Investigating the mechanism of nutrient acquisition by SusCD transport systems using cryoEM

Joshua Brian Roy White

Submitted in accordance with the requirements of the degree of Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences, School of Molecular and Cellular Biology

March 2022

Intellectual property and publication statement

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 3 contains data from one jointly authored publication:

Madej, M.*, **White, J. B. R**.*, Nowakowska, Z., Rawson, S., Scavenius, C., Enghild, J. J., Bereta, G. P., Pothula, K., Kleinekathoefer, U., Baslé, A., Ranson, N. A., Potempa, J. and van den Berg, B. (2020) 'Structural and functional insights into oligopeptide acquisition by the RagAB transporter from Porphyromonas gingivalis', *Nature Microbiology,* 5(8), pp. 1016-1025. (Equal contribution *)

The candidate performed the cryoEM experiments, including data collection, processing and interpretation supervised by S.R. and N.A.R. The candidate also wrote the corresponding section of the manuscript, generated related figures and was heavily involved in reviewing and editing the manuscript.

M.M., J.P. and B.v.d.B. initiated the project. M.M. cultured cells, purified and crystallised proteins and performed MST binding experiments. Z.N. performed cloning and strain construction. G.P.B. carried out quantitative PCR experiments, and C.S. and J.J.E. performed peptidomics analysis. K.P. performed the molecular dynamics simulations, supervised by U.K. B.v.d.B purified and crystallised proteins and determined RagAB crystal structures. A.B. collected crystallography data. The manuscript was written by B.v.d.B. with input from the candidate, M.M., N.A.R. and J.P.

Chapter 4 contains data from one jointly authored publication:

Gray, D. A., **White, J. B. R.**, Oluwole, A. O., Rath, P., Glenwright, A. J., Mazur, A., Zahn, M., Baslé, A., Morland, C., Evans, S. L., Cartmell, A., Robinson, C. V., Hiller, S., Ranson, N. A., Bolam, D. N. and van den Berg, B. (2021) 'Insights into SusCD-mediated glycan import by a prominent gut symbiont', *Nat Commun,* 12(1), pp. 44.

The candidate performed the cryoEM experiments, including data collection, processing and interpretation supervised by N.A.R. The candidate also wrote the corresponding section of the manuscript, generated related figures and was heavily involved in reviewing and editing the manuscript.

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

Additional data presented in Chapter 4 includes SDS-PAGE analysis and downstream protein identification mass spectrometry. SDS-PAGE analysis was carried out by the candidate and mass spectrometry was performed by Rachel George (Mass Spectrometry Facility, University of Leeds).

Chapter 5 contains data from collaborative work and preparation of a jointly authored manuscript is ongoing. The candidate performed all negative stain and cryo TEM experiments including data collection, processing and interpretation. Protein purification was carried out by the van den Berg Group and proteomic analysis of levan and dextran PUL components was performed by Matthias Trost, Newcastle University.

Acknowledgements

First and foremost, I would like to thank my supervisors Neil Ranson and Bert van den Berg for providing me with the opportunity to work on this project and for their support and guidance throughout. A special thanks to Neil for his mentorship and impartial career advice which have been invaluable throughout my PhD and for which I am extremely grateful.

I am indebted to my amazing collaborators without whom, this project would not have been possible. Special thanks go to Mariusz Madej, Declan Gray, Augustinas Silale, Bert van den Berg and all the others who have provided me with samples and biochemical/biophysical evidence to support my microscopy endeavours.

Huge thanks to Becky Thompson, Emma Hesketh, Dan Maskell, Charlie Scarff and Martin Fuller for their endless patience and unwavering microscopy support. Special thanks to Becky for fostering my career aspirations and for always offering impartial advice. Massive thanks to Shaun Rawson for his teaching and mentorship at the start of my project, and for reminding me to always be pessimistic so as to never be disappointed.

I would like to thank all members of the Ranson lab, both past and present, for the insightful discussions, advice, and assistance over the course of my project and more importantly for just being an amazing group of people to spend time with. Special mentions to Dave Nicholson, Matt Byrne, Rodrigo Gallardo, Joe Snowden, Sam Haysom and Matt Iadanza for making office life so enjoyable. I would also like to take the opportunity to thank the wider cryoEM community at Leeds for the countless friendships that made PhD life that much easier. Big shout-out to Oliver Debski-Antoniak for being the best flatmate and lockdown companion.

Last but not least, I owe a huge debt of gratitude to my family and friends who have helped me to navigate the ups and downs of PhD life. A very special thank you to Aimie for her endless patience (which I tested), love and support, without which this work would not have been possible.

Abstract

The Gram-negative Bacteroidetes are abundant in the human microbiota, occupying niches devoid of easily accessible nutrients. Uptake of complex nutrient molecules, especially glycans, by these bacteria is facilitated by outer membrane SusCD protein complexes, comprising a barrel-shaped integral membrane transporter component (SusC) and a substrate-binding lipoprotein (SusD). Recent X-ray crystal structures show that these complexes exist as $SusC_2D_2$ dimers with SusD subunits tightly capping the exterior face of the SusC barrels, trapping substrate in a solvent excluded cavity. However, the mechanistic details of substrate capture and translocation remain unclear, as does the role and organisation of additional surface-exposed lipoproteins that are prevalent in glycan utilisation systems. Here we present cryoEM structures of three SusCD systems from important members of the human microbiota. Using the SusCD-like peptide-transporting RagAB complex from the keystone pathogen in human periodontitis, *Porphyromonas gingivalis,* SusD subunits are shown to function as lids, capable of opening and closing to alternately expose and occlude the substrate binding cavity to the extracellular space. For the levan utilisation system of the prominent model gut symbiont *Bacteroides thetaiotaomicron*, we demonstrate that the additional lipoproteins, a levan binding protein and an endo-levanase, assemble on the core transporter. The resulting octameric complex constitutes a stable outer membrane machine, which we name a utilisome, that presents all of the activities required for efficient levan capture. Such utilisomes are not restricted to levan transport, as demonstrated by accompanying structures of the equivalent complex for the dextran transporter of the same bacterium. Finally, cryoEM structures of the levan utilisome in the absence and presence of substrate reveal concerted conformational changes that rationalise the position of each component for efficient nutrient capture, as well as providing a direct demonstration of the mechanism of levan capture at the OM for the first time. All Bacteroidetes sequenced to date contain SusCD homologues and we anticipate that utilisome assemblies represent a general mechanism for the translocation of large nutrients across the outer membrane of these bacteria.

Table of contents

List of figures

List of tables

Abbreviations

ABC	ATP=binding cassette
ATP	Adenosine triphosphate
AFIS	aberration free image shift
BAM	beta-barrel assembly machinery
BSA	Bovine serum albumin
CC	closed-closed state
CCD	charge coupled device
cryoEM	cryo-electron microscopy
CSS	complexation significance score
CTF	contrast transfer function
DALI	distance matrix alignment
DBP	dextran binding protein
DDM	n-Dodecyl-β-D-Maltopyranoside
DED	direct electron detector
DP	degree of polymerisation
ECF	extracytoplasmic function
EMDB	electron microscopy data bank
EPR	electron paramagnetic resonance spectroscopy
FEG	field emission gun
FOS	fructo-oligosaccharide/s
FSC	Fourier shell correlation
Glc	Glucose
GlcN	Glucoseamine
Hep	Heptose
HTCS	hybrid two component system
iBAQ	intensity-based absolute quantification
IM	inner membrane
IMAC	immobilised metal affinity chromatography
ITC	isothermal titration calorimetry

Kdo	3-deoxy-D-*manno*-oct-2-uronic acid
LBP	levan binding protein
LC-MS/MS	liquid chromatography - tandem mass spectroscopy
LDAO	Lauryldimethylamine oxide or N,N-dimethyldodecylamine N-oxide
Lol	Localisation of lipoprotein
LOS	lipooligosaccharide
LPS	lipopolysaccharide
Lpt	LPS transport system
Man	Mannose
MAPS	monolithic active pixel sensor
MD	molecular dynamics
ML	maximum likelihood
MM	minimal medium
MS	mass spectrometry
MSP	membrane scaffold protein
MST	microscale thermophoresis
MW	molecular weight
N	'narrow open' state
NBD	nucleotide binding domain
NOESY	nuclear Overhauser effect spectroscopy
NTE	N-terminal extension
OC	open-closed state
OM	outer membrane
OMV	outer membrane vesicle
OO	open-open state
PDB	protein data bank
PE	phosphatidyl ethanolamine
PG	phosphatidyl glycerol
PMF	proton motive force
POTRA	polypeptide transport associated
PUL	polysaccharide utilisation locus
RMSD	root-mean-square deviation

SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SEC-MALS	Size exclusion chromatography and multi-angle light scattering
SGBP	surface glycan binding protein
SNR	signal to noise ratio
SPA	single particle analysis
STN	Secretin and TonB N-terminus domain
Sus	starch utilisation system
TBDT	TonB-dependent transporter
TCA	trichloroacetic acid
TEM	transmission electron microscopy
TLC	thin-layer chromatography
TM	transmembrane
W	wide open state
WT	wild type

Chapter 1 Introduction

Biological lipid bilayers act as barriers. They delineate cells and, in eukaryotic organisms, organelles. This compartmentalisation allows cells to establish environments that are distinct from their external medium, and specialised to facilitate different cellular processes. However, no cellular process operates in isolation and therefore translocation of substrates across these membranes is essential.

Whether prokaryotic or eukaryotic, cell survival is dependent on the uptake of nutrients across the cell membrane. For Gram negative bacteria this presents an interesting challenge, as an additional outer-membrane (OM) protects these organisms from the often harsh and rapidly changing environment of their ecological niche. They must therefore import nutrients without compromising the integrity of this outer protective layer.

1.1 The Gram-negative bacterial cell envelope

The Gram-negative bacterial cell envelope has a tripartite structure. First observed using transmission electron microscopy (TEM) in 1958 (Kellenberger and Ryter, 1958) and more comprehensively described ten years later (Glauert and Thornley, 1969), the three principle layers are now known to correspond to the inner membrane (IM), the peptidoglycan sacculus and the OM, with the architecture of each understood in molecular detail (Nikaido, 2003, Vollmer et al., 2008, Furse and Scott, 2016).

The inner membrane is an energised, symmetrical, phospholipid bilayer which, in *E. coli*, is comprised mainly of the glycerophospholipids phosphatidyl ethanolamine (PE) and phosphatidyl glycerol (PG) (Lugtenberg and Peters, 1976). Functionality is predominantly conferred by its complement of integral and associated proteins, which include transporters, receptor-signalling systems, as well as the components required for a number of essential processes such as energy transduction and lipid synthesis (Silhavy et al., 2010).

The presence of the OM creates an additional cellular compartment called the periplasm. Unlike the cytoplasm, the periplasm is an oxidising environment, harbouring oxidoreductases,

disulphide isomerases and chaperones, making it an important site for protein folding and quality control (Goemans et al., 2014). Potentially harmful hydrolytic enzymes, including nucleases and phosphatases, are also sequestered here, ensuring vulnerable components within the cell are protected (Miller and Salama, 2018). The periplasmic space also harbours the mixed saccharide-peptide structure known as the peptidoglycan sacculus. This scaffold-like assembly provides the cell with the structural rigidity required to withstand turgor pressure (Doyle and Marquis, 1994) and is tethered to the OM via Braun's lipoprotein, or LPP, the length of which determines the periplasmic width (Cohen et al., 2017; Asmar et al., 2017). Lastly, the periplasm is home to the assembly platforms and trans-envelope machines required for biogenesis of the OM. Remarkably, these processes must proceed in the absence of ATP since the periplasm is devoid of this nucleotide. As such, trafficking and assembly pathways have evolved to be energetically favourable and, in some cases, are coupled to energy transduction at the IM.

The OM serves as a selective permeability barrier, providing bacteria with a means of defence against noxious compounds in the environment, such as antibiotics, whilst permitting the influx of nutrients required for growth (Nikaido, 2003). The bilayer itself is an inherently effective barrier owing to its construction (see **1.2.1**), whilst selective permeability is conferred by the outer membrane proteins and lipoproteins embedded within it. Given that the work presented in this thesis is focussed on protein-mediated transport of nutrients across the OM, the architecture of the OM, its assembly, and the molecular basis of outer membrane permeability will be reviewed here.

1.2 The outer membrane – an asymmetric permeability barrier

The composition and organisation of the OM of Gram-negative bacteria differs markedly from that of the IM. Although still a lipid bilayer, the OM is asymmetric, possessing an inner leaflet of glycerophospholipids (similar to the IM) and an outer leaflet comprised almost exclusively of lipopolysaccharide (LPS) (Kamio and Nikaido, 1976). The hydrophobic core of the bilayer excludes hydrophilic solutes, which includes most nutrients, however the translocation of lipophilic solutes across the OM is also extremely slow (Nikaido, 1976). The explanation for this is found in the structure of LPS.

1.2.1 Lipopolysaccharide

LPS is an amphipathic molecule produced by most Gram-negative bacteria and is the major structural component of the outer leaflet of the outer membrane. Its structure can be divided into three domains: lipid A, the core oligosaccharide and the O-antigen.

Lipid A is an acylated β 1`-6-linked glucosamine disaccharide and its hydrophobic acyl chains form the outer leaflet of the OM (Raetz and Whitfield, 2002). The 'classical' example produced by *E. coli* and *Salmonella* is hexa-acylated and bisphosphorylated (see **Figure 1.1**). Lipid A is the most conserved region of LPS and, as a result, host immune systems have evolved exquisite sensitivity to it (Bryant et al., 2010).

The core oligosaccharide has a non-repeating structure and comprises two halves, the inner core which is proximal to lipid A, and the outer core to which the O-antigen is attached. The inner core usually consists of 3-deoxy-D-*manno*-oct-2-ulosonic acid (Kdo) residues and heptose units and is well conserved at the genus/family level, while the outer core is built from hexose units and exhibits greater structural diversity (Raetz and Whitfield, 2002).

The O-antigen is synthesised independently from the rest of the LPS molecule and exhibits enormous structural variability. It consists of repeating oligosaccharide units of 2-8 sugars, and with over 60 monosaccharides and 30 non-carbohydrate substituents identified as building blocks these units are very diverse. Repeating units from different structures may differ in the number of monosaccharides as well as in their branching pattern. On top of this, modifications such as glucosylation (Bogomolnaya et al., 2008), acetylation (Kim et al., 1999), and the addition of acidic repeats such as colanic acid (Meredith et al., 2007) have been observed, resulting in an incredible variety of O-antigen structures.

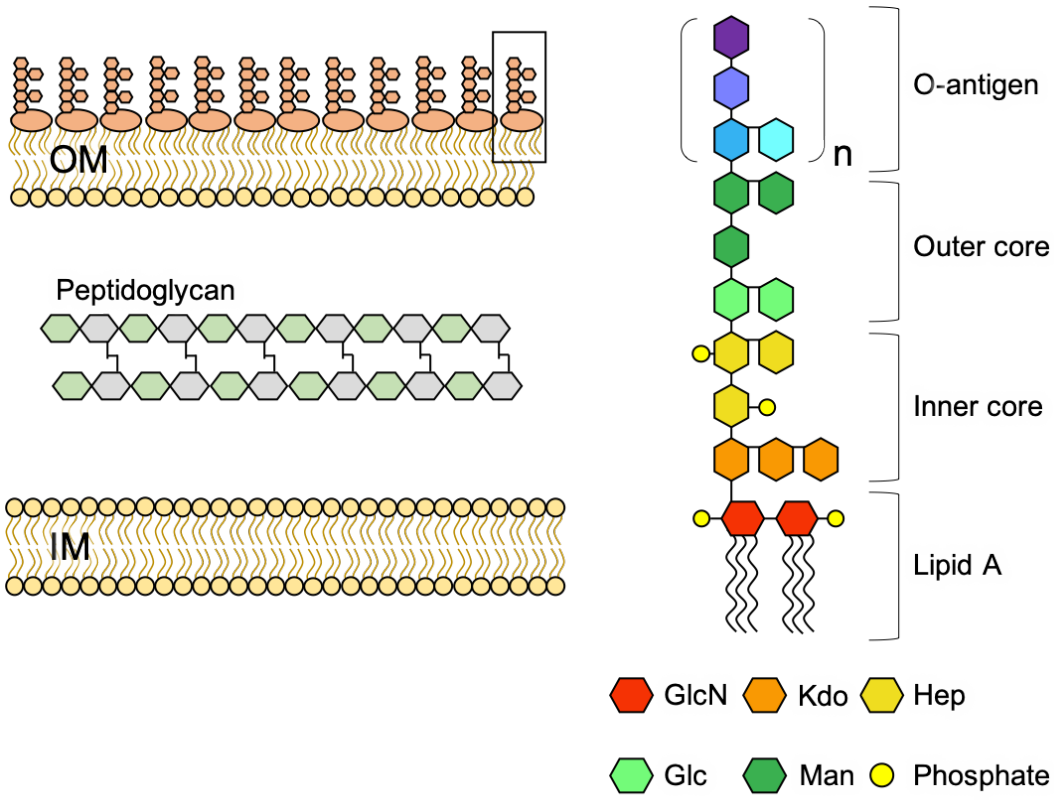


Figure 1.1: Schematics of the Gram-negative bacterial cell envelope and LPS structure. The three principle layers of the Gram-negative bacterial cell envelope are shown (left). The outer leaflet of the OM is predominately comprised of LPS molecules (boxed region). A simplified structure of LPS is shown (right). Individual regions of the LPS molecule are labelled from Lipid A, to the O-antigen. A single repeating oligosaccharide unit is shown for the O-antigen. GlcN = Glucosamine; Kdo = 3-deoxy-D-*manno*-oct-2-ulosonic acid; Hep = Heptose; Glc = Glucose; Man = Mannose. Adapted from Figure1 Lin et al., 2020.

What is it about LPS that makes the OM such an effective barrier? The high number of saturated fatty acid moieties on lipid A promotes tight packing in the outer leaflet. This reduces the fluidity of the OM and creates a more rigid, gel-like state which is retained even at temperatures of ~60 °C (Melchior and Steim, 1976). Electrostatic repulsion between the phosphate groups at positions 1 and 4' of the glucosamine disaccharide, and those decorating the core oligosaccharide, is prevented by intercalation of divalent cations. These instead promote packing via polyionic interactions and, as a result, the cell presents a dense, hydrophilic LPS 'forest' to the extracellular milieu, offering innate protection against toxic molecules such as antibiotics and bile salts (steroid-like), while the bilayer effectively excludes polar molecules (Nikaido, 2003; Carpenter et al., 2016). Kdo_2-lipid A, i.e. lipid A and an inner core lacking the usual heptose units, is the minimal structure required for growth of *E. coli.* This so called 'deep-rough' phenotype exhibits hypersensitivity to hydrophobic agents such

as detergents, dyes and some antibiotics, whilst also leaking periplasmic proteins into the external medium, reaffirming the importance of LPS structure to the integrity of the OM (Raetz and Whitfield, 2002). Notably, some mucosal pathogens (e.g. *Neisseria spp*. (Plant et al., 2006)) and members of the human gut microbiome (e.g. *Bacteroides spp.* (Jacobson et al., 2018**)**) possess lipooligosaccharides (LOS) rather than LPS. These analogous molecules lack the O-antigen entirely and instead have an extended oligosaccharide core that, in some cases, mimics host antigen structure (Moran, 2010).

1.3 OM Biogenesis – Bridges, Shuttles and Tunnels

Assembly of the OM requires trafficking of LPS, lipids, lipoproteins and OM proteins (OMPs) across one or, often, two membranes, as well as the intervening aqueous environment of the periplasm. The amphipathic nature of these molecules mean that these processes are non-trivial and require sophisticated cellular machinery.

1.3.1 LPS synthesis and trafficking

Synthesis of LPS begins in the cytoplasm with assembly of the lipid A-core anchored in the inner leaflet of the IM (Whitfield and Trent, 2014). Flipping of LPS to the outer leaflet of the IM and subsequent trafficking to, and flipping across, the OM is facilitated by the concerted function of the MsbA flippase, and the proteins of the Lpt (LPS transport) system respectively (see **Figure 1.2**). MsbA is a multi-drug resistance ABC-transporter located in the IM. Assembled as a homodimer, each subunit possesses a cytoplasmic nucleotide binding domain (NBD) and a transmembrane domain. Substrate recognition and transport is carried out by the transmembrane domains which together form an LPS binding cavity, whilst ATP binding and hydrolysis occurs in the NBDs. Conformational changes involving these domains are intimately linked, such that ATP binding and hydrolysis is coupled to LPS flipping to the outer leaflet. A combination of crystal structures (including trapped intermediates) and cryo-electron microscopy structures of MsbA in a bilayer environment have culminated in a 'trap and flip' model for MsbA-mediated LPS transport (Ward et al., 2007; Mi et al., 2017).

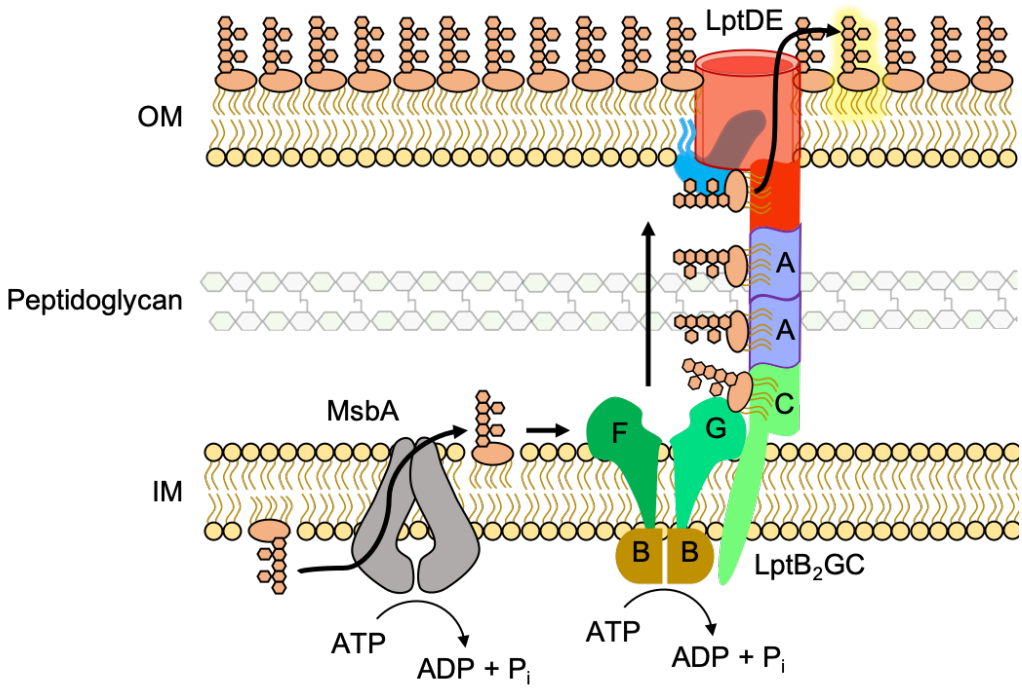


Figure 1.2: Role of MsbA and the Lpt system in the 'PEZ' model of LPS trafficking. Schematic depicting the transport of LPS from the inner leaflet of the inner membrane to the outer leaflet of the outer membrane. Flipping of LPS across the IM is facilitated by the ATP-binding cassette transporter MsbA (grey). $LptB_2FG$ then uproots LPS from the periplasmic leaflet of the IM and loads it onto the LptCAD bridge. This process is repeated, with each newly extracted LPS molecule being pushed onto the bridge, forcing the previous molecule towards LptDE complex in the OM. Once LPS reaches the end of the bridge it is 'dispensed' to the cell surface. The whole process is reminiscent of a PEZ dispenser where a sweet at the top of a stack is pushed out by the action of a spring at the bottom of the stack. Ferrying LPS along a bridge in this way is a highly efficient transport mechanism, allowing maintenance of membrane integrity during cell growth and division.

Once flipped to the outer (periplasmic-facing) leaflet of the IM, LPS is handed to the Lpt system; a protein bridge that traverses the entire cell envelope (Sherman et al., 2018) (see **Figure 1.2**). Made up of seven unique components (LptA-G), the Lpt system comprises two sub-assemblies: the $LptB_2FGC$ complex at the IM, and a heterodimer of LptD and LptE at the OM. The periplasmic protein LptA links these sub-assemblies and completes the trans-envelope bridge (Sperandeo et al., 2019). A wealth of structural information is available for Lpt components, including complete structures of the IM and OM sub-assemblies (Tang et al., 2019; Qiao et al., 2014; Dong et al., 2014; Botos et al., 2016). We therefore have a near-complete picture of the LPS trafficking pathway which is described by the so-called 'PEZ' model (see **Figure 1.2**) (Sherman et al., 2018; Okuda et al., 2016).

$LptB_2FG$ is a member of the ABC-transporter family responsible for 'uprooting' its membrane anchored LPS substrate (Tang et al., 2019). Post extraction, LPS is received by LptC, an auxiliary lipoprotein associated with the $LptB_2FG$ complex (Okuda et al., 2012, Narita and Tokuda, 2009). LptC, together with LptA and the periplasmic domain of LptD, forms a bridge across the periplasm. The bridge presents a concave hydrophobic surface that facilitates the movement of the lipid-A portion of LPS across the aqueous environment of the periplasm (Freinkman et al., 2012; Okuda et al., 2012; Sperandeo et al., 2019). Translocation of the LPS molecule across the OM and consequent presentation on the cell-surface is facilitated by the LptDE complex. The transmembrane domain of LptD is comprised of a large, 26-stranded β-barrel that is plugged by the lipoprotein LptE (Qiao et al., 2014; Dong et al., 2014; Botos et al., 2016). A two-portal mechanism is proposed whereby the lipid A portion of LPS moves straight into the membrane from the N-terminal periplasmic domain of LptD, whilst the hydrophilic oligosaccharide region moves through the lumen of the LptD/E complex. This is permitted by the opening of a lateral gate between the first and last β-strands of LptD, the cross-linking of which produces a lethal phenotype (Dong et al, 2014; Botte et al., 2021). An exit pore, gated by extracellular loop L4, permits escape of the oligosaccharide region, driven by the presence of divalent cations intercalated between existing LPS molecules on the cell surface (Dong et al., 2014; Gu et al., 2015).

1.3.2 Lipid trafficking

In comparison to the LPS transport system, very little is known about the mechanism of glycerophospholipid transport from their site of synthesis at the IM, to the inner leaflet of the OM. MCE proteins, so named because the *mce1A* gene was found to be important for mammalian cell entry in *Mycobacterium tuberculosis* (Arruda et al., 1993), contain one or more MCE domains, and oligomerise to from hexameric rings with a central channel. Members of this superfamily have been implicated in the transport of lipids (Malinverni and Silhavy, 2009; Thong et al., 2016), cholesterol (Pandey et al., 2008) and steroids (Mohn et al., 2008). However, MCE-containing systems are generally considered to mediate lipid transport from the OM to the IM (retrograde transport) and have important roles in the maintenance of OM asymmetry rather than supplying lipids to the OM (Malinverni and Silhavy 2009, Chong et al., 2015; Ekiert et al.,2017).

The Mla system, containing the single MCE domain protein MlaD, is one of the best studied and is proposed to function as part of a shuttle system for retrograde lipid transport, with MlaD receiving phospholipids from a periplasmic carrier MlaC for reinsertion into the IM (see **Figure 1.3**). However, recent structure function studies on the multi MCE domain-containing proteins PqiB and LetB (formerly YebT) have revealed novel strategies for moving lipids. These proteins form elongated tunnel structures, albeit with strikingly different architectures, that have the potential to span the periplasm and therefore directly facilitate lipid transport between membranes (see **Figure 1.3**) (Ekiert et al., 2017; Isom et al., 2020). However, directionality of transport has not been demonstrated.

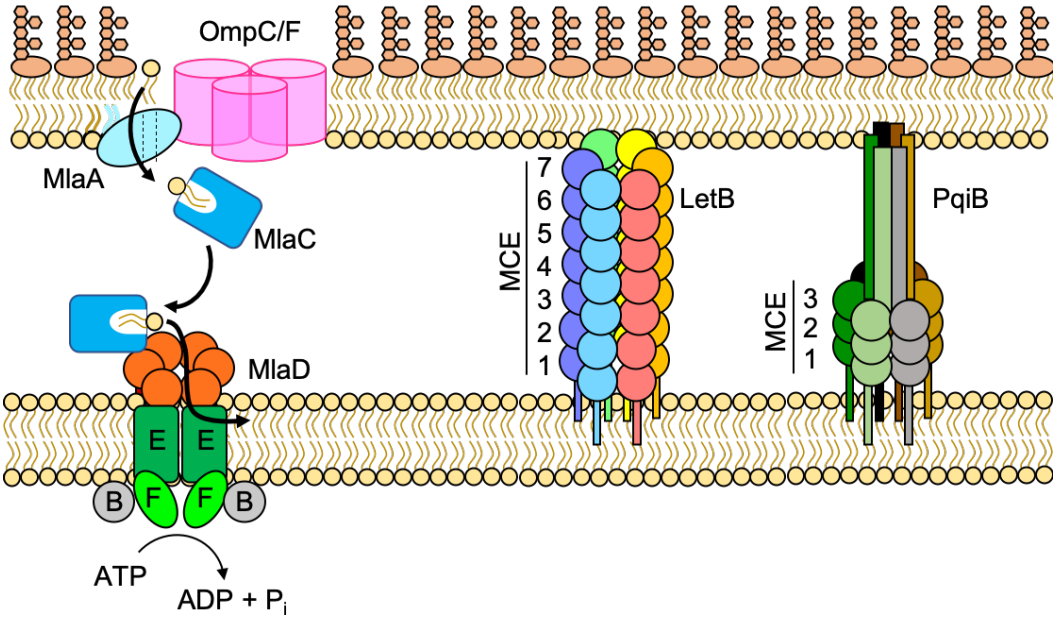


Figure 1.3: MCE-mediated phospholipid transport across the periplasm in *E. coli*. Schematic of various mechanisms of lipid transport including the shuttle-like Mla system (left) and the putative trans-envelope tunnels of LetB (centre) and PqiB (right). The Mla system is considered a retrograde lipid transport system and is responsible for maintaining asymmetry of the OM by removing misplaced phospholipids. A complex of MlaA-OmpC/F removes phospholipids from the OM and passes them to the periplasmic carrier protein MlaC. MlaC then delivers the lipid to the MlaFEDB complex at the IM for re-insertion into the IM. The MlaFEDB complex functions as an ABC transporter. MlaD contains a single MCE domain resulting in a hexameric ring assembly with a central hydrophobic pore capable of receiving the lipid substrate. MlaE is an ABC-transporter permease and MlaF is an ABC transporter nucleotide binding component. MlaB is important for assembly and activity of the complex. LetB and PqiB proteins contain 7 and 3 MCE domains respectively and this determines the number of stacked rings in the final hexameric assembly. In the case of PqiB, elongated C-terminal helical regions come together in the hexamer to give a hollow syringe-like structure. Both structures have the potential to span the periplasmic space and therefore facilitate lipid transport between membranes. Adapted from Figure 7 Ekiert et al., 2017.

1.4 Lipoproteins

Another unique feature of the OM is its low lipid to protein ratio (LPR). Proteins, including integral OMPs and OM-associated lipoproteins (Koebnik et al., 2000), account for ~50 % of the mass of the OM. The latter play important roles in a diverse range of cellular processes including: OM biogenesis and maintenance, OMP folding, signal transduction, nutrient import, biofilm formation and cell division (Braun and Hantke, 2019). Like all proteins they are synthesised in the cytoplasm and must therefore cross the IM and periplasm before insertion into the OM. Translocation across the IM occurs post-translationally and is facilitated by the Sec translocon; a heterotrimer of the integral membrane proteins SecY, SecE and SecG. SecY has 10 transmembrane helices arranged in two pseudo-symmetric halves and is the channel-forming component of the complex (van den Berg et al., 2004). Targeting of nascent pre-proteins to the translocon is achieved via a signal peptide at their N-terminus (Dreissen and Nouwen, 2008). Upon arrival, the substrate protein is threaded through SecY in an unfolded state, and translocation is powered by the associated ATPase SecA. SecYE and the SecA motor protein constitute the minimal translocase, although the presence of SecG increases the efficiency of translocation (Brundage et al., 1990; Nishiyama et al., 1994).

Pro-lipoproteins remain anchored in the IM via their signal peptide. This region contains a consensus sequence known as the lipobox; Leu-(Ala/Ser)-(Gly/Ala)-Cys (Hayashi and Wu, 1990; Braun and Wu, 1994). The N-terminal cysteine of the lipobox is covalently linked to diacyl glycerol in the IM by the enzymatic activity of Lgt. Once modified, the pro-lipoprotein is targeted by LspA (or signal peptidase II) which cleaves the signal peptide (Narita and Tokuda, 2017). The resulting α-amino group is then acylated by the N-acyltransferase Lnt, generating a mature, tri-acylated lipoprotein; the acyl chains of which are embedded in the membrane. Whether the mature lipoprotein remains at the IM or is trafficked to the OM is governed by the so-called '+2' rule. An aspartate residue following the N-terminal Cys (+2), is the signal for IM retention, although the identity of the residues at the +3 and, outside the enterobacteria, +4 positions can also play a role in differentiating IM and OM lipoproteins (Yamaguchi et al., 1988; Konovalova and Silhavy, 2015).

Lipoproteins without a retention signal are handled by the localisation of lipoproteins (Lol) system (see **Figure 1.4**), which in *E. coli* comprises five proteins (LolA-E). A subcomplex of $LolCD_2E$ at the IM extracts the substrate lipoprotein from the outer leaflet of the IM before passing it to the soluble periplasmic chaperone LolA (Tang et al., 2021; Sharma et al., 2021). Although homologous, LolC and LolE perform distinct functions, with the former recruiting the shuttle protein LolA and the latter binding the substrate lipoprotein. Lipoprotein extraction is energy-dependent and is coupled to ATP hydrolysis by the cytoplasmic ATPase LolD, which drives the handover from LolE to LolA. A hydrophobic cavity within LolA shields the acyl chains of the substrate whilst it is delivered to the final component of the Lol system, the OM lipoprotein LolB. Like LolA, LolB contains a hydrophobic cavity and a 'mouth-to-mouth' transfer event from LolA to LolB is driven by a higher affinity of the latter for the acyl chains of the substrate (Taniguchi et al., 2005). LolB then facilitates lipoprotein insertion into the inner leaflet of the OM. It should be noted that LolCDE has been implicated in an alternative trafficking pathway that is independent of the LolA and LolB components (Grabowicz and Silhavy, 2017).

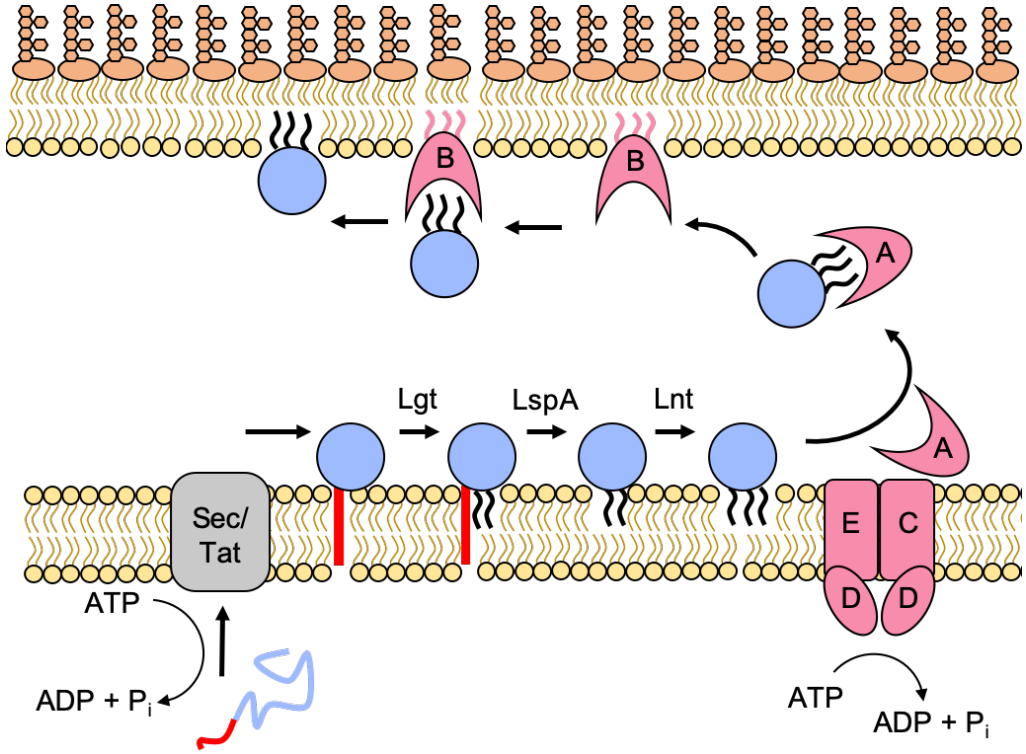


Figure 1.4: Lipoprotein trafficking by the Lol system. Schematic of lipoprotein maturation and transport to the OM. The nascent lipoprotein is targeted to the Sec translocon via its N-terminal signal sequence. The signal sequence anchors the pro-lipoprotein in the IM during the maturation process. Lgt catalyses the addition of diacylglycerol to the thiol group of the N-terminal cysteine of the lipobox. The signal sequence is then cleaved by LspA and Lnt acylates the new N-terminus. An ABC transporter complex, LolCDE, provides the energy to uproot the mature lipoprotein, extracting it from the IM and handing it off to an associated chaperone, LolA. LolA ferries the lipoprotein across the periplasm to the LolB acceptor at the OM. LolB then deposits the lipoprotein in the inner leaflet of the IM. Adapted from Figure 1 Konovalova and Silhavy, 2015.

