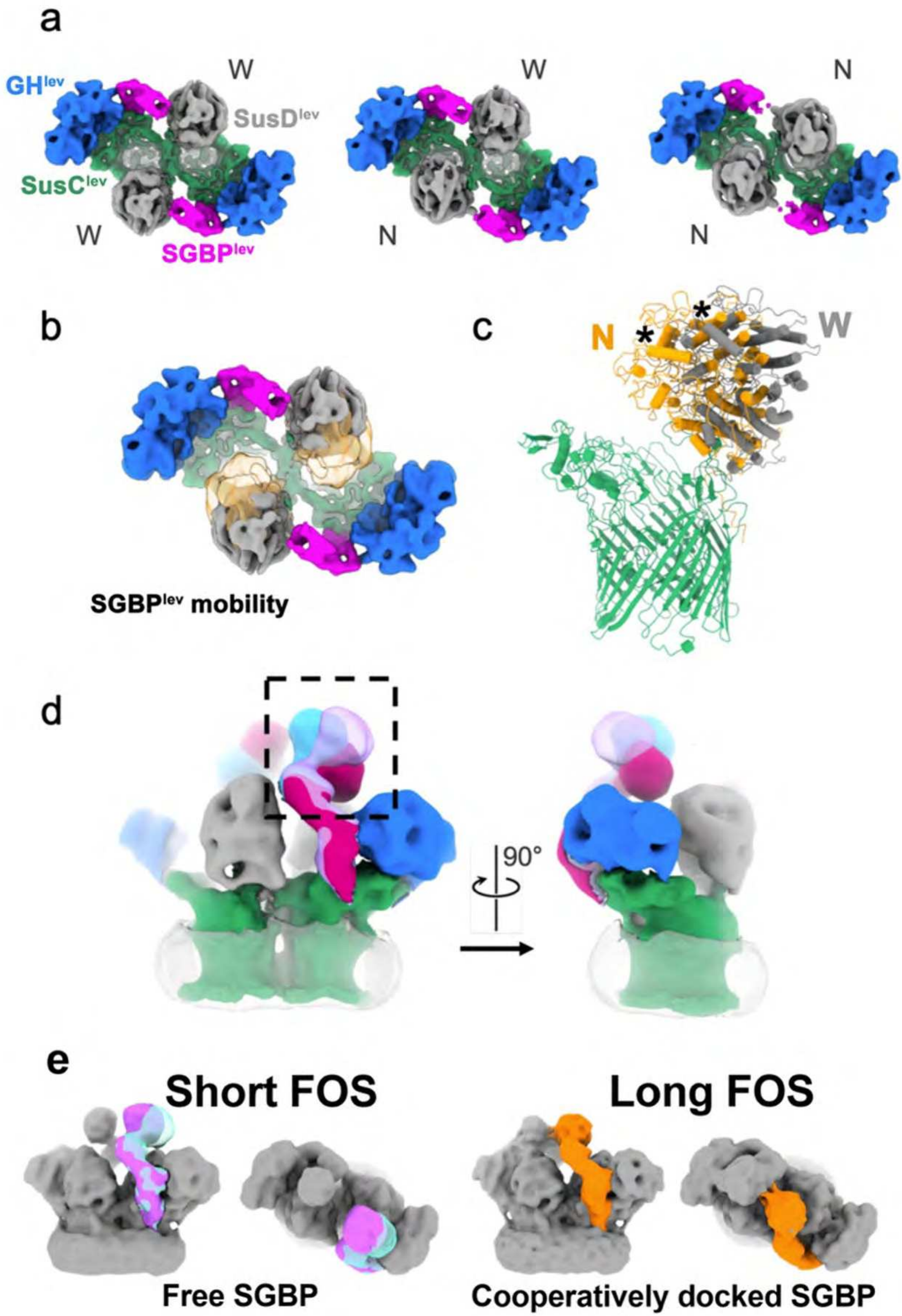

a
GH^lev
SusD^lev
2.9Å
+substrate
(lids
closed)
SGBP^lev
SusC^lev
b
~6Å
APO
(lids
open)
c
APO
+FOS
D406
N901
F649
(loop 8
ONLY)
R407
D383
FOS1
E379
E481
FOS2
K454
Q372
E170
H169
APO
+FOS
d
to W685
Loop 8
F649
APO
+FOS
e
to F649
APO
+FOS
F583
W685
S193
f
Y191
(plug)
F558
(barrel)
Y89
APO
g
S193
F558
Y191
+FOS