Importantly, some lipoproteins are presented on the cell surface, and until recently this phenomenon was under-appreciated. The previous general consensus was that the vast majority of lipoproteins resided in the inner leaflet of the OM, facing the periplasm (Wilson and Bernstein, 2016). Whilst this holds true for *E. coli*, in other organisms such as *Bacteroides spp.* (abundant in the gut microbiome) surface exposed lipoproteins make up ~35% of the lipoprotein complement under laboratory growth conditions (Wilson et al., 2015). It is now generally accepted that surface exposure of lipoproteins is widespread. Despite this, for the majority of surface-exposed lipoproteins, the mechanism by which they reach their final destination is unknown.

1.5 Outer membrane proteins – architecture, folding and function

Although lipoproteins are anchored in the OM, the term outer-membrane protein is reserved for integral β-barrel proteins i.e. proteins that traverse the OM using varying numbers of β-strands folded such that they form a closed 'barrel' structure. As discussed above, the OM is an effective barrier to hydrophilic solutes, including most essential nutrients. For this reason, many OMPs are channel-forming proteins that confer selective permeability to the OM. These can be grouped into porins, substrate-specific channels, and active transporters (Koebnik et al., 2000). The following sections will cover OMP architecture, folding and function.

1.5.1 OMP structure

All outer membrane proteins traverse the membrane by forming β-barrel structures (Schulz et al., 2002). These barrels are made up of all-next-neighbour anti-parallel strands tilted 30-60 ° in relation to the barrel axis. This generates a right-handed twist and higher degrees of tilt result in barrels with larger diameters. Strands are generally connected by short turns (1-12 residues) at the periplasmic face and longer loops on the extracellular side.

The number of strands per barrel varies considerably, from the minimal construction of 8 in OmpA and OmpX to the huge 36-stranded barrel structure of SprA: the OM translocon of the type 9 secretion system (Pautsch and Schulz, 1998; Vogt and Schulz, 1999; Lauber et al., 2018) (see **Figures 1.5** and **1.6**). Multimeric barrels are larger still, with the secretin of the type II secretion system possessing 60 strands, contributed by 15 copies of GspD (Yan et al., 2017) (see **Figure 1.6**). Regardless of their strand complement, extensive inter-strand hydrogen bonding makes β-barrels highly stable (Bishop et al., 2001), with *in vitro* estimates of folding free energy values in the range of -13 to -32 kcal mol^{-1}. As a result, OMPs are resistant to unfolding by chaotropic agents and thermal denaturation (Moon et al., 2013; Chaturvedi and Mahalakshmi, 2017). Within the hydrophobic core of the bilayer, the outer surface of the barrel presents non-polar and aromatic side-chains to the membrane environment (Koebnik et al., 2000). Belts of aromatic residues, specifically Trp and Tyr, are found at the membrane-water interfaces where they can interact with the lipid headgroups, fixing the orientation of the barrel within the membrane (Schulz, 1992; Yao et al., 1998; Schulz et al., 2002; Wimley, 2003; Killian and von Heijne, 2000).

Residues presented to the lumen of these β-barrels are predominantly polar. Even the 8-stranded non-channel-forming OmpA and OmpX proteins possess a polar core containing intricate hydrogen bonding networks (Pautsch and Schulz, 1998; Vogt and Schulz, 1999). This is in contrast to water-soluble β-barrels which possess nonpolar interiors (Schulz, 2002). Interactions in the interior of the barrel can have a significant impact on their cross-sectional shape. OMP barrels are often not circular: OmpX has an ellipsoidal cross-section whilst OmpLA, FhuA and LptD have half-moon-, bean- and kidney-shaped cross-sections respectively. These distortions result from internal sidechain interactions as well as inserted domains, loops and independent protein subunits that occupy the barrel (Koebnik et al.,2000) (see **Figures 1.5** and **1.6**). Interestingly, transitions in cross-sectional shape are suggested to be important in the mechanism of substrate export by SprA of the type 9 secretion system, where two distinct barrel conformers (circular and kidney-like) are linked to the mutually exclusive binding of two partner proteins on the basis of shape complementarity (see **Figure 1.6**) (Lauber et al., 2018).

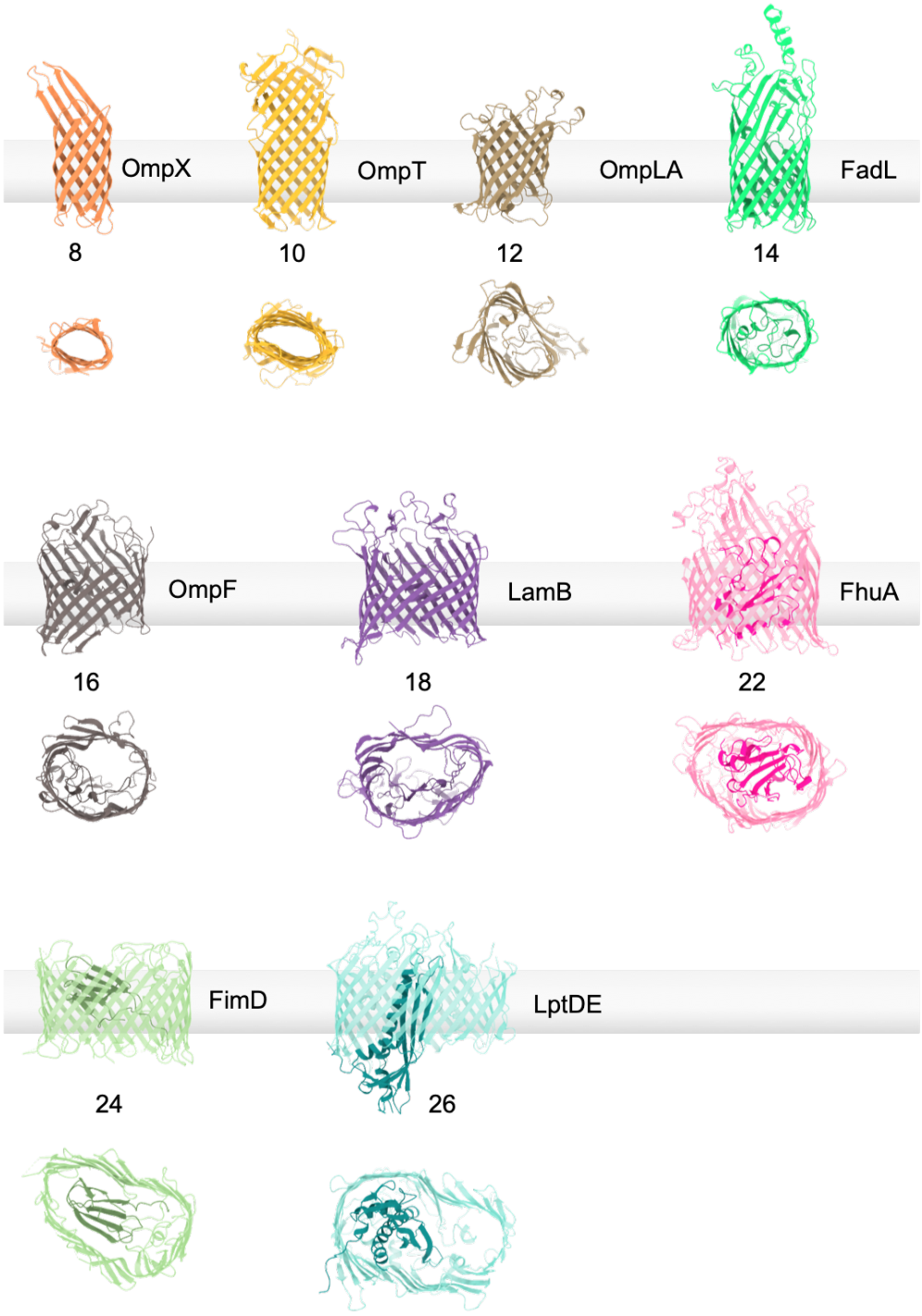


Figure 1.5: Variation in strand complement of outer membrane proteins. Selection of OMP structures demonstrating differing β-strand complements (8-26). OMPs are shown in the plane of the the membrane and from the periplasm to demonstrate the variation in cross-sectional shape. The number of strands in each barrel is indicated. Note that barrels with more than 18 strands possess a plug domain e.g. FhuA or separate plugging protein e.g. LptDE that occludes the barrel lumen. Models are on presented on the same scale as those in **Figure 1.6** for comparison.

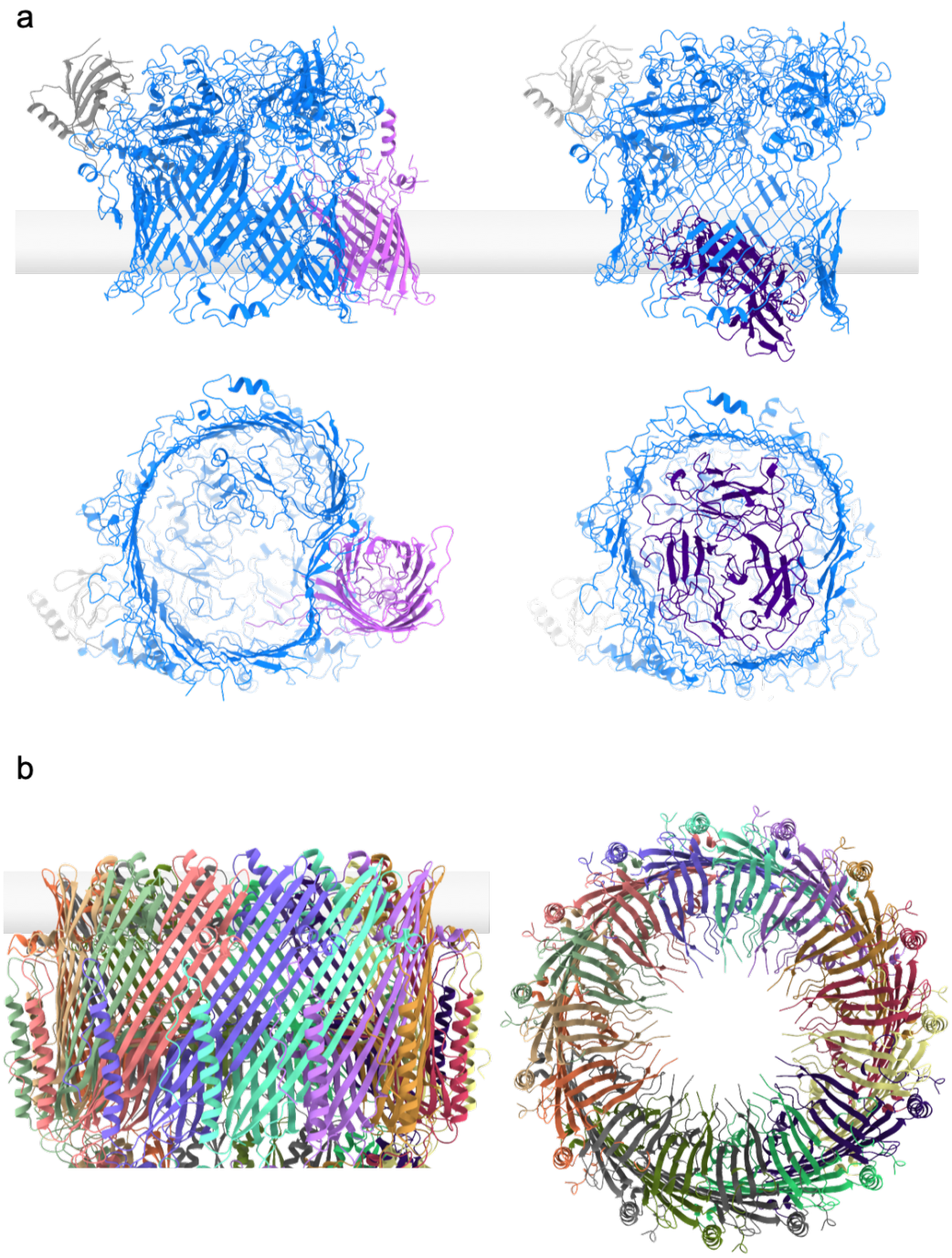


Figure 1.6: Type9 and Type2 secretion systems as examples of large outer membrane proteins. **a**,The 36 stranded SprA protein of the Type9 secretion system (blue) shown with a partner protein PorV (pink) (left) or a plug protein (purple) (right). The binding of PorV and the plug protein are mutually exclusive. The kidney-like conformer of the barrel complements PorV binding whilst the circular conformer complements plug protein binding. **b,** An example of a multimeric barrel comprising 60 strands contributed by 15 copies of GspD component of the type-II secretion system. Left is a view in the plane of the membrane (N.B. much of the periplasmic domain is cropped from the image). A clipped view from the periplasm is shown on the right. Models are presented on the same scale as those in **Figure 1.5** for comparison.

1.5.2 OMP trafficking and folding

Like lipoproteins, OMPs are translocated across the IM by the Sec translocon. OMPs are then processed by signal peptidase-I, which cleaves the N-terminal signal sequence to release the pre-protein into the periplasm (Auclair et al., 2012). Aggregation and misfolding are prevented by the concerted action of the periplasmic chaperones SurA and Skp, as well as the protease DegP (Rollauer et al., 2015). SurA is thought to be the principle OMP chaperone, responsible for protecting nascent OMPs from aggregation and delivering them to the β-barrel assembly machinery (BAM complex) at the OM (Sklar et al., 2007; Bennion et al., 2010) (see **Figure 1.7**). Cross-linking experiments, hydrogen-deuterium exchange mass spectrometry and single molecule FRET suggest that SurA adopts a dynamic and responsive structure that cradles clients to prevent misfolding (Calabrese et al., 2020). These dynamics are likely critical to the ability of SurA to hand-off clients to the BAM complex in the absence of ATP.

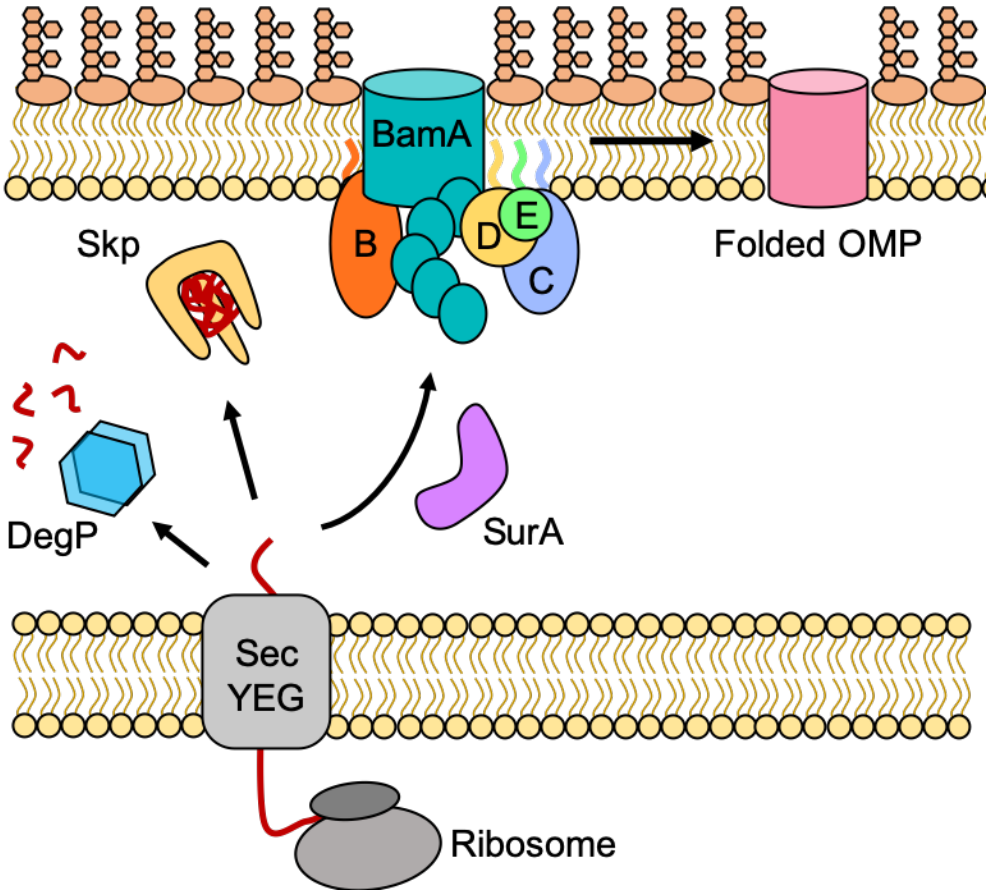


Figure 1.7: OMP biogenesis and quality control schematic. Nascent OMPs are translocated across the IM by the Sec translocon. Once in the periplasm the OMP remains unfolded and is handled by the chaperones SurA and Skp for delivery to the Bam complex. The Bam complex catalyses folding of the nascent OMP into the OM. Misfolded OMPs are targeted by the protease DegP to prevent toxic accumulation of aggregates. Adapted from Hagan et al., 2011.

surA knockouts show OM defects whilst knockouts of *skp* and *degP* have a lesser effect; even double knockouts of *skp* and *degP* are viable. It is therefore suggested that Skp and DegP act in secondary pathways, parallel to SurA (Rizzitello et al., 2001). Skp functions as a 'jellyfish'-like homotrimer (Walton et al., 2004; Korndörfer et al., 2004). Three α-helical 'legs' provide a hydrophobic cavity that cages the prospective β-barrel region of the unfolded client (Walton et al., 2009). For clients that exceed the capacity of this cavity, evidence indicates that multiple Skp assemblies can act cooperatively to maintain the unfolded state (Schiffrin et al., 2016). In contrast, DegP functions primarily as a serine protease (Rollauer et al., 2015). Essential in Gram negative bacteria under heat shock conditions (Strauch et al., 1989), it breaks down misfolded OMPs avoiding toxic accumulation of aggregates in the periplasm (Konovalova et al., 2018). Inactive DegP is hexameric but binding to misfolded substrates

triggers oligomerisation resulting in 12- and 24-mer cages with proteolytic activity (Krojer et al., 2008).

After successfully navigating the periplasm, nascent OMPs must insert and fold into the OM. Since the constituent β-strands of OMPs are amphipathic, this poses an interesting folding challenge. Accordingly, this process is catalysed by a sophisticated OMP folding machine, the BAM complex. In *E. coli*, the BAM complex comprises five components: BamA, which is itself a 16-stranded OMP possessing five periplasmic polypeptide transport-associated (POTRA) domains, and four associated lipoproteins (BamB-E). BamA and BamD are essential for cell viability and, accordingly, homologues of these proteins are present in all Gram-negative bacteria (Voulhoux et al., 2003, Wu et al., 2005; Malinverni et al., 2006). Whilst not essential, deletion of any one of the remaining components causes defects in OMP folding (Sklar et al., 2007; Hagan et al., 2010).

Several models for Bam-catalysed folding of OMPs have been proposed over the last decade, with two of the most prominent being the BamA-assisted model and the budding model (Konovalova et al., 2017). In the assisted model, chaperones deliver the unfolded protein to the Bam complex. Initiation of barrel formation is suggested to occur on the periplasmic side of the OM, driven by the intrinsically high free energy of folding for β-barrels. The Bam complex, namely BamA, is proposed to facilitate insertion of the nascent OMP into the OM by locally destabilising the lipid bilayer. In contrast, the budding model describes a more systematic mode of catalysis. BamA possesses a lateral gate or seam between β-strands 1 and 16 (Gu et al., 2016; Han et al., 2016; Iadanza et al., 2016) and this seam permits templating of β-strand formation in the nascent OMP (Noinaj et al., 2014). The C-terminal strand of the nascent OMP is proposed to interact with the N-terminal strand of BamA through β-augmentation, and sequential templating of β-hairpins leads to the formation of a hybrid BamA-substrate barrel, satisfying the amphipathic requirements of the β-strands. Once complete, the substrate barrel is hypothesised to 'bud-off' into the OM. Recent work has revealed that the BAM complex likely utilises aspects of both models to catalyse OMP folding. Molecular dynamics simulations show disordering of lipids and membrane thinning at the lateral gate, while cryoEM BAM-nanodisc structures provide evidence for bilayer deformation

in the form of a perturbed membrane scaffold (Iadanza et al., 2020). High resolution structure determination of a trapped BamA-substrate intermediate by cryoEM provides direct evidence for the hybrid barrel predicted by this budding model and also gives insight into the mechanism of substrate release (Tomasek et al., 2020). Moreover, structural studies of BamA in complex with the natural product darobactin confirm the importance of the lateral gate for initial strand templating of the nascent OMP (Kaur et al., 2021).

1.6 OMP-mediated transport across the OM

As discussed above, the OM is an effective barrier to hydrophilic solutes, including most essential nutrients. For this reason, the OM contains integral, channel-forming proteins that provide selective permeability. These include porins which permit passive diffusion of ions or small polar molecules, substrate specific channels, and TonB-dependent transporters. This latter class is involved in the active transport of rare nutrients or nutrients too large to pass through diffusion channels (Nikaido, 2003).

1.6.1 Porins

Structural studies on the classical porins of *E.* coli, namely OmpC, OmpF and PhoE, revealed that these proteins traverse the membrane via a 16-stranded β-barrel (Nikaido et al., 2003). These barrels exist as trimeric assemblies in the OM, stabilised by a combination of hydrophobic interactions at the barrel interfaces and inter-subunit interactions provided by extracellular loop 2 which folds away from the barrel interior and contacts the neighbouring barrel (see **Figure 1.8**). The lumen of porin barrels are generally water-filled, and therefore facilitate the non-specific diffusion of small hydrophilic solutes i.e. the direction of solute movement is dependent on the concentration gradient that exists across the OM. A characteristic feature of porins is an extended extracellular loop 3 (known as the 'eyelet') which folds into the channel and restricts the diameter of the pore, resulting in a size limit of ~600 Da for this family of transporters (Vergalli et al., 2019). Importantly, this makes porins the main route of entry for small antibiotics such as β-lactams and tetracycline but also means that many larger antibiotics are excluded. Loop 3 is also largely responsible for the selective behaviour of porins; OmpF and OmpC exhibit a preference for cations while PhoE prefers

anions. Interactions of acidic residues in loop 3 with basic residues lining the opposing barrel wall creates a transverse electrostatic field, the nature of which determines the unique properties of the pore. The electrostatic potential around the mouth of the pore also contributes to this selective behaviour (Weiss et al., 1991; Karshikoff et al., 1994; Van Gelder et al., 1997).

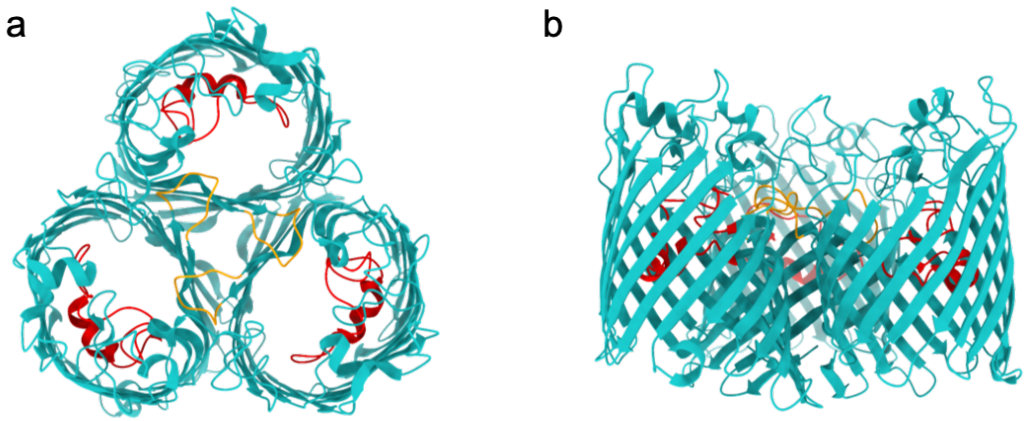


Figure 1.8: X-ray crystal structure of OmpC. a, Structure of the trimeric assembly of OmpC viewed from the extracellular space. Loop 2, important in trimerisation, is shown in orange and loop 3 that constricts the pore is shown in red. **b**, The same structure as in (**a**) viewed in the plane of the membrane.

1.6.2 Channels

Porins have limited utility to the cell when the extracellular concentration of a solute is low. The rate of diffusion through porins is proportional to the concentration gradient across the membrane, and where extracellular substrate concentrations are sub millimolar, the rate of diffusion is not sufficient to satisfy cellular demand. Substrate-specific channels exist to circumvent this problem. Unlike porins, these proteins possess specific binding sites which serve to accelerate diffusion of a particular solute (or group of solutes) at low concentrations. The archetypal channel protein is LamB (Maltoporin) which transports maltose and maltose-containing oligosaccharides (Kullman et al., 2002). LamB traverses the membrane via an 18-stranded β-barrel (see **Figure 1.9**). Like the porins, LamB is trimeric with each subunit possessing its own water-filled channel. Loop 3 folds into the barrel interior, again creating a central constriction (Schirmer et al., 1995). Structure determination in the presence of

maltodextrins revealed the mechanism of substrate binding (Dutzler et al., 1996). A helical arrangement of six aromatic residues accommodates the hydrophobic faces of the sugar residues through van der Waals' contacts (see **Figure 1.9**). Meanwhile polar residues on loop 3 were expected to participate in hydrogen bonding with sugar hydroxyl groups, a hypothesis later confirmed by site-directed mutagenesis (Dumas et al., 2000). Notably, LamB also permits translocation of other sugars (Luckey and Nikaido, 1980; Death et al., 1993), although the translocation of sucrose is severely hindered by the presence of a bulky fructosyl residue whose angle relative to the pore axis inhibits passage through the constriction point (Wang et al., 1997).

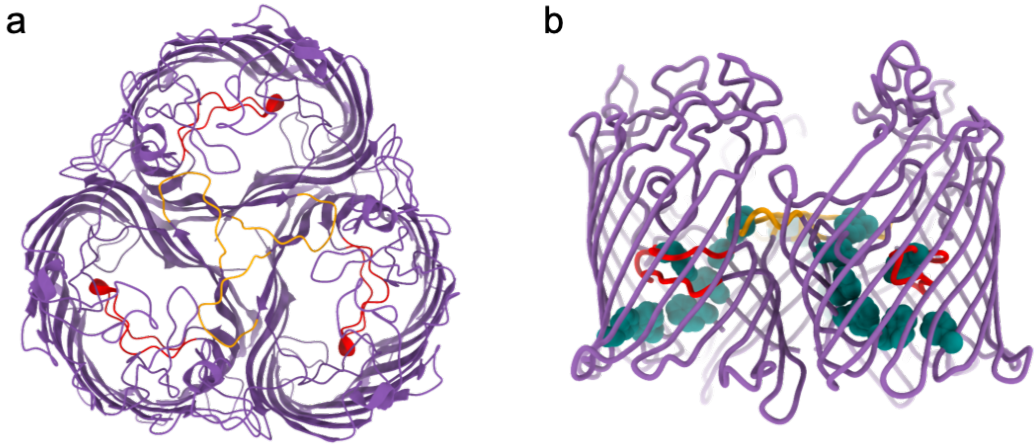


Figure 1.9: **X-ray crystal structure of LamB**. **a**, Structure of the trimeric assembly of LamB viewed from the extracellular space. Loop 2, important in trimerisation, is shown in orange and loop 3 that constricts the pore is shown in red. **b**, The same structure as in (**a**) viewed in the plane of the membrane. The helical assembly of aromatic residues (greasy slide) important for substrate specificity is highlighted; sidechains are shown with atoms as spheres (teal). Labelled sidechains of the greasy slide are W74; contributed by loop 2 of neighbouring subunit, Y41, Y6, W420, W358, F227 and Y118 (on loop 3).

1.6.3 Active transporters – TonB dependent transporters

Collectively, substrate-specific channels and porins permit the energy independent diffusion of ions and small polar molecules across the OM. However, the need to preserve the permeability barrier of the OM means that these channels are not suitable for the import of larger substrates. Moreover, the lack of high affinity binding sites makes them inefficient where a concentration gradient favouring nutrient uptake does not exist. To overcome these limitations, Gram negative bacteria utilise TonB-dependent active transport systems.

Given that 1) ATP is not present in the periplasm and 2) that the permeability of the OM to ions occludes the establishment of an energising ion gradient across the outer-membrane, these systems must transduce energy in the form of the proton motive force (PMF) across the IM. They therefore comprise a specific TonB-dependent transporter (TBDT) located in the OM, and an IM-associated Ton complex made up of TonB, ExbB and ExbD that is responsible for energy transduction (Noinaj et al., 2010).

1.7 Structure and function of TonB-dependent transporters

Unlike passive diffusion channels which show variation in their strand complement (8-18), TBDTs invariantly possess 22-stranded barrels (Noinaj et al., 2010) (see **Figure 1.10**). The first crystal structures of TBDTs (apo and ligand-bound) were solved in the late nineties and revealed the archetypal features of this transporter family (Ferguson et al., 1998; Locher et al., 1998). The lumen of their barrels is completely occluded by a globular 'plug' domain which is inserted from the periplasmic side, and ligand binding sites lie on the extracellular face of the transporter (see **Figure 1.10**). 'Unplugging', be it partial or complete, is therefore essential for substrate transport and is dependent on the presence of a semi-conserved peptide sequence (~5 residues) known as the TonB box; located just N-terminal of the plug (Schramm et al., 1987). This region interacts directly with TonB, thus permitting coupling of the OM transporter to the energy transducing Ton complex at the inner membrane (Gudmunsdottir et al. 1989; Cadieux and Kadner, 1999; Freed et al. 2013).

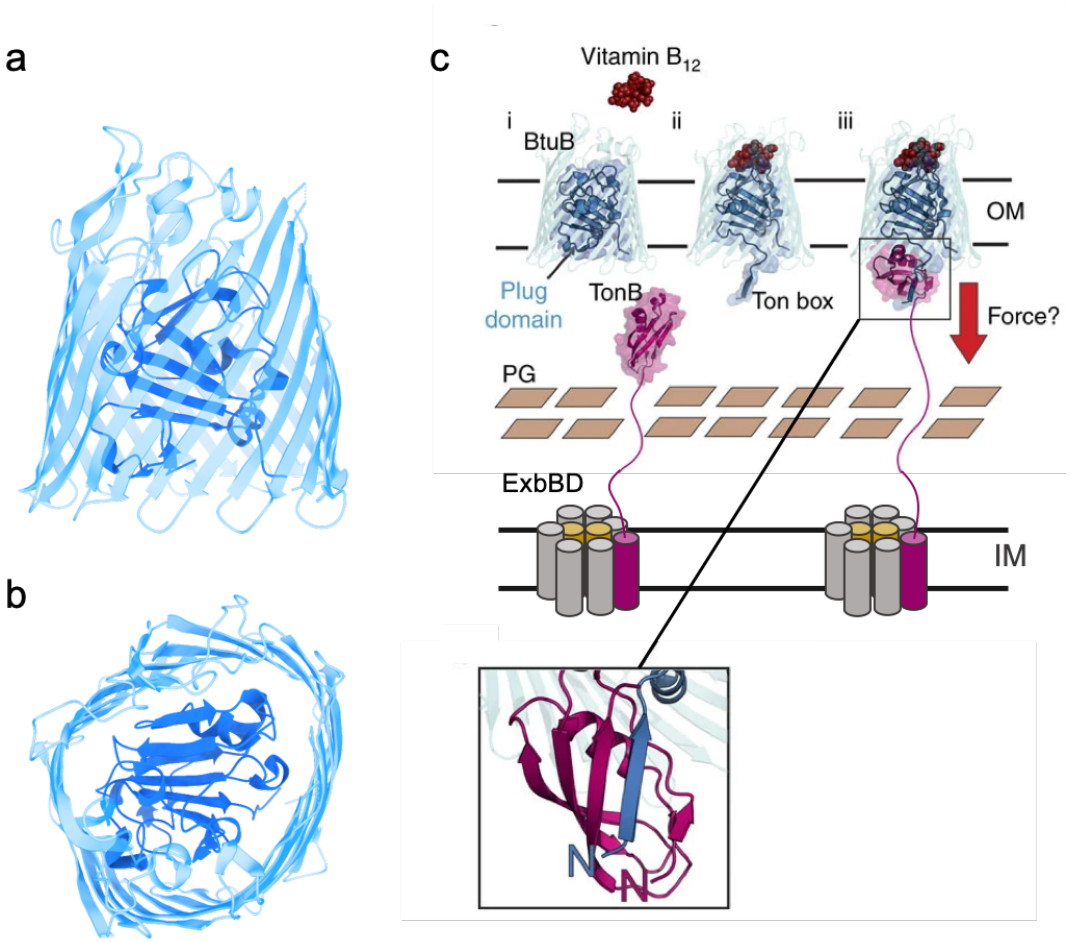


Figure 1.10: **Overview of the model TonB-dependent transporter BtuB**. **a**, Crystal structure of BtuB as viewed from the plane of the membrane. The 22-stranded barrel structure is clear. **b,** BtuB as viewed from the periplasm. The plug domain can clearly be seen to occlude the pore of the barrel. **c,** An overview of the proposed mechanism for TonB-dependent substrate translocation. Substrate binding (in this case Vitamin B12) is proposed to induce a conformational change that is transmitted through the plug domain resulting in disordering of the TonB box. This renders the TonB box accessible to TonB (part of the TonB subcomplex at the inner membrane) which exerts a force on the plug, disrupting it and therefore permitting substrate translocation. The inset depicts the interprotein β-sheet that is the basis of the TonB box TonB interaction. Adapted from Figure 1 Hickman et al. 2017

In accordance with the apparent need to 'unplug' the barrel, characterisation of the plug-barrel interfaces shows that they are extensively hydrated, resembling those of transient protein complexes, and therefore suggesting that they are conducive to conformational change (Chimento et al., 2005). A pulling mechanism has been proposed whereby coupling of TonB to the transporter results in the transfer of force, and consequently the disruption of the plug. This hypothesis is based on insights from single molecule unfolding experiments which show that modest forces cause unfolding of β-sheet structures when applied perpendicularly to the direction of the strands (Brockwell et al., 2003). The core 4-stranded

sheet of the TBDT plug domain is oriented such that a force exerted by TonB would also be perpendicular, thus facilitating force-mediated disruption of the plug.

Crystal structures of BtuB, FhuA and FoxA TBDTs in complex with the C-terminal domain of TonB show that the TonB box-TonB interaction involves formation of an interprotein β-sheet via strand exchange (Pawelek et al., 2006; Shultis et al., 2006; Josts et al., 2019). Building on this insight, molecular dynamics experiments and, more recently, single molecule force spectroscopy have demonstrated that this interaction is of sufficient strength to facilitate force-mediated unfolding of the plug (Gumbart et al., 2007; Hickman et al., 2017). Moreover, they reveal a preference for partial unfolding of the N-terminal region of the plug domain over its complete removal from the barrel. Unfolding of this force labile region of the plug opens a channel within the transporter, the size of which is proposed to be tailored to the specific substrate of the transporter. In contrast, the C-terminal region of the plug domain is described as 'mechanically strong', remaining native-like during pulling experiments and simulations. However, a recent study using EPR to investigate the behaviour of BtuB in intact *E. coli* cells revealed a substrate induced conformational change in the C-terminal region of the plug, making a case for a less 'static' role of this region in proposed mechanisms (Nilaweera et al., 2021). Remarkably, these changes were not observed for BtuB reconstituted into liposomes, indicating an important role for native features of the OM e.g. LPS. Indeed, molecular dynamics simulations predict LPS-mediated stabilisation of extracellular loops of BtuB (Balusek and Gumbart, 2016).

Unproductive transport cycles i.e. expending energy to unplug the barrel in the absence of substrate would be detrimental to cell growth. Substrate binding must therefore be signalled across the outer membrane to the periplasm. Available crystal structures of various TBDTs show TonB boxes can adopt both ordered and disordered conformations (Noinaj et al., 2010). When ordered, the TonB box is tucked up against the plug and is thought to be inaccessible to TonB. Current models propose that substrate binding induces a conformational change that is propagated through the plug domain resulting in the disordering and consequent exposure of the TonB box in the periplasm. Thus, TonB box exposure is thought to act as a signal, identifying the transporter as substrate bound, and permitting interaction with TonB (see **Figure 1.10**).

Electron paramagnetic resonance spectroscopy (EPR) has shown this substrate-induced order to disorder transition for the model TBDTs BtuB and FecA (Fanucci et al., 2003; Kim et al., 2007). Notably however, site-directed spin labelling failed to identify this behaviour in the ferrichrome transporter FhuA. Instead, the TonB box was shown to be constitutively disordered, suggesting either that it is always available for interaction with TonB, or that the activity of this transporter is subject to some unknown form of regulation (Kim et al., 2007). Further to this, studies on the ferrioxamine B transporter FoxA, demonstrate that TBDT-TonB interactions are multifaceted in some cases (Josts et al., 2019). The C-terminal domain of TonB is able to interact with the FoxA transporter in the absence of substrate via a region preceding the TonB box. However, substrate binding is still required to generate the transport competent state of the complex where the TonB box is exposed and participates in the characteristic β-augmentation interaction with the C-terminal domain of TonB. However, the functional relevance of constitutive FoxA-TonB interactions in the absence of ferrioxamine B remains unclear.

Interestingly, the available crystal structures of apo and ligand-bound TBDTs provide little support for TonB box signalling, exhibiting a poor correlation between the presence of substrate and the state of the TonB box. A potential explanation for this has been provided by EPR studies which demonstrate that many crystallisation reagents, as well as the crystal lattice itself, inhibit the order-disorder transition of the TonB box associated with substrate binding (Fanucci et al., 2003; Kim et al., 2006; Freed et al., 2010). Thus, x-ray crystallography may not be the ideal tool to probe the mechanisms of these transporters.

Crucially, the details of the signal transduction and substrate translocation mechanisms are unknown. Even where apo and ligand-bound structures are available, identifying the allosteric pathways responsible for substrate-mediated signal transduction across the OM is difficult and few studies have addressed this question directly (Ferguson et al., 2007). Moreover, despite having observed the direct interaction between TonB and the TonB box, the mechanism(s) of force generation and plug disruption remain enigmatic.