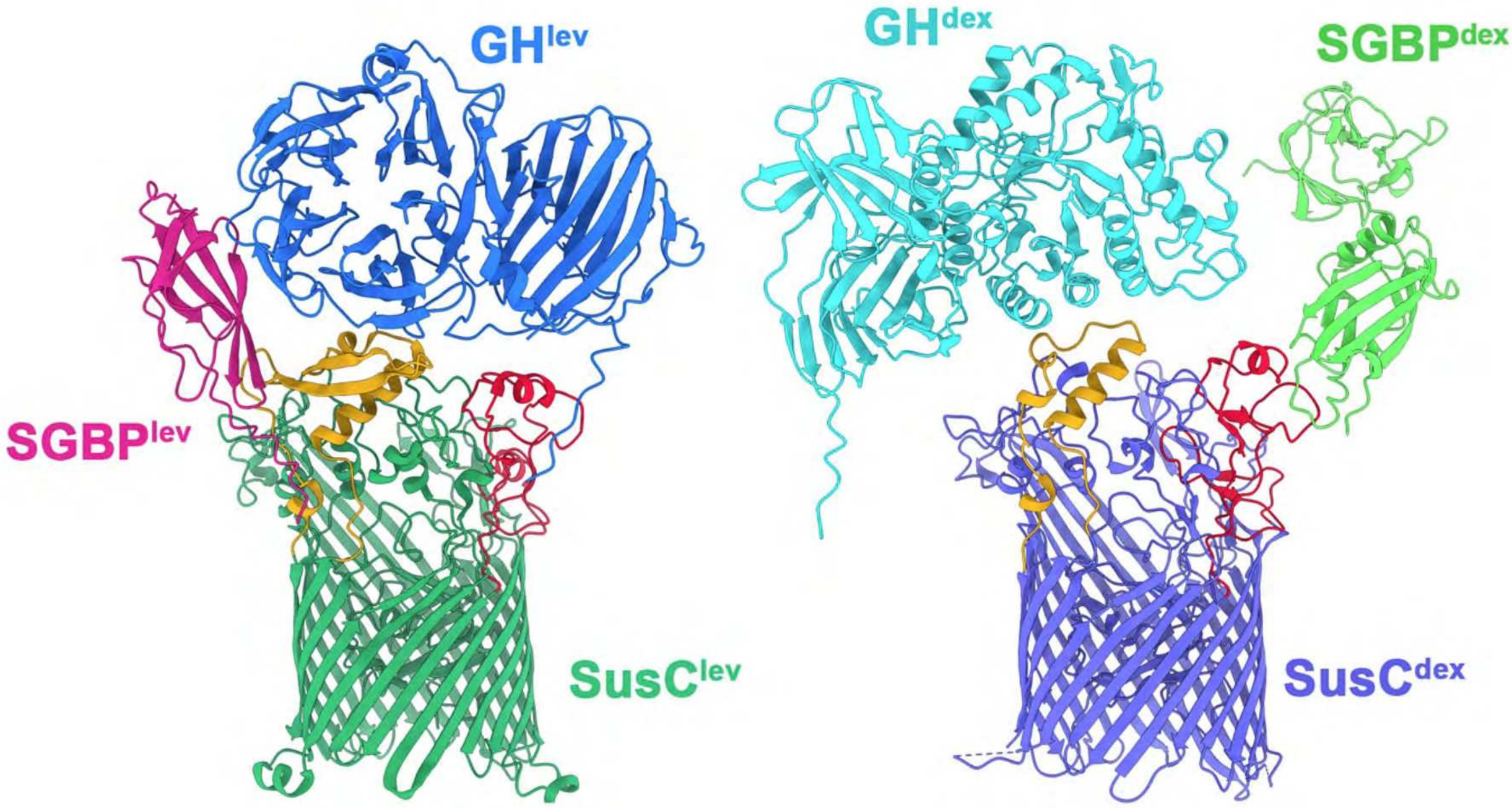

a
FOS
b
SusD^lev
SGBP^lev
GH^lev
SusC^lev
90°
c
90°
d
e
FOS3
W318
Y70
Ala42
f
g
Frc12
Trp311
Trp297
Trp359
Trp217
FOS4
Trp269
Frc1
Tyr437
h
ΔH (kcal.mol⁻¹)
WT
DP~15
Kd=22.6±4.0 μM
Molar Ratio
ΔH (kcal.mol⁻¹)
WAWA
DP~15
Molar Ratio


Levan
GH
SusD
SGBP
OM
(1)
SusC
(2)
(3)
TonB Box
(4)
periplasm

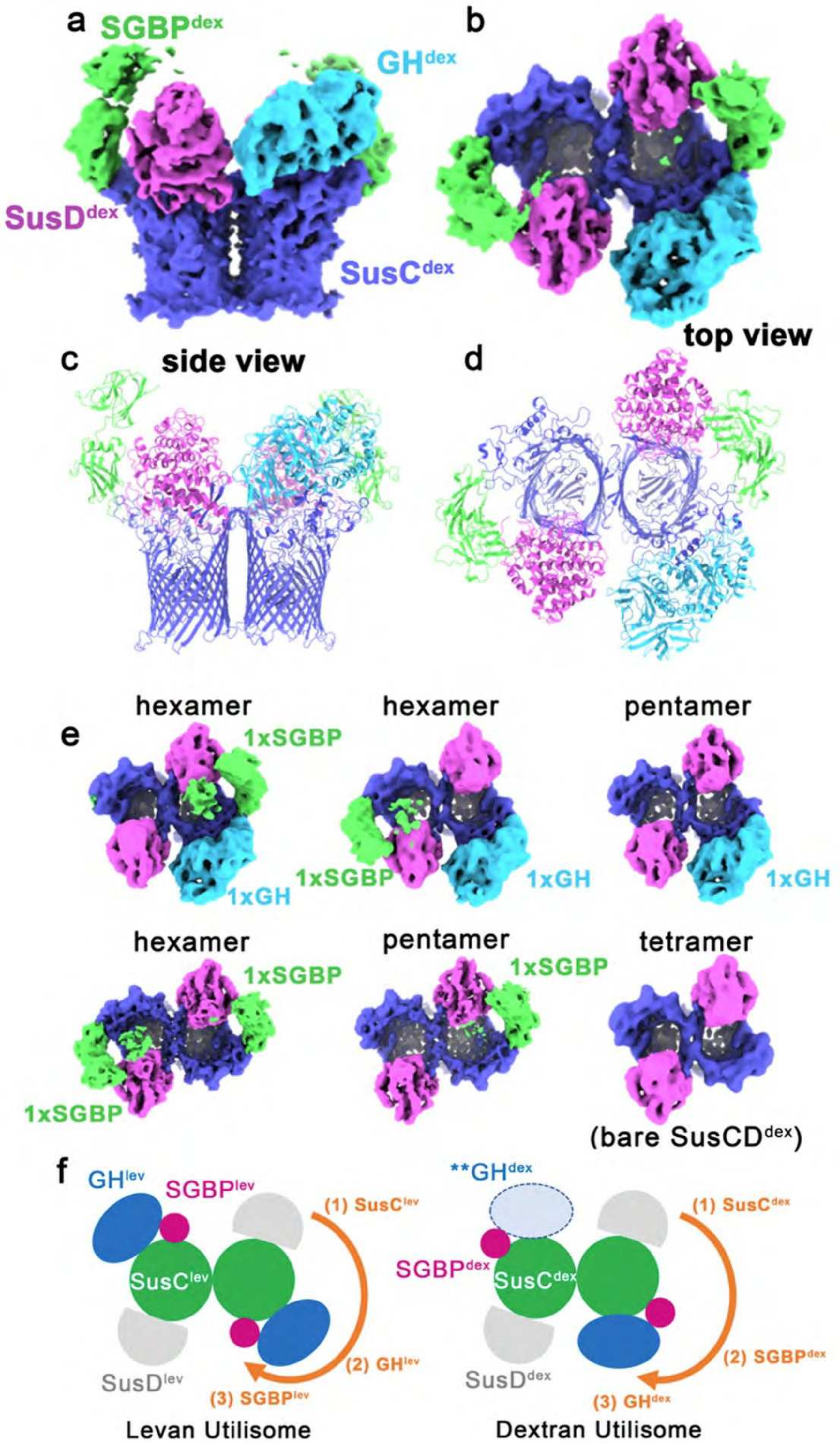

a
The LEVAN Polysaccharide Utilisation Locus (PUL)
Bt1754
Bt1757/58
Bt1759
Bt1760
Bt1761
Bt1762
Bt1763
Bt1765
HTCS Sensor/regulator
Fructokinase
IM monosaccharide import
GH32 β-fructosidase
GHlev
SGBPlev
SusClev
SusDlev
GH32 β-fructosidase
β-sandwich domain
GHlev
N'
C'
β-propeller domain
X-ray structure
C'
C-terminal domain
Ig-like
SGBPlev
Ig-like
N'
Alphafold prediction
SusDlev
SusClev
CryoEM structure
b
The DEXTRAN Polysaccharide Utilisation Locus (PUL)
Bt3086
Bt3087
Bt3088
Bt3089
Bt3090
Bt3091
GHdex
SusDdex
SusCdex
SusR-like Sensor/regulator
GH31 α-glucosidase
Putative SGBPdex
SusDdex
C'
N'
GHdex
Alphafold prediction
C'
N'
SGBPdex
Alphafold prediction
SusClev
Alphafold prediction
c
*
180
130
100
70
55
40
35
25
15
10
kDa
SusClev
SusDlev
putative GHlev
putative SGBPlev
d
SusDlev
Additional Density
SusClev
e
GHlev
SGBPlev

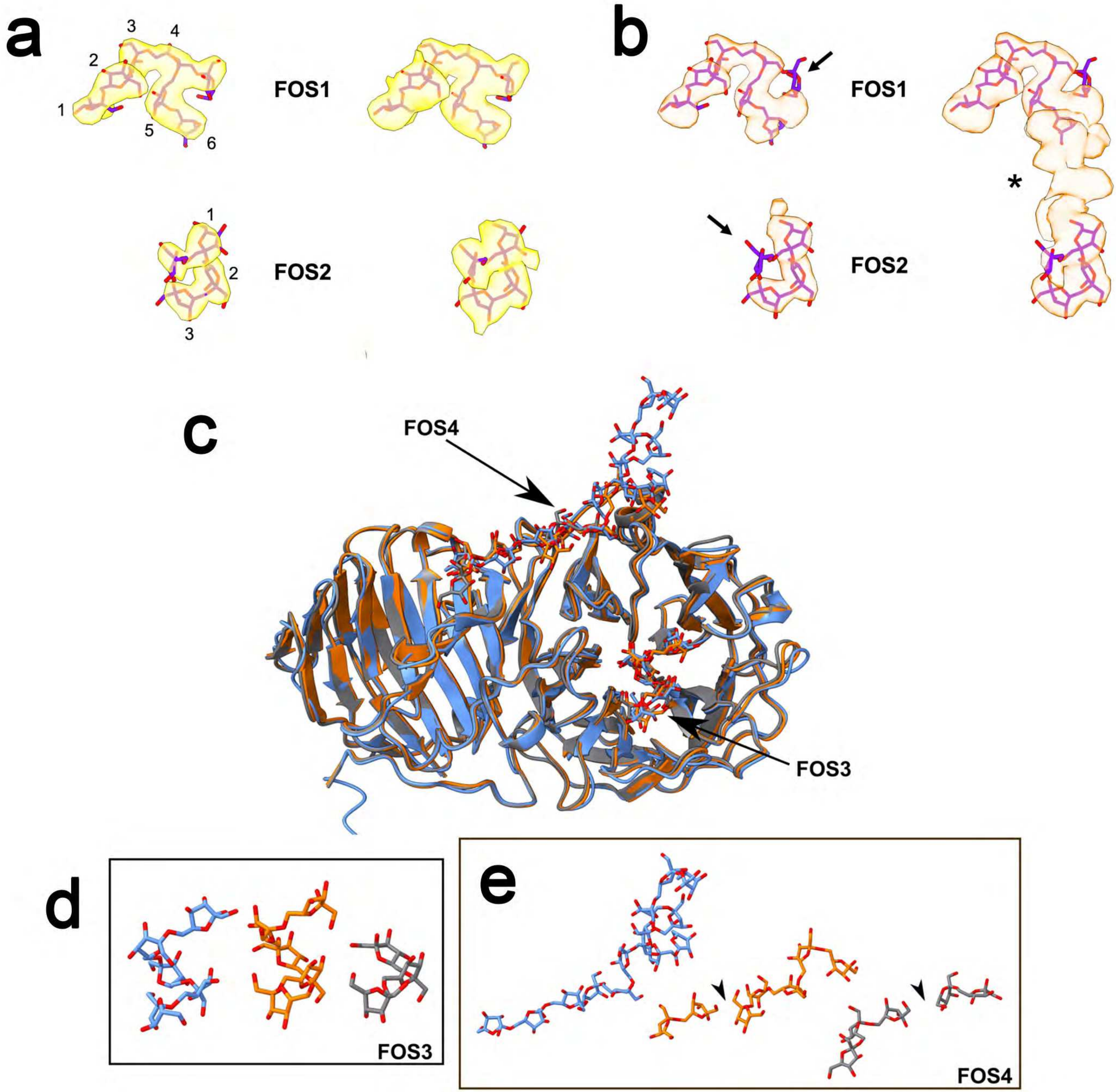

a
Apo Utilisome
SusDlev
SGBPlev
GHlev
SusClev
8 %
17 %
18 %
28 %
12 %
17 %
Short FOS (DP8-12)
b
c
Docked
Free
Long FOS (DP15-25)
d
SusDlev
SGBPlev
GHlev
SusClev
e
f
focused classification

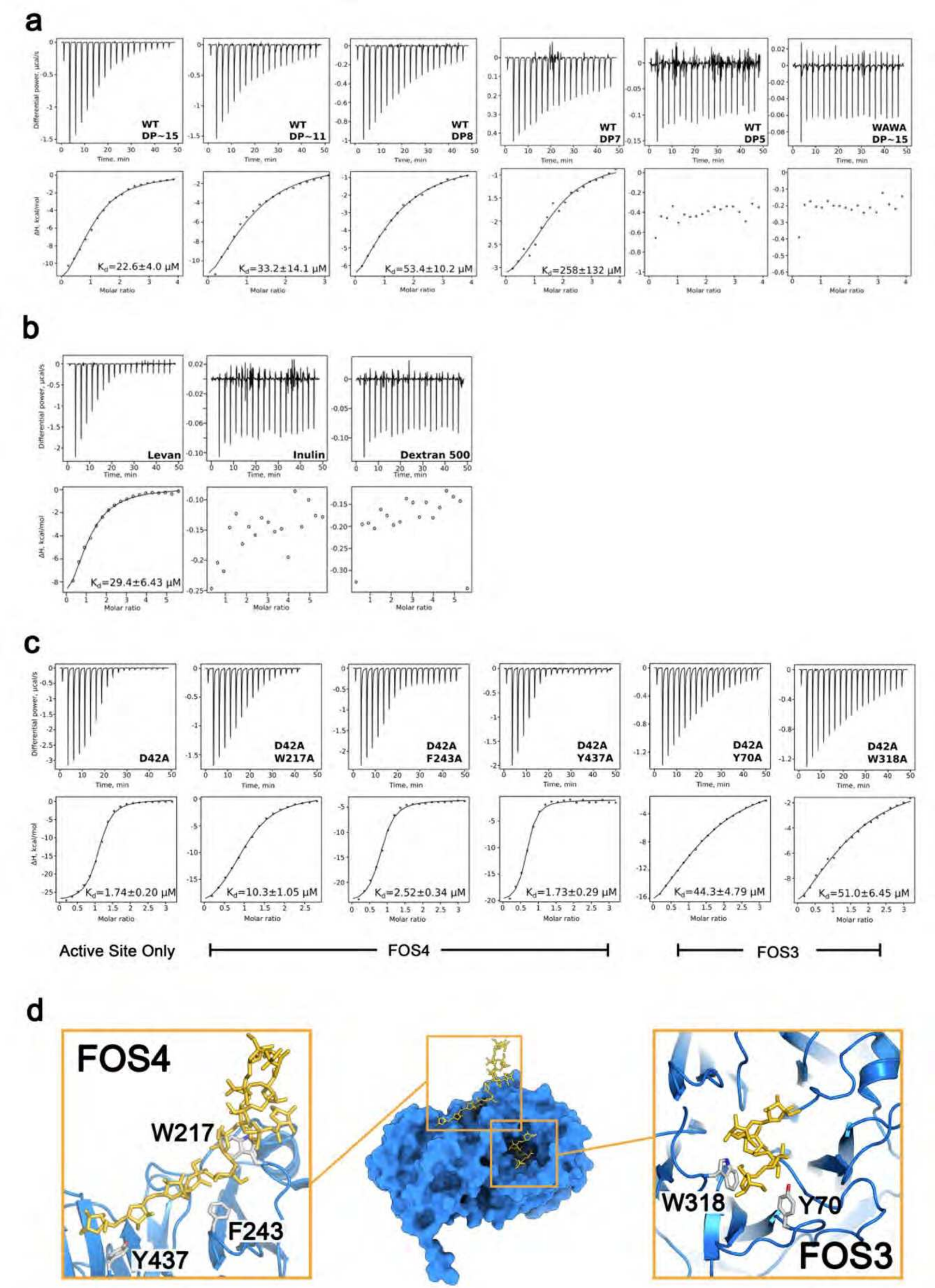

a
GH^lev
W
SusD^lev
SusC^lev
W
SGBP^lev
W
N
N
N
b
SGBP^lev mobility
c
N
*
*
W
d
90°
e
Short FOS
Long FOS
Free SGBP
Cooperatively docked SGBP