Regarding force generation, live cell fluorescence studies using GFP-tagged TonB showed that this component exhibits energy-dependent motion and that this motion is coupled to the

PMF by ExbB-ExbD (Jordan et al 2013). The motion was proposed to be rotational, given the homology of ExbB-ExbD to MotA-MotB of the flagellar motor, a torque-generating proton pump (Zhai et al., 2003). Indeed, recent cryoEM structures of the ExbB-ExbD complex and of MotA-MotB flagellar motor systems from various species show that these complexes have similar architectures and are thus assumed to function using a common mechanism (Celia et al., 2019; Deme et al., 2020; Santiveri et al., 2020; Ratliff et al., 2021) (see **Figure 1.11**). The ExbB-ExbD complex has a stoichiometry of 5:2 where ExbB assembles as a pentamer with a hydrophobic central pore that is occupied by a dimer of ExbD. The location of TonB is not immediately apparent in the data, but weak density is attributed to a single TonB subunit associated at the periphery of the complex (Deme et al., 2019). The structures of the MotA-MotB motor complex led to the proposal of a rotary mechanism. A conserved aspartate residue on MotB plays a central role, acting to accept a proton from the periplasm and deliver it to the cytoplasm. Alternating protonation and deprotonation of equivalent aspartate residues in each copy of MotB is coupled to rotation of the pentameric ring of MotA. In the case of the TonB-ExbB-ExbD system, how the putative rotary motion is relayed to TBDTs at the OM to facilitate disruption of the plug is unknown. A rotational surveillance and energy transfer (ROSET) model proposes that rotation of the TM region of TonB at the IM is transmitted, via a rigid periplasmic domain, to the C-terminal domain. This rotational motion is hypothesised to move the TonB-ExbB-ExbD complex laterally through the IM, allowing the C-terminal domain to 'scan' the periplasmic surface of the OM for exposed TonB box regions. Upon engaging a transport competent complex, the kinetic energy would be transferred to the transporter, facilitating disruption of the plug (Klebba, 2016). This model remains to be tested experimentally and a great deal more work is required to fully elucidate the mechanism of force transfer and channel opening. Ultimately, structures of the complete IM Ton complex bound to a TBDT at the outer membrane will be required.

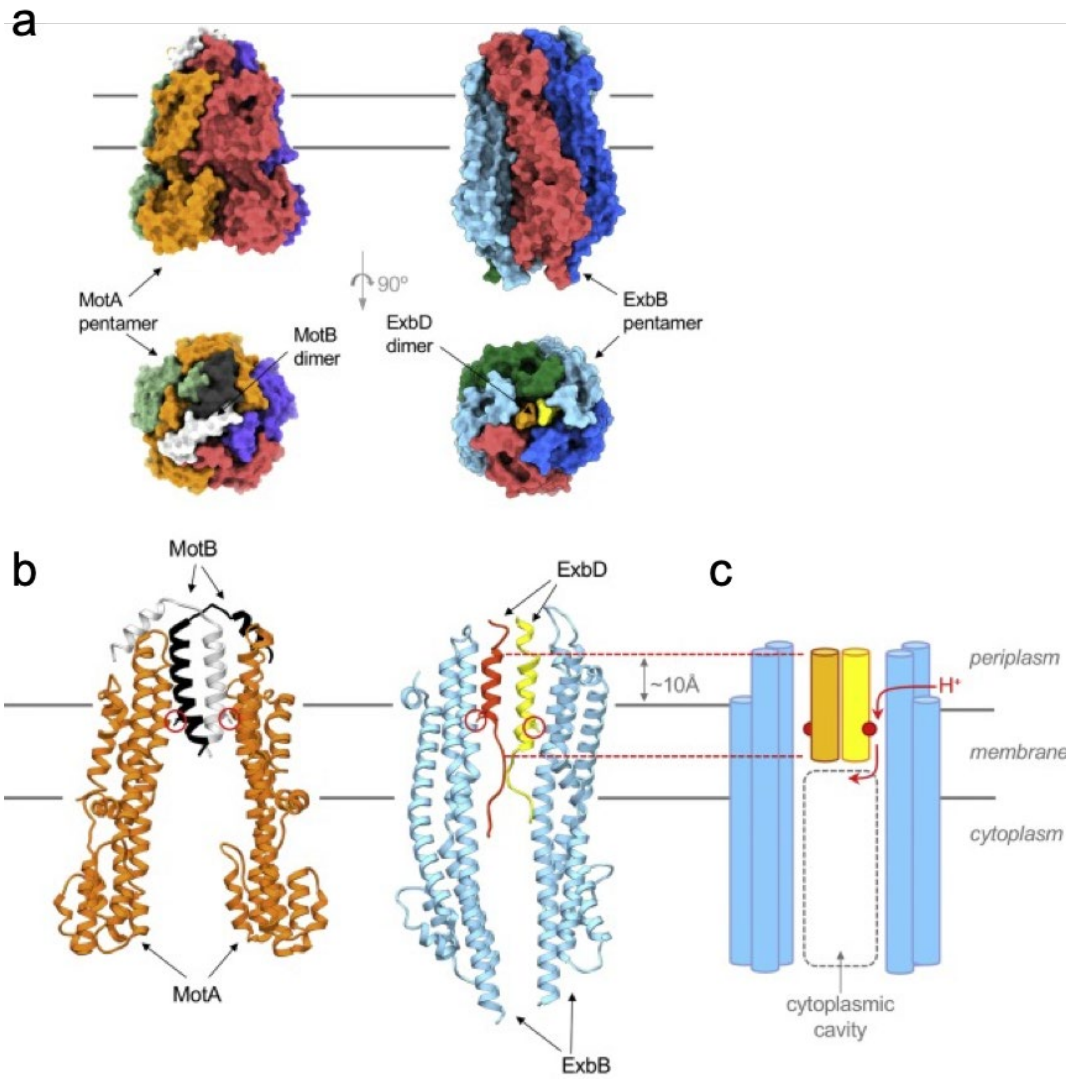


Figure 1.11: Structural homology between the Ton motor complex ExbBD and the flagellar motor complex MotAB. a, Surface representations of the MotAB (left) and ExbBD (right) motor complexes viewed in the plane of the membrane (top) and from the periplasmic space (bottom). The pentameric assemblies of MotA (orange, purple, red and green) and ExbB (cyan, blue, red and green) are shown to encompass the dimeric assemblies of MotB (black and white) and ExbD (orange and yellow) respectively. **b**, Cartoon representation of MotAB (left) and ExbBD (right) structures. For clarity only 2 subunits are shown from each pentamer. The essential aspartate residue on the MotB and ExbD transmembrane helices is circled red. The distance the substrate proton has to travel is minimised as the central hydrophobic pore of both complexes is shifted ~10 Å towards the periplasm (dashed red lines). **c**, Schematic representation of ExbBD with conserved Asp residues shown as red circles. Red arrows depict the putative proton path. A lateral opening between two ExbB subunits permits passage to the acceptor Asp on ExbD which, after large scale rotation of the complex, subsequently releases the proton to the cytoplasmic cavity. Reproduced from Figure 2 Ratliff et al., 2021.

1.8 SusCD-like TonB-dependent transporters

The requirement for TBDTs is dependent on the ecological niche occupied. The human large intestine possesses an extremely high bacterial cell density (>10^{11} cell/ml), dominated by the Gram-positive Firmicutes and the Gram-negative Bacteroidetes. The Bacteroidetes are composed largely of *Bacteroides spp.* which make up 10-15% of all bacteria in the large intestine (Arumgam et al., 2011). This environment is devoid of easily accessible nutrients such as proteins and simple sugars, since these are absorbed in the small intestine (Marchesi et al., 2016). Instead, complex dietary glycans, inaccessible to the enzymes of the human digestive tract are the primary nutrient source in this environment. Collectively, these factors suggest that nutrient uptake in *Bacteroides spp.* is likely to be fundamentally different to that of bacteria, such as *E. coli*, that inhabit the small intestine. Indeed, a comparison of the complement of membrane proteins in *E. coli* and *Bacteroides thetaiotaomicron* (*B. theta*, an important model gut symbiont) revealed far fewer porins and an abundance of TBDTs in the latter (Glenwright et al., 2017).

The Bacteroidetes employ a common strategy for glycan utilisation, where TBDTs are often expressed as part of a co-regulated gene cluster known as a polysaccharide utilisation locus (PUL) (Bolam and Koropatkin, 2012). A PUL encodes all of the machinery required for the uptake, processing and metabolism of a specific glycan, and multiple PULs endow *Bacteroides spp.* with the ability to process almost all dietary polysaccharides. *B. theta* has 88 predicted PULs, corresponding to around 20% of its total genome, of which only ~20 have been characterised to date. The remarkable size of this tool kit and the diversity of substrate utilisation it supports establishes *B. theta* as a glycan generalist and demonstrates the importance of these systems to the fitness of this organism in the complex environment of the distal gut.

The first of these systems to be characterised was the starch utilisation system (Sus). This system will be used to exemplify the characteristic features of PUL systems; a schematic overview is shown in **Figure 1.12**. Central to transport is a SusCD complex where SusC is a TBDT which, unlike 'classical' TBDTs, is dependent on the additional lipoprotein SusD for function (Cho and Salyers, 2001). An authentic PUL also minimally encodes an OM glycosidase

(SusG), responsible for processing the target glycan into transport-competent substrates at the cell surface, and an IM sensor/regulator responsible for substrate-dependent regulation of PUL expression (SusR) (van den Berg and Bolam, 2018). Often, PULs also encode other OM-associated proteins with complementary function such as surface glycan binding proteins that have a role in recruiting target glycans to the cell surface (SGBPs). The Sus system contains two SGBPs, SusE and SusF. The remaining components of the system are periplasmic glycosidases responsible for further breakdown of the imported oligosaccharides prior to import into the cytoplasm.

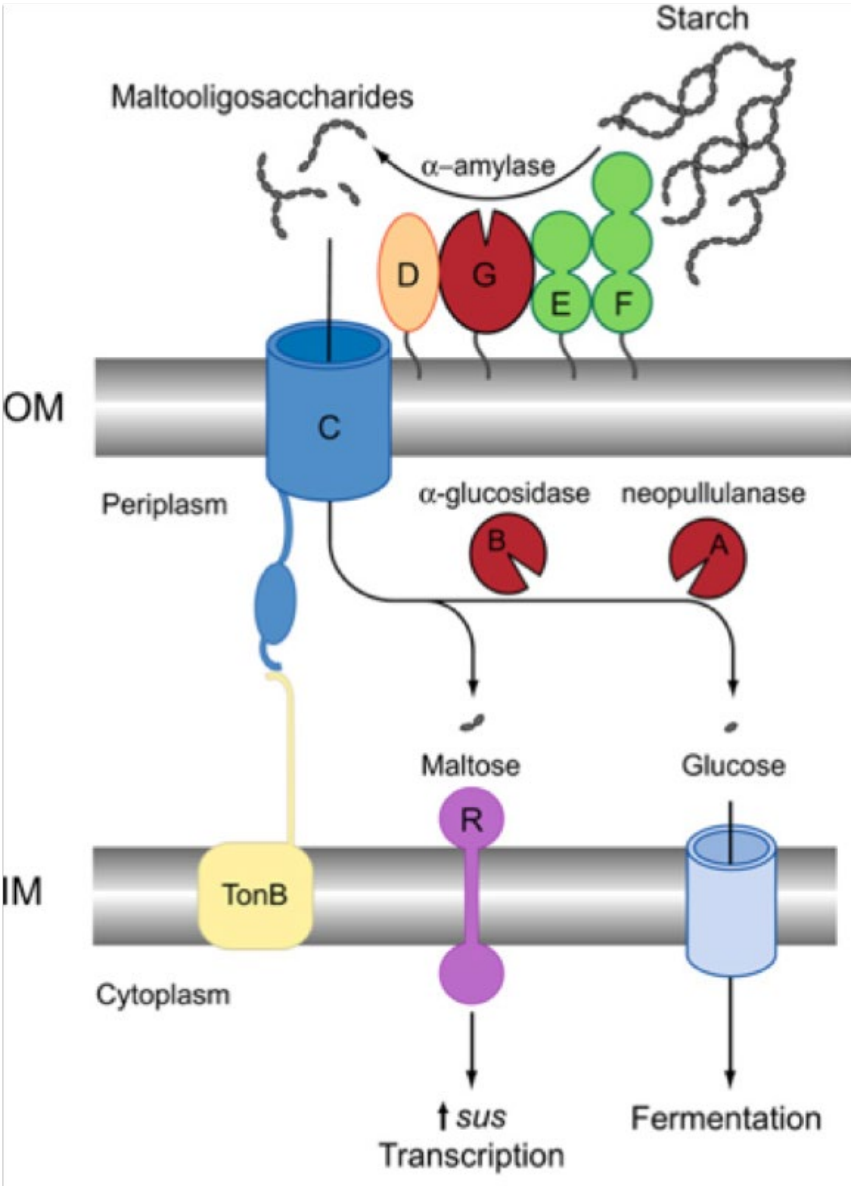


Figure 1.12: **Schematic overview of the starch utilisation system.** The sus polysaccharide utilisation locus encodes SusA-G. Together these proteins act to facilitate the uptake and processing of starch. The lipoproteins SusE and SusF bind starch to the surface of the cell. SusG is an α-amylase responsible for hydrolysing starch to yield oligosaccharides. These are bound by SusD and are small enough to be translocated through the TBDT SusC. Post-translocation, SusA and SusB process the oligosaccharides into glucose. The complex of SusC and SusD is essential for transport and is the central feature of all PUL systems. The sensor regulator SusR is also encoded by the sus PUL and upregulates transcription of the PUL in response to maltose in the periplasm. Adapted from Figure 1 Foley, Cockburn and Koropatkin. 2016.

PULs targeting a diverse range of substrates have been characterised including xyloglucan (Larsbrink et al., 2014), xylans (Rogowski et al., 2015), α-mannan (Cuskin et al., 2015), inulin, levan (Sonnenburg et al., 2010) and rhamnoglacturonan-II (Ndeh et al., 2017). Despite this, relatively little is known about the mechanism of nutrient uptake by SusCD-like transport

systems and this represents an important gap in our knowledge. These transporters are the primary mechanism of nutrient acquisition for the predominant flora of the large intestine. The ability of *Bacteroides spp.* to utilise complex sugars is essential for their survival and, crucially, is integral to mutualism between host and bacteria, via the generation of short-chain fatty acids that are accessible to the host. These metabolites contribute substantially to our daily calorific intake (McNeil, 1984) and their availability is associated with normal gastrointestinal physiology and systemic health benefits (Koh et al., 2016; Morrison and Preston, 2016).

1.8.1 Mechanistic insight into SusCD systems from structural studies

Our mechanistic understanding of these nutrient import systems was recently bolstered by the first X-ray crystal structures of SusCD transporters: The SusCD transporter of the levan utilisation locus (Bt1762-Bt1763), and a SusCD complex from an uncharacterised PUL (Bt2263-Bt2264) (Glenwright et al., 2017) (see **Figure 1.13**). Strikingly, these SusCD complexes were shown to be dimeric, existing as $SusC_2D_2$ tetramers. This is a unique feature amongst TBDTs although the functional relevance remains unknown. The SusC-like components possess the archetypal TBDT structure described above, but for the two complexes studied, these transporters were observed to be tightly capped at the extracellular face by their corresponding SusD-like lipoproteins.

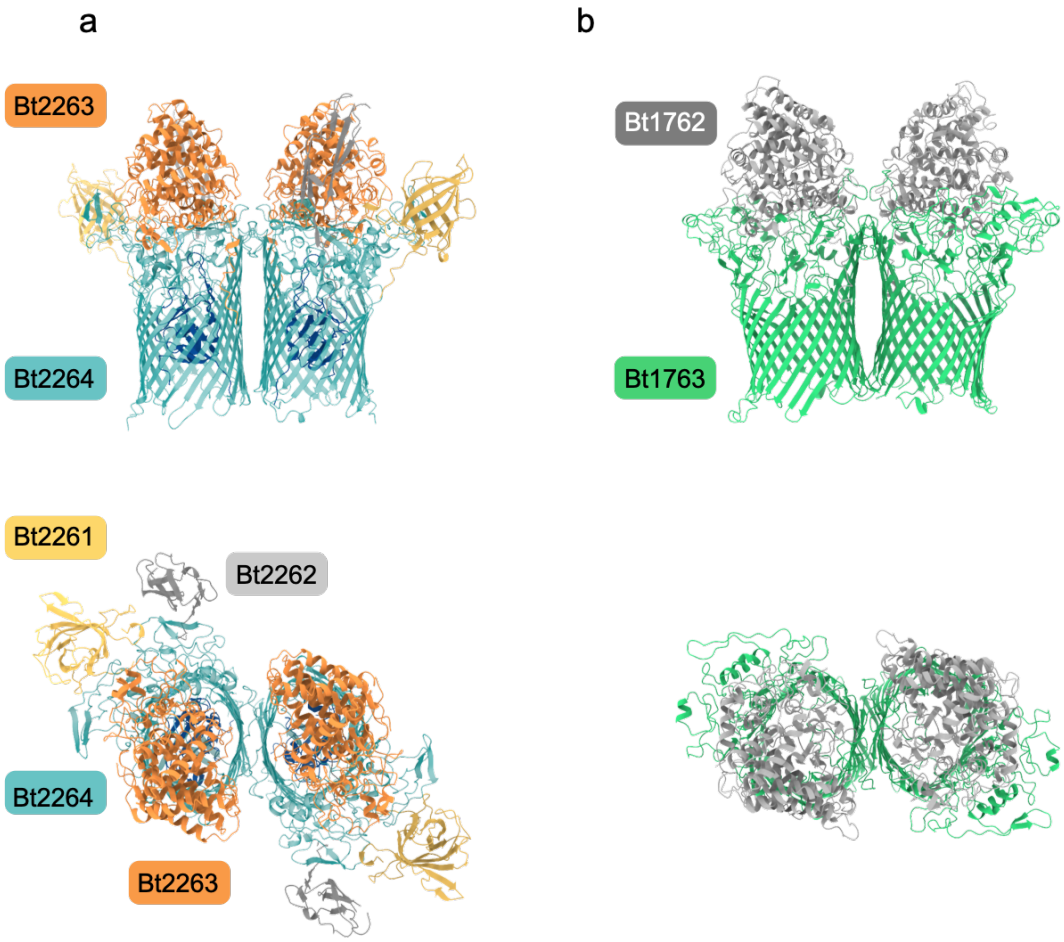


Figure 1.13: **Overview of the first structures of SusCD-like systems solved by Glenwright and colleagues**. The uncharacterised Bt2261-2264 system (**a**) and the Levan transporting Bt1762-1763 system (**b**) as viewed from the plane of the membrane (top) and from the extracellular side (bottom). The dimeric assembly of the core SusCD-like units can be seen for both systems and the SusC-like barrels are clearly capped by their SusD-like counterparts. Two auxiliary lipoproteins co-purify with the BT2263-BT2264 core complex and are easily observable in the lower left panel.

Intriguingly, the SusCD complex from the uncharacterised PUL was found to co-purify with its ligand which does not appear to be a glycan, but was instead tentatively assigned as a short peptide or peptide ensemble; a novel class of substrate for SusCD-like systems. The ligand occupies a large solvent excluded cavity between the SusC and SusD components, both of which contribute to substrate binding. The initial binding event therefore requires a conformational change in the complex that exposes the binding site to the extracellular environment.

Molecular dynamics simulations in the absence of a modelled peptide substrate predicted that SusD could undergo a rigid-body hinge-like motion; moving away from SusC to expose the binding site within the lumen of the SusC barrel. This same conformational change was not observed when substrate was modelled in, with the SusD 'lid' remaining closed for the duration of the simulation (Glenwright et al. 2017). Further to this, the crystal structure of the levan-transporting SusCD complex revealed that the plug domain of SusC was missing due to proteolytic cleavage. These plug-less transporters were used in patch-clamp channel conductance experiments that also indicated dynamic opening and closing of the SusD lid. These observations led to the proposal of a 'pedal-bin' mechanism for substrate capture and translocation by SusCD-like systems (**see Figure 1.14**). In the apo state, the SusD lid of the transporter is dynamic and explores a range of conformations, pivoting around the hinge region. Substrate binding, via interaction with SusD and SusC, stabilises the closed state, thereby locking the lid shut. Finally, energy-dependent TonB-mediated disruption of the plug is proposed to facilitate substrate translocation and, consequently, restore the dynamic state of SusD. This final step has been likened to 'pressing the pedal' to 'open the SusCD bin'.

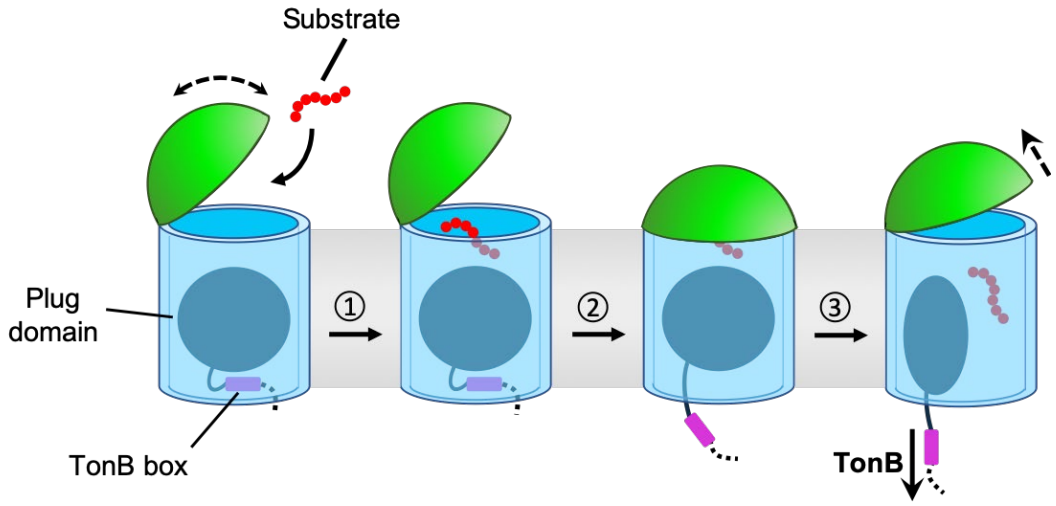


Figure 1.14: A schematic model of substrate acquisition by SusCD-like transporters based on crystal structures and MD simulations. In the absence of substrate, SusD (green) samples a range of conformations permitted by the hinge region. 1. The lid open state of the transporter permits peptide binding. 2. Contributions from both the SusC and SusD components to substrate binding induces closing of the lid, forming the transport state of the complex. This is signalled across the OM by perturbation of the TonB box region on the periplasmic side of the plug domain, making it accessible to TonB. 3. According to the literature consensus, TonB-mediated disruption of the plug permits substrate translocation and restores the SusD lid to a dynamic state.

Notably, the SusCD complex from the uncharacterised PUL co-purifies with two additional lipoproteins of unknown function (Bt2261 and Bt2262) such that the whole complex is octameric. The levan PUL also encodes two additional surface exposed lipoproteins: an endo-acting glycosidase and a levan-binding SGBP. These lipoproteins have complementary functions and the glycosidase is required for growth on levan (Sonnenberg et al., 2010), yet they did not co-purify with the core transporter, suggesting that they perhaps associate transiently with the transporter components. Indeed, recent studies on the starch utilisation system argue in favour of transient, substrate-induced complex formation and describe a role for the SGBPs SusE and SusF as immobile starch binding centres around which the SusCD transporter and SusG glycosidase can assemble (Tuson et al.,2018).

1.8.2 PUL-based strategies are successfully employed throughout the Bacteroidetes

To date, research efforts have primarily focused on characterising PULs from Bacteroides spp. occupying the human distal gut. As expected, the majority of these PULs contribute to the impressive capacity of Bacteroides spp. to degrade complex dietary and host glycans that are

inaccessible to the enzymes of the human digestive tract. However, evidence suggests that other nutrient sources, namely peptides, are also targeted by PULs (Glenwright et al., 2017). Outside of the gut, this theme is continued with Bacteroidetes from marine environments employing suites of PULs to target complex algal polysaccharides, and thus playing an important role in carbon cycling, especially in deep sea sediments (Arnosti et al., 2021; Zheng et al., 2021). Moreover, *Porphyromonas gingivalis,* a pathogenic member of the Bacteroidetes that constitutes part of the human oral microbiome possesses a SusCD-like system which has been implicated in peptide uptake by this asaccharolytic bacterium (Nagano et al., 2007). Finally, PUL-like cassettes including SusCD-like components and phosphatases have been identified in Bacteroidetes species occupying rhizosphere environments where they are implicated in the uptake of organic phosphates (Lidbury et al., 2020). Thus, the PUL-based strategy for nutrient acquisition is associated with a diverse range of complex substrates.

Given that Bacteroidetes occupying different environmental niches have adopted this common strategy for the acquisition of complex nutrients, it is important to consider how the constituent components are adapted to their biological role. Firstly, considering the transporter itself, the requirement of a SusD 'lid' distinguishes SusC-like TBDTs from their classical counterparts. As described above, structural data indicate that the lid component acts synergistically with the transporter through cooperative binding of the substrate, and in doing so, sequesters the substrate away from the extracellular environment (Glenwright et al., 2017). In environments where bacterial cell density is high, such substrate encapsulation may offer a competitive advantage by reducing non-productive binding events i.e. substrate binding and release in the absence of translocation. It is interesting to note the similarity between the proposed lid-closing event in SusCD-like systems and the behaviour of extracellular loops of some classical TBDTs such as FecA, which fold in over the top of the transporter in response to substrate binding (Ferguson et al., 2002). Both mechanisms likely act to improve efficiency of transport by favouring substrate translocation into the cell.

Addressing the other OM-localised PUL components, SGBPs are hypothesised to be the first to encounter substrate. A wealth of structural information is available for SGBPs from various glycan targeting PULs. Invariantly, these lipoproteins contain N-terminal Ig-like domains that

act to project carbohydrate binding modules away from the cell surface and towards the extracellular space (Tamura and Brumer, 2021). This likely facilitates interaction with the target glycan by positioning substrate binding sites above the dense LPS/LOS 'forests' that constitute the outer leaflet of the OM. Notably, the number of these Ig-like domains varies between SGBPs. For example, the starch binding proteins SusE and SusF possesses just a single N-terminal Ig-like domain with two and three C-terminal amylose binding domains, respectively (Cameron et al., 2012), while the SGBP from the heparin/heparan sulfate-targeting PUL comprises five Ig-like domains with a C-terminal carbohydrate binding domain (Cartmell et al., 2017). The presence of several substrate binding sites as seen in SusE and SusF appears, thus far, to be a unique feature of the starch utilisation system, with the remainder of characterised SGBPs possessing a single distal substrate binding site (Tamura and Brumer, 2021). Regardless of the aforementioned diversity, the apparent fundamental function of these lipoproteins is to increase the local concentration of the target molecule at the cell surface so as to promote its efficient breakdown by PUL-encoded, cell surface-exposed glycosidase components.

Given the close functional relationship of these OM-localised PUL components, further spatial co-localisation e.g. by complexation, would seem biologically advantageous. Indeed, although functionally uncharacterised, the Bt2262-Bt2264 system of *B. theta* comprises a complex of a SusCD transporter, onto which the additional lipoprotein components of the PUL are assembled (Glenwright et al., 2017). In contrast, the nature of the interactions between OM-localised components of glycan targeting systems is unclear, nonetheless the functional significance of putative higher order complexes in these systems is self-evident. SGBPs would act to recruit large glycan chains to the complex, increasing the local concentration of these substrates around the glycosidase components and thus promoting efficient breakdown into transport-competent nutrients. These breakdown products would in turn be released proximal to the transporter components, promoting efficient loading of the transporter and consequently, efficient translocation of the substrate into the cell.

1.8.2 New tools for structural biology of SusCD systems

Detailed structural information on both SusCD systems and classical TBDTs has so far come solely from X-ray crystal structures. Whilst these have been invaluable, there are important limitations to this technique. As discussed above, osmolytes and the crystal lattice itself can bias the conformation of the TonB box, hindering interpretation of the structure-function relationship (Freed et al., 2010). Furthermore, direct comparisons of apo and ligand bound structures are not straightforward given that their respective crystallisation conditions and crystal packing are different. The crystal structure of the levan SusCD revealed a lid closed conformation despite the apo state of the transporter (Glenwright et al., 2017). This observation is in conflict with the proposed 'pedal bin' hypothesis and has been attributed to the preference of crystal packing for less hydrated and more compact conformations i.e. 'lids closed'. If true, providing structural evidence in support for lid opening requires a complementary structural technique.

Whilst 'classical' TBDTs such as BtuB are small (<90 kDa), the transporter components of SusCD like systems are ~40% larger and their novel dimeric structure brings the molecular weight of the core complex (SusCD-like dimer) to ~320 kDa. This makes them readily accessible to cryo-electron microscopy (cryoEM), which allows sampling of the conformational heterogeneity present in solution (Murata and Wolf, 2018). Studying SusCD systems by cryoEM may therefore reveal novel conformational changes that would be important for understanding the mechanism of substrate capture. More generally, these systems provide a valuable opportunity to explore the mechanistic basis of TonB-dependent transport in the absence of a crystal lattice, overcoming many of the aforementioned limitations associated with crystallography.

1.9 Project aims

This project will study SusCD-like transport systems using cryoEM to answer some of the major open questions in the field. Our fundamental understanding of substrate capture is based largely on MD simulations. A principle aim is therefore to provide structural evidence in support of these, and capture the opening and closing events in solution at high resolution. The functional relevance of the dimeric assembly of these transporters is also unknown. MD simulations predict that the two symmetrically-related units of these transport systems are

able to operate independently (at least in terms of lid opening), whether this observation can be recapitulated experimentally remains to be seen. The observation that Bt2263-Bt2264 purifies as part of an octameric complex whilst the levan SusCD does not is intriguing. It is possible that, in the case of the levan system, the complex is less stable, or that association of the additional lipoproteins is a transient substrate-induced event. Regardless, structural information on the arrangement of these additional lipoproteins in a functionally characterised system is required to understand how they might operate in a concerted way to process, capture and transport substrate across the OM.

Chapter 2 Materials and Methods

2.1 Background and Instrumentation

Transmission electron microscopy (TEM) and, more specifically, cryo-electron microscopy (cryoEM) are the principle methodologies used to generate the results presented in this thesis. This chapter will provide an overview of image formation in the electron microscope before considering the general work-flows associated with negative stain TEM and single particle cryoEM of biological samples. Finally, the specific methods associated with each results chapter will be presented.

2.1.1 Anatomy of the TEM

All electron microscopes possess a fundamental arrangement of components that are housed in a column held at high vacuum (see **Figure 2.1**). An electron source (or gun) emits electrons that are accelerated across an electric potential and focused by a series of electromagnetic lenses. Accelerating voltages range from 100-300 kV with the upper end of this range yielding electrons with a wavelength ~0.02 Å (Milne et al., 2014). A specimen holder is positioned in the path of the electron beam such that electrons are scattered by the sample before arriving at a detector or fluorescent screen located at the bottom of the column. Scattering interactions can be elastic, in which case the energy of the electron is unchanged but its trajectory is altered, or inelastic, where the electron transfers energy to the sample. Only the former contributes usefully to image formation, with the latter resulting in noise. Inelastic scattering also causes radiation damage to the sample, in the form of bond rearrangement, free radical formation and secondary scattering events, all of which perturb the native structure of the specimen (Orlova and Saibil, 2011). Unfortunately, the ratio of elastic to inelastic scattering events is ~1:4 and thus these detrimental effects impose limitations on imaging of biological samples, forcing the adoption of low electron doses and consequently reducing the signal to noise ratio (SNR) in the images (Henderson, 1995).

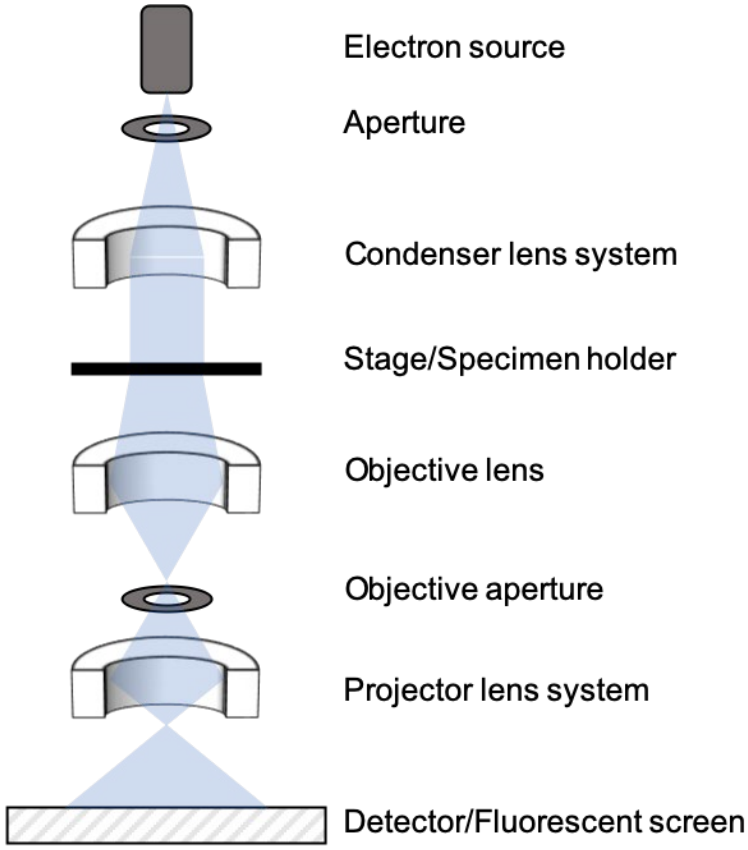


Figure 2.1 Schematic of a transmission electron microscope. The electron beam is represented by the shaded blue area. Figure adapted from Orlova and Saibil (2011).

Considering the main components of the TEM from top to bottom in more detail, electron sources were traditionally thermionic, such as filaments of tungsten or lanthanum hexaboride. However, for high-resolution structural biology applications, and invariably in cryo-electron microscopes, thermionic sources have been superseded by field emission guns (FEGs) that boast higher coherence of emitted electrons (Thompson et al., 2016). A condenser lens system ensures parallel illumination of the specimen which is critical for high-resolution cryoEM. The objective lens provides the primary magnification and the specimen itself sits within it. An objective aperture, positioned at the back focal plane of the objective lens, serves to block electrons scattered at high angles by the sample, resulting in improved contrast (Orlova and Saibil, 2011). Projector lenses magnify the image further and finally the electrons arrive at the detector.

Samples for imaging by TEM are loaded onto grids that are ~3 mm in diameter, and comprise a metal mesh with a regular spacing (most commonly made of copper or gold) and a support

layer of amorphous carbon or gold that sits on top. Support layers contain arrays of punched-out holes with well-defined hole size and spacing. For carbon support films, a random lacy pattern of holes is also available. In addition to the basic structure described above, a further film/coating such as an ultrathin (~2 nm) layer of continuous carbon or graphene oxide is sometimes applied on top of the support layer. The choice of support material, additional layer, and hole size and pattern can all influence the way which the sample behaves on the grid (Drulyte et al., 2018; Naydenova et al., 2021).

2.1.2 Detecting electrons

The method of electron detection has an enormous influence on the quality of data obtainable. A perfect detector would add no noise to the image but in practice all detection methods do, and this noise can be expressed as the detective quantum efficiency (DQE) which equates to the square of the output SNR over the square of the input SNR. The perfect detector would have a DQE of 1 across all spatial frequencies. Early work utilised photographic film where image formation is dependent on the interactions of scattered electrons with silver halide, causing a chemical reaction and altering the surface of the film at the site of the event. The small grain size of silver halide crystals allowed fine sampling over a large field of view and the DQE of this detection medium at high spatial frequencies was better than that of typical charge-coupled device (CCD) cameras that were available at the time (Thompson et al., 2016; McMullan et al., 2009). As a result, most of the highest resolution single-particle cryoEM maps determined prior to 2012 (~3-4 Å) were generated from images collected on film (Grigorieff and Harrison, 2011; Zhang et al., 2010). More recently, the advent of direct electron detectors (DEDs) has revolutionised biological TEM (McMullan et al., 2016). These detectors utilise monolithic active pixel sensors (MAPS) which directly translate incident electrons into an electric signal, as opposed to CCD cameras which use a scintillator layer to convert electrons to photons for detection (indirect). As a result, they possess an improved DQE and, importantly, they offer the advantage of a high frame rate rolling readout mechanism i.e. each frame is read out continuously to generate a ‘movie’ of the exposure. The availability of these frames is invaluable as it allows movement in the image, caused by stage drift or beam-induced specimen motion, to be tracked and corrected computationally (discussed further in **section 2.4.1**). DEDs can be operated in one of two modes: counting

mode or integrating mode. In counting mode each electron event is recorded independently whereas in integrating mode the signal is measured over a fixed period of time and, as a result, multiple incident electrons contribute to the recorded signal. Historically, integrating mode offered the advantages of higher dose rates and shorter exposure times but at the cost of a reduction in SNR. In contrast, operating in counting mode allows for more accurate localisation of electron events. Indeed, sub-pixel localisation is also possible with some detectors meaning that recorded images contain information beyond the physical Nyquist limit ('super-resolution') that is imposed by the pixel size of the detector (Feathers et al., 2021). The downside of counting mode is that extremely low dose rates are required to avoid saturation of the detector and consequently exposure times are typically much longer, slowing the rate of data acquisition. However, while this was a necessary consideration on detectors such as the Falcon III (Thermo Fisher Scientific), next generation detectors such as the Falcon IV (Thermo Fisher Scientific) and K3 (GATAN) offer increased internal frame rates that have massively reduced the required exposure times (up to 10-fold reduction for the Falcon IV). Combined with advances in automated data acquisition approaches such as aberration-free image shift (AFIS), collecting data in counting mode is now fast enough to be standard practice.