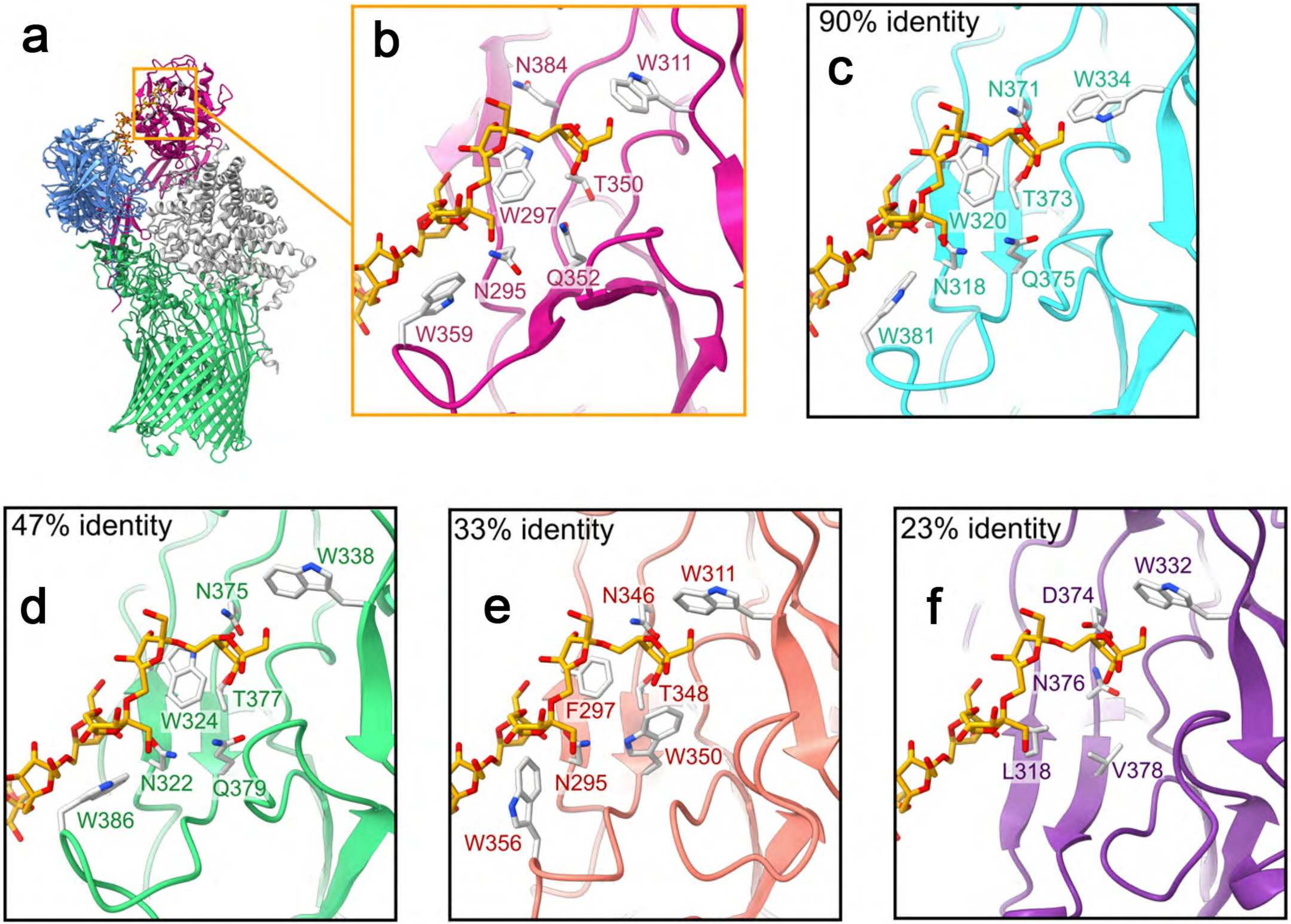


Levan **Dextran**


a
SGBPdex
GHdex
SusDdex
SusCdex
b
top view
c
side view
d
e
hexamer
1xSGBP
1xGH
hexamer
1xSGBP
1xGH
pentamer
1xGH
hexamer
1xSGBP
1xSGBP
pentamer
1xSGBP
tetramer
(bare SusCDdex)
f
GHlev
SGBPlev
SusClev
SusDlev
(1) SusClev
(2) GHlev
(3) SGBPlev
Levan Utilisome
**GHdex
SGBPdex
SusCdex
SusDdex
(1) SusCdex
(2) SGBPdex
(3) GHdex
Dextran Utilisome


a
1
2
3
4
5
6
FOS1
1
2
3
FOS2
b
FOS1
FOS2
*
c
FOS4
FOS3
d
FOS3
e
FOS4


a
WT DP~15
WT DP~11
WT DP8
WT DP7
WT DP5
WAWA DP~15
Differential power, µcal/s
Time, min
ΔH, kcal/mol
Molar ratio
Kd=22.6±4.0 µM
Kd=33.2±14.1 µM
Kd=53.4±10.2 µM
Kd=258±132 µM
b
Levan
Inulin
Dextran 500
Kd=29.4±6.43 µM
c
D42A
D42A W217A
D42A F243A
D42A Y437A
D42A Y70A
D42A W318A
Kd=1.74±0.20 µM
Kd=10.3±1.05 µM
Kd=2.52±0.34 µM
Kd=1.73±0.29 µM
Kd=44.3±4.79 µM
Kd=51.0±6.45 µM
Active Site Only
FOS4
FOS3
d
FOS4
W217
F243
Y437
W318
Y70
FOS3


a
b
N384
W311
T350
W297
N295
Q352
W359
90% identity
c
N371
W334
T373
W320
N318
Q375
W381
47% identity
d
W338
N375
T377
W324
N322
Q379
W386
33% identity
e
W311
N346
T348
F297
W350
N295
W356
23% identity
f
W332
D374
N376
L318
V378

Data statistics*			
	BT1760 SeMet	BT1760 D42N	BT1760 D42N
Beamline	I03	I02	I04
Date	05/08/08	27/09/09	31/05/10
Wavelength (Å)	0.97630	0.97930	0.97930
Resolution (Å)	61.92-2.80 (2.85-2.80)	67.87-2.65 (2.74-2.65)	73.71-2.30 (2.35-2.30)
Space group	$P4_12_12$	$I4_122$	$I4_122$
Unit-cell parameters			
a (Å)	175.14	174.72	175.50
b (Å)	175.14	174.72	175.50
c (Å)	221.50	215.59	214.53
α (°)	90.0	90.0	90.0
β (°)	90.0	90.0	90.0
γ (°)	90.0	90.0	90.0
Unit-cell volume ($Å^3$)	6794525	6581332	6607577
Solvent content (%)	49	47	47
No. of measured reflections	4649806 (166168)	788798 (66908)	512272 (18602)
No. of independent reflections	83001 (3468)	48504 (4407)	73764 (4314)
Completeness (%)	97.7 (76.0)	99.9 (99.7)	99.6 (95.4)
Redundancy	56.0 (47.9)	16.3 (15.2)	6.9 (4.3)
$CC_{1/2}$ (%)	99.8 (87.9)	99.7 (70.8)	99.5 (54.9)
$I/\sigma(I)$	22.7 (4.3)	12.1 (1.2)	9.4 (1.1)
Anomalous completeness (%)	97.8 (76.3)		
Anomalous redundancy	29.3 (24.6)		
Refinement Statistics			
Rwork (%)		19.13	20.89
Rfree# (%)		25.55	26.75
No. of non-H atoms			
No. of protein atoms		7800	7800
No. of solvent atoms		5	201
No. of ligand atoms		284	230
R.m.s. deviation from ideal			
Bond angle (°)		2.14	1.89
Bond length (Å)		0.0099	0.0094
Average B factor ($Å^2$)			
Protein		59	62
Solvent		38	54
Ligand		81	77
Ramachandran plot+, residues in most favoured regions (%)		91.2	92.9
PDB file code		7ZNR	7ZNS

	Levan Utilisome		Levan Utilisome (FOS DP8-12)		Inactive Levan Utilisome - long FOS		Dextran Utilisome
	Dataset 1	**Dataset 2**					
Data collection and processing							
Magnification	75,000 x	75,000 x	75,000 x		75,000 x		96,000 x
Voltage (kV)	300	300	300		300		300
Electron exposure (e^-/Å^2)	35.4	37	38.5		37.8		38.78
Defocus range (µm)	-1.5 to -3.3	-1.5 to -3.3	-1.5 to -3.0		-1.5 to -3.0		-1.2 to -3.0
Pixel size (Å)	1.065	1.065	1.065		1.065		0.86
Initial particle images pre-3D classification (no.)	89,305	280,696	63,789		146,056		477,707
Particle stack	Utilisome	Utilisome	Utilisome	$SusC_2D_2$ core	Post focused 3D classification	$SusC_2D_2$ core	All SusCD containing particles
Final particle images post-classification (no.)	45,594	146,512	15,012	54,736	27,310	120,957	305,372
Total particle images after datasets combined (no.)	192,106		n/a	n/a	n/a	n/a	n/a
Symmetry imposed	C2		C2	C2	C1	C2	C1
Map resolution (Å) (FSC 0.143)	3.5		3.2	2.9	3.1	2.7	3.1
Map sharpening B-factor (Å^2)	-149		-90.63	-96.42	-76.41	-91.93	-79.2
Refinement							
Model composition							
Non-hydrogen atoms	23,688		33,058	23,634	35,942	23,634	13,616
Protein residues	2988		4134	2950	4472	2950	1742
R.m.s. deviations							
Bond lengths (Å)	0.002		0.005	0.002	0.003	0.001	0.007
Bond angles (°)	0.479		0.555	0.459	0.486	0.397	0.672
Validation							
Molprobity score	1.75		1.75	1.38	1.62	1.38	2.24
Clashscore	8.48		7.80	6.98	7.25	6.90	4.8
Poor rotamers (%)	0		0	0	0	0	4.69
Ramachandran plot							
Favoured (%)	95.76		95.34	98.37	96.57	98.37	92.08
Allowed (%)	4.17		4.47	1.56	3.37	1.63	7.92
Disallowed (%)	0.07		0.19	0.07	0.07	0	0
Deposition ID							
PDB	8A9Y		8AA0	8AA1	8AA2	8AA3	8AA4
EMDB	EMD-15288		EMD-15289	EMD-15290	EMD-15291	EMD-1592	EMD-15293

SI Guide:

Supplementary Figure 1. Uncropped Gels from Figure 1 and Extended Data Figure 1.

Supplementary discussion

Supplementary Figure 2. The TonB-dependent transport cycle. (i) The b-barrel of the TonB-dependent transporter (green) is occluded by the plug domain (blue/cyan) when no substrate is bound. The TonB box is poorly accessible from the periplasm. (ii) Substrate binding induces conformational changes in the TonB-dependent transporter that lead to increased periplasmic accessibility of the TonB box. The TonB CTD binds the exposed TonB box. The transmembrane segment of TonB is associated with the proton-conducting IM ExbBD complex. (iii) TonB transduces the energy stored in the proton gradient to exert a force on the mechanically-labile subdomain of the plug (dark blue), which is pulled out of the b-barrel lumen. The substrate can diffuse into the periplasm via the open transport channel. (iv) TonB disengages the TonB box, the mechanically labile subdomain of the plug re-inserts into the b-barrel lumen, and the substrate is further processed in the periplasm (in the case of *e.g.* oligosaccharides; not shown) and imported into the cytoplasm via ATP binding cassette (ABC) or major facilitator superfamily (MFS) transporters. A "classical" TBDT such as *E. coli* BtuB is shown for simplicity. As such, lipoprotein components of the utilisome are not shown. Likewise, the N-terminal extension (NTE) domain, present N-terminal to the TonB box in many SusC transporters, is not shown.

Supplementary Figure 3. SGBP^{lev} C-terminal domain homologue amino acid sequence alignment. Only the C-terminal levan binding domain sequences of the SGBP^{lev} were aligned. UniProt accession numbers correspond to: *B. theta* VPI-5482 SGBP^{lev} (Q8A6W5); *Bacteroides* sp. D2 (E5CCB3); *Prevotella oralis* ATCC 33269 (E7RM14); *Flavobacterium commune* (A0A1D9P8I4); *F. cellulosilyticum* (A0A4R5CJN9). Residue numbering is for *B. theta* SGBP^{lev}. Arrowheads indicate positions in the *B. theta* SGBP^{lev} sequence that bind to FOS in the cryo-EM structure. The alignment was made with Clustal Omega[14], and visualised in ESPript 3.0[15].

SI Guide:

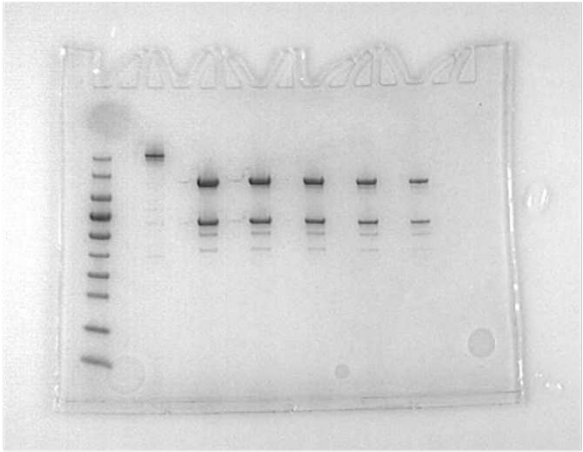
Supplementary Figure 1. Uncropped Gels from Figure 1 and Extended Data Figure 1.