2.1.3 Generating contrast in TEM

A major concern regarding the imaging of biological samples by TEM is contrast. Contrast in TEM images has two sources, amplitude contrast and phase contrast. Amplitude contrast is fundamental to negative stain TEM. Here, the protein sample is coated in a layer of a heavy metal salt (often uranium based) which forms a shell around the protein. The electron dense uranium atoms within this stain layer give rise to high angle scattering of electrons which are consequently lost from the image. In comparison, proteins are comprised of light atoms, mostly H, C, N and O, thus where protein is present in the stain layer, scattering is reduced and more incident electrons are detected, resulting in an intensity difference between regions of stain and regions of protein (De Carlo and Harris, 2011; Orlova and Saibil, 2011). In cryoEM, the absence of a staining material means that amplitude contrast is minimal, with the protein sample possessing a similar electron density to the surrounding solvent. This problem is exacerbated by the requirement for low-dose imaging and, under these

conditions, we depend on the phase contrast that results from interference between electrons that pass straight through the sample, and those that are scattered by the sample. Unfortunately, thin electron transparent samples like those suitable for cryoEM scatter electrons through small angles, generating only small phase shifts which means contrast in the images is still poor (Orlova and Saibil, 2011).

In practice, the amount of phase shift and in turn the amount of contrast in the image can be increased by altering the optics of the microscope to introduce small amounts of defocus. Importantly, together with spherical aberration that is inherent to the microscope lens systems, the defocus value affects the contrast-transfer function (CTF); a mathematical description of the imaging process in TEM expressed in Fourier space (Erickson and Klug, 1970; Cheng, 2015). A perfect TEM image would represent a true projection of the Coulomb potential of the sample, however imperfect microscope optics mean that the real image is convoluted with a point-spread function or the CTF (the point spread function in Fourier space). The CTF is an oscillating function with decreasing amplitude and increasing frequency which measures contrast as a function of spatial frequency, and its effects are visualised when looking at the Fourier transform of an electron micrograph (see **Figure 2.2**). The amplitudes decay with an envelope function at high spatial frequencies, attenuating their signal in the image. Other effects of the CTF include spatial frequencies at which no information is recovered, as well as frequencies where contrast is reversed. The greater the defocus value the higher the frequency of oscillation and the more severe the envelope function. High defocus values also result in increased signal delocalisation. Collectively, the effects of the CTF mean that generating contrast by the introduction of defocus requires several considerations prior to data collection. 1) The minimum amount of defocus at which the particles of interest are visible should be used such that high-resolution information is preserved. 2) Data should be collected at a range of defocus values around this optimum to compensate for the spatial frequencies at which there is no contrast for a particular defocus value 3) When extracting particle images from micrographs (see **section 2.4.2**) the box size should be large enough to account for the significant delocalisation of high-resolution signal (Cheng et al., 2015; Bhella, 2019).

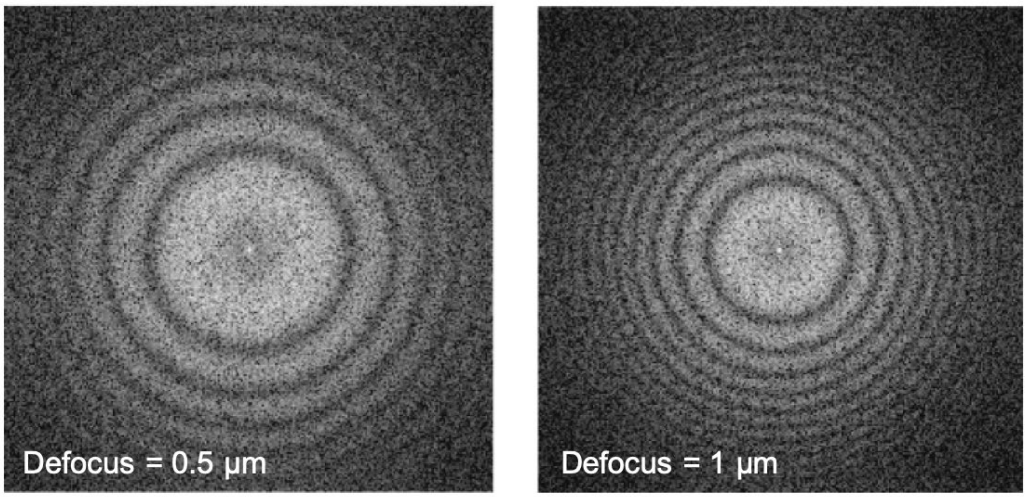


Figure 2.2: **Fourier transforms/power spectra of TEM images of a carbon support film acquired with different defocus values.** Left was taken at 0.5 µm defocus and right was taken at 1 µm defocus. The oscillating CTF function is visible as alternating rings of positive and negative contrast. These Thon rings are closer to the origin at higher defocus values and oscillate more rapidly. Figure adapted from Orlova and Saibil (2011).

2.2 Negative stain TEM

Before embarking on a cryoEM project, the sample of interest must first be characterised biochemically. For a protein sample this would minimally include demonstration of purity by size exclusion chromatography (SEC) and/or SDS-PAGE analysis. However, these techniques provide little insight into any structural heterogeneity of the sample, be it compositional or conformational. Negative stain TEM is a fast and powerful approach for assessment of structural heterogeneity and requires only small amounts of protein (picomoles), making it suitable for screening sample behaviour in different buffers (Ohi et al., 2004; Scarff et al., 2018). Additional complementary techniques such as SEC-MALS and mass photometry can also be advantageous for assessing compositional heterogeneity of protein complexes (Some et al., 2019; Sonn-Segev et al., 2020).

Grids used for negative stain TEM, typically possess a continuous carbon support. The surface of the grid is rendered hydrophilic by glow-discharge, during which ionised gas molecules are deposited onto the surface of the grid, imparting a small negative charge and consequently improving the wettability of the grid to ensure that the sample solution spreads evenly across the surface (Dubochet et al., 1971). The sample is applied to the grid and given time to adsorb to the surface, after which excess solution is blotted away. Protocols for staining vary but

often include one or more wash steps, each followed by blotting of excess liquid, before one or more staining steps which ultimately culminate in the drying down of a thin stain layer that envelops the protein and provides high contrast TEM images as described above (Scarff et al., 2018).

The small grain size of uranyl salts (4-5 Å) means that it is possible to resolve structural features on the level of protein subunits and domains e.g. relative domain orientations. Thus, negative stain data can provide important biological insights in its own right, such as providing information on the inherent flexibility of a molecule or complex (Burgess et al., 2004). More commonly however, negative stain images are used to demonstrate particle homogeneity and monodispersity on the grid as a prerequisite to cryoEM studies. An important caveat is that good behaviour of a sample observed with negative stain does not always translate to a well-behaved sample in cryoEM; samples that look excellent in stain screening can aggregate under the conditions required for cryoEM and, conversely, samples that give poor results in stain may show improved behaviour in cryoEM.

2.3 CryoEM

Cryo electron microscopy is so named because it involves imaging the sample of interest in a thin film of vitreous ice at temperatures around -180°C. Several structural biology techniques utilise cryoEM including tomography and sub-tomogram averaging as well as single particle analysis (SPA) and helical reconstruction. Tomography is capable of providing nanometre scale *in vivo* structural information on biological samples from cells, tissues and organisms that are thinned to an appropriate thickness. Sub-tomogram averaging builds on tomography, taking advantage of repeating structures within a tomogram by averaging them together to improve resolution. In contrast, SPA and helical reconstruction are capable of providing structural information at near atomic resolution for purified proteins and macromolecular complexes ranging in size from tens of kilodaltons to several megadaltons (Wu and Lander, 2020; Sauer et al., 2021), as well as virus structures that reach hundreds of megadaltons in size (Dai and Zhou, 2018). As the structural data presented in this thesis were generated through SPA, the sample preparation and data processing pipelines of this technique will be outlined here.

2.3.1 Sample preparation for single particle cryoEM

Sample preparation for cryoEM refers to taking an aqueous solution of the protein or complex of interest, applying it to an EM grid and generating a thin film (10-80 nm) that can be frozen quickly enough that crystalline ice does not have time to form (Passmore and Russo, 2016). The resulting vitreous ice is electron transparent and protects the sample from the high vacuum environment of the microscope column. Moreover, imaging at cryo-temperatures imparts some protection against radiation damage (Bammes et al., 2009). Achieving an ice layer of appropriate thickness is crucial for high-resolution structural studies; ice that is too thick leads to reduced contrast and increased noise, while ice that is too thin may exclude protein particles completely (Drulyte et al., 2018). As for negative stain TEM, grids for cryoEM must first undergo a glow-discharge treatment to render the surface hydrophilic. Often glow-discharge treatment is performed in air, but many devices allow bleeding in of a defined gas mixture that is more reproducible such as argon and oxygen (Passmore and Russo, 2016).

From here, grid preparation is commonly facilitated by plunge-freezing devices such as the Leica EM GP or the Thermo Fisher Scientific Vitrobot. These devices possess a temperature- and humidity-controlled chamber that will hold the grid at a desired set of conditions prior to, and immediately after sample application. The grid is first picked up with tweezers that are then attached to the plunger of the plunge freezing device. Next, the tweezes are retracted such that the grid is held in the environmentally controlled chamber. Small ports allow entry of a pipette into the chamber and typically ~3 µL of sample is dispensed onto the grid surface. Excess liquid is automatically blotted with filter paper, leaving behind a thin film of the sample in aqueous solution, and the plunger then rapidly descends, immersing the grid in a cryogen, usually liquid nitrogen-cooled liquid ethane or an ethane-propane mixture (Dubochet et al., 1988; Dobro et al., 2010).

Ideally, inspection of a grid prepared as described above by cryoEM would reveal a uniformly thin film of vitreous ice across the grid squares, in which the particles are evenly distributed, intact and visible within the holes of the support film, thus providing a snapshot of the sample randomly oriented in solution. However, during the process of grid preparation, the thin film

environment exposes the sample to a range of conditions that are markedly different from those encountered either *in vivo* or in more standard *in vitro* environments such as a test tube. These conditions can have deleterious effects on proteins and complexes, rendering them unsuitable for structural studies due to dissociation or partial denaturation. Mitigating these effects is non-trivial and, since it is difficult to predict how a particular protein will behave *a priori*, this stage in the pipeline represents a significant hurdle for many projects (Passmore and Russo, 2016). However, novel approaches to grid preparation that limit detrimental exposure to the air-water interface constitute an exciting and active area of research (Razinkov et al., 2016; Feng et al., 2017; Nobel et al., 2018; Klebl et al., 2020).

Even for samples that remain intact, their behaviour on the grid often requires optimisation. Particles can adopt a preferred orientation, they may adhere strongly to the carbon support and therefore not localise to the holes, or they may no longer be monodispersed as a result of the preparation conditions. It is often possible to improve particle behaviour by exploring grid types (Drulyte et al., 2018). For example, choosing a grid with a continuous ultrathin carbon or graphene oxide support can modify particle behaviour as they provide an interaction surface to which particles can adsorb. In this way continuous supports can improve localisation of particles into holes and, if the affinity for of the particle for the support is stronger than the affinity of the particles for each other, self-association of particles can be alleviated. It should however be noted that these supports come at the cost of reduced SNR in the images and therefore may not be suitable where particles are small with low signal, or where buffer components already attenuate contrast in the image (e.g. glycerol or sucrose).

2.3.2 Data collection for single particle cryoEM

In order to obtain a dataset for single particle analysis, hundreds to thousands of micrograph movies must be collected on a DED. Fortunately, this is carried out in a semi-automated fashion. Low magnification images of the entire grid are acquired and automatically montaged to provide an overview of potential areas for data acquisition. Desired areas are selected by the user on the basis of ice quality and particle distribution (determined through screening of representative areas) (Thompson et al., 2019). For grids that possess a regular array of holes, a template can be defined such that holes within a grid square are

automatically identified. The user is then able to use a combination of automatic filters and manual selection to leave an ideal set of hole locations to be imaged. The electron dose rate and desired defocus range are then set appropriately and data collection proceeds automatically, periodically checking and correcting the applied defocus while monitoring drift to ensure that it does not exceed a pre-defined threshold (Thompson et al., 2019). As a result, data collections can continue autonomously until all specified acquisition areas have been visited, often over a period of several days.

2.4 Single particle image processing workflow

2.4.1 Image pre-processing – Motion correction and CTF estimation

The general steps required to go from raw micrograph movies to a 3D reconstruction using SPA are shown in **Figure 2.3** and each step will be described briefly below. Before single particle analysis can begin, micrograph movies must undergo several pre-processing steps which serve to mitigate problems inherent to cryoEM imaging. As described above, data acquisition with DEDs generates micrograph movies comprising multiple frames from a single exposure. Using software such as MotionCor2 (Zheng et al., 2017), these frames are aligned and averaged to correct for micrograph-wide motions that occur as a result of both mechanical stage drift and beam-induced specimen motion during the exposure. A dose-weighting scheme is also applied to the frames such that high-resolution information in early and late frames is down-weighted to mitigate the effects of sample movement (most severe immediately after exposure) and radiation damage respectively.

Once movie frames have been aligned and averaged, the defocus and CTF parameters of the resulting micrographs must be estimated such that they can be used for CTF-correction later in the processing pipeline. Programs such as CTFFIND4 (Rohou and Grigorieff, 2015) and Gctf (Zhang, 2016) estimate the defocus and astigmatism parameters of the image by fitting a model of the theoretical CTF to the power spectrum of the micrograph.

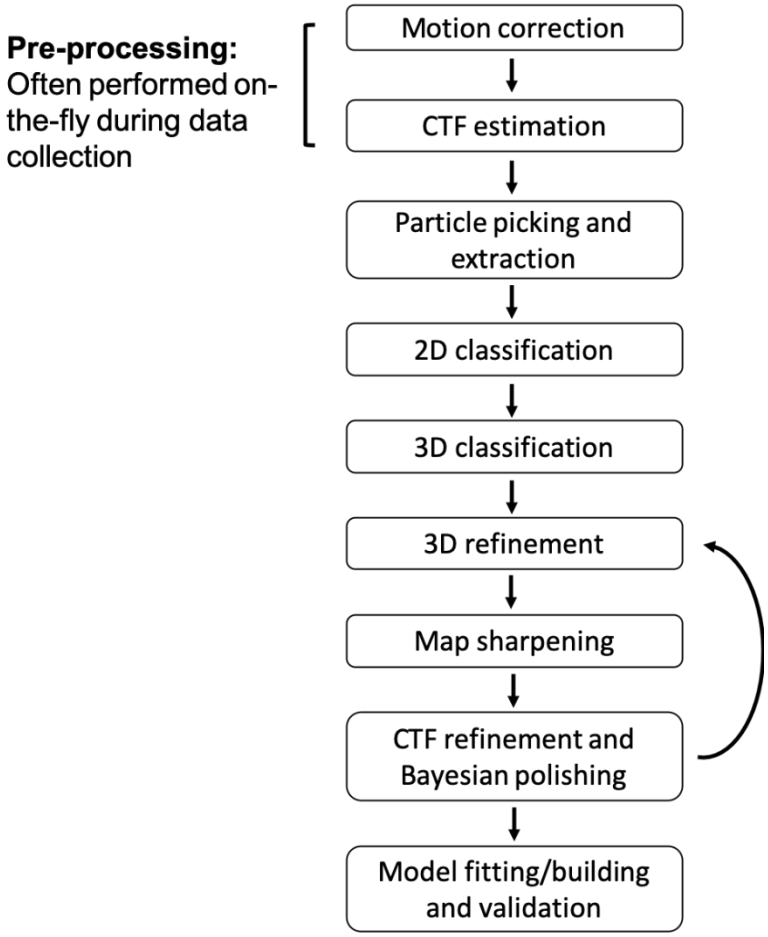


Figure 2.3: CryoEM processing workflow. Flow-chart depicting the typical steps taken for a single particle cryoEM project.

2.4.2 Particle picking

Next, the particle locations within each micrograph must be identified. Historically, particle picking was done by hand. However, the process is now performed at least semi-automatically in most cases. Some of the most popular approaches are implemented in the RELION software suite, including template-based picking and Laplacian-of-Gaussian picking (Zivanov et al., 2018). In template-based picking, the user manually selects ~2000 particles from a small subset of the data which are then classified and averaged. These averages, representing different particle views, are then used as references or templates in an automated picking procedure which looks to identify regions of density in the micrographs that match the references. Alternatively, picking with a Laplacian-of-Gaussian filter avoids the manual picking step completely, instead it works to pick all 'blobs' of density that fall within a size range specified by the user. More recently, intelligent neural network-based pickers

such as crYOLO and TOPAZ have been widely adopted by the cryoEM community (Wagner et al., 2019; Bepler et al., 2019). These algorithms offer improved picking accuracy and speeds while still requiring minimal user input. The models themselves have been trained on hundreds of cryoEM datasets containing a diverse range of samples and as a result are widely applicable without the need for additional training. If performance on a particular dataset is poor, perhaps owing to exotic particle shapes, the network model can be fine-tuned by the user with minimal effort.

Once particle coordinates have been defined by one of the aforementioned methods, the particles must then be extracted from the micrographs. Particles are extracted with a bounding box and collated to give a particle stack for further processing. Each particle image has a low SNR and for this reason SPA depends on the alignment and averaging of thousands of projection images of identical particles present in a range of orientations relative to the electron beam (Cheng et al., 2015). In this way, the SNR is improved and information at near-atomic resolution can be retrieved.

The particles are first subject to reference-free classification in 2D. Modern SPA software suites such as RELION and cryoSPARC employ a maximum likelihood (ML) algorithm for 2D class averaging (Zivanov et al., 2018; Punjani et al., 2017). This approach is advantageous as all particles contribute to all classes but with various probability weightings. The result of this is that no hard decisions are made in early iterations of classification where references are noisy, alignments are poor and thus the uncertainty of alignment is high. Towards the end of classification, true particles end up grouped together, contributing heavily to a single class or subset of similar classes to which they were assigned with a high degree of certainty while making a negligible contribution to all other classes. ‘Junk’ particles that are inevitably present in the particle stack are sorted into less populated, noisy classes on the basis that they exhibit poor alignment with all true particle classes, or align equally well with all. These classes and their constituent particles can then be removed from the dataset, leaving a cleaner particle stack to work with.

2.4.3 3D reconstruction – Classification and refinement

The way in which 2D projections of an object can be used to obtain a 3D reconstruction of the same object is described by the Fourier slice theorem. This states that the Fourier transform of a 2D projection image of an object is identical to a central slice through the 3D Fourier transform of the same object. The central slice is normal to the projection vector and thus if the projection vectors of 2D projection images are known, the slices can be assembled in 3D, and an inverse Fourier transform would then give a real-space 3D reconstruction of the object (Sigworth, 2016). The difficulty with the SPA approach is that it presents an incomplete data problem. Particle images representing different projections can be averaged to improve the SNR, but the projection vectors are unknown. A technique called projection matching is used to circumvent this problem (see **Figure 2.4**). Here, an initial low-pass filtered model possessing roughly the expected size and shape of the target is back projected from all directions. The experimental images can then be compared to these computed projections allowing estimation of the projection vectors for each particle image. With these estimates, the experimental data can be used in an inverse Fourier transform, generating a new and slightly more accurate reference model. Iterating this process leads to gradually improved angular assignments for the particles and, consequently, a reconstruction that moves away from the starting model, towards a better representation of the true structure based solely on the experimental data (Nogales and Scheres, 2015). This process of angular assignment also implements a ML approach which helps account for uncertainty of alignments early in the process.

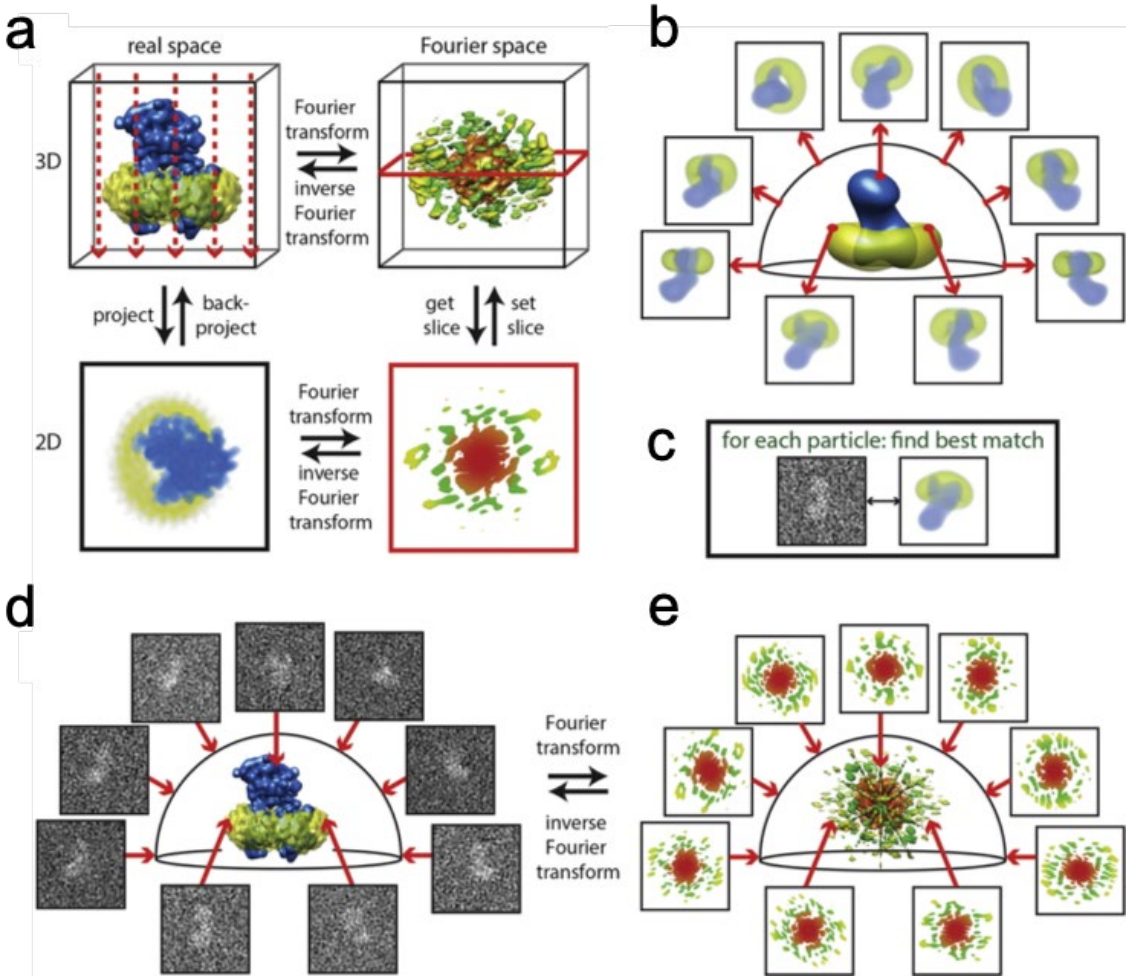


Figure 2.4: Fourier slice theorem and the projection matching approach. a, Schematic demonstration of the Fourier-slice theorem. The Fourier transform of a projection image of an object is identical to a central slice through the 3D Fourier transform of the same object. The slice in Fourier space is perpendicular to the real-space projection vector. **b-e**, Process of generating a 3D reconstruction of an object from 2D projections of that same object where projection vectors are unknown (projection matching approach). First, reference projections are calculated of an initial model of similar shape and size to the expected structure (**b**). Experimental projection images are compared against all reference projections to find the best match (**c**) allowing initial angles to be assigned to experimental images (**d**). Using the Fourier slice theorem, Fourier transforms of aligned projection images are positioned to give a 3D transform (**e**). The inverse Fourier transform then gives a 3D reconstruction of the object in real space. Iterating steps **b** through **e** using the 3D reconstruction obtained from the experimental data results in gradual improvement in the assignment of orientations of particle images and in this way the resolution of the reconstruction is improved. Figure reproduced from Figure 1 Nogales and Scheres (2015).

Returning to the SPA pipeline, once a clean particle stack has been obtained by several rounds of 2D classification, the next step is to assess the extent of structural heterogeneity present as a result of conformational or compositional differences in the particles. As an example, RELION tackles this problem with unsupervised 3D classification which, like 2D classification, is based on an ML approach (Scheres, 2016). No prior knowledge of structural heterogeneity

is required and the user simply selects the desired number of classes (optimal number is empirically determined). These are initialised using a low-pass filtered initial model and, in the first instance, each particle is assigned to a random class resulting in small variations in the class averages after one round of projection matching. The classification process then proceeds iteratively with all particle images being compared against all 3D class averages. Through this process, the class averages become more distinct from one another and, rather than differing as a result of differences in the initially assigned random particle subsets, they show biologically relevant differences that correspond to conformational and compositional differences in the experimental particles (Scheres, 2016). Several rounds of 3D classification may be required to generate truly homogenous particle stacks and these subsequent rounds are conducted in a hierarchical manner where the particles assigned to a particular class in one job are then used as the input to the next job. Furthermore, where information is required about heterogeneity in a particular subunit or region of a particle, a focused classification approach can be employed. Here the user provides a mask that encompasses the area of interest and the classification is performed using only the signal within this region (Scheres, 2016).

Once a homogeneous particle stack has been obtained, it is then possible to generate a high-resolution reconstruction of the particle using 3D refinement. In this job, the particles are split into two random subsets and each is refined separately. The refinement process involves the projection matching approach described above and, as it proceeds, the angular and translational sampling is increased until no further improvements in the estimated resolution or estimated accuracy of angular assignments are observed (convergence). Refining the dataset as two independent halves is considered the gold-standard approach. It permits accurate estimates of resolution via the Fourier Shell Correlation (FSC) which measures the agreement between the independent half maps at different spatial frequencies. An FSC value of 0.143 is commonly accepted as the resolution cut-off of the reconstruction (Rosenthal and Henderson, 2003). Treating the two half maps independently improves the accuracy of resolution estimates since high resolution noise is unlikely to correlate across the reconstructions and, as such, the risk of overfitting is reduced (Scheres and Chen, 2012).

2.4.4 Map sharpening

Maps generated by 3D refinement suffer from loss of information at high spatial frequencies partially as a result of the imaging conditions, including radiation damage and specimen movement, but also as a result of uncertainties in the computational processing. Estimating the rate of decay at resolutions beyond 10 Å gives rise to a B-factor which can then be used to correct for this dampening by up-weighting the high-resolution components (Rosenthal and Henderson, 2003). Maps from 3D refinement should always be subject to post-processing in this way, not only for the reason described above, but because resolution is underestimated until this point. Resolution estimates in 3D refinement come from comparisons of unmasked half maps and therefore contain contributions from noisy solvent regions. In post-processing, a mask is used that excludes the surrounding solvent, removing the noise and consequently increasing the FSCs and thus the estimated resolution as well.

Importantly, the resolution of many reconstructions can be improved further through the use of CTF refinement and Bayesian polishing in RELION (Zivanov et al., 2018). CTF refinement involves the estimation and correction of several imaging aberrations (beam tilt, symmetrical aberrations and asymmetrical aberrations). Furthermore, CTF parameters can be estimated on a per-particle level, improving accuracy especially where particles are present at different heights within the ice layer, or when the ice layer is tilted. Similarly, particle polishing involves per-particle motion tracking allowing more complex specimen motions to be mapped and corrected. Since both of these algorithms benefit from higher resolution input models, these jobs benefit from an iterative approach (CTF-refinement → Bayesian polishing → 3D auto-refinement → Post processing → …loop) until no further improvement in map resolution is observed.

2.4.5 Improving map interpretability

Once all efforts have been made to obtain the highest resolution reconstruction achievable from the data, it is often advantageous to optimise the interpretability of the map. As described above, B-factor sharpening in post-processing is important for retrieving high resolution features. However, the resolution of a reconstruction is rarely consistent across the entire map and, as such, sharpening using a single B-factor value will often lead to some

regions that are over-sharpened and some that are under-sharpened. This effect is especially pronounced where reconstructions contain flexible regions, and can hinder interpretation and model building. One option to alleviate this problem is to estimate and filter by local resolution. This can be achieved by considering small sub-volumes of the half-maps, calculating FSCs of these sub-volumes and using the estimated resolution to inform local low-pass filtering as is implemented in RELION. RELION also provides a wrapper for the ResMap algorithm which estimates resolution by testing for the presence of a 3D sinusoid of a particular wave-length at each voxel and filters accordingly (Kucukelbir et al., 2014). Alternative approaches involve local map sharpening e.g. LocScale which performs localised amplitude scaling based on a fitted atomic model (Jakobi et al., 2017), and the deep learning-based method DeepEMhancer which, having been trained on pairs of experimental maps and LocScale sharpened maps, performs masking and sharpening operations simultaneously, without the need for an atomic model (Sanchez-Garcia et al., 2021).

2.5 Model Building and Refinement

Where the resolution of a reconstruction is ~3.8 Å or better, it is possible to refine an atomic model into the cryoEM density. The starting point for modelling varies depending on the prior information available. If a previously determined model is available from the Protein Data Bank (PDB) or Electron Microscopy Data Bank (EMDB), it can be rigid fit to the density. Alternatively, in cases where there is no pre-existing model, it is possible to use structure prediction algorithms such as SWISS-MODEL (Waterhouse et al., 2018) or AlphaFold2 (Jumper et al., 2021) to provide a useful starting point. If the resolution is good enough, the model can also be built *de novo*. The typical model building workflow includes manual editing in Coot to improve the fit of the model to the density while at the same time satisfying backbone geometry restraints (Emsley and Cowtan, 2004). The model is then subject to automatic real-space refinement in software packages such as Phenix (Adams et al., 2010) or Refmac (Murshudov et al., 2011). Model optimisation typically requires several passes of manual adjustment and automatic refinement, with model quality assessed by validation tools such as MolProbity (Williams et al., 2018). Once the model is as good as reasonably achievable, model building can be considered complete.

2.6 Materials

2.6.1 Chemicals and biological reagents

Detergents were purchased from Anatrace and all other chemicals and reagents were purchased from either Sigma-Aldrich or Thermo Fisher Scientific unless otherwise stated.

2.6.2 Electron microscopy consumables and equipment

Grids used for negative stain TEM were 100 mesh continuous carbon-coated copper grids that were purchased from, and coated by, the Astbury Biostructure Laboratory facility. Lacy carbon grids with and without an ultrathin carbon support were purchased from Agar Scientifc. All holey/Quantifoil grids were purchased from Quantifoil. Glow discharging of negative stain grids was carried out with a PELCO easiGlow and glow-discharging of grids for cryoEM was carried out using either a Cressington 208 or the Quorum GloQube. Plunge freezing was performed using the Vitrobot IV (Thermo Fisher Scientific). All cryoEM grids were stored in dewars under liquid nitrogen (LN_2).

All microscopes used are located within the Astbury Biostructure Laboratory. Imaging of negative stain grids was performed on a Tecnai F20 (FEI) operated at 120 kV with an FEI Ceta, 4k x 4k CMOS CCD camera. Imaging of cryoEM grids was performed on a Titan Krios microscope (Thermo Fisher Scientific) with an accelerating voltage of 300 kV. Direct electron detectors used for data collection include the K2 summit (GATAN), Falcon III and Falcon IV (Thermo Fisher Scientific).

2.7 Peptide transporter RagAB project

2.7.1 Protein production and purification

2.7.1.1 RagAB

P.gingivalis RagAB W83 (KRAB) was generously provided by Bert van den Berg (Newcastle University). RagAB sample concentration was 3.5 mg/ml in 10 mM HEPES buffer pH 7.5 supplemented with 100 mM NaCl and 0.03% DDM.

2.7.1.2 MSP1E3D1

Membrane scaffold protein MSP1E3D1 was kindly provided by Antonio Calabrese (University of Leeds). The protein stock was in a standard MSP buffer (20 mM Tris-HCl, 0.1M NaCl, 0.5 mM EDTA, pH 7.4)

2.7.1.2 MSP2N2

BL21 (DE3) cells were transformed with the pET28a pMSP2N2 plasmid (addgene) that encodes His-tagged MSP2N2. Transformants were selected by overnight growth on LB-agar plates containing kanamycin (50 µg/ml). Two 50 ml volumes of LB supplemented with kanamycin (50 µg/ml) were inoculated with single colonies and incubated overnight at 37 °C, 200 rpm.

10 ml starter culture was added to 1 L of LB supplemented with kanamycin (50 µg/ml) (2 L culture was grown up in total). Cells were grown at 37 °C, 200 rpm to an OD_{600} of ~0.6. MSP expression was induced by the addition of 1 mM IPTG.

Cells were harvested after 3 h. The contents of each flask were transferred to a 1 L centrifuge bottle and spun at 6000 xg for 15 min. Pellets were resuspended in 30 ml 20mM phosphate buffer (pH 7.4) supplemented with cOmplete™ EDTA-free protease inhibitor cocktail (Roche). After resuspension, a stock of 10% Triton X-100 was added to a final concentration of 1 %. Finally, a small amount (~5mg) of deoxyribonuclease I was added. Cells were lysed by sonication and the lysate was spun at 50,000 xg for 10 min to remove cell debris.

Lysate supernatant was loaded onto a HiTrap™ 5 ml Chelating sepharose HP column (GE healthcare) equilibrated with 20 mM Phosphate buffer pH 7.4. The column was then washed with 100 ml of each of the following:

- 40 mM Tris-HCl, 0.3 M NaCl, 1% Triton X-100, pH 8.0
- 40 mM Tris-HCl, 0.3 M NaCl, 50 mM sodium cholate, 20 mM imidazole, pH 8.0
- 40 mM Tris-HCl, 0.3 M NaCl, 50 mM imidazole, pH 8.0

Finally, the MSP was eluted in 25 ml of 40 mM Tris-HCl, 0.3 M NaCl, 0.4 M imidazole, pH 8.0. The sample was placed in 3500 MWCO snakeskin dialysis tubing (ThermoFisher Scientific) and dialysed overnight (4 °C) into 20 mM Tris-HCl, 0.1M NaCl, 0.5 mM EDTA, pH 7.4. The final preparation was concentrated, syringe filtered (0.22 µm) and stored at -80 °C.

2.7.2 Nanodisc assembly

2.7.2.1 Preparation of Bio-Beads™

Bio-Beads ™ (Bio-Rad) were placed in a 50 ml Falcon tube. Methanol was added to cover the beads, and the tube inverted to ensure resuspension of the beads. The tube was centrifuged at 4000 rpm for 5 min and the methanol was decanted. This process was repeated twice over with methanol and a further three times with ddH_2O. Two final washes were performed using the stock RagAB buffer without the DDM. The required mass of beads was patted dry on a paper towel immediately prior to use.

2.7.2.2 Preparation of *E. coli* polar lipid extract

E. coli polar lipid extract in chloroform (25 mg/ml) was purchased from Avanti Polar Lipids. Samples were aliquoted into a clean, dry test tube and chloroform was evaporated off under a stream of nitrogen. To ensure complete drying, lipid films were left in a vacuum desiccator overnight. The lipid film was resuspended in an equivalent volume of 10 mM HEPES, 100 mM NaCl, 100 mM sodium cholate, pH 7.5. The sample was vortexed and sonicated until the film was fully dissolved. The lipid suspension was then aliquoted out and stored at -20 °C.

2.7.2.3 Reconstitution of RagAB into MSP1E3D1 or MSP2N2

Cholate-lipid mixtures, MSP and RagAB were mixed in a 1.5 ml Eppendorf such that the final volume (100 µl) would contain the appropriate ratio of components (estimated according to **Equation 1**). The final concentration of RagAB in all preparations was 1.75 mg/ml (5 µM). The final concentration of sodium cholate in the mixture was fixed at 14 mM as this facilitates disc formation. 50 mg of prewashed Bio-Beads™ were added, and the sample was incubated for 2 hrs, 4 °C on a rocking platform. To separate the sample from the beads, a hole was pierced in the top and bottom of the Eppendorf and the tube was placed through the lid of 15 ml Falcon tube (hole had been bored previously). The lid containing the Eppendorf was placed on a 15 ml falcon tube and spun at 1500 rpm for 2 min such that the sample collected in the falcon tube while the beads remained in the Eppendorf tube. The sample was then collected from the bottom of the 15 ml Falcon and placed in a clean Eppendorf tube. A fresh aliquot of beads was added, and the mixture was incubated overnight 4 °C with agitation. Finally, the sample was placed into a fresh aliquot of beads and incubated for a further 2 hrs. This last step was repeated once more to produce the final sample used in experiments.

$$N_{lip} = 2 * (S_{ND} - S_{prot})/S_{lip}$$

Equation 1: Where N_{lip} is the number of lipid molecules per disc, S_{ND} is the surface area of the disc (values are known), S_{prot} is the area of bilayer occupied by the transmembrane region of the protein of interest and S_{lip} is the area occupied by a single lipid molecule.

2.7.3 Negative stain TEM of RagAB samples

Continuous carbon-coated copper grids were glow discharged for 30 s (10 mA) using a PELCO easiGlow immediately prior to use. RagAB samples were diluted to ~3-7 µg/ml prior to application to the grid. 3 µl of sample was loaded and after ~30 s blotted off with filter paper. Three wash steps were performed with water, with immediate blotting after each. Uranyl acetate (1% w/v) was then applied to the grid and immediately blotted, followed by a second application of uranyl acetate (1% w/v) which was left for 30 s before being partially blotted to leave a visible meniscus that was then left to dry prior to imaging. Imaging was carried out on a TF20 operated at 120 KeV with an FEI Ceta, 4k x 4k CMOS camera.

2.7.4 High-resolution cryoEM of RagAB, principle dataset

A quantifoil grid (1.2/1.3, 300 mesh) was glow discharged for 30 s (10 mA) using a PELCO easiGlow (Ted Pella). Immediately after, the grid was loaded into the chamber of the FEI Vitrobot Mark IV (Thermo Fisher Scientific) that had been cooled to 6 °C, 100% relative humidity. A volume of 3.5 µL of the RagAB sample at 1.75 mg/ml was applied to the grid. Excess sample was blotted immediately using a blot force of 6 and a blot time of 6 s before plunging the grid into liquid nitrogen-cooled liquid ethane.

Data was collected on an FEI Titan Krios (Thermo Fisher Scientific) (ABSL, University of Leeds) TEM operated at 300 keV with a GIF energy filter (Gatan) and a Gatan K2 Summit detector operated in counting mode. Data acquisition parameters are shown in **Table 1**.

3605 micrograph movies were collected and image processing was carried out within Relion2.1 and, subsequently, Relion3.0 (Scheres et al., 2012; Zivanov et al., 2018). Drift and dose-corrected averages for each exposure were generated using MOTIONCOR2 (Zheng et

al., 2017). CTF estimation of motion corrected micrographs was done using Gctf (Zhang, 2016). Micrographs with estimated resolutions poorer than 5 Å and defocus values >4 µm were discarded using a Python script resulting in a final stack of 2843 micrographs (Thompson et al., 2019). Particles were picked using a Gaussian blob and subject to several rounds of 2D classification to remove 'junk' particles. Particles contributing to classes containing high-resolution structural information were carried forwards. An initial model was generated *de novo* from the data using the stochastic gradient descent algorithm implemented in RELION. Reference-free 3D classification was used to assess heterogeneity within the data and three conformational states – representing the closed-closed (CC), open-closed (OC) and open-open states (OO) – were apparent after the first round. The corresponding particle stacks, containing 87,897, 218,052 and 52,754 particles respectively, were treated independently in further processing. C2 symmetry was applied in refinements of both the CC and OO states other than where peptide occupancy was being assessed. Post-processing was performed using soft, extended masks and yielded sharpened reconstructions for the closed-closed, open-closed and open-open states with resolutions of 3.7 Å, 3.7 Å and 3.9 Å respectively, as estimated by the gold-standard Fourier shell correlations using the 0.143 criterion. The original micrographs movies were later motion-corrected in RELION 3.0. Particles contributing to the final reconstructions were re-extracted from the resulting micrographs. Following reconstruction, iterative rounds of per-particle CTF refinement and Bayesian polishing were employed. A final round of 3D classification without alignment was conducted on each stack to remove poorly aligned particles. After refinement and post-processing, final resolutions for CC, OC and OO reconstructions were resolved to 3.3 Å, 3.3 Å and 3.4 Å, respectively.

For assessment of NTE domain flexibility in the OC state, a mask was generated that encompassed only the plug and NTE domain on the open side of the transporter. This mask was used to identify the signal to be retained in a particle subtraction job in RELION. In this way, signal outside of the mask was removed from the experimental particle images allowing consistent comparison of projections from masked regions with the experimental projections in downstream classification jobs (Scheres, 2016). Classification of subtracted particles was performed without alignment. Selecting one resulting class and reverting to the original particle images before 3D refinement resulted in improved density for the NTE domain and improved interpretability of density for an interdomain linker.

2.7.5 Model building of RagAB, principle dataset

Examination of the maps revealed that their handedness was incorrect. Maps were therefore z-flipped in UCSF chimera (Pettersen et al., 2004). The RagAB W83 KRAB crystal structure was fit as a rigid body into the CC density map and subjected to several iterations of manual refinement in COOT and real space refinement in Phenix (Adams et al., 2010). The asymmetric unit was symmetrised in Chimera after each iteration. Starting models for the OC and OO states of the complex were obtained from the CC structure by rigid-body fitting one or both RagB subunits into their cognate open density in the OC and OO maps, respectively. These too were subjected to several iterations of manual refinement in COOT and real space refinement in Phenix. For C2 symmetric structures, BIOMT records containing matrices for generating the biological assembly from the asymmetric unit were generated using the sym command in Chimera. Molprobity was used for model validation model statistics are shown in **Table 2.1**.

	OO	OC	CC
Data collection and processing			
Magnification	130,000 x	130,000 x	130,000 x
Voltage (kV)	300	300	300
Electron exposure (e-/Å^2)	77.88	77.88	77.88
Defocus range (μm)	-1.2 to -3.0	-1.2 to -3.0	-1.2 to -3.0
Pixel size (Å)	1.07	1.07	1.07
Symmetry imposed	C2	C1	C2
Initial particle images pre-3D classification (no.)		491,870	
Initial particle images post-classification (no.)	87,897	218,052	52,754
Final particle images (no.)	86,877	213,143	51,849
Map resolution (Å) (FSC 0.143)	3.3	3.3	3.4
Map sharpening B factor (Å)	-120	-114.1	-119.9
Refinement			
Initial model used (PDB code)	6SLI	6SLI	6SLI
Model composition			
Non-hydrogen atoms protein	11,020	22,136	11,076
Non-hydrogen atoms peptide ligand	79	82	65
Protein residues	1,398	2,791	1,405
R.m.s deviations			
Bond lengths (Å)	0.009	0.011	0.009
Bond angles (°)	0.938	0.998	0.996
Validation			
Molprobity score	1.5	1.59	1.59
Clashscore	2.45	3.32	3.67
Poor rotatmers (%)	0.26	0.34	0.09
Ramachandran plot			
Favoured (%)	92.39	92.44	93.27
Allowed (%)	7.61	7.38	6.66
Disallowed (%)	0	0.18	0.07
Deposition			
EMDB	10243	10245	10241
PDB	6SML	6SMQ	6SM3

Table 2.1: Data acquisition parameters and refinement statistics for models of the RagAB complex from the principle RagAB dataset

2.7.6 CryoEM of RagAB – flushing and peptide addition experiments

Quantifoil 1.2/1.3 300 mesh grids were glow discharged using the Cressington 208 for 30 s (10 mA). RagAB was loaded onto the grids at a concentration of 3 mg/ml in all cases. In the first of two experiments conducted, samples included RagAB washed extensively on an anion-exchange column, RagAB doped with 100x molar excess of the P4 peptide, and RagAB as purified to act as a control. All RagAB samples came from the same purified stock for consistency. In all cases grids were prepared using an FEI Vitrobot Mark IV (Thermo Fisher Scientific) that had been cooled to 6 °C, 100% relative humidity. A sample volume of 3.5 µL was applied to the grid and excess sample was blotted immediately using a blot force of 6 and a blot time of 6 s before plunging the grid into liquid nitrogen-cooled liquid ethane.

Data acquisition parameters are shown in **Table 2.2**. Image processing was performed using Relion3.0. Micrographs were pre-processed using Motioncor2 and Gctf. Around 2000 particles were picked manually for each dataset and subjected to 2D classification. Representative 2D class averages were selected for use in template-based particle picking. Processing of each dataset was taken as far as 3D classification to allow comparison of populations of the different conformational states of RagAB present in the data. Contributing particle numbers are shown in **Table 2.2**.

In the P21-addition experiment, control and P21-doped grids (50-fold molar excess of P21 peptide) were prepared in the same way as described above. Data acquisition parameters are shown in **Table 2.2**. Pre-processing of micrographs was performed as described above, however particle picking was carried out using crYOLO (Wagner et al., 2019). Both datasets were processed up until 3D classification to allow comparison of the conformational states present. In an attempt to identify peptide density unique to P21, the CC state of the transporter was refined, post-processed and subject to CTF refinement and Bayesian polishing, resulting in a sharpened reconstruction of 3.8 Å (B factor -125.5) from 23,045 particles.

	RagAB flushed	RagAB P4	RagAB control	RagAB P21	RagAB control
Microscope	Krios	Krios	Krios	Krios	Krios
Camera	Falcon III	Falcon III	Falcon III	K2 summit	K2 summit
Imaging mode	Counting	Integrating	Counting	Counting	Counting
Voltage (kV)	300	300	300	300	300
Pixel Size (Å)	1.065	1.065	1.065	1.07	1.07
Total Dose (e-/Å^2)	46.2	85.43	42.7	75	73
Number of Frames	45	79	45	50	50
Defocus Range (μm)	-1.5 to -3.3	-1.5 to -3.3	-1.5 to -3.3	-1.5 to -3.3	-1.5 to -3.3
Micrographs	511	1515	1609	1100	1464
Particles in good 3D classes	92,785	120,449	145,035	49,873	60,544

Table 2.2: Data acquisition parameters for flushed and peptide-doped RagAB datasets. Columns of the same colour indicate that the corresponding grids were prepared from a single aliquot of purified RagAB.

2.7.7 Density analysis and figure making

Figures of electron micrographs were made using ImageJ (Schneider et al., 2012). Investigation and comparison of EM density maps was performed using Chimera (Pettersen et al., 2004) and COOT (Emsley and Cowtan, 2004). Figures of maps and models were generated using Chimera, ChimeraX (Goddard et al., 2018) and PyMol. Interface analysis was carried out using PDBePISA (Krissinel and Henrick, 2007).

2.8 Bt1762-Bt1763, levan transporting SusCD complex project

2.8.1 Protein acquisition

The levan transporting SusCD (Bt1762-Bt1763) complex was generously provided by Bert van den Berg (Newcastle University). SusCD was at a concentration of 3 mg/ml in 10 mM HEPES buffer pH 7.5 supplemented with 100 mM NaCl and 0.05% LDAO.

2.8.2 SDS-PAGE and protein ID mass spectrometry

Protein samples were prepared in a 3:1 ratio with 4x Laemmli Sample Buffer (BioRad). Where samples were heat treated, the tube containing the aforementioned mix was placed in a heat block at 95 °C for ~5 min. Samples were loaded into a 4-15% Mini-PROTEAN® TGX™ Precast Protein Gel (Bio-Rad) next to PageRuler™ Prestained Protein Ladder (180 kDa to 10 kDa)

(Thermo Fisher Scientific). Electrophoresis was conducted in a running buffer of 25 mM Tris, 192 mM glycine, pH 8.3 at 180 V for ~45 min. Protein was stained using InstantBlue® Coomassie protein stain (abcam) until clear bands were visible (~20-30 min). For protein ID mass spectrometry, the intact gel was supplied to the University of Leeds Mass Spectrometry Facility who carried out band excision, downstream sample preparation, mass spectrometry and data analysis.

2.8.3 Negative stain TEM of the levan SusCD sample

Continuous carbon-coated copper grids were glow discharged for 30 s (10 mA) using a PELCO easiGlow immediately prior to use. Levan SusCD was diluted to 5 μg/ml prior to application to the grid. 3 μl of sample was loaded and after ~30 s blotted off with filter paper. Three wash steps were performed with water, with immediate blotting after each. Uranyl acetate (1% w/v) was then applied to the grid and immediately blotted, followed by a second application of uranyl acetate (1% w/v) which was left for 30 s before being partially blotted to leave a visible meniscus that was then left to dry prior to imaging. Imaging was carried out on a TF20 operated at 120 KeV with an FEI Ceta, 4k x 4k CMOS camera.

2.8.4 High-resolution cryoEM of the levan transporting SusCD complex

A sample of purified Bt1762-Bt1763 solubilised in an LDAO-containing buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.05% LDAO) was prepared at 0.02 mg/ml. Lacy carbon 300-mesh copper grids coated with a <3 nm continuous carbon film (Agar Scientific) were glow-discharged in air (10 mA, 30 s) using a Cressington 208. Grids were used immediately after glow-discharging and were prepared using the FEI Vitrobot Mark IV (Thermo Fisher Scientific) pre-cooled to a chamber temperature of 4 °C and 100% relative humidity. A sample volume of 3.5 μL was loaded onto the grid and allowed to sit for 10 s before blotting (6 s blot time) and plunge-freezing into liquid nitrogen-cooled liquid ethane.

Data was collected on an FEI Titan Krios (Thermo Fisher Scientific) TEM operated at 300 kV with a GIF energy filter (Gatan) and K2 Summit direct electron detector (Gatan) operating in counting mode. Data acquisition parameters can be found in **Table 2.3**.

A dataset comprising 3276 micrograph movies was collected and image processing was carried out within Relion3.1 (Zivanov et al., 2018). Drift correction and dose-weighting was carried out using MOTIONCOR2 (Zheng et al., 2017). CTF estimation of motion corrected micrographs was performed using Gctf (Zhang, 2016). Manual picking of ~2000 particles was carried out prior to an initial round of 2D classification to generate templates for automated template-based particle picking within Relion. The initial particle stack contained ~775,000 particles which were 2x binned upon extraction. Assessment of the picked coordinates showed that picking was not stringent and therefore the first round of 2D classification was prepared using the --maxsig 25 additional argument to reduce the computational time required and separate 'false' particles such as carbon edges, which were picked heavily despite attempts to optimise the picking parameters.

Around 575,000 particles were carried forward to more stringent rounds of 2D classification after which ~177,000 particles contributed to classes containing high-resolution structural features. A 3D starting model was generated *de novo* from the data using the stochastic gradient decent algorithm within RELION. Several rounds of 3D classification revealed significant compositional heterogeneity in the data. Particles from classes representing the complete $SusC_2D_2$ complex alone (115,332 particles) were subjected to subsequent rounds of 3D classification to assess conformational heterogeneity. Three predominant conformational states of the levan transport system were identified: open-open (OO), open-closed (OC) and closed-closed (CC). Invariably, the closed position of SusD was associated with an absence of density for the plug domain of Bt1763. When visualised at high threshold levels, the closed-closed structure showed evidence of plug density in one of the SusC barrels, indicating that global classification was ineffective at generating a homogeneous particle stack. As a result, a masked classification approach was employed. A mask was generated that encompassed the barrel with weak plug density and 3D classification was performed without alignment. Given that the region of interest was relatively small and without alignment there is no risk of overfitting, the regularisation parameter, T, was set to 20. This classification successfully separated 'true' plug-less closed structures from a conformational state where a marginally open position of SusD is associated with a plugged SusC barrel. Clean particle stacks for the OO, OC and CC conformations contained 32,190, 22,205 and 17,416 particles respectively. These stacks were subject to multiple rounds of CTF refinement and Bayesian polishing until

no further improvement in resolution was observed. C2 symmetry was applied in refinements of the open-open and closed-closed states. Post-processing was performed using soft, extended masks and yielded sharpened reconstructions for the OO, OC and CC states with resolutions of 3.9, 4.7 and 4.2 Å, respectively, as estimated by gold standard Fourier Shell correlations using the 0.143 criterion. Maps were then filtered by local resolution using the local resolution filtering algorithm within RELION. The number of particles assigned to the open-open state with one SusD unit possessing a wide-open position, and the open-marginally open state discovered by focused 3D classification are 9,903 and 12,637 respectively.

2.8.5 Model building of the levan transporting SusCD complex

Assessment of the cryoEM density for the three predominant conformational states revealed that map handedness was incorrect. Maps were therefore z-flipped in UCSF Chimera (Pettersen et al., 2004). The reconstruction of the OO state was of sufficient resolution for model building and refinement. SusC and SusD subunits from the closed-closed crystal structure were independently rigid body fit to the local resolution filtered map and later subjected to several iterations of manual refinement in COOT (Emsley and Kowtan, 2004) and 'real-space refinement' in Phenix (Adams et al., 2010). The asymmetric unit was symmetrised in Chimera after each iteration. Molprobity (Chen et al., 2010) was used for model validation. The reconstructions of OC and CC states were of insufficient resolution to permit model building and refinement owing to low particle numbers and a sub-optimal distribution of viewing angles. Instead, the crystal structure of Bt1762-1763 was rigid body fit to the CC state. The ligand was removed from the model and inspection in COOT showed that no density extended past Lys213 in the direction of the N-terminus. All residues N-terminal of Lys213 were therefore removed from the model before a second round of rigid body fitting. The open state from the OO EM structure and the closed state from the crystal structure (modified as described above) were rigid fit to their corresponding densities in the OC state. Rigid body fitting was performed in Phenix. For C2 symmetric structures, BIOMT records containing matrices for generating the biological assembly from the asymmetric unit were generated using the sym command in Chimera. Model validation statistics can be found in **Table 2.3**.

	OO	OC	CC
Data collection and processing			
Magnification	130,000 x	130,000 x	130,000 x
Voltage (kV)	300	300	300
Electron exposure	63.84	63.84	63.84
Defocus range (μm)	-1.5 to -3.3	-1.5 to -3.3	-1.5 to -3.3
Pixel size (Å)	1.07	1.07	1.07
Symmetry imposed	C2	C1	C2
Initial particle images pre-3D classification (no.)	_____	203,450	_____
Final particle images (no.)	32,190	22,205	17,416
Map resolution (Å) (FSC 0.143)	3.9	4.7	4.2
Map sharpening B-factor	-110	-138.78	-92.95
Refinement			
Initial model used (PDB code)	6ZAZ	6ZAZ	6Z8I
Model composition			
Non-hydrogen atoms protein	11,425	22,275	10,860
Protein residues	1,475	2,844	1,359
R.m.s deviations			
Bond lengths	0.005	0.007	0.008
Bond angles	0.867	0.929	0.988
Validation			
Molprobity score	1.93	2.06	2.08
Clashscore	4.76	13.46	17.04
Poor rotatmers (%)	1.90	0.95	0
Ramachandran plot			
Favoured (%)	92.30	93.55	94.91
Allowed (%)	7.70	6.24	4.65
Disallowed (%)	0	0.21	0.44
Deposition			
EMDB	11273	11277	11274
PDB	6ZLT	6ZM1	6ZLU

Table 2.3: Data acquisition parameters and refinement statistics for models of the levan transporting SusCD complex in the absence of substrate.

2.8.6 Density analysis and figure making

Figures of electron micrographs were made using ImageJ (Schneider et al., 2012). Investigation and comparison of EM density maps was performed using Chimera (Pettersen

et al., 2004) and COOT (Emsley and Cowtan, 2004). Figures of maps and models were generated using Chimera, ChimeraX (Goddard et al., 2018) and PyMol.

2.9 Levan and dextran four-component complex projects

2.9.1 Protein acquisition

The levan and dextran utilising four-component complexes were generously provided by Bert van den Berg (Newcastle University). Sample concentrations were ~3 mg/ml for the levan utilising complex and ~2.5 mg/ml for the dextran utilising complex. The solubilising buffer was identical for both systems and comprised 10 mM HEPES buffer pH 7.5 supplemented with 100 mM NaCl and 0.03 % DDM.

2.9.2 Negative stain TEM of the levan utilising four-component complex

Continuous carbon-coated copper grids were glow discharged for 30 s (10 mA) using a PELCO easiGlow immediately prior to use. The sample was diluted to 3.75 μg/ml prior to application to the grid. 3 μl of sample was loaded and after ~30 s blotted off with filter paper. Three wash steps were performed with water, with immediate blotting after each. Uranyl acetate (1% w/v) was then applied to the grid and immediately blotted, followed by a second application of uranyl acetate (1% w/v) which was left for 30 s before being partially blotted to leave a visible meniscus that was then left to dry prior to imaging. Imaging was carried out on a TF20 operated at 120 kV with an FEI Ceta, 4k x 4k CMOS camera.

2.9.3 Negative stain TEM of the dextran utilising four-component complex

Continuous carbon-coated copper grids were glow discharged for 30 s (10 mA) using a PELCO easiGlow immediately prior to use. The sample was diluted to 12 μg/ml prior to application to the grid. 3 μl of sample was loaded and after ~30 s blotted off with filter paper. Three wash steps were performed with water, with immediate blotting after each. Either uranyl acetate (1% w/v) or uranyl formate (1% w/v) was then applied to the grid and immediately blotted, followed by a second application of the same stain which was left for 30 s before being partially blotted to leave a visible meniscus that was then left to dry prior to imaging. Imaging

was carried out on a TF20 operated at 120 kV with an FEI Ceta, 4k x 4k CMOS camera. Datasets of uranyl formate-stained dextran utilising four-component complex collected with 0°, 35° and 50° stage tilts contained 59, 63 and 55 micrographs, respectively. Micrographs were acquired manually at 39,886 x magnification with estimated defocus values ~2-4 μm. Particles were picked using the Laplacian of Gaussian picker within RELION and extracted images were processed as far as 2D classification.

2.9.4 High-resolution cryoEM of the levan utilising four-component complex without substrate

A sample of the purified levan utilisome complex (Bt1760-Bt1763) solubilised in DDM-containing buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.03% DDM) was prepared at 3 mg/ml. Lacy carbon 300-mesh copper grids (Agar Scientific) were glow-discharged in air (10 mA, 30s, Cressington 208). A sample volume of 3.5 μL was applied to the grid. Blotting and plunge freezing into liquid nitrogen-cooled liquid ethane were carried out using an FEI Vitrobot Mark IV (Thermo Fisher Scientific) with chamber conditions set at a temperature of 4 °C and 100% relative humidity. Micrograph movies were collected on a Titan Krios Microscope (Thermo Fisher Scientific) operating at 300 kV with a Falcon III direct electron detector operating in counting mode. Data acquisition parameters can be found in **Table 2.4**.

An initial dataset comprising 2057 micrograph movies was collected and image processing was carried out in Relion3.1 (Zivanov et al., 2018). Drift correction and dose-weighting was carried out using MOTIONCOR2 (Zheng et al., 2017). CTF estimation of motion corrected micrographs was performed using Gctf (Zhang, 2016). Template-based particle picking within Relion was hindered by the large amount of carbon present in many micrographs. The micrograph stack was therefore manually culled to remove micrographs containing >50% carbon, leaving 1093 micrographs for further processing. Final particle picking was performed using the crYOLO general model (Wagner et al., 2019) and yielded 96,639 particles. The particle stack was subject to several rounds of 2D classification, after which 89,305 particles remained. A 3D starting model was generated *de novo* from the data using the initial model job within RELION. 3D classification was used to isolate particles corresponding to the complete octameric utilisome complex (45,594). These particles were carried forward and

subjected to further rounds of classification in 3D to assess the conformational heterogeneity of the complex. Classification revealed considerable heterogeneity in the position of the SusD lids, with positions described as ‘wide open’ (W) and ‘narrow open’ (N) identified in all possible combinations. WW, WN and NN states contained 16,155, 22,452 and 6,987 particles respectively. To increase particle numbers and improve the results of downstream processing a second dataset was collected.

The second dataset comprised 3142 micrographs movies. These were processed in the same way as described for the initial dataset, with particles picked using crYOLO. Classification in 3D yielded 146,512 particles that corresponded to the complete octameric utilisome complex. These particles were carried forward to further rounds of 3D classification that revealed extensive heterogeneity in position and quality of density for the SusD subunit. WW and WN states consistent with the early dataset were identified and contained 76,065 and 33,263 particles, respectively. Particles contributing to these consistent states (WW and WN) from the two datasets were combined and refined together with unique optics groups identifying the parent datasets. Particles were subject to CTF-refinement and Bayesian polishing (run separately for each dataset) until no further improvement in resolution was seen. Note that C2 symmetry was applied in refinement of the WW state. Post-processing was performed using soft, extended masks and yielded global sharpened reconstructions of the WW and WN states of 3.7 Å and 4.0 Å respectively (applied B factors were -144 and -128 respectively), as estimated by gold standard Fourier Shell correlations using the 0.143 criterion. The local resolution algorithm within RELION was used to assess variation in resolution across the map and showed relatively poor resolution for SusD subunits.

To improve the resolution for the more stable, C2 symmetric portions of the utilisome (SusC, levanase and N-terminal region of the levan-binding protein subunits) all particles corresponding to the complete octameric complex from both datasets (192,106 total) were combined and subject to focused refinement with C2 symmetry. The mask applied in focused refinement excluded the SusD subunits. The resolution after sharpening (B factor -149) was 3.5 Å.

2.9.5 Model building of the levan utilising four-component complex

The sharpened reconstruction from the focused refinement described above was of sufficient resolution for model building and refinement of the SusC and levanase subunits, as well as the most N-terminal domain of the levan binding proteins. The SusC subunit from the open-open cryoEM structure of the levan $SusC_2D_2$ complex in isolation was fit into the density as a rigid body and used as a starting model. The same approach was used for a crystal structure of the levanase in isolation (6R3R). The N-terminal domain of the levan binding protein was built using the corresponding region of an AlphaFold2 (Jumper et al., 2021) predicted model as a starting point. The model was subjected to several iterations of manual refinement in COOT (Emsley and Kowtan, 2004) and 'real-space refinement' in Phenix (Adams et al., 2010). The asymmetric unit was symmetrised in Chimera after each iteration. BIOMT records containing matrices for generating the biological assembly from the asymmetric unit were generated using the sym command in Chimera. Molprobity (Chen et al., 2010) was used for model validation (**Table 2.4**). The model was prepared for deposition by Augustinas Silale.

	Dataset 1	Dataset 2
Data collection and processing		
Magnification	75,000 x	75,000 x
Voltage (kV)	300	300
Electron exposure (e-/Å^2)	35.4	37
Defocus range (μm)	-1.5 to -3.3	-1.5 to -3.3
Pixel size (Å)	1.065	1.065
Initial particle images pre-3D classification (no.)	89,305	280,696
Final utilisome particles post-classification (no.)	45,594	146,512
Total utilisome particles after datasets combined (no.)	192,106	
Symmetry imposed	C2	
Map resolution (Å) (FSC 0.143)	3.5	
Map sharpening B-factor	-149	
Refinement		
Initial model used (PDB code or source)		
SusC (Bt1763)	6ZLT	
Levanase (Bt1760)	6R3R	
Levan binding protein (Bt1761)	AlphaFold2	
Model composition		
Non-hydrogen atoms	23,688	
Protein residues	2988	
R.m.s deviations		
Bond lengths	0.002	
Bond angles	0.442	
Validation		
Molprobity score	1.75	
Clashscore	8.51	
Poor rotatmers (%)	0	
Ramachandran plot		
Favoured (%)	95.76	
Allowed (%)	4.17	
Disallowed (%)	0	

Table 2.4 Data acquisition parameters for the two levan four-component complex datasets obtained in the absence of substrate. Datasets were combined and refinement statistics for the model of the levan utilisome complex are shown. Note that SusD subunits were not included in this model.

2.9.6 High-resolution cryoEM of the levan utilising four-component complex in the presence of FOS (DP8-12)

A sample of the levan utilisome containing an active levanase solubilised in a DDM-containing buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.03% DDM) was prepared at 3 mg/ml and incubated with 0.5 mM levan FOS with a degree of polymerisation of ~8-12 for at least one hour before grid preparation. Quantifoil carbon grids (R1.2/1.3, 300 mesh) were glow discharged (30 mA, 60 s, Quorum GloQube) in the presence of amylamine vapour. A sample volume of 3.5 μL was applied to the grid. Blotting and plunge-freezing into liquid nitrogen-cooled liquid ethane were carried out using an FEI Vitrobot Mark IV (Thermo Fisher Scientific) with chamber conditions set at a temperature of 4 °C and 100% relative humidity. Micrograph movies were collected on a Titan Krios Microscope (Thermo Fisher Scientific) operating at 300 kV with a Falcon III direct electron detector operating in counting mode. Data acquisition parameters can be found in **Table 2.5**.

A dataset comprising 974 micrograph movies was collected and image processing was carried out in RELION 3.1 (Zivanov et al., 2018). Drift correction and dose-weighting was done using MOTIONCOR2 (Zheng et al., 2017). CTF estimation of motion corrected micrographs was performed using Gctf (Zhang, 2016). Particle picking was performed using the general model of crYOLO and yielded 72,373 particles (Wagner et al., 2019). Unwanted particles/contamination were removed from the particle stack through two rounds of 2D classification, after which 63,789 remained. Classification in 3D was used to address compositional heterogeneity. Good classes containing clear $SusC_2D_2$ density represented the complete octameric utilisome, a hexameric assembly which lacked one levanase and one LBP subunit and a naked $SusC_2D_2$ complex. Contributing particles numbers were 17,045, 31,789 and 7390, respectively. SusD lids invariantly occupied a closed position and conformational heterogeneity was limited to the levan binding protein.

Particle subtraction was performed to remove signal for additional lipoprotein components for all experimental images contributing to good classes, as defined above, leaving just the $SusC_2D_2$ core complex. A mask encompassing only the $SusC_2D_2$ core was generated using the volume eraser tool within Chimera (Pettersen et al., 2004) before using the resulting carved

volume as an input for mask creation within RELION. Subtracted particles were used in a focused refinement with the same mask applied while enforcing C2 symmetry. Iterative rounds of CTF-refinement and Bayesian polishing were employed until no further improvement in resolution was observed. Post-processing resulted in a final sharpened reconstruction at 2.9 Å.

2.9.7 Model building of the core $SusC_2D_2$ complex in the closed-closed state with bound FOS (DP8-12)

The sharpened reconstruction from focused refinement was of sufficient resolution for model building and refinement of both the $SusC_2D_2$ complex and the FOS ligands present. The crystal structure of the isolated core complex in the presence of FOS (DP6-12) (6ZAZ) was used as a starting model and fit as a rigid body to the cryoEM density. Restraints for the FOS ligand were generated in JLigand (Lebedev et al., 2012) and link definitions were manually added to the pdb file. The model was subjected to several iterations of manual refinement in COOT (Emsley and Kowtan, 2004) and real-space refinement in Phenix (Adams et al., 2010). The asymmetric unit was symmetrised in Chimera after each iteration. BIOMT records containing matrices for generating the biological assembly from the asymmetric unit were generated using the sym command in Chimera. Molprobity (Chen et al., 2010) was used for model validation (**Table 2.5**). The model was prepared for deposition by Augustinas Silale.

Data collection and processing		
Magnification	75,000 x	
Voltage (kV)	300	
Electron exposure (e-/Å^2)	38.5	
Defocus range (μm)	-1.5 to -3.0	
Pixel size (Å)	1.065	
Initial particle images pre-3D classification (no.)	63,789	
Particle stack	**Utilisome**	**SusC$_2$D$_2$ core**
Final particle images post-classification (no.)	15,012	54,736
Symmetry imposed	C2	C2
Map resolution (Å) (FSC 0.143)	3.2	2.9
Map sharpening B-factor	-90.63	-96.42
Refinement		
Initial model used (PDB code or source)		
SusC (Bt1763)	6ZAZ	6ZAZ
SusD (Bt1762)	6ZAZ	6ZAZ
Levanase (Bt1760)	6R3R	-
Levan binding protein (Bt1761)	Alphafold2	-
Model composition		
Non-hydrogen atoms	33,128	23,704
Protein residues	4142	2958
R.m.s deviations		
Bond lengths	0.003	0.002
Bond angles	0.498	0.411
Validation		
Molprobity score	1.71	1.36
Clashscore	8.33	6.62
Poor rotatmers (%)	0.03	0
Ramachandran plot		
Favoured (%)	96.17	98.17
Allowed (%)	3.73	1.76
Disallowed (%)	0.1	0.07

Table 2.5: Data acquisition parameters for the levan four-component complex in the presence of FOS DP8-12, and refinement statistics for models of the levan utilisome complex and SusC$_2$D$_2$ core.

2.9.8 High-resolution cryoEM of the levan utilising four-component complex with inactivated levanase in the presence of FOS (DP15-25)

A sample of the levan utilisome containing inactivated levanase (D41A) solubilised in a DDM-containing buffer (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.03% DDM) was prepared at 3 mg/ml and incubated with 0.5 mM levan FOS with a degree of polymerisation 15-25 for at least one hour before grid preparation. Grid type, preparation, microscope and detector were the same as for the active levan utilisome described above. Data acquisition parameters can be found in **Table 2.6**.

A dataset comprising 1388 micrographs movies was collected and image processing was carried out in RELION 3.1 (Zivanov et al., 2018). Drift correction and dose-weighting were carried out using MOTIONCOR2 (Zheng et al., 2017). Particle picking was performed using the general model of crYOLO and yielded 157,953 particles (Wagner et al., 2018). Unwanted particles/contamination were removed from the particle stack through two rounds of 2D classification, after which 146,056 particles remained. Classification in 3D was used to address compositional heterogeneity. Good classes representing the complete octameric utilisome and the hexameric assembly lacking one levan binding protein and one levanase subunit were observed and contained 78,469 and 42,488 particles, respectively. Conformational heterogeneity in the position of the levan binding protein was observed with some classes possessing a conformation where this subunit was held close to the levanase. Particles contributing to all classes with evidence of this docked conformation of the levan binding protein were pooled (98,755) and a consensus refinement was carried out. CTF refinement and Bayesian polishing were performed iteratively until no further improvement in resolution was observed. The resulting reconstruction possessed relatively weak density for the C-terminal domain of the levan binding protein. To improve this, a focused classification approach without alignment was used.

A mask encompassing only the docked position of the levan binding protein with some surrounding density was created using the volume eraser tool in Chimera followed by mask creation in RELION. A 3D classification job was run with the aforementioned mask without alignment. Notably, the initial model was minimally low pass filtered to just above the

resolution reported in refinement (in this case 3.5 Å) thus allowing classification on high resolution features. Several T values were empirically tested and a T value of 40 was found to give the best results. A single class, containing 27,310 particles, was identified that possessed well resolved density for the C-terminal domain of the levan binding protein. A particle star file containing information for particles contributing to this class was created manually via command line arguments. From this, two new star files were generated that contained random half sets of the selected data, making use of the random subset value (1 or 2) that RELION automatically assigns to each particle. Each star file was then used as the input for relion_reconstruct, and in this way two independent half maps corresponding to the unmasked structure were generated from the data. These half maps were then used in a post-processing job and yielded a sharpened reconstruction of 3.0 Å, as estimated by gold standard Fourier Shell correlations using the 0.143 criterion. The density for the C-terminal domain of the levan binding protein was improved, and density corresponding to levan chain that links the levan binding protein to the levanase was also visible.

To obtain the best resolution for bound FOS molecules all $SusC_2D_2$-containing particles were considered regardless of lipoprotein complement or conformation. A particle subtraction and focused refinement strategy targeting the $SusC_2D_2$ core of the complex was used as previously described in **2.9.6**. Post-processing the final C2 symmetrised refinement resulted in a sharpened reconstruction at 2.7 Å.

2.9.9 Model building into maps of the levan utilising four-component complex with inactivated levanase in the presence of FOS (DP15-25)

Considering the utilisome map containing well resolved density for the C-terminal domain of the levan binding protein associated with one half of the complex, an adhoc lowpass filter of 3.2 Å was applied in post-processing and the resulting map was used for model building. The $SusC_2D_2$ model built into the cryoEM reconstruction obtained for the core complex in the presence of FOS DP8-12 was docked as a rigid body into the corresponding density in the complete utilisome map and used as a starting model for the SusC and SusD components. The cryoEM structure of the levanase that was refined into the apo utilisome structure was rigid body fit to its corresponding density and used as a starting model. For the docked levan

binding protein, an AlphaFold2 (Jumper et al., 2021) predicted model was segmented into the three distinct domains of the protein (two Ig-like domains and the C-terminal domain) and each was fit individually as a rigid body to the corresponding cryoEM density, which allowed differences in relative domain orientations between the predicted and experimental structures to be easily accommodated. The individual domains were later merged in COOT allowing building of a contiguous model. For the levan binding protein associated with the alternate SusC subunit, only the N-terminal Ig-like domain was well resolved and thus the rest of the model was truncated in this case. FOS substrates were modelled by the addition of one fructose monomer at a time within COOT. For β2,6 linkages the O6 atom was deleted and for β2,1 linkages the O2 atom was deleted to maintain correct valences. The fructose monomers corresponding to each levan chain were merged into one model within COOT and link definitions were manually added to the pdb files prior to refinement. Restraints for these linkages were generated in JLigand (Ledbedev et al., 2012). The model was subjected to several iterations of manual refinement in COOT (Emsley and Kowtan, 2004) and real-space refinement in Phenix (Adams et al., 2010). Molprobity (Chen et al., 2010) was used for model validation (**Table 2.6**). The model was prepared for deposition by Augustinas Silale.

Data collection and processing		
Magnification	75,000 x	
Voltage (kV)	300	
Electron exposure (e-/Å^2)	37.8	
Defocus range (μm)	-1.5 to -3.0	
Pixel size (Å)	1.065	
Initial particle images pre-3D classification (no.)	146,056	
Particle stack	**From focused 3D incl. resolved Bt1761**	**Containing $SusC_2D_2$ core**
Final particle images post-classification (no.)	27,310	120,957
Symmetry imposed	C1	C2
Map resolution (Å) (FSC 0.143)	3.1	2.7
Map sharpening B-factor	-76.41	-91.93
Refinement		
Initial model used (PDB code or source)		
SusC (Bt1763)	6ZAZ	6ZAZ
SusD (Bt1762)	6ZAZ	6ZAZ
Levanase (Bt1760)	6R3R	-
Levan binding protein (Bt1761)	Alphafold2	-
Model composition		
Non-hydrogen atoms	36,012	23,704
Protein residues	4480	2958
R.m.s deviations		
Bond lengths	0.003	0.001
Bond angles	0.499	0.398
Validation		
Molprobity score	1.65	1.39
Clashscore	7.37	7.09
Poor rotatmers (%)	0	0
Ramachandran plot		
Favoured (%)	96.37	98.58
Allowed (%)	3.56	1.42
Disallowed (%)	0.07	0

Table 2.6: Data acquisition parameters for the levan four-component complex with inactivated levanase in the presence of FOS DP15-25 and refinement statistics for models of the utilisome with a tethered conformation Bt1761, and of the $SusC_2D_2$ core.

2.9.10 CryoEM of the dextran utilising four-component complex

A sample of the dextran utilisome complex (Bt3087-Bt3090) solubilised in a DDM-containing buffer (10 mM HEPES, 100 mM NaCl, pH 7.5, 0.03 % DDM) was prepared at 0.05 mg/ml. Lacy carbon 300-mesh copper grids coated with a <3 nm continuous carbon film (Agar Scientific) were glow-discharged in air (10 mA, 30 s). A sample volume of 3.5 μL was applied to the grid. Blotting and plunge-freezing into liquid nitrogen-cooled liquid ethane were carried out using an FEI Vitrobot Mark IV (Thermo Fisher Scientific) with chamber conditions set at a temperature of 4 °C and 100% relative humidity. Blotting and plunge freezing were performed 10 seconds after loading the sample on the grid. Micrograph movies were collected on a Titan Krios Microscope (Thermo Fisher Scientific) operating at 300 kV with a Falcon IV direct electron detector operating in counting mode. Data acquisition parameters can be found in **Table 2.7**.

A dataset comprising 6,331 micrographs was collected and image processing was carried out in RELION 3.1 (Zivanov et al, 2018). Drift correction and dose-weighting were performed using RELION's own implementation of MOTIONCOR2. CTF estimation of motion corrected micrographs was done using CTFFIND4 (Rohou and Grigorieff, 2015). Particle picking was done using the crYOLO general model which identified 820,184 particles in the micrographs. Junk particles and contaminants were removed through several rounds of 2D classification, after which 477,707 particles remained. Particles were extracted at 2x down-sampling to reduce the computational demand early in the processing pipeline. An initial model was generated *de novo* from the data. Extensive 3D classification was used to address the considerable compositional heterogeneity that was present in the data (see **Table 2.8**). Each unique composition was refined and sharpened independently. Particles were later unbinned for a consensus refinement, and subject to iterative rounds of CTF-refinement and Bayesian polishing until no further improvement in resolution was seen. A final, sharpened reconstruction was obtained at 3.1 Å.