Supplementary discussion

Supplementary Figure 2. The TonB-dependent transport cycle. (i) The b-barrel of the TonB-dependent transporter (green) is occluded by the plug domain (blue/cyan) when no substrate is bound. The TonB box is poorly accessible from the periplasm. (ii) Substrate binding induces conformational changes in the TonB-dependent transporter that lead to increased periplasmic accessibility of the TonB box. The TonB CTD binds the exposed TonB box. The transmembrane segment of TonB is associated with the proton-conducting IM ExbBD complex. (iii) TonB transduces the energy stored in the proton gradient to exert a force on the mechanically-labile subdomain of the plug (dark blue), which is pulled out of the b-barrel lumen. The substrate can diffuse into the periplasm via the open transport channel. (iv) TonB disengages the TonB box, the mechanically labile subdomain of the plug re-inserts into the b-barrel lumen, and the substrate is further processed in the periplasm (in the case of *e.g.* oligosaccharides; not shown) and imported into the cytoplasm via ATP binding cassette (ABC) or major facilitator superfamily (MFS) transporters. A "classical" TBDT such as *E. coli* BtuB is shown for simplicity. As such, lipoprotein components of the utilisome are not shown. Likewise, the N-terminal extension (NTE) domain, present N-terminal to the TonB box in many SusC transporters, is not shown.

Supplementary Figure 3. SGBP^{lev} C-terminal domain homologue amino acid sequence alignment. Only the C-terminal levan binding domain sequences of the SGBP^{lev} were aligned. UniProt accession numbers correspond to: *B. theta* VPI-5482 SGBP^{lev} (Q8A6W5); *Bacteroides* sp. D2 (E5CCB3); *Prevotella oralis* ATCC 33269 (E7RM14); *Flavobacterium commune* (A0A1D9P8I4); *F. cellulosilyticum* (A0A4R5CJN9). Residue numbering is for *B. theta* SGBP^{lev}. Arrowheads indicate positions in the *B. theta* SGBP^{lev} sequence that bind to FOS in the cryo-EM structure. The alignment was made with Clustal Omega[14], and visualised in ESPript 3.0[15].

Supplementary Figure 1. Uncropped Gels from Figure 1 and Extended Data Figure 1.



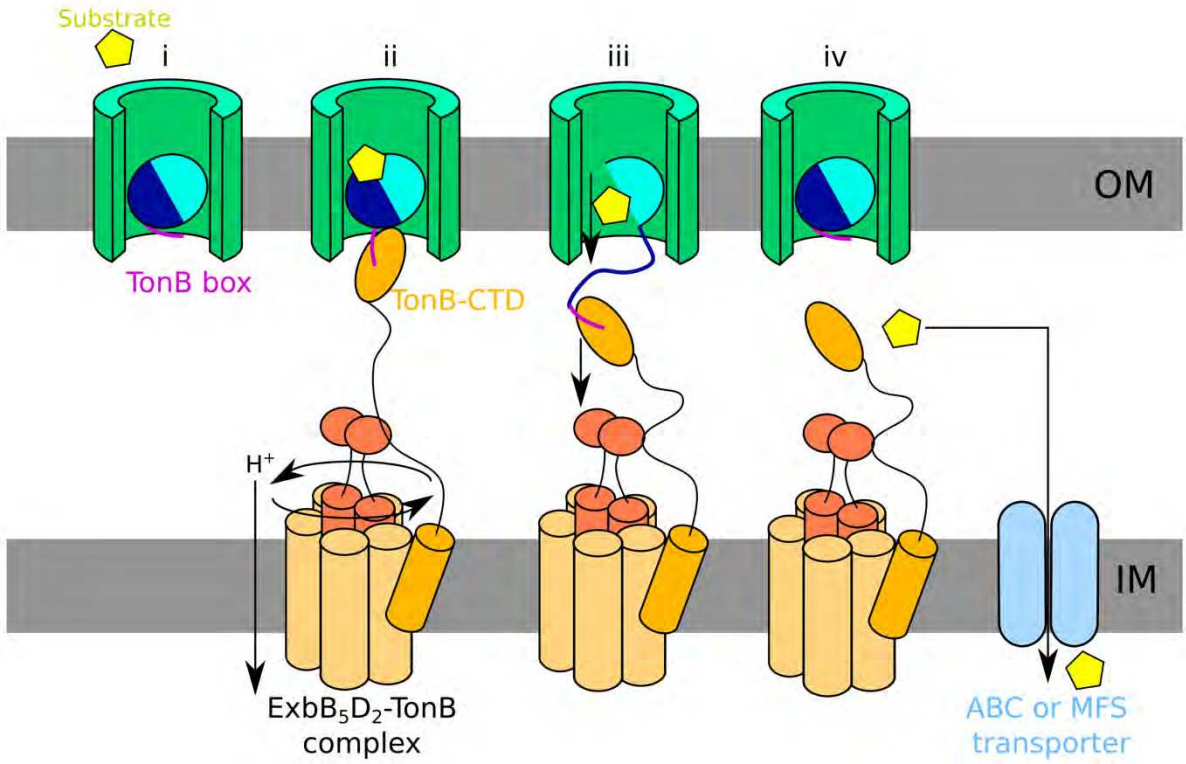
Supplementary Discussion

TonB-dependent transport and signalling substrate binding across the OM

TonB-dependent transporters (TBDTs), including SusC-type TBDTs, require energy for their transport activity. Because there are no ion gradients across the OM nor ATP in the periplasm, TBDTs couple to an inner membrane (IM) complex that harnesses the energy stored in the proton gradient across the IM. In the latest model for transport via TBDTs[1] (Supplementary Figure 2), protons are conducted by the ExbBD complex into the cytoplasmic space, which induces rotation or movement of the IM TonB protein that is most likely bound to the outside of the ExbBD complex[1]. The C-terminal domain of TonB binds to the TonB box on the periplasmic face of the TBDT in the presence of substrate. In this way, TonB transduces mechanical energy from the inner membrane to the TBDT in a poorly-understood process that results in partial unfolding of the TBDT barrel plug and, consequently, the formation of a channel via which the substrate can diffuse into the periplasm (Supplementary Figure 2)[2].

Following transport, the TonB C-terminal domain (CTD) interaction with the TonB box is broken via an unknown mechanism, after which the plug re-folds and closes the substrate channel, resetting the transporter for another cycle.

A key feature of the TBDT transport mechanism is prevention of unproductive transport cycles, *i.e*. engagement of TonB with TBDTs that do not have any substrate bound. Information about substrate binding on the extracellular side of the TBDT must be relayed across the OM to the periplasm. This is achieved via substrate binding-induced conformational changes that are propagated through the barrel and the plug of the TBDT. This ultimately leads to increased exposure of the TonB box in the periplasmic space, which is a pre-requisite for interaction with the TonB CTD, disruption of the plug domain and consequent channel formation. In the *E. coli* vitamin B12 transporter BtuB for example, a salt bridge, or 'ionic lock', is present between the barrel wall and a plug residue downstream from the TonB box in the absence of substrate. Binding of vitamin B12 by extracellular loops and the apex of the plug causes allosteric conformational changes, breaking the ionic lock and increasing accessibility of the TonB box[3,4]. However, while statistical coupling analysis (SCA) and structure-based analyses have provided useful insights[5,6], identifying the exact residues involved in these conformational changes upon substrate binding has been challenging, as TBDT structures with and without substrate are usually very similar beyond the TonB box[7].