2.9.11 Model building of the levan utilising four-component complex

The sharpened reconstruction from the consensus refinement described above was of sufficient resolution for model building of the SusC subunit. An AlphaFold2 (Jumper et al.,

2021) predicted model was fit as a rigid body into the density and used as a starting point. The model was subjected to several iterations of manual refinement in COOT (Emsley and Kowtan, 2004) and 'real-space refinement' in Phenix (Adams et al., 2010). Molprobity (Chen et al., 2010) was used for model validation (**Table 2.7**).

To better understand the organisation of the dextran utilisome components, the map corresponding to the heptameric complex was used for rigid-body fitting of the lipoprotein subunits. X-ray crystal structures of SusD and the dextranase solved by collaborators at Newcastle University were docked into their corresponding densities. For the dextran binding protein, the two N-terminal domains of an AlphaFold2 (Jumper et al., 2021) predicted model were found to fit well to the density and as such were also docked. The C-terminal domain was truncated from this predicted model on the basis that density in the cryoEM map was of insufficient quality to allow confident fitting.

Data collection and processing	
Magnification	96,000 x
Voltage (kV)	300
Electron exposure (e-/Å^2)	38.78
Defocus range (μm)	-1.2 to -3.0
Pixel size (Å)	0.86
Initial particle images pre-3D classification (no.)	477,707
Final particle images post-classification (no.)	305,372
Symmetry imposed	C1
Map resolution (FSC 0.143)	3.1
Map sharpening B-factor	-79.2
Refinement	
Initial model used	AlphaFold2
Model composition	
Non-hydrogen atoms	13,616
Protein residues	1742
R.m.s deviations	
Bond lengths	0.001
Bond angles	0.407
Validation	
Molprobity score	1.29
Clashscore	3.52
Poor rotatmers (%)	0
Ramachandran plot	
Favoured (%)	97.23
Allowed (%)	2.66
Disallowed (%)	0.12

Table 2.7: Data acquisition parameters for the dextran four-component complex and refinement statistics for a model of the dimeric dextran transporting SusC (Bt3090). Map anisotropy owing to suboptimal angular distribution of particles meant density was of insufficient quality to build the SusD components.

Assembly	Particles (no.)	Resolution (Å)	B factor
Naked (SusCD alone)	30,107	6.7	-129
1x DBP	60,803	4.4	-149
2x DBP	68,748	4.4	-144
1x Dextranase	27,517	5.7	-129
1x Dextranase + 1x DBP (alt. sides)	16,158	7.37	-170
1x Dextranase + 1x DBP (same side)	27,484	6.1	-147
1x Dextranase + 2x DBP	43,766	4.6	-137

Table 2.8: Assemblies present in the dextran four-component complex dataset. Resolutions provided are for sharpened reconstructions.

2.9.12 Density analysis and figure making

Figures of electron micrographs were made using ImageJ (Schneider et al., 2012). Investigation and comparison of EM density maps was performed using Chimera (Pettersen et al., 2004) and COOT (Emsley and Cowtan, 2004). Figures of maps and models were generated using Chimera and ChimeraX (Goddard et al., 2018). Interface analysis was carried out using PDBePISA (Krissinel and Henrick, 2007).

Chapter 3 Structural and functional insights into oligopeptide acquisition by the RagAB transporter from *Porphyromonas gingivalis*

3.1 Introduction and build-up work

Our current understanding of the structure and function of SusCD-like transport systems is based almost exclusively on examples from the generalist glycan degrader *Bacteroides thetaiotaomicron* (*B. theta*). However, polysaccharide utilisation loci (PUL) encoding SusCD-like transporters are found throughout the Bacteroidetes phylum. An interesting example is the *ragAB* locus from the oral pathogen *P. gingivalis*. RagA is a predicted TonB-dependent transporter (TBDT) (PG_0185) and RagB is a surface-anchored substrate-binding lipoprotein (PG_0186). The two interact to form a complex localised to the OM, consistent with the characteristics of SusCD-like complexes (Nagano et al., 2007). However, unlike many human gut Bacteroidetes that specialise in the degradation of complex sugars, *P. gingivalis* is asaccharolytic, relying exclusively on the uptake and metabolism of peptides as a carbon source (Mayrand and Holt, 1988). Crucially, the mechanism of peptide uptake is unclear, but RagAB has been implicated in the process and is important for the fitness and virulence of *P. gingivalis in vivo* (Nagano et al., 2007; Shi et al., 2007).

P. gingivalis possesses a range of endo- and exopeptidases in order to digest extracellular proteins to oligopeptides (Potempa et al., 2000) and, later, di- and tripeptides (Nemoto and Ohara-Nemoto., 2016). The best characterised of these are the Lys- and Arg-gingipains, large and abundant surface-anchored cysteine endoproteases with cumulative trypsin-like activity. Arg-gingipains are essential in media where proteins are available but short peptides or amino acids are not, demonstrating their importance in generating accessible oligopeptide nutrients from protein substrates (Grenier et al., 2001). As the putative uptake machinery for these gingipain-generated peptides, RagAB appears to play a central role in nutrient acquisition by *P. gingivalis* (see **Figure 3.1**). It should however be noted that a contradictory role for RagAB in polysaccharide uptake has been proposed on the basis of an X-ray crystal structure of RagB (Goulas et al., 2016). The structure revealed considerable homology to known sugar-binding SusD-family members, leading Goulas and colleagues to attribute additional electron density to bound monosaccharides.

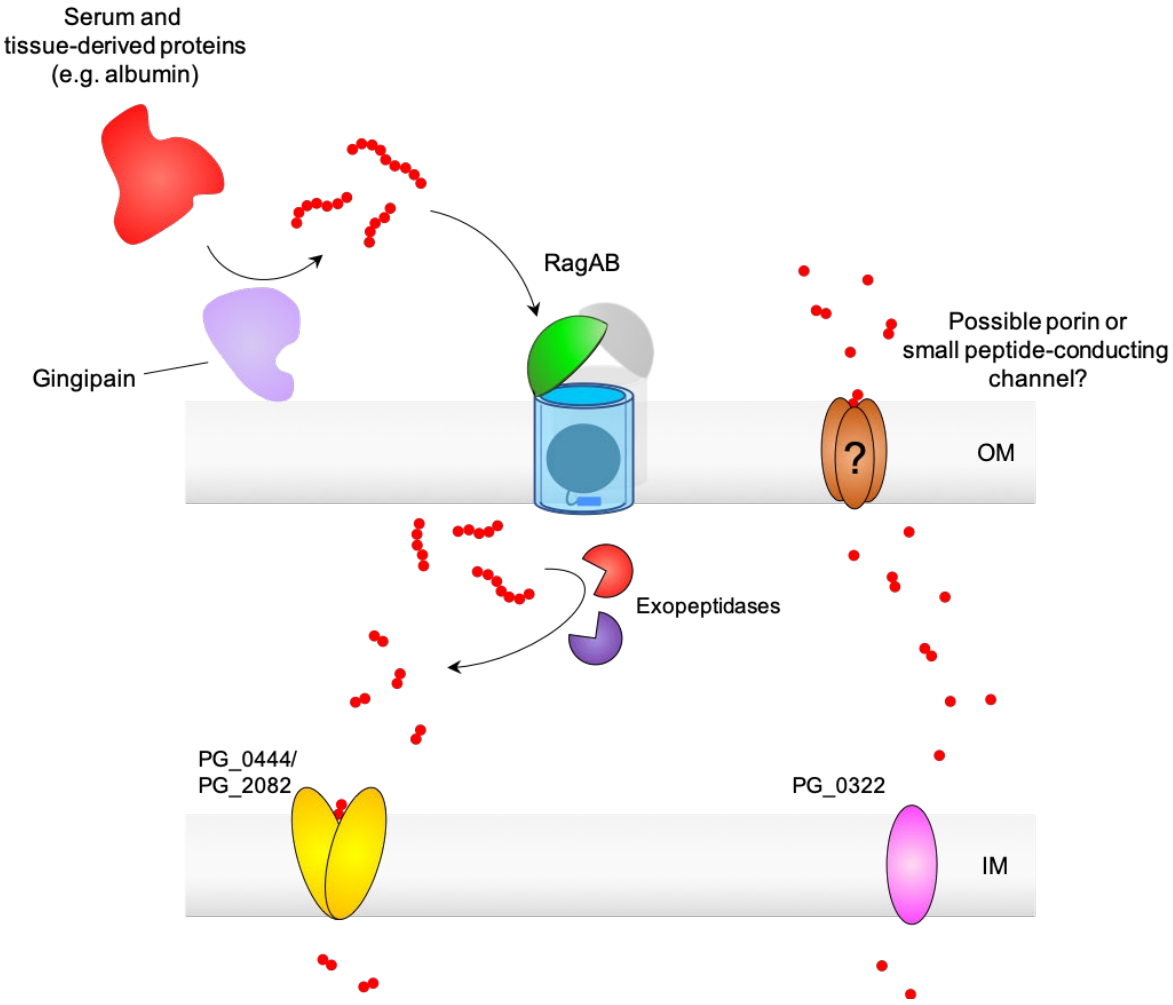


Figure 3.1 Schematic of peptide/amino acid acquisition by *P. gingivalis*. The action of gingipains on serum and tissue derived proteins generates oligopeptides that are proposed to be transported across the outer membrane by the RagAB complex. These peptides are targeted by an arsenal of exopeptidases (Nemoto et al., 2015), the cumulative action of which results in di- and tri-peptide production. Putative inner membrane peptide transporters (PG_0444 and PG_2082) allow translocation of these short peptides across the IM (Nelson et al., 2003). A functionally characterised sodium-ion driven serine/threonine transporter (PG_0322) permits assimilation of free serine/threonine amino acids (Dashper et al., 2001). Given that *P.gingivalis* is able to grow on very small peptides in a RagAB-independent manner, the existence of some small peptide-conducting porin/channel in the OM has been proposed.

In comparison to *B. theta*, *P. gingivalis* provides a simpler model system for the study of SusCD transporters, expressing RagAB as its only SusCD-like pair and just three paralogues of TonB (c.f. >90 and 10 respectively in *B. theta*). Moreover, *P. gingivalis* plays a critical role in the development of periodontitis (gum disease), and is therefore of high interest in its own right.

3.1.1 *P. gingivalis*: an aetiological agent in periodontal disease

Outside of the gut, the Gram-negative Bacteroidetes often cause disease. Two of the best characterised are the oral pathogens *Porphyromonas gingivalis* and *Tanneralla forsythia*, both of which have an important role in the development of periodontitis (gum disease). Together with *Treponema denticola* these bacteria constitute the so-called 'red complex', a group of bacteria routinely isolated together from periodontal pockets and a hallmark of periodontal disease (Socransky et al., 1998; Holt and Ebersole, 2005). The term 'red complex' resulted from early work which sought to implicate specific species, or groups of species, in the aetiology of periodontitis i.e. species present in disease-associated microbiota that are almost undetectable in health (Socransky et al., 1998). Since then, the disease model has been refined with more recent evidence suggesting that *P. gingivalis* facilitates the onset of disease indirectly (Hajishengallis and Lamont, 2012; Hajishengallis et al., 2012). By subverting the host's immune response, *P. gingivalis* promotes dysbiosis of the oral microbiota leading to chronic inflammation with consequent soft-tissue destruction and progressive bone loss at the periodontal pocket (periodontitis) (see **Figure 3.2**). The disproportionate effect of this organism on the bacterial community relative to its abundance has led to it being dubbed a 'keystone' pathogen in the onset and progression of periodontal disease. Notably, accumulating evidence suggests a link between this chronic inflammatory disease and others including: rheumatoid arthritis (Potempa et al., 2017), cardiovascular disease (Tonetti et al., 2007), chronic obstructive pulmonary disease (Usher and Stockley, 2013), Type-I diabetes (Preshaw et al., 2012) and even Alzheimer's disease (Dominy et al., 2019; Bui et al., 2019). Given these links, it is important to understand the molecular basis of nutrient acquisition that allows this oral pathogen to thrive below the gum line.

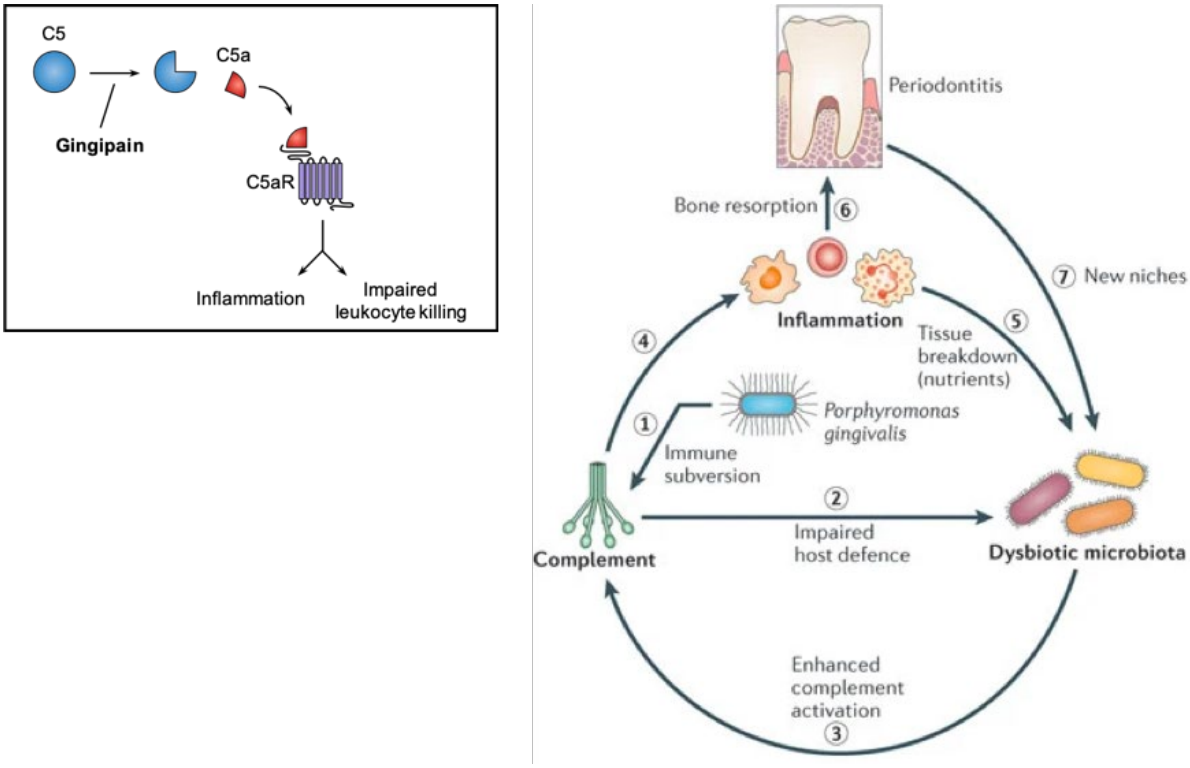


Figure 3.2 *P. gingivalis*-mediated dysbiosis of the oral microbiota. 1) *P. gingivalis* subverts the complement pathway leading to inflammation and impaired host defence. The inset (left) depicts the mechanism of complement subversion. Gingipains (cysteine-like endoproteases from *P. gingivalis*) possess C5 convertase-like activity and are capable of cleaving complement component C5, resulting in a local increase of the pro-inflammatory mediator C5a. Activation of the C5a receptor (C5aR) not only triggers an inflammatory response but is involved in subversive cross-talk with Toll-like receptor 2 (TLR2) (Wang et al., 2010). The downstream effect is impaired leukocyte killing via the cyclic AMP-protein kinase A pathway. 2) Impaired host defences allow unchecked growth of oral commensals resulting in further complement-dependent inflammation (steps 3 and 4). 5) Consequent tissue destruction provides additional nutrients that favour the growth of proteolytic and asaccharolytic bacteria, altering the demographics of the bacterial community. 6) Inflammatory bone resorption provides new niches (7), amplifying the disease state. Figure adapted from Hajishengallis et al., 2012.

3.1.2 Structural insights into RagAB provided by X-ray crystallography

Collaborators at Newcastle University purified the RagAB complex directly from *P. gingivalis* W83 KRAB (Δ*kgp*/Δ*rgpA*/Δ*rgpB*), a strain lacking gingipains used to reduce proteolysis of OM proteins. Aside from Hemaglutinin A (HagA), no co-purifying proteins were present. Diffracting crystals were obtained by hanging drop vapour diffusion and the structure was solved by molecular replacement to 3.4 Å (**see Figure 3.3**).

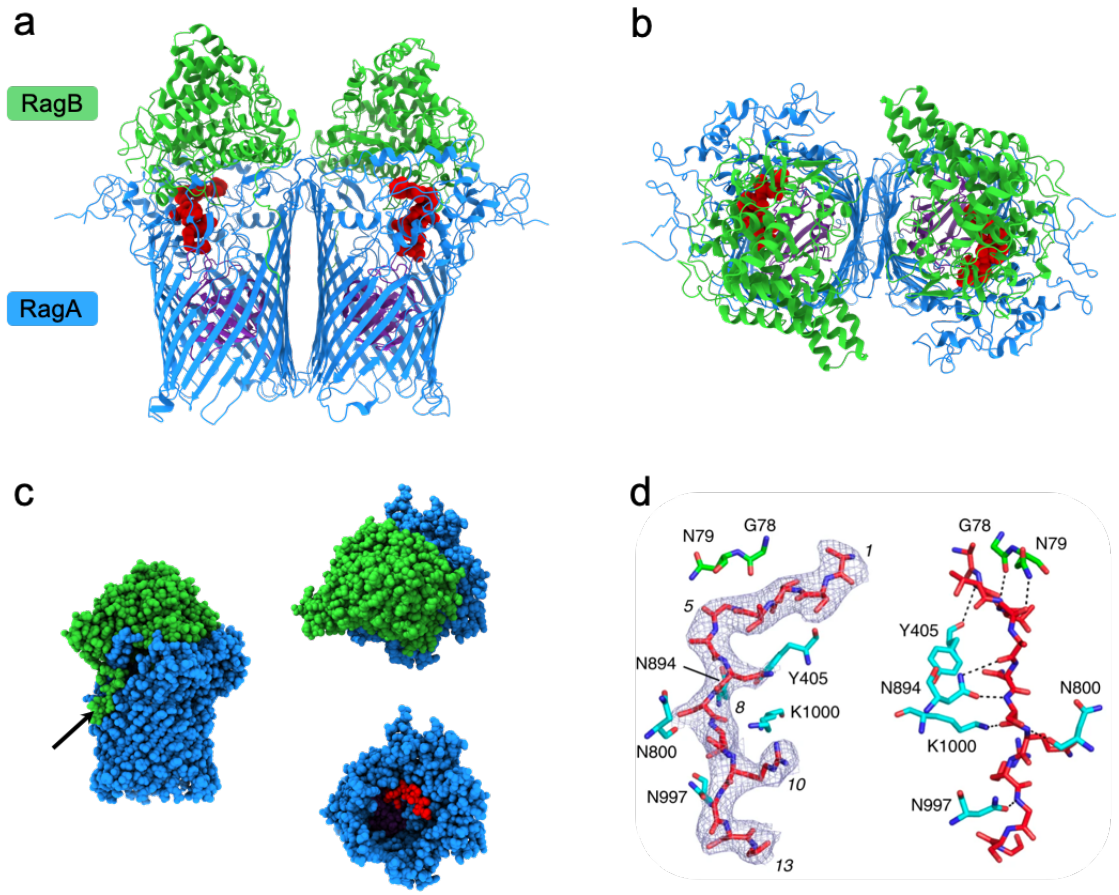


Figure 3.3 Crystal structure of RagAB from W83 (KRAB) *P. gingivalis*. a, RagAB complex viewed in the plane of the membrane. The peptide is shown as a red space-filling model and the plug doman of RagA is coloured purple. **b**, RagAB viewed from the extracellular space. RagB subunits clearly occlude the binding site of the substrate. **c**, Space-filling model of RagAB shown as a monomer for clarity. The lipid anchor of the RagB lipoprotein is indicated by an arrow. Views from the extracellular space show the complex with and without RagB. **d**, $2F_o$-F_c density (1.0 Å, carve = 1.8) of the peptide (left). Potential hydrogen bonds between RagAB and the backbone of the peptide are shown (right). X-ray crystal structures were determined by collaborators at Newcastle University.

The structure is dimeric, with RagA subunits tightly capped by their RagB counterparts, consistent with the architecture recently observed for other SusCD-like transporter complexes (Glenwright et al., 2017). The RagB subunits occlude the barrel interiors of RagA from the extracellular space and in doing so create a large solvent-excluded cavity. At this interface, there is an extended region ~40 Å long of well-resolved density distinct from that of either RagA or RagB (see **Figure 3.3**). The geometry of this density led to its assignment as

a peptide around 13 residues in length. Aside from putative residues 9 and 10, sidechain densities are truncated after Cβ or Cγ, and B-factors for the main chain are slightly higher than for the adjacent RagA. It is therefore hypothesised that the observed density corresponds to an ensemble of peptides, possessing a common backbone conformation, which will be referred to as 'the peptide' throughout.

The peptide, arbitrarily modelled as $A^{1}STTG^{5}ANSQR^{10}GSG^{13}$, possesses a defined ~90 ° kink, indicating stable binding for the region that is visible. Whilst it is not possible to comment on specific side-chain interactions due to sequence ambiguity, the model indicates potential hydrogen bonding between residues in RagAB and the backbone of 8 of the 13 modelled peptide residues. The majority of interactions are contributed by RagA. The carbonyls of Y405 and N800 form hydrogen bonds with peptide residues 3 and 9 respectively. The sidechain of N894 may be important for substrate binding as it forms hydrogen bonds to both residues 6 and 8 of the peptide, while the sidechains of N997 and K1000 likely interact with peptide residues 11 and 8 respectively. From RagB, the carbonyl of G78 and the side chain of N78 interact with peptide residues 4 and 5 respectively.

The greater contribution of RagA to substrate binding is emphasised by the results of a PISA interface analysis of the complex. RagA contributes 26 peptide-interfacing residues compared with only 6 from RagB, corresponding to interface areas of 620 and 240 $Å^2$ respectively. The CSS (complexation significance score), defined as 'the maximal fraction of the total free energy of binding that belongs to the interface', indicates how significant an interface is for assembly of the complex. The RagA-peptide interface scores 1.0 (maximal) whilst the RagB-peptide interface scores only 0.014. These results indicate that the RagA-peptide complex is likely stable in the absence of RagB, and, in the context of the 'pedal-bin' model, suggests that the observed co-crystal structure may correspond to a state in which the substrate has been passed from a low(er) affinity binding site on RagB to a high(er) affinity binding site on RagA.

Not all of RagA is visible in the crystal structure. Residues N-terminal of Leu115, including the TonB box and an N-terminal extension (NTE) (annotated in the KEGG database as a carboxypeptidase-D regulatory-like domain), are missing, perhaps owing to increased mobility in these regions. Indeed, evidence suggests that, at least for some classical TBDTs,

the substrate-loaded state of the transporter is associated with greater mobility in the region of the TonB box relative to apo states (Fanucci et al., 2003; Kim et al., 2007).

Crucially, the question 'How does the substrate access the binding site?' remains unanswered. As discussed in **section 1.8.1**, the working hypothesis is that SusCD-like systems function by a 'pedal-bin' mechanism of action. However, in the crystal structure of BT1762-BT1763, the SusD subunits occlude the substrate binding site from the extracellular space, despite the absence of substrate, which is problematic for a model where closure is promoted by substrate binding. This discrepancy has been attributed to the preference for less hydrated and more compact conformations, i.e. 'lids closed', in crystal packing. Therefore, in this chapter, cryo-electron microscopy was employed as a complementary structural technique to probe this hypothesis and understand the structural changes necessary for substrate capture.

3.2 Assessing sample quality by negative stain TEM

RagAB solubilised in DDM (3.5 mg/ml) was provided by collaborators at Newcastle University. Sample quality was assessed using negative-stain transmission electron microscopy (TEM) (see **Figure 3.4**). An 800-fold dilution of the stock sample was necessary to provide a useful distribution of particles on the grid. Particles are clearly visible, homogeneous and monodisperse. Particles are preferentially oriented 'top-down' (perpendicular to the plane of the membrane) such that the characteristic dimer of β-barrels is identifiable, although side-on views can also be seen. The particles are around 10 nm in diameter, consistent with the crystal structure.

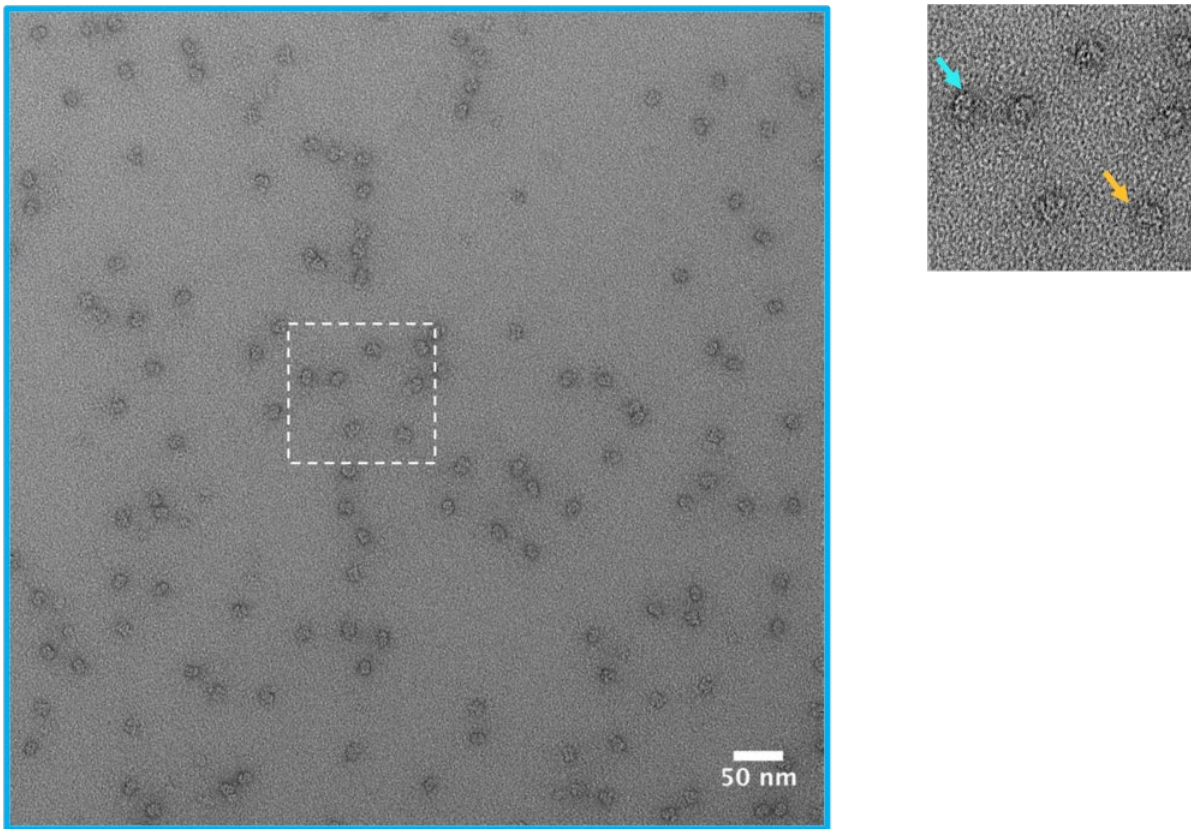


Figure 3.4 TEM micrograph of uranyl acetate stained RagAB solubilised in DDM at 50k x magnification. Particles are homogeneous and monodisperse. The zoomed inset corresponds to the area within the dashed box. Two distinct views of the RagAB complex are easily identifiable and marked by orange and blue arrows. These correspond to views in the plane of the membrane and perpendicular to the membrane respectively.

3.3 Reconstitution of RagAB into nanodiscs

Whilst detergents are an excellent way of stabilising membrane proteins after their extraction from the native lipid environment, the functionality of these proteins is often compromised. In many cases, membrane protein integrity and/or functionality is dependent on native protein-lipid interactions (Dawaliby et al., 2016; Gupta et al., 2017; Laganowsky et al., 2014; Martens et al., 2016; Martens et al., 2018). These are often lost upon extraction with detergent, compromising protein function. In addition, NMR studies on OMPs have shown that the dynamics of these proteins are suppressed in detergent micelles compared to systems that mimic lipid-bilayers (Frey et al., 2016). Nanodiscs are a popular choice for this mimicking role, with >169 cryoEM structures of nanodiscs-stabilised membrane proteins deposited in the EMDB to date (July 2020).

Nanodiscs are circular patches of lipid bilayer enveloped by two amphipathic alpha-helical proteins known as membrane scaffold proteins (MSPs) (Denisov and Sligar, 2016) (see **Figure 3.5**). MSPs possessing helical regions of various lengths are available allowing the size of the disc to be tailored to specific applications. Detergent-solubilised membrane proteins can be reconstituted into nanodiscs according to the protocol described in **section 2.7.2**. Briefly, a detergent solubilised protein of interest is mixed with MSP and a lipid/lipid mixture of choice. Detergent is then removed using Bio-Beads™ and a self-assembly process ensues. Successful reconstitution requires these components to be mixed in the correct ratio; estimated according to **Equation 1**. Given the goal of trying to capture RagAB complexes in what are proposed to be highly dynamic states (Glenwright et al., 2017), nanodiscs were an attractive choice.

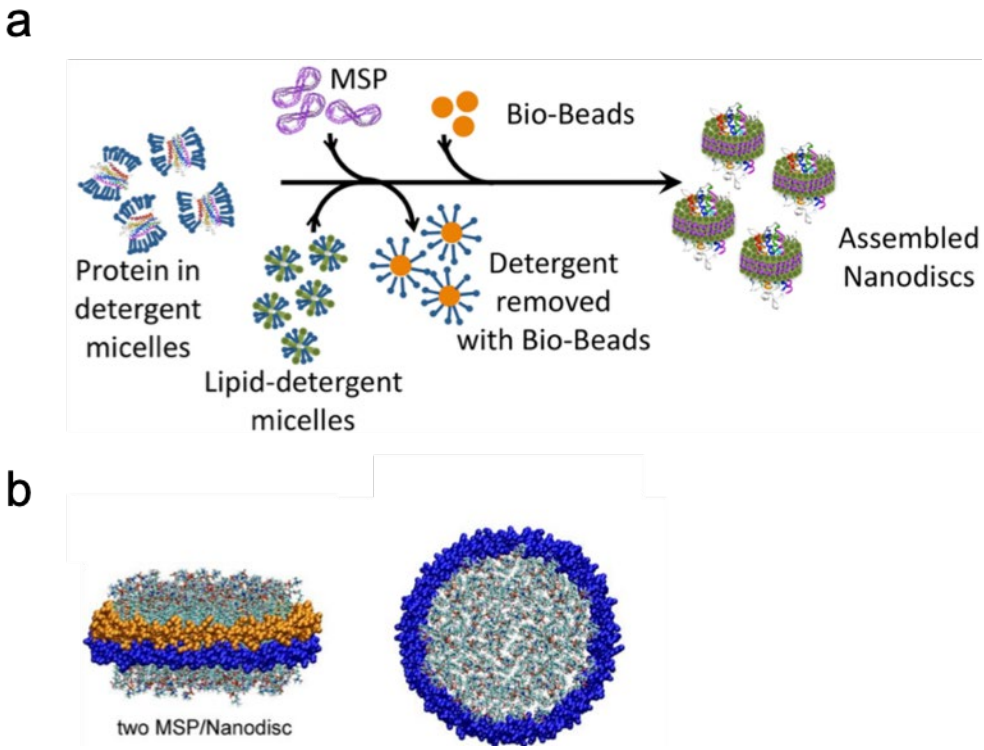


Figure 3.5 An overview of nanodisc technology. a, Schematic of nanodiscs assembly for reconstitution of detergent solubilised membrane proteins. The protein of interest is mixed with MSP and a lipid detergent mixture. Bio-Beads are added to remove the detergent and facilitate MSP-bordered bilayer formation. **b**, Illustration of an empty nanodiscs showing side and top views. The two MSPs are distinguished (orange and blue). This figure was adapted from Bayburt and Sligar, 2010.

3.3.1 Reconstitution of RagAB into MSP1E3D1 discs

Given the dimensions of the RagAB complex (~85 x 50 Å), MSP1E3D1 was selected as the most appropriate MSP. This MSP forms discs with a diameter of 13 nm (± 1) which should comfortably accommodate the dimeric RagAB complex. MSP1E3D1 was mixed with *E. coli* polar lipid extract and RagAB such that the molar ratio was 1:2:160 (RagAB dimer : MSP : lipid). Negative stain TEM was used to characterise the nanodisc preparation.

As shown in **Figure 3.6a**, particles were heterogeneous and large aggregates were also observed, indicating poor reconstitution into discs. The same results were seen upon repeating this preparation. A discussion with colleagues revealed that the addition of a small amount of glycerol (2% v/v) was reported to reduce aggregation. A third reconstitution including glycerol was therefore trialled and a representative micrograph is shown in **Figure 3.6b**. Particle homogeneity looked to be marginally improved and aggregation reduced but not eliminated. A small negative stain data set of 100 micrographs was collected from which 3023 particles were picked and processed using RELION 2.1 (Scheres 2012). The 2D class averages indicated partial reconstitution into nanodiscs. Some classes (**Figure 3.6c**) appear to show RagAB dimers surrounded by additional circular density that could be attributed to the nanodisc. RagAB tetramers were also identified (not seen in detergent) and these too show signs of circular density encompassing the protein. Collectively, these dimeric and tetrameric complexes constitute ~25% of picked particles and, given the small dataset and limited range of views, generating a 3D reconstruction was not possible.

Notably, a class possessing an 'hourglass' morphology was identified in the data (**Figure 3.6c**). Structurally distinct from RagAB, this class may represent a population of naked MSPs formed as a result of insufficient lipid in the reconstitution mixture. In the absence of lipid, the hydrophobic stretches of the scaffold proteins plausibly self-associate, giving rise to the twisted structure observed.

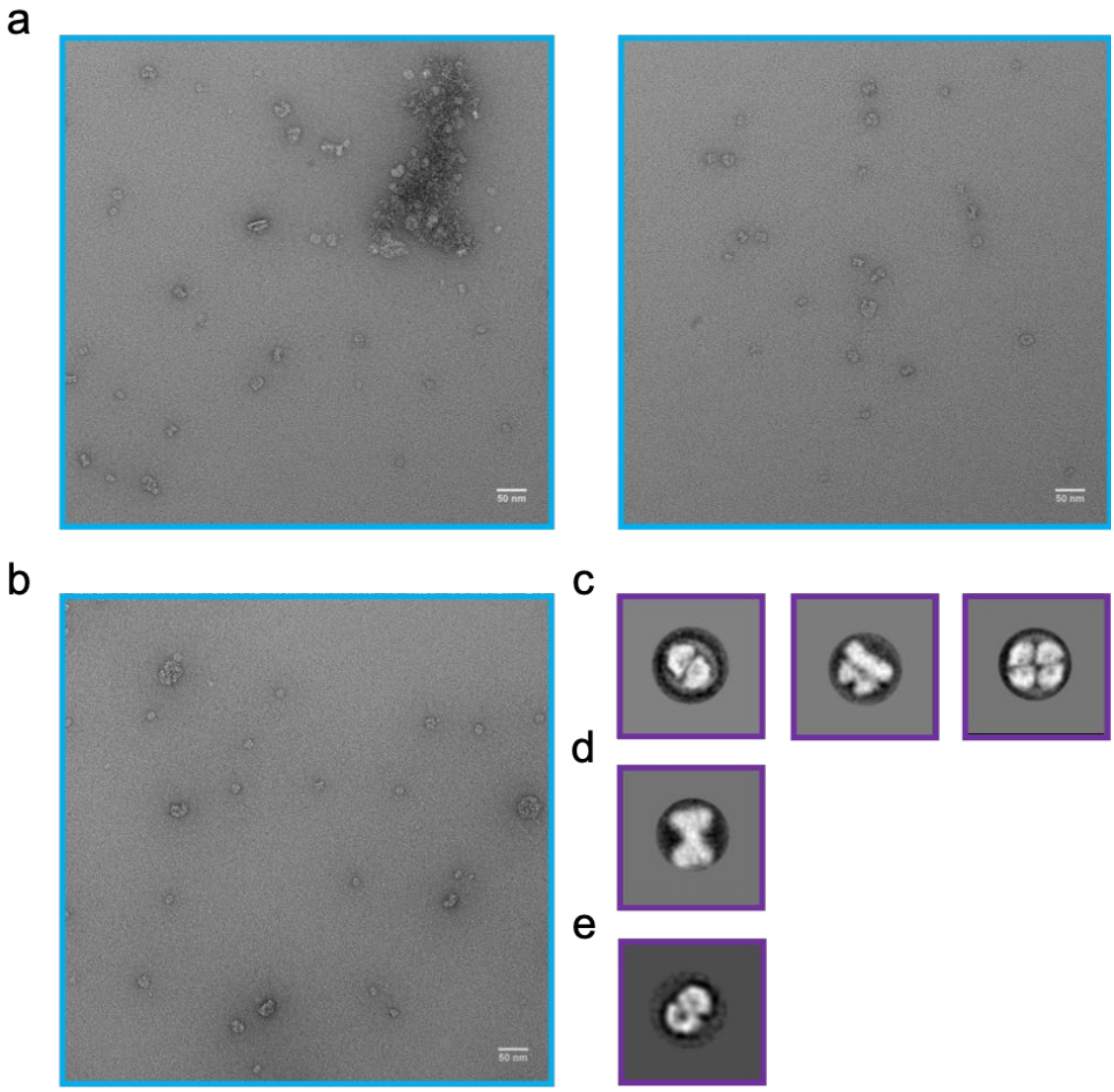


Figure 3.6 Negative stain data on RagAB MSP1E3D1 preparations at 50k x magnification. a, Representative micrographs of the RagAB MSP1E3D1 preparation. Particles are heterogeneous and large aggregates were present (left). **b**, Representative micrograph of a RagAB MSP1E3D1 preparation containing 2% glycerol. Small aggregates are still present but large aggregates like those seen in (**a**) were not observed. **c**, 2D class averages displaying evidence of disc density, generated using particles from the RagAB MSP1E3D1 preparation containing 2% glycerol. Dimeric and tetrameric RagAB complexes are clearly visible encompassed by a halo of density not observed in 2D classes of detergent-solubilised RagAB (**e**). **d**, A class average of particles with an hourglass morphology suspected to represent naked MSPs. Scale bars are 50 nm.

Despite these sub-optimal results in negative stain, a small cryoEM dataset was collected (772 micrographs). 2D classification of the extracted particles showed no evidence of disc density (see **Figure 3.7**). Some class averages show the expected dimer of RagAB, presumably partially stabilised by the presence of lipid or residual detergent. Others show higher order aggregates formed by association of the hydrophobic barrel exteriors. Large concatemers were also visible, even from the raw micrographs. 3D reconstructions were unsuccessful owing to the sub-optimal distribution of viewing angles.

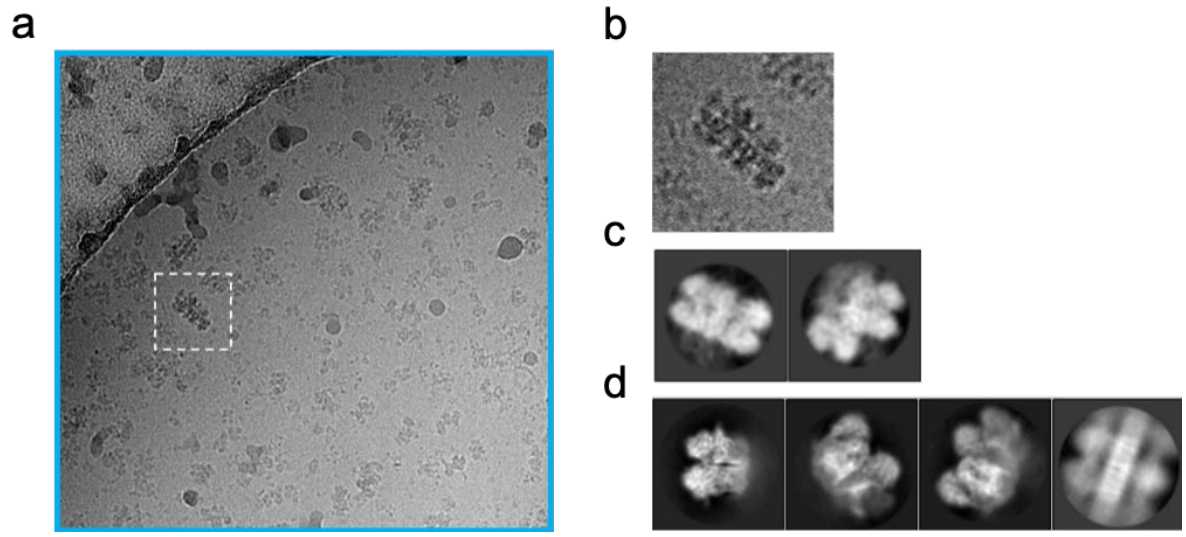


Figure 3.7 CryoEM of RagAB MSP1E3D1 preparation with 2% glycerol. a, Representative micrograph at 130k x magnification. Particles look heterogeneous and large concatemers are visible (highlighted by the dashed region). **b**, Zoomed inset of area marked in (**a**). **c**, Manual picking and 2D classification of larger concatemers show that they are formed by end-to-end association of RagAB. **d**, Representative 2D class averages from auto-picked particles. Dimeric RagAB is clearly visible along with higher order aggregates.

3.3.2 Reconstitution of RagAB into MSP2N2 discs

MSP2N2 is a membrane scaffold protein that facilitates the assembly of larger nanodiscs (~16 nm diameter). Work by Gao et al. (2016) on the ion channel TRPV1 involved reconstitution of this tetrameric protein into both MSP1E3D1 and MSP2N2 nanodiscs. Despite both preparations showing positive results by negative stain TEM, only the latter was taken forward to cryo-EM studies. Since TRPV1 and dimeric RagAB possess similar footprints in a bilayer, it was proposed that the larger disc may permit better reconstitution.

A pET 28a vector encoding a His-tagged variant of MSP2N2 was transformed into *E. coli* BL21 (DE3) cells, expressed, and purified by affinity chromatography. Ratios of components in the reconstitution mixture were selected based on the optimal ratios reported for TRPV1 (1:1:150 -1:1.5:225) (Gao et al., 2016). Reconstitutions were attempted at the upper and lower ends of this range with the former providing the most promising results (as determined by negative stain screening) (see **Figure 3.8**). A small stain dataset (144 micrographs) was collected from which 6775 particles were picked and processed using RELION 2.1 (Scheres 2012); 2D class averages are shown in **Figure 3.8**. Evidence of discs is arguably more convincing than for MSP1E3D1 nanodiscs, with classes containing clear and recognisable signal for both dimeric RagAB and the surrounding disc structure accounting for ~20% of the data. Again, attempts to generate a 3D reconstruction were unsuccessful. Despite this, another small cryoEM data set was acquired (939 images) but, as for the MSP1E3D1 preparations, packing of hydrophobic RagA barrels facilitated aggregate formation (data not shown).

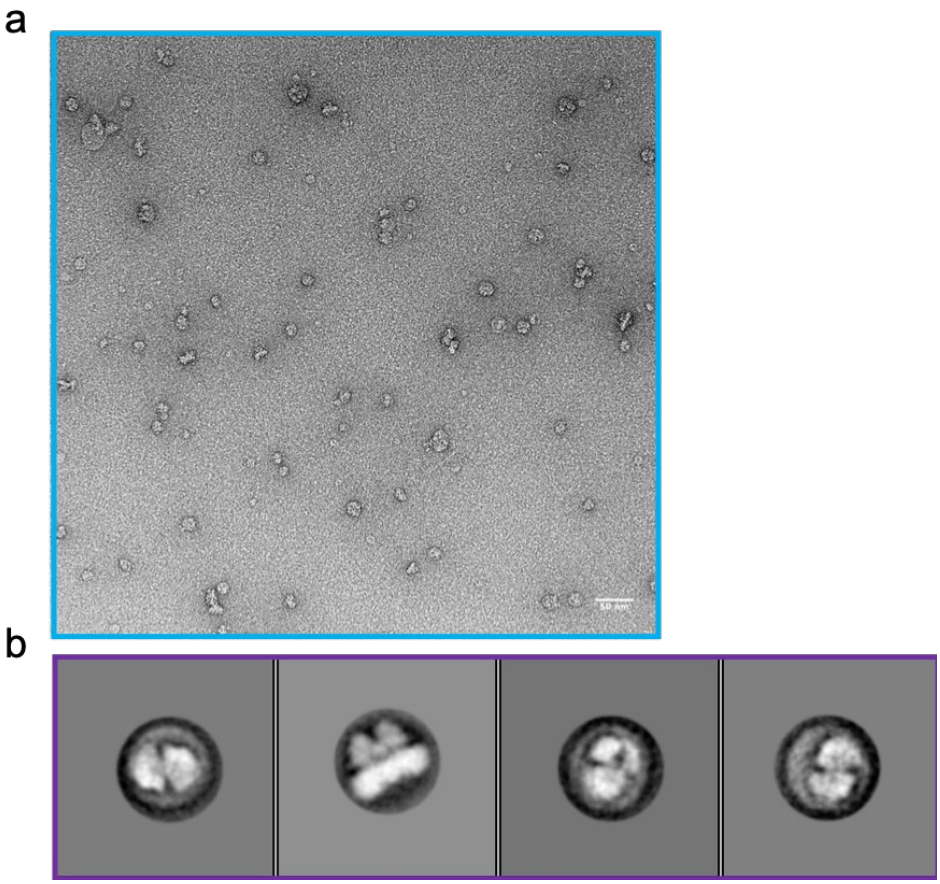


Figure 3.8 Negative stain TEM of RagAB MSP2N2 preparation. a, Representative micrograph of RagAB MSP2N2 preparation at 50k x magnification. Particles are slightly heterogeneous with small aggregates also visible. **b**, 2D class averages displaying evidence of disc density. Clear halos of density are visible around the dimeric structure of RagAB. Scale bar is 50 nm.

3.4 Single particle cryo-electron microscopy of RagAB solubilised in DDM

In addition to work with nanodiscs, structural studies of RagAB solubilised in DDM were carried out using single particle cryoEM. An overview of screening for best ice quality is shown in **Figure 3.9**. Cryo-EM data showed monodispersed, intact particles (see **Figure 3.10**), and a dataset comprising 3605 images was collected. After motion correction and CTF estimation, in-house scripts were used to remove micrographs with defocus values greater than 4 μm and estimated resolutions poorer than 5 Å (Thompson et al., 2018). The remaining 2843 micrographs were processed according to the RELION v2.1 pipeline (Scheres, 2012) (**Figure 3.10**). 2D classes showed evidence of secondary structure, indicative of high-quality data.

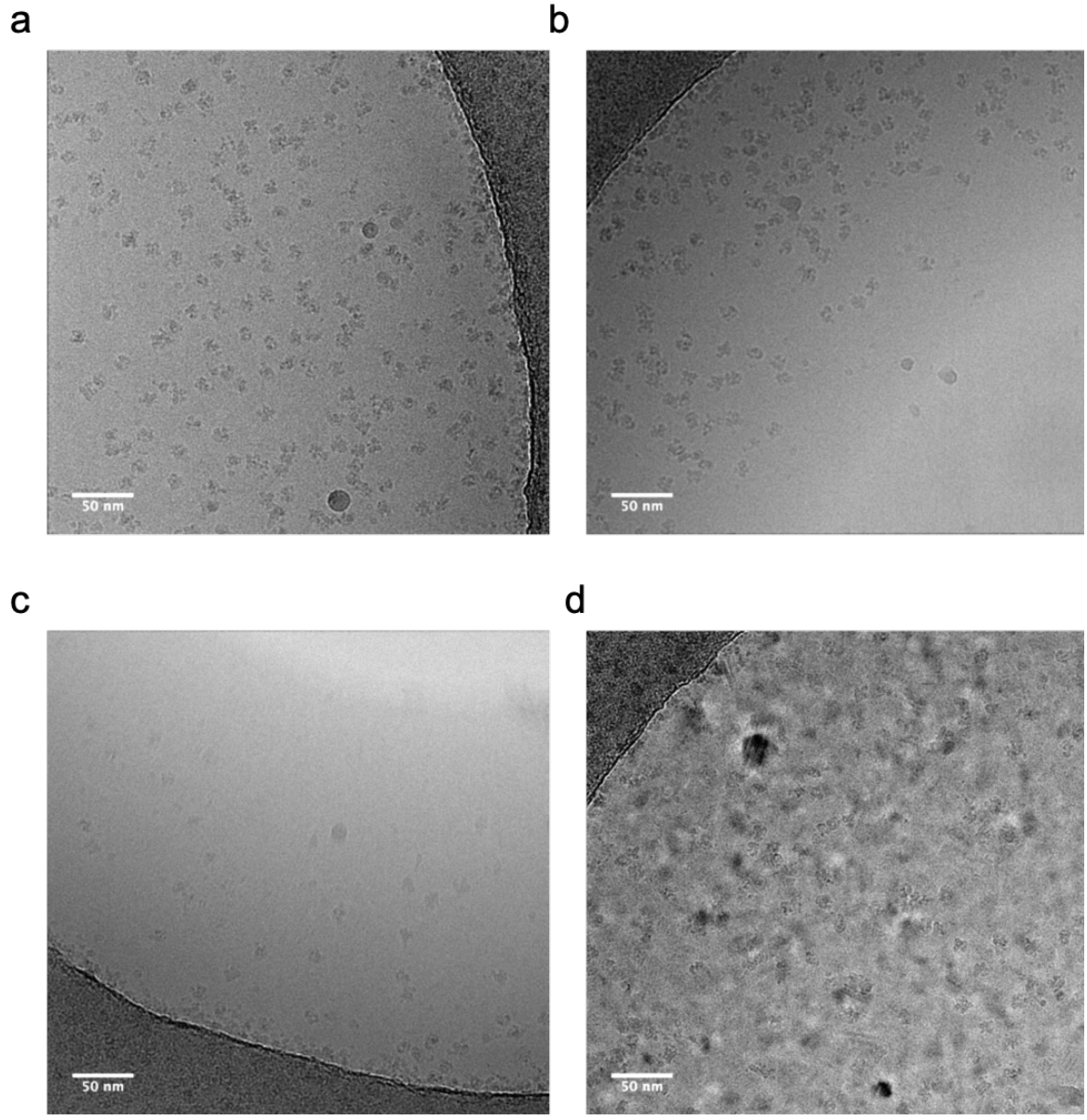


Figure 3.9 Representative micrographs showing the relationship between ice thickness and particle distribution. a, Good ice thickness showing a good distribution of particles adopting a range of orientations. **b**, Thinner ice occurs towards the centre of the hole resulting in physical exclusion of particles from the centre and crowding at the hole edge. In most cases these regions were still suitable for data collection. **c**, Ice that is too thin in the centre of the hole. Exposure to the electron beam results in deformation or a 'blow-out' causing excessive motion in the micrograph. **d**, Non-vitreous ice. Areas in **c** and **d** are not suitable for data collection. Scale bars are 50 nm.

3.4.1 Structural evidence in support of hinge-like opening of RagB 'lids'

An initial model was generated *de novo* from the data by stochastic gradient descent (SGD). This model was used as a reference for unsupervised 3D classification, after which three unique conformational states of the complex were observed. One of these was consistent with the previously described crystal structure in which both RagA subunits are tightly capped

by their RagB counterparts. However, the remaining classes correspond to novel conformational states in which either one or both RagB lids occupies an open position. Independent refinement of these classes yielded 3.7 Å, 3.7 Å and 3.9 Å reconstructions for the closed-closed (CC), open-closed (OC) and open-open (OO) states respectively (resolutions are for post-processed maps). The CC and OO states possess C2 symmetry and this was imposed in generating the final reconstructions.

The OC state was predominant, accounting for 47% of the particles used in the final reconstructions. CC and OO states represent 37% and 16% respectively. Intriguingly, additional rounds of 3D classification provided no evidence of intermediate states i.e. conformations falling between lid open and lid closed. This perhaps suggests that the energy landscape for opening and closing is very rugged, with stable minima existing only for the observed states.

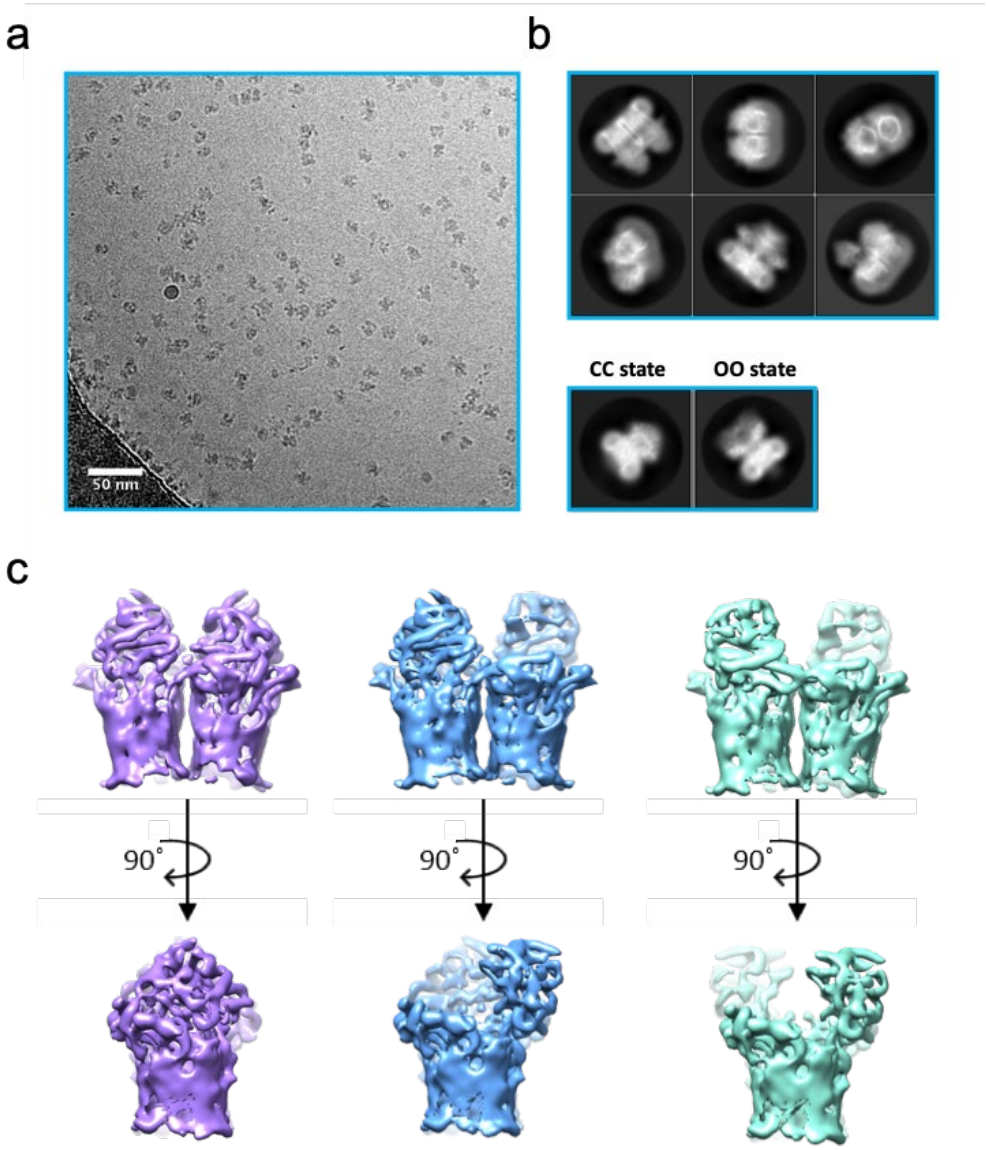


Figure 3.10 Results of 2D and 3D classification for RagAB solubilised in DDM. a, Representative micrograph demonstrating good particle distribution and concentration for the RagAB complex in ice. **b**, Representative 2D class averages. The plug domain can be observed within the TM barrels of RagAB. Conformational heterogeneity is identifiable even at this stage with clear open-open (OO) and closed-closed (CC) separated into distinct 2D classes. **c**, Output of unsupervised 3D classification. Three conformational states are identifiable. A closed-closed state consistent with the crystal structure (purple), an open-closed state (blue) and an open-open state (green).

The original micrograph movies were later motion corrected in RELION 3.0 (Zivanov et al., 2018). Particles contributing to the final reconstructions were re-extracted from their respective micrographs. Following reconstruction using 3D auto-refinement, iterative rounds of per-particle CTF refinement and Bayesian polishing were employed. The final, sharpened maps for the CC, OC and OO states were resolved to 3.3 Å, 3.3 Å and 3.4 Å respectively (see

Figure 3.11). Individual β-strands of the RagA barrels are easily discernible at these resolutions, as is side-chain density for bulky aromatic residues.

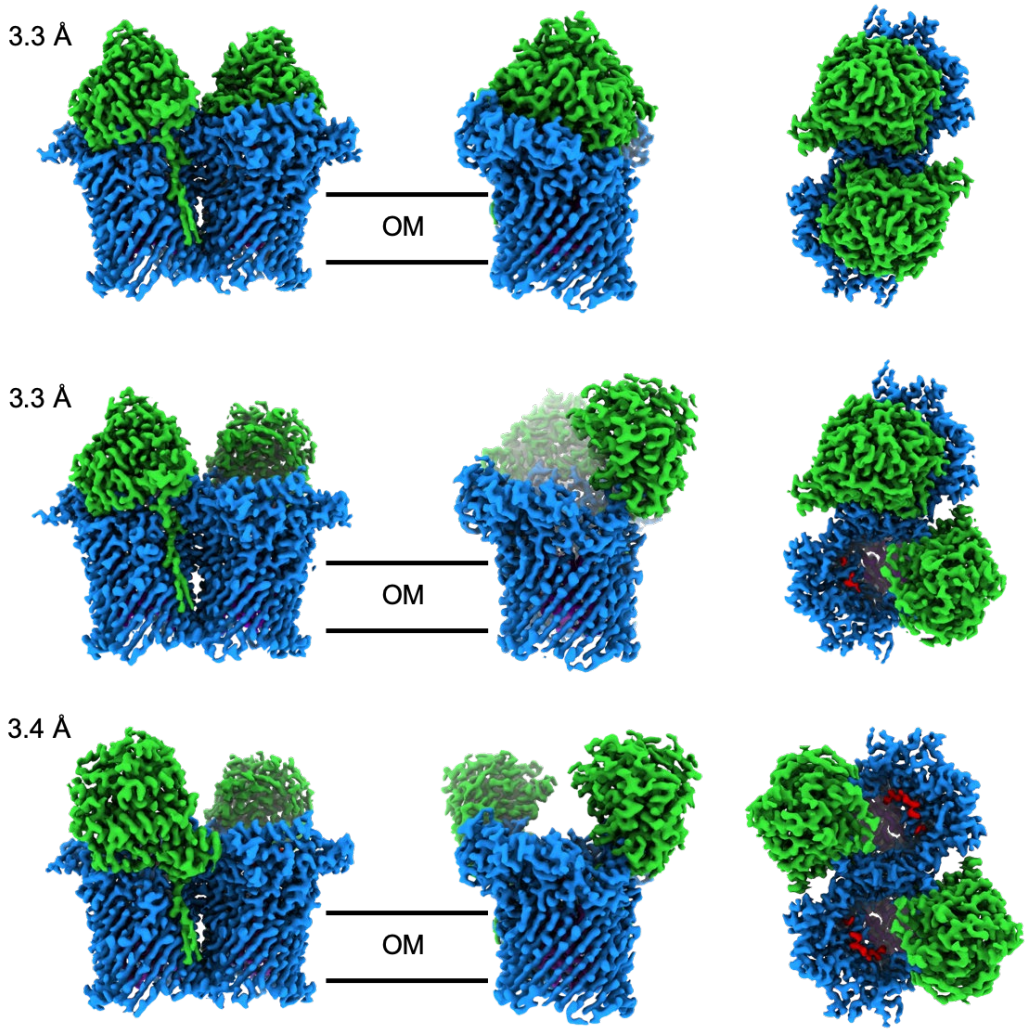


Figure 3.11 Different conformations of the $RagA_2B_2$ transporter revealed by cryoEM. Final reconstructions of the closed-closed (top), open-closed (middle) and open-open (bottom) states of the $RagA_2B_2$ transporter. Maps have been filtered by local resolution and coloured to match Figure 3. Views in the plane of the OM (left and middle panels) and from the extracellular space (right panel) are shown.

The crystal structures of the RagA and RagB subunits were fitted as rigid bodies into the EM density and subjected to multiple rounds of real space refinement in COOT and Phenix. The closed-closed state is almost identical to the crystal structure with backbone rmsd values of ~0.6 Å. Comparing the open and closed states of the transporter, RagB moves as a rigid body through a complex arc-like trajectory. It pivots about a hinge region close to its N-terminus whilst rotating around an axis through the RagB subunit. The movement is substantial with

displacements of up to 45 Å for main-chain atoms located furthest from the hinge region (see **Figure 3.12**). Whilst most RagA-RagB interfacing interactions are lost upon opening of the RagB lid, interactions with extracellular loops L7 and L8 of RagA are retained (designated hinge-loops 1 and 2, respectively). As expected, these loops also undergo substantial conformational changes with main chain atom movements of up to ~13 Å (see **Figure 3.12**).

The observed opening of the RagB lid exposes the binding site and neatly substantiates predictions derived from MD simulations of other SusCD-like transporters (Glenwright et al., 2017). The open-closed state of the transporter is particularly interesting as the movement of a single lid breaks the C2 symmetry of the molecule, suggesting that the two transport hubs may function independently. Such independent lid opening was also predicted by MD simulations.

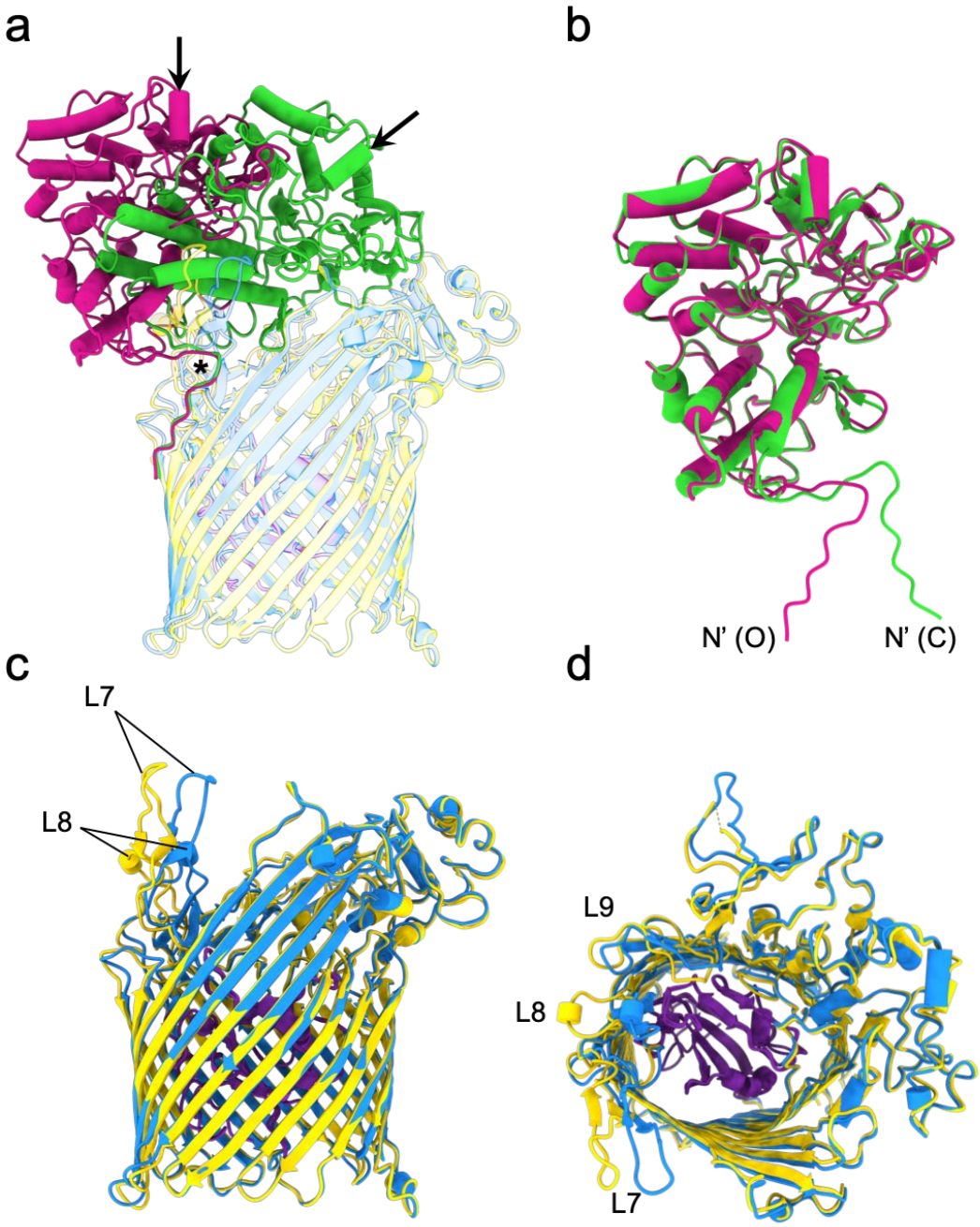


Figure 3.12 Opening of the RagB lid. a, Superposition of the open (yellow and pink) and closed (blue and green) states of the RagAB complex aligned on RagA. Arrows highlight the equivalent helix in the open and closed position of the lid. The hinge region towards the N-terminus of RagB is indicated with an asterisk. **b**, Superposition of RagB subunits from the open (pink) and closed (green) state of the transporter. The overall fold of RagB is unchanged, demonstrating that this subunit moves as a rigid body. The N-termini are labelled **c**, Superposition of the RagA subunits shown in the absence of RagB. Conformational changes of the hinge-loops L7 and L8 associated with lid opening can be observed. **d**, As in (**c**) but viewed from the extracellular space.

3.4.2 Assessing occupancy of the substrate binding sites

In agreement with the crystal structure, density corresponding to bound substrate was evident in the closed-closed state of the transporter (see **Figure 3.13**). Substrate density was also observed in the closed side of the single open structure. In both cases, substrate density was of comparable strength to that of the surrounding protein backbone, suggesting near complete occupancy of a rigidly-held conformation. The density resembles a peptide of ~13 residues in length that possesses a 90 ° kink, as was the case in the crystal structure. Whilst density for the backbone of the bound peptide is strong, sidechain density is poor and provided little insight into the nature of the sequence. However, given its putative function in nutrient scavenging, it is conceivable that RagAB does not exhibit stringent sequence-specific peptide binding behaviour. Instead, an ensemble of bound peptides likely contributes to the observed density, in which case well-defined sidechain density would not be expected. Similarly, Glenwright and colleagues (2017) observed poor sidechain density for a putative peptide substrate bound by the Bt2763-Bt2764 SusCD transporter, and this too was attributed to an ensemble of peptides being bound.

Strikingly, substrate density was also observed in the open sites of the transporter, albeit weaker than in the closed sites (see **Figure 3.13**). The density is most prominent in either site of the OO state where it unambiguously adopts the same ~90° turn as observed in the closed sites. Intriguingly, signal corresponding to substrate is much weaker in the open site of the OC state, and is almost invisible at a threshold appropriate for viewing the RagAB backbone. Assuming the two halves of the transporter function independently, it would be expected that substrate density in the open sites should be comparable between the OO and OC structures. The observed discrepancy suggests that binding may exhibit some form of cooperativity.

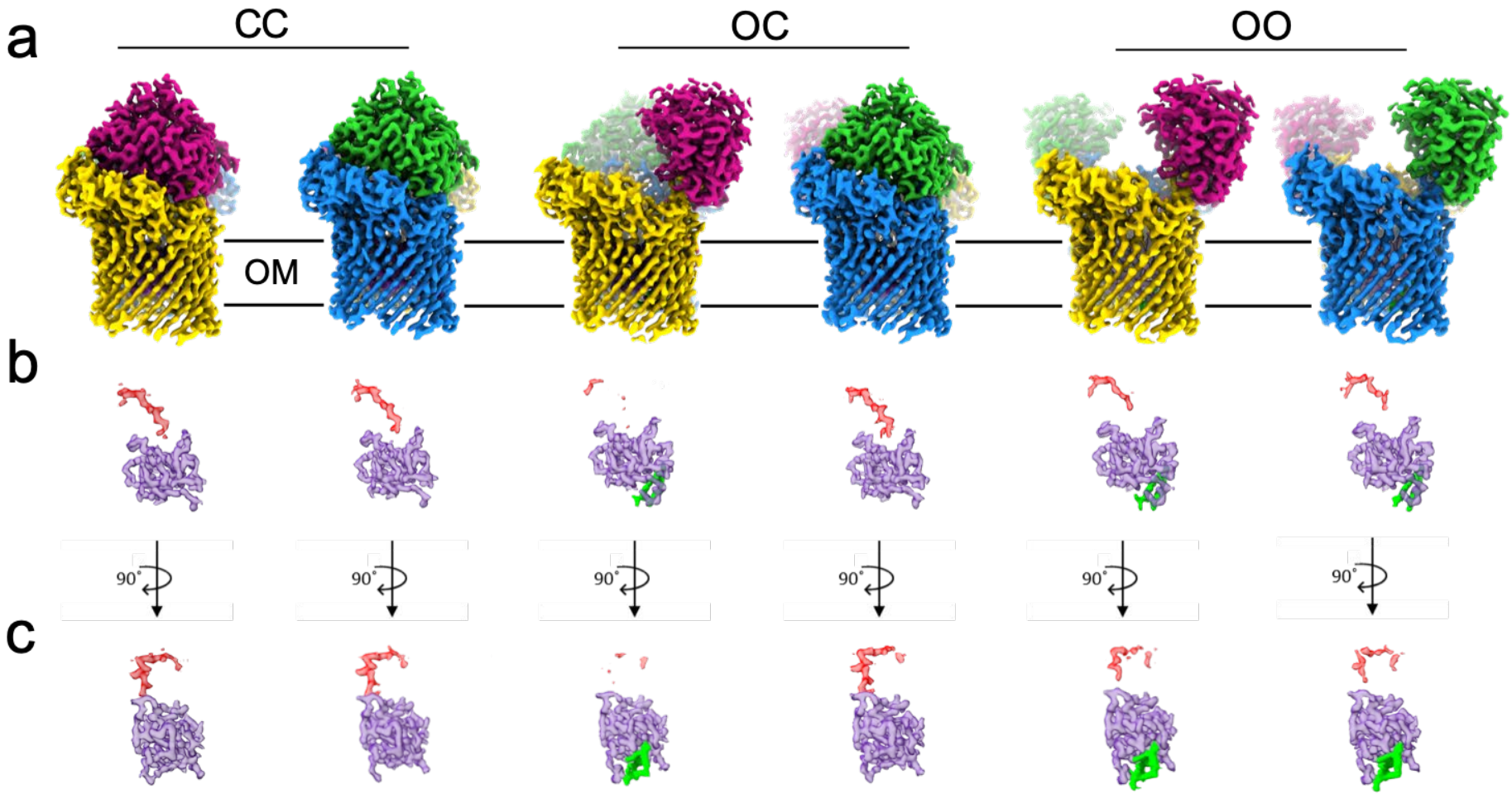


Figure 3.13 Variations in plug and substrate density between the different conformational states of RagAB. a, Density maps for the closed-closed, open-closed and open-open states of the RagAB transporter filtered by local resolution (top). **b,** Isolated density for the plug domain and bound peptide viewed as in (**a**) and with a 90°clockwise rotation (**c**). Additional density unique to plug domains of open RagAB complexes is shown in green. Densities corresponding to plug domains and peptides from the closed RagAB transporters are very similar, irrespective of whether the partner complex is open or closed.

3.4.3 The open-closed complex reveals changes in the RagA plug domain

Current models for TonB-dependent transport propose that substrate binding induces a conformational change that is propagated through the plug domain and results in a disordering of the TonB box at the periplasmic face of the transporter, exposing it for interaction with TonB (Kim et al., 2007). The disordered state of the Ton box can therefore be considered as a signal communicating substrate binding across the membrane. This would ensure that only substrate-loaded transporters form productive complexes with TonB, avoiding futile transport cycles. In crystal structures of TBDTs, the conformation of the Ton box varies. Where visible, this region interacts with other parts of the plug domain and is assumed to be inaccessible to TonB. In other structures, electron density for the Ton box is absent, indicating increased mobility. Notably, the correlation between ligand binding and visibility of the Ton box is poor (Noinaj et al., 2010) and is likely influenced by crystallisation reagents as well as the crystal lattice itself (Fanucci et al., 2003; Kim et al., 2007; Freed et al., 2010). In contrast, the open-closed state of the transporter observed by cryo-EM allows the relationship between occupancy of the binding site and the conformation of the plug domain to be examined outside of a crystal lattice. Considering this, we looked for differences at the N-terminus of RagA in the closed versus the open conformation.

Despite the weak substrate density observed in the open states, lid closure likely represents an essential step in the translocation process. As such the OC state permits an unbiased structural comparison of two conformational snapshots of the translocation cycle. Alignment reveals that the overall architecture of the plug is not grossly affected by lid closure (see **Figure 3.14**). However, a stretch of density accounting for 12 residues is visible at the N-terminus of the plug domain in the open state but not the closed, suggesting that lid closure induces disorder in this region (observable N' = res. 115 in the closed state c.f. res. 103 in the open state). This observation is mechanistically important given that the stretch of residues involved includes the majority of the TonB box (res. 102-108). In the open state, the visible N-terminus turns sharply towards the plug, positioning the TonB box well within the barrel where it would presumably be inaccessible to TonB (see **Figure 3.14**). In contrast, the TonB box is not observed in the closed state and is therefore assumed to be mobile and accessible.