Supplementary Figure 2. The TonB-dependent transport cycle. (i) The β-barrel of the TonB-dependent transporter (green) is occluded by the plug domain (blue/cyan) when no substrate is bound. The TonB box is poorly accessible from the periplasm. (ii) Substrate binding induces conformational changes in the TonB-dependent transporter that lead to increased periplasmic accessibility of the TonB box. The TonB CTD binds the exposed TonB box. The transmembrane segment of TonB is associated with the proton-conducting IM ExbBD complex. (iii) TonB transduces the energy stored in the proton gradient to exert a force on the mechanically-labile subdomain of the plug (dark blue), which is pulled out of the β-barrel lumen. The substrate can diffuse into the periplasm via the open transport channel. (iv) TonB disengages the TonB box, the mechanically labile subdomain of the plug re-inserts into the β-barrel lumen, and the substrate is further processed in the periplasm (in the case of *e.g.* oligosaccharides; not shown) and imported into the cytoplasm via ATP binding cassette (ABC) or major facilitator superfamily (MFS) transporters. A "classical" TBDT such as *E. coli* BtuB is shown for simplicity. As such, lipoprotein components of the utilisome are not shown. Likewise, the N-terminal extension (NTE) domain, present N-terminal to the TonB box in many SusC transporters, is not shown.

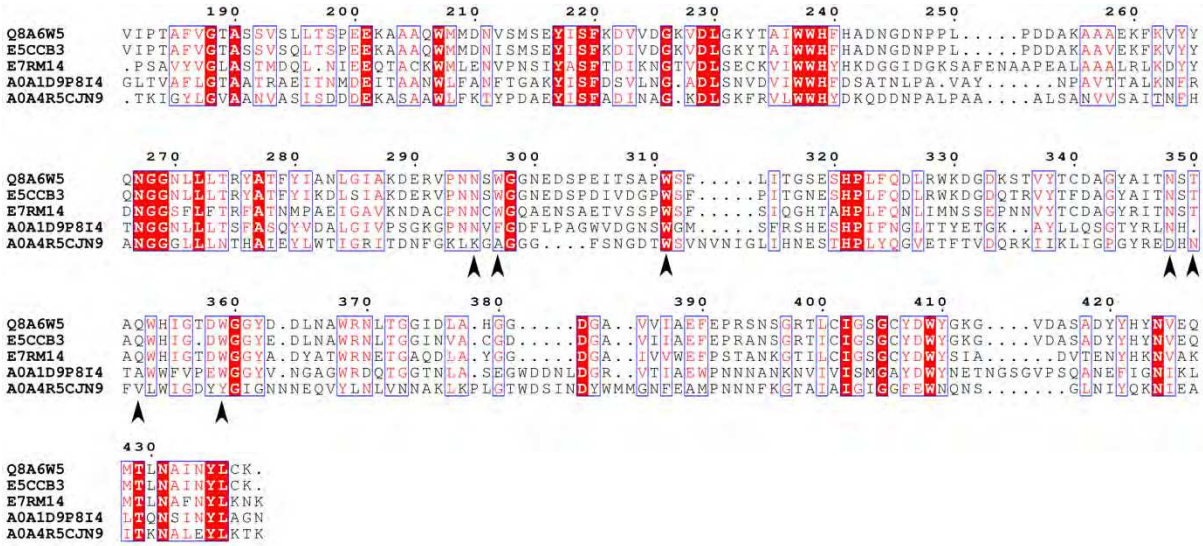
While our cryo-EM structures of the apo- and FOS-bound levan utilisomes shed some light on the allosteric signalling pathway within $SusC^{lev}$, experimentally proving that the residues that undergo conformational changes upon substrate binding indeed form an allosteric network remains profoundly challenging. Part of this stems from the fact that, so far, it has not been possible to reconstitute TonB-dependent transport *in vitro*, leaving growth assays as the only way to test the effect of mutations within the allosteric network. Moreover, such mutations would generate negative results (*i.e.*, no growth on levan). In principle, progress might be made via methods that allow interrogation of protein dynamics on relevant timescales, *e.g.* in vivo electron paramagnetic resonance or hydrogen-deuterium exchange mass spectrometry which have only recently been applied to studying BtuB[3,4]. Adapting such methods to the anaerobic *B. theta* will be a challenge.

The presence of one, or sometimes two additional N-terminal domains in SusC-type TBDTs represents an additional layer of complexity compared to classical TBDTs such as *E. coli* FhuA or BtuB. STN domains, absent in both $SusC^{lev}$ and $SusC^{dex}$, are involved in cell surface signalling by interacting with IM-embedded anti-sigma regulators[8,9]. An N-terminal extension (NTE) domain is present N-terminal and adjacent to the TonB box in most *B. theta* TBDTs, including $SusC^{lev}$ and $SusC^{dex}$. It is approximately 7 kDa in size and has an Ig-like fold[10]. The function of the NTE is not known, although its location suggests it could plausibly interact with the TonB CTD in the periplasmic space. Intriguingly, deletion of the NTE in $SusC^{lev}$ and *P. gingivalis* RagA generates a growth defect that is much more severe than a deletion of the TonB box, suggesting a role for the NTE that goes beyond the transport process itself[10].

Glycan specificity of the $SGBP^{lev}$

Investigation of the levan binding site of SGBP^{lev} (ED Fig. 10) reveals three tryptophan residues (W297, W311 and W359) that potentially form stacking interactions with the β-fructofuranose units of the β2,6-linked FOS chain, consistent with previous SGBP-ligand complex structures[11,12]. Aromatic stacking makes a crucial contribution to binding affinity as demonstrated by ITC with alanine substitutions of the SGBP^{lev} (ED Fig. 9a). Importantly, the orientation of the aromatic side chains affects the shape of the binding site. This results in specificity by enabling stacking interactions between the aromatic side chains and the β-fructofuranose faces only in a β2,6-linked FOS chain. Indeed, our ITC data clearly show that SGBP^{lev} has a relatively high affinity for levan but does not bind another fructan, β2,1-linked inulin (ED Fig. 9d).

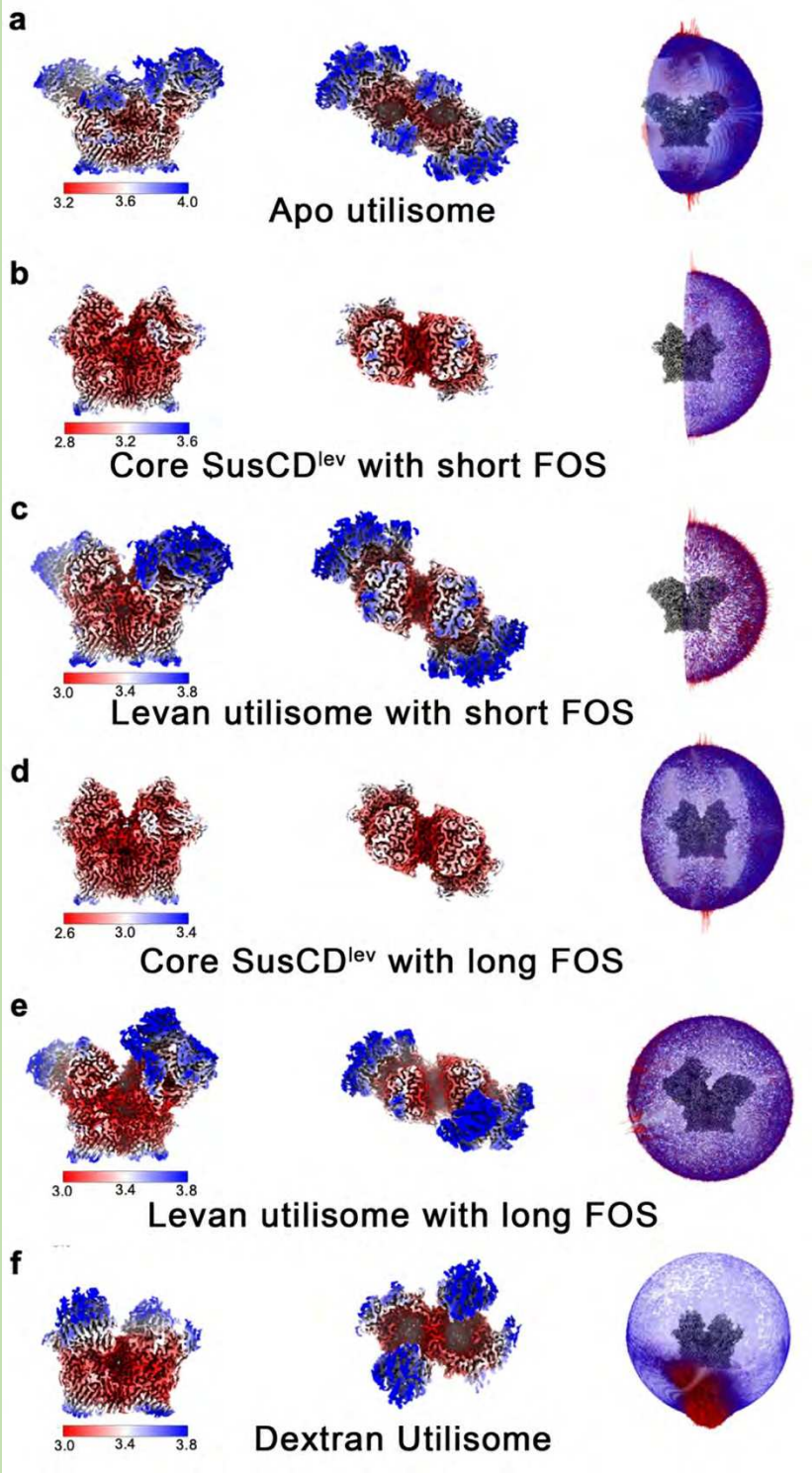
Given that the local resolution of our SGBP^{lev} maps was insufficient to assign hydrogen bonds with confidence, we performed a BLAST search with the SGBP^{lev} C-terminal domain amino acid sequence and made an alignment with sequences that were 90%, 47%, 33% and 23% identical to SGBP^{lev} (Supplementary Figure 3). We confirmed that all chosen sequences are likely to be genuine levan SGBPs by looking at their genomic context: all of the sequences are part of a PUL that also has a gene predicted to encode a GH32 family enzyme (endo-levanase). Through analysis of AlphaFold2-predicted models[13] of these SGBP^{lev} homologues, we infer that a number of residues are also involved in specific hydrogen-bonding interactions with FOS (N295, T350, Q352, N384) (ED Fig. 10). These binding site residues are mostly conserved, even in proteins with only ~33% sequence identity to that of SGBP^{lev} (ED Fig 10, Supplementary Figure 3). It is likely that these hydrogen-bonding residues, as well as the stacking tryptophan residues, are configured within the binding site in such a way that allows them to bind only glycans with the correct chemical composition as well as the correct geometry or secondary structure imposed by a specific glycosidic linkage. Other highly conserved residues that do not bind levan are likely involved in maintaining the fold of the domain.



Supplementary Figure 3. SGBP^{lev} C-terminal domain homologue amino acid sequence alignment. Only the C-terminal levan binding domain sequences of the SGBP^{lev} were aligned. UniProt accession numbers correspond to: *B. theta* VPI-5482 SGBP^{lev} (Q8A6W5); *Bacteroides* sp. D2 (E5CCB3); *Prevotella oralis* ATCC 33269 (E7RM14); *Flavobacterium commune* (A0A1D9P8I4); *F. cellulosilyticum* (A0A4R5CJN9). Residue numbering is for *B. theta* SGBP^{lev}. Arrowheads indicate positions in the *B. theta* SGBP^{lev} sequence that bind to FOS in the cryo-EM structure. The alignment was made with Clustal Omega[14], and visualised in ESPript 3.0[15].

References

1. Ratliff, A. C., Buchanan, S. K. & Celia, H. The Ton Motor. *Front. Microbiol.* **13**, 1240 (2022).
2. Hickman, S. J., Cooper, R. E. M., Bellucci, L., Paci, E. & Brockwell, D. J. Gating of TonB-dependent transporters by substrate-specific forced remodelling. *Nat. Commun. 2017 81* **8**, 1–12 (2017).
3. Nilaweera, T. D., Nyenhuis, D. A. & Cafiso, D. S. Structural intermediates observed only in intact Escherichia coli indicate a mechanism for TonB-dependent transport. *Elife* **10**, (2021).
4. Zmyslowski, A. M., Baxa, M. C., Gagnon, I. A. & Sosnick, T. R. HDX-MS performed on BtuB in E. coli outer membranes delineates the luminal domain's allostery and unfolding upon B12 and TonB binding. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).
5. Ferguson, A. D. *et al.* Signal transduction pathway of TonB-dependent transporters. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 513–518 (2007).
6. Chimento, D. P., Kadner, R. J. & Wiener, M. C. Comparative structural analysis of TonB-dependent outer membrane transporters: Implications for the transport cycle. *Proteins Struct. Funct. Genet.* **59**, 240–251 (2005).
7. Noinaj, N., Guillier, M., Barnard, T. J. & Buchanan, S. K. TonB-Dependent Transporters: Regulation, Structure, and Function. *Annu. Rev. Microbiol.* **64**, 43–60 (2010).
8. Malki, I. *et al.* Interaction of a Partially Disordered Antisigma Factor with Its Partner, the Signaling Domain of the TonB-Dependent Transporter HasR. *PLoS One* **9**, e89502 (2014).
9. Jensen, J. L., Jernberg, B. D., Sinha, S. & Colbert, C. L. Structural basis of cell surface signaling by a conserved sigma regulator in Gram-negative bacteria. *J. Biol. Chem.* jbc.RA119.010697 (2020). doi:10.1074/jbc.RA119.010697
10. Gray, D. A. *et al.* Insights into SusCD-mediated glycan import by a prominent gut symbiont. *Nat. Commun.* **12**, 1–14 (2021).
11. Tamura, K. *et al.* Surface glycan-binding proteins are essential for cereal beta-glucan utilization by the human gut symbiont Bacteroides ovatus. *Cell. Mol. Life Sci.* **76**, 4319–4340 (2019).
12. Tamura, K., Dejean, G., Van Petegem, F. & Brumer, H. Distinct protein architectures mediate species-specific beta-glucan binding and metabolism in the human gut microbiota. *J. Biol. Chem.* **296**, (2021).
13. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nat. 2021 5967873* **596**, 583–589 (2021).
14. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
15. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, (2014).



Supplementary Figure 4. **Reconstructions of levan and dextran systems filtered and coloured by local resolution.** Side view (left column), and top view (centre column) of the various utilisome maps (and sub-complexes) presented in this work. All maps are coloured according to the colour key accompanying each. Right hand column shows angular distribution plots for each reconstruction, and highlights the strongly preferred orientation of the Dextran Utilisome data.