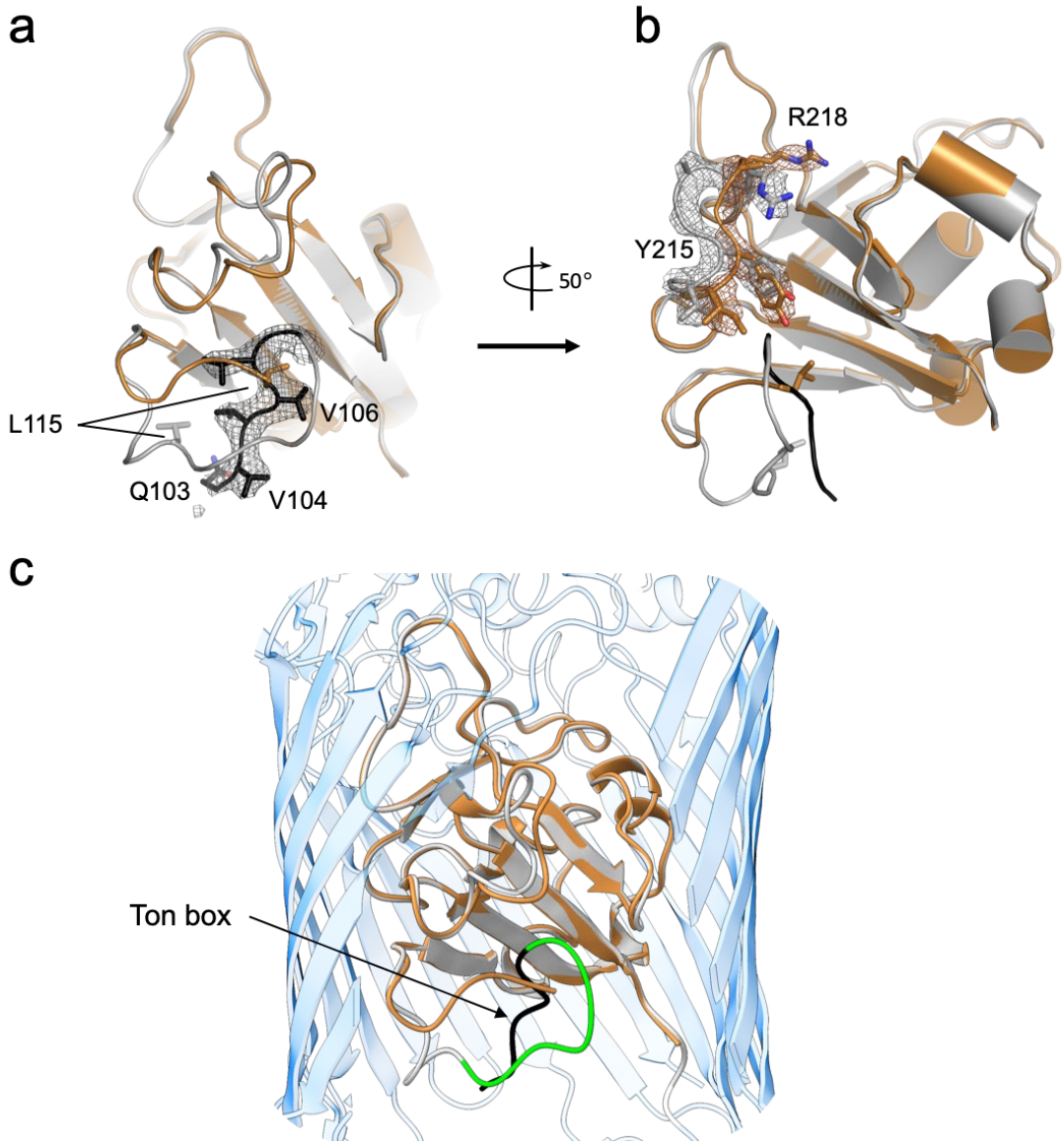


Figure 3.14 A structural comparison of plug domains from each side of the open-closed state of the transporter. a, alignment of RagA plug domains from the open (silver) and closed (bronze) states of the complex. The most N-terminal residue of RagA visible in the closed state is L115 whilst in the open state it is Q103 (both shown as sticks). The Ton box region (Q103-G108), visible only in the open state, is coloured black (black) and its associated density is shown. **b,** Rotated ~50° clockwise from (**a**). Residues exhibiting pronounced shifts between the two states are labelled and shown as sticks. Densities shown are for the I214-R218 region. **c**, Plug overlay in the context of the RagA barrel. Residues visible only in the open state are coloured green and black, where black again denotes the Ton box region which can be seen to reside within the barrel.

How lid closure facilitates disordering of the visible N-terminus of RagA is not immediately obvious. Pronounced shifts are visible for L115-S119, especially L115 which is displaced ~10 Å from its position in the open state. Shifts are also seen in the A211-A219 region, notably for Tyr215 and Arg218, the latter exhibiting a ~9 Å shift of its head group (see **Figure 3.14**). Arg218

lies at the bottom of the substrate binding cavity and A211-219 contacts the region just C-terminal of the TonB box. This observation prompts speculation on an allosteric route via which adoption of a transport competent state of the complex (i.e. lid closed) may affect the conformation and dynamics of the TonB box. However, the substrate is positioned well away from the bottom of this binding pocket (minimum distance 8 Å), perhaps indicating an important role for as yet unidentified residues lining the inside of the barrel.

The assumed increase in mobility of the Ton box region in the closed state is supported by observations made when viewing maps at a high contour level (see **Figure 3.15**). Here, diffuse density corresponding to the NTE is visible immediately N-terminal of the plug on the open side, but not on the closed side. The globular density of the NTE is connected to the plug via a linker region, the flexibility of which likely explains the weak density of the NTE relative to the main body of the complex. The absence of NTE density when the lid is closed suggests that this mobility is increased, likely as a result of the disordering of res. 103-114 which would effectively increase the length of the linker.

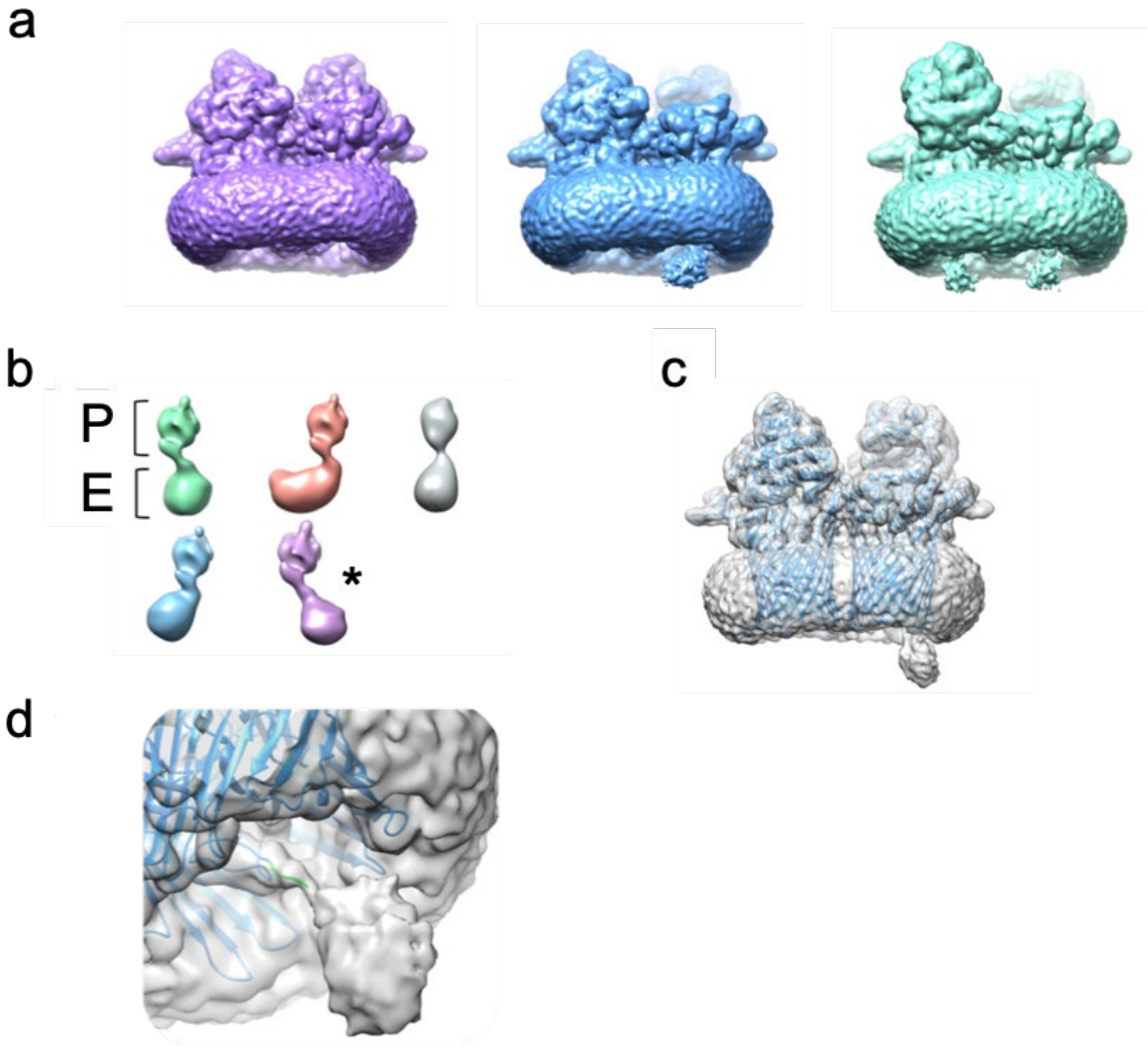


Figure 3.15 Observation and assessment of putative density for the NTE domain. a, Unsharpened maps contoured to show the relatively weak NTE domain density. This density is exclusively associated with the open state of the transporter. Structures are coloured by conformational state where purple, blue and green correspond to closed-closed, open-closed, and open-open states respectively. Particle subtraction allowed isolation of plug (P) and NTE (E) domain density from the open side of the single open structure. 3D classification without alignment revealed differing positions of the NTE domain (**b**). Generating a reconstruction from the particles contributing to the asterisked class (**c**) provided marginally improved density for the NTE domain such that a linker region consisting of a single chain could be traced from the N-terminus of RagA (**d**).

3.5 RagAB is important for growth of *P. gingivalis* on proteins as a carbon source

Four well-defined allelic variants of RagAB (locus-1, -2, -3 and -4) are present among strains of *P. gingivalis* (Hall et al., 2005). Two of the best characterised strains are W83 and ATCC 33277 possessing the *rag-1* and *rag-4* variants respectively. To examine the role of RagAB in the uptake of proteinaceous substrates, collaborators at Newcastle University assessed the

growth of both strains on minimal medium supplemented with bovine serum albumin (BSA-MM). Whilst these strains exhibit identical growth on rich medium, only the *rag-1* W83 strain showed robust growth on BSA-MM (see **Figure 3.16**). The ATCC strain grows more slowly, and only after a substantial lag phase. As such, it appears that RagAB possesses some degree of substrate specificity and that this specificity is strain-dependent.

To conclusively demonstrate that variation in RagAB was responsible for the difference in growth behaviour, collaborators at Newcastle University expressed W83 *ragAB* from a single-copy plasmid in a RagAB-null ATCC strain. This was sufficient to confer W83-like growth on BSA-MM. Further to this, an ATCC strain in which the genomic copy of *ragB* was replaced by W83 *ragB* was constructed. This too exhibited W83-like growth on BSA-MM (see **Figure 3.16**). These results confirm a central role for RagAB in peptide uptake by *P. gingivalis*. Furthermore, they demonstrate that RagA and RagB proteins originating from different strains can assemble to form a functional complex, and, more precisely, that RagB is responsible for the observed strain-dependent substrate specificity. Sequence similarity of both RagA and RagB between W83 and ATCC strains is high. Perhaps the most notable difference is the insertion of an acidic loop in W83 RagB that is not present in ATCC RagB. This loop protrudes into the binding site and plausibly interacts with sidechains of the peptide substrate (see **Figure 3.17**). We hypothesised that its position and acidic nature may be sufficient to alter substrate specificity, giving rise to the difference in growth observed for W83 and ATCC on BSA-MM.

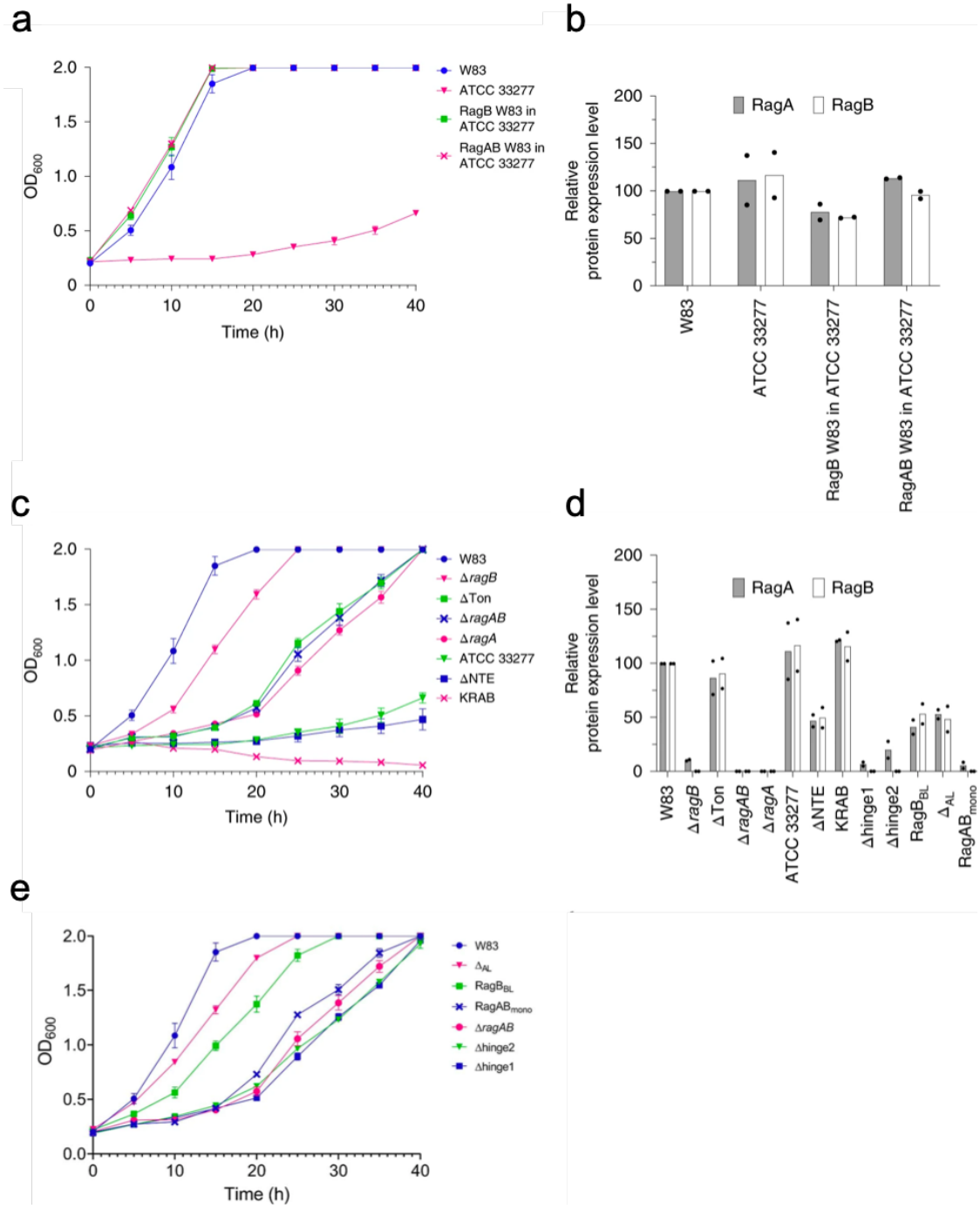


Figure 3.16 RagAB is important for growth on BSA. a, Representative growth curves (n=3, mean ± s.e.m.) for growth on BSA-MM for wild-type *P. gingivalis* W83 and ATCC strains, as well as ATCC strains in which either RagAB or RagB was replaced with the corresponding orthologue from W83. **b**, Corresponding outer-membrane protein expression levels of RagAB determined by SDS-PAGE densitometry. **c**, Representative growth curves (n=3, mean ± s.e.m.) for different *P. gingivalis* W83 and ATCC strains, and mutant *ragAB* variants on BSA-MM. **d**, Outer-membrane protein expression levels for strains and mutants. **e**, Representative growth curves (n=3, mean ± s.e.m.) for mutant W83 *ragAB* variants on BSA-MM. W83 WT and Δ*ragAB* strains are shown for comparison. Graphs in **b** and **d** show the mean of two independent replicates. Dots show the individual replicates. Figure adapted from Madej and White et al., 2020. Cloning and strain construction were performed by Zuzanna Nowakowska, Jagiellonian University, Krakow, Poland. Growth assays and SDS-PAGE densitometry were performed by Mariusz Madej, Newcastle University.

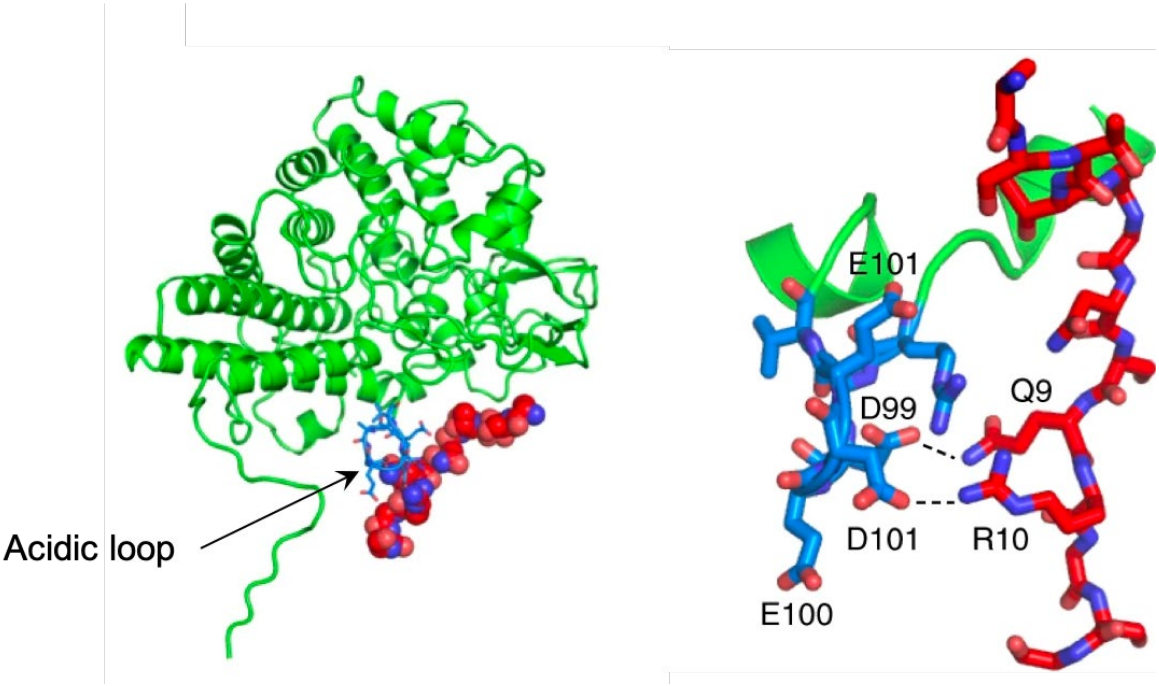


Figure 3.17 Acidic loop of RagB protrudes into the substrate binding cavity. a, Position of the RagB acidic loop relative to the bound peptide. RagB is shown in green with the acidic loop (Asp^{99}-Glu-Asp-Glu^{102}) in blue. The peptide substrate is shown as spheres. **b**, Loop residues Asp99 and Asp101 are within hydrogen bonding distance of modelled peptide sidechains. Potential hydrogen bonds are shown as dashed lines. Figure adapted from Madej and White et al., 2020. X-ray crystallography experiments were performed by the van den Berg group, Newcastle University.

To define the role of the RagAB complex in peptide uptake more fundamentally, collaborators assessed the growth of W83 strains possessing clean deletions of *ragA, ragB* and *ragAB* on BSA-MM (see **Figure 3.16**). All deletion mutants show impaired growth, characterised by a long (15-20 h) lag phase. Notably, this behaviour is different to the gingipain-free W83 KRAB strain which shows no growth even at 40 h. These results demonstrate the essential role of gingipains in generating nutritious peptides from BSA, and, secondly, suggest that the continued action of gingipains likely generates short peptides that can be taken up by a mechanism independent of RagAB, supporting the slow growth observed for the Δ*ragAB* mutant at later timepoints. Notably, the Δ*ragB* strain grows better than Δ*ragAB*. Protein expression levels show that a small fraction of RagA is still present in the OM in the absence of RagB. Thus, it appears that RagB, in contrast to RagA, is not necessary for the uptake of BSA-derived peptides *in vivo*, as was suggested previously (Nagano et al., 2007).

Collaborators at Jagiellonian University (Poland) constructed several mutant strains for structure-function studies of the RagAB complex. For RagA these included: a ton box deletion, an NTE deletion, two hinge loop deletions (L7 and L8) and a putatively monomeric version of RagAB in which a His6-tag was introduced between residues Q570 and G571 at the dimer interface. Unfortunately, most mutants showed low OM expression levels precluding further analysis (see **Figure 3.16**). However, the Ton box (ΔTon) deletion and the NTE (ΔNTE) deletion were reasonably well expressed. The strain lacking the Ton box region has a phenotype similar to ΔRagAB strain, suggesting that the transporter is inactive and thus showing RagAB to be a 'true' TBDT. Surprisingly the ΔNTE strain possessed a similar growth phenotype to the KRAB strain. Given that the ΔRagAB strain also lacks the NTE yet grows after a lag phase, this observation is difficult to explain. Regardless, the NTE appears to play an important role, warranting further investigation in future work.

Mutations in RagB were limited to the acidic loop: a deletion of the loop R97-S104 (Δ_{AL}), and a modification of the loop such that acidic residues were substituted for basic residues (Δ_{BL}, D99-E100-D101-E102 substituted for R99-K100-R101-K102). OM expression levels for these mutants are reasonable but 2-fold less than for W83, precluding unambiguous analysis of the growth phenotypes. However, both mutants grow reasonably on BSA-MM despite the reduced expression levels (see **Figure 3.16**). This indicates that, contrary to our original hypothesis, the acidic loop is not responsible for the strong growth of W83 relative to ATCC and is dispensable in RagAB-mediated uptake of BSA-derived peptides *in vitro*. Instead, small differences in RagAB are likely responsible for the strain-dependent growth behaviour. Given the relatively small pool of peptides that can be generated from BSA, small changes in specificity may be sufficient to severely limit peptide uptake.

3.6 Characterising peptide binding behaviour of RagAB *in vitro*

Having observed different conformational states of the RagAB complex by cryoEM, a primary goal was to see if the proportion of these states could be perturbed through the addition of substrate, providing direct evidence for the pedal-bin mechanism of substrate capture. Collaborators at the Newcastle University used microscale thermophoresis (MST) to assess peptide binding to RagAB. However, given that RagAB co-purifies with an ensemble of

peptides, determination of true dissociation constants was not possible. Instead, measurements provide apparent dissociation constants as a result of added peptides competing with bound peptides. The best choice of substrate for these experiments was unclear. A literature search revealed that an 11-residue peptide from the arginine deiminase ArcA of *Streptococcus cristatus* had been shown by peptide micro-array to bind ATCC RagB (Ho et al., 2017). This peptide, dubbed P4, has the sequence NIFKKNVGFKK. Given its basic nature, it was reasoned that this peptide may also be a good substrate for W83 RagB owing to the presence of the acidic loop in this variant. An N-terminally fluorescein-labelled P4 peptide (P4-FAM) was incubated with serial dilutions of unlabelled RagAB to provide good signal to noise. The apparent dissociation constant for ATCC RagAB was ~0.2 μM, whilst binding to W83 RagAB was weaker (~2 μM). A negative control containing the OM protein Omp40-41 confirmed that signal was not due to non-specific partitioning of the fluorescent label into the stabilising detergent micelles (see **Figure 3.18**). Thus, P4 appears to be a better substrate for ATCC RagAB than W83 RagAB.

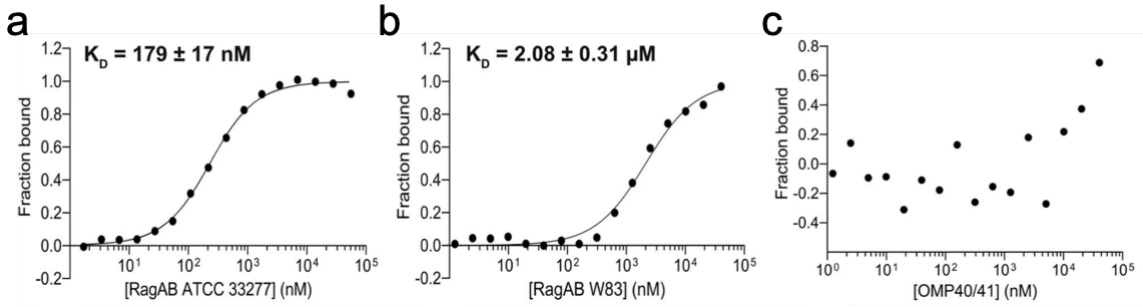


Figure 3.18 MST assessment of P4-FAM binding by RagAB. MST titration curves for the P4-FAM peptide with RagAB from ATCC 33277 (**a**), RagAB from W83 (**b**) and Omp40-41 from W83 (negative control). Experiments and listed K_d values represent the mean of three independent experiments ± SD. Figure adapted from Madej and White et al., 2020. All MST experiments were performed by Mariusz Madej, Newcastle University.

3.7 Identification of peptides co-purified with RagAB by mass spectrometry

The identity of the peptides contributing to the observed density in both the crystal and cryo-EM structures is unknown. It is plausible that these peptides are better substrates for RagAB than P4, and, if identifiable, could be used in downstream experiments. They bind strongly enough to have survived multiple purification steps. Indeed, extensive washing on an anion exchange column was ineffective at removing bound peptide (see section 3.9). However,

these peptides can be released upon denaturation of the complex with trichloroacetic acid (TCA) after which LC-MS/MS was employed to detect, and positively identify them. LC-MS/MS was carried out by collaborators at Aarhus University, Denmark.

Mass spectrometry is inherently non-quantitative, but the data provide an overview of the properties of the peptides bound (see **Figure 3.19**). Several hundred unique peptides were found to be associated with both W83 WT RagAB and W83 (KRAB) RagAB, confirming our original hypothesis that the observed substrate density represents an ensemble of bound peptides. All were endogenous to *P. gingivalis* and, interestingly, all cellular compartments were represented, with peptides belonging to ribosomal proteins, OM and IM proteins having been detected. Importantly, no BSA-derived peptides were identified in either sample. As expected, peptides associated with the WT strain almost invariably possessed a C-terminal Lys or Arg residue, a hallmark identifying them as products of gingipain-mediated proteolysis; this was not the case for the KRAB strain. Peptides bound to W83 KRAB ranged in length from 7-29 amino acids, with the majority of peptides being 11-13 residues long, consistent with the buildable density obtained by X-ray crystallography and cryo-EM. Uncharged or singularly charged (positive or negative with an assumed pH of 7.0) peptides were preferentially bound and the distribution of peptides based on their isoelectric point (pI) shows two maxima at ~4-5 and ~8-9. In contrast, W83 WT RagAB showed less stringent specificity for peptide length with a range of 5-36 residues, but the distribution of peptides by charge and by pI was similar to that observed for W83 KRAB. Considering both peptide pools together, Ala, Glu, Lys and Val appear over-represented in RagAB-bound peptides relative to the *P. gingivalis* genome whilst Phe, Trp and Leu are under-represented.

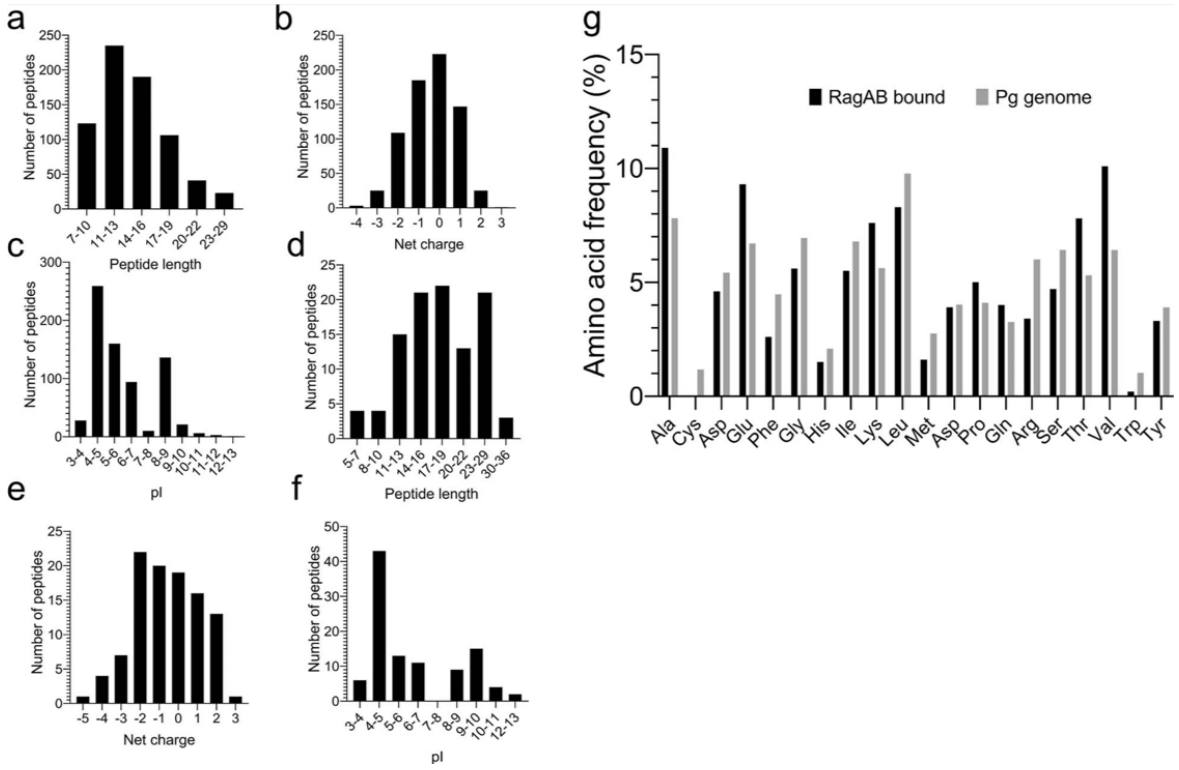


Figure 3.19 RagAB binds a diverse array of oligopeptides. a-c, LC-MS/MS anaylsis of peptides bound to RagAB W83 KRAB, showing length distribution (**a**), total charge (**b**) and pI (**c**). **d-f**, Analysis of peptides bound to RagAB W83 wild-type showing length distribution (**d**), total charge (**e**) and pI (**f**). For charge calculations, pH was assumed to be 7.0 and contributions of any His residues were ignored. **g**, Amino acid frequency of RagAB-bound peptides (KRAB and WT combined; black) vs. the amino acid composition of the *P. gingivalis* proteome (grey). Figure adapted from Madej and White et al., 2020. All peptidomics experiments and analyses were performed by Carsten Scanvenius and Jan Enghild, Aarhus University, Denmark.

A semi-quantitative approach to analysing the MS data, based on chromatogram peak heights and the number of MS/MS spectra observed for a particular peptide, revealed the 21-residue peptide KATAEALKKALEEAGAEVELK (P21) from ribosomal protein L7 to be highly abundant in W83 KRAB RagAB. Moreover, the three most abundant peptides share a common 10-residue core (KATAEALKKA). As such, these were proposed to be good candidate ligands for use in further binding and structural studies. Given that a 10-residue peptide is slightly smaller than the apparent 'optimal length' for RagAB KRAB peptides, an amino acid was added at both the N- and C-terminus of the common core to yield the so-called P12 peptide (DKATAEALKKAL). Both P21 and P12 were commercially synthesised.

3.8 Characterising binding behaviour for peptides identified in MS

The peptides identified by MS were plausibly better substrates for RagAB than P4. Collaborators at Newcastle University again characterised binding behaviour by MST,

however, this time, the peptide was unlabelled and RagAB labelled instead. In this way, any uncharacterised effects that the fluorescein label may have on binding are avoided. All three peptides (P4, P12 and P21) were tested against W83 RagAB. Only P21 produced a good binding curve in these experiments (K_d ~0.4 μM) (see **Figure 3.20**). The lack of P4 binding is interesting considering that the previous assay involving P4-FAM produced a strong binding curve with a dissociation constant ~2 μM; this discrepancy is addressed in detail in the discussion section.

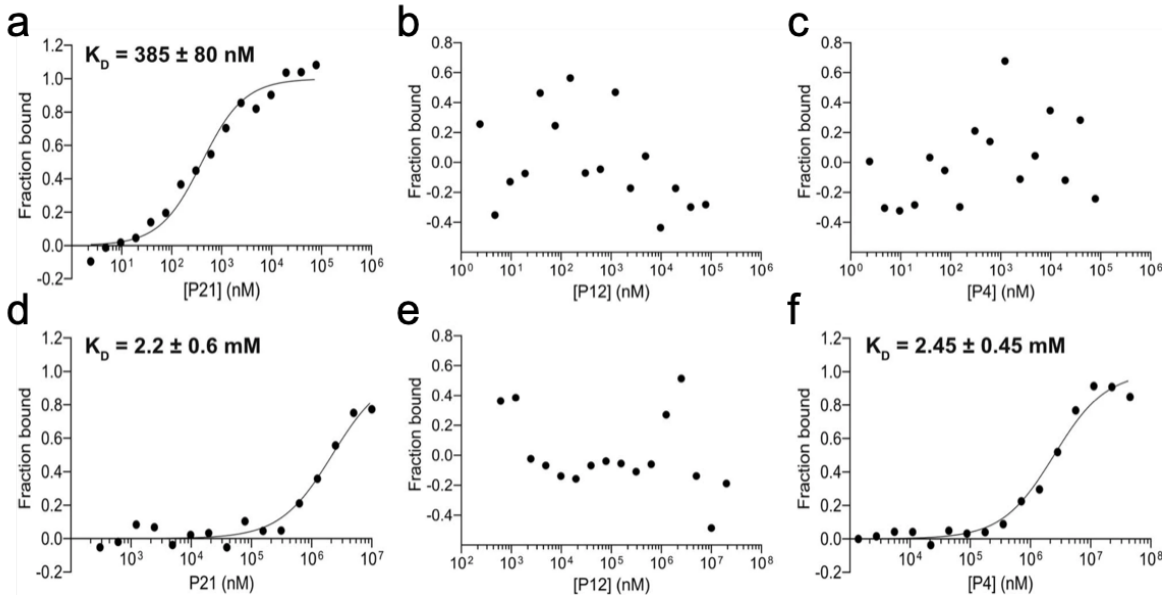


Figure 3.20 Binding behaviour of peptides identified by LC-MS/MS. a-c, MST profiles for unlabelled P21, P12 and P4 binding to His-tag labelled W83 RagAB. **d-f**, MST profiles for unlabelled P21, P12 and P4 binding to His-tag labelled W83 RagB. Experiments and listed K_d values represent the mean of three independent experiments ±SD. Figre adapted from Madej and White et al., 2020. All MST experiments were performed by Mariusz Madej, Newcastle University.

Binding of these peptides to RagB in isolation was also assessed. In this case, binding was observed for both P21 (K_d ~2.2 mM) and P4 (K_d ~2.5 mM). Considering the data for P21, binding affinity of substrates for RagB is lower than binding affinity for the complete RagAB complex (~2 mM c.f. ~0.4 μM). This observation is consistent with the idea that RagB plays an assistive role in binding, guiding the substrate to the principal binding site. The lack of binding for P12 suggests that this peptide either doesn't bind, or is unable to outcompete those peptides that are already bound, indicating that it is not a good substrate for RagAB.

3.9 Perturbation of the proportion of lid open and lid closed states observed by cryo-EM

The presence of substrate density in the open states of the transporter observed by cryo-EM may indicate that not all peptides are capable of inducing lid closure. It may also indicate that RagB lids are not 'locked shut' by the presence of substrate, diverging from the pedal-bin hypothesis. As such, we aimed to probe the effect of substrate on conformation further. A new batch of RagAB was provided by collaborators at Newcastle University and was used for all following experiments to ensure consistent comparison.

A dataset comprising 1609 micrographs was collected for the sample as purified and processed through to 3D classification (145,035 particles in 'good' 3D classes). The three expected conformational states emerged and the relative proportion of each state is shown in **Table 3.1**. An aliquot from the same batch was subjected to extensive washing on an anion exchange column in an attempt to flush out co-purified peptide from the open state. If successful, this would provide a sample in which open states were homogeneously apo, simplifying downstream experiments involving the addition of putative peptide substrates. Moreover, since it is possible that RagB lids can resume an open state without the input of energy, a secondary aim of this experiment was to assess whether excessive washing increases the proportion of open states. A dataset of 511 micrographs for the flushed sample was therefore collected and processed (92,785 particles in 'good' 3D classes).

RagAB Sample	**CC** (%)	**OC** (%)	**OO** (%)
As purified	18	56	26
Flushed	36	46	18
P4 treated	-	48	52
As purified	25	49	23
P21 treated	53	27	-

Table 3.1 Distribution of conformational states observed for RagAB under differing experimental conditions

A comparison of the population distributions revealed a shift towards the closed conformation in the flushed sample. This is difficult to reconcile and the simplest explanation is that we are observing substrate-independent variability. Support for this came from retrospective analysis of previous RagAB data collections that also showed considerable

variation in the proportion of states occupied; despite grids being prepared from the same protein stock (see **Table 3.2**).

Rag Dataset	**CC** (%)	**OC** (%)	**OO** (%)
Preliminary	37	47	16
Final	24	61	15

Table 3.2 Distribution of conformational states observed for RagAB in two independent data collections

The substrate density observed in all states of the flushed sample was equivalent to that observed in the corresponding states of the untreated sample (data not shown). Thus, extensive washing is ineffective at removing co-purified peptide(s) from RagAB. Despite this, it remained possible that the addition of a suitable substrate might still promote closure of open lids. Indeed, the ability of MST to detect peptide binding to RagAB indicates that added peptides are capable of competing with co-purified peptide at the binding cavity. To probe this hypothesis, RagAB was incubated overnight with an excess of either P4 or P21. In each case, untreated RagAB was used as a control.

In accordance with previous datasets, both control cases contained the three expected conformational states, albeit with some variation in population distribution (see **Table 3.1**). The OC state was invariably most abundant in the untreated RagAB datasets, constituting around half of the particles in good 3D classes. The proportions of OO and CC are similar and each constitute 18-26% of the data.

1515 micrographs were collected for the P4 treated sample (100-fold molar excess) and yielded 120,449 particles in 'good' 3D classes. It should be noted that resolution, even at the 2D classification stage, was poor considering the particle number (see **Figure 3.21**). Strikingly, no CC states were observed, even after several rounds of 3D classification. The pedal-bin hypothesis proposed that the closed state of the transporter cannot open in the absence of an energy input. In the context of this hypothesis, the loss of the CC state is particularly interesting. If it is assumed that P4 exerts its effect by competing at the binding cavity, our results indicate that closed complexes are periodically open and that in this open state P4 is

perhaps able to displace bound peptide. Taken further, this suggests that P4 does not trigger closing of the lid and thus the proportion of open states is increased.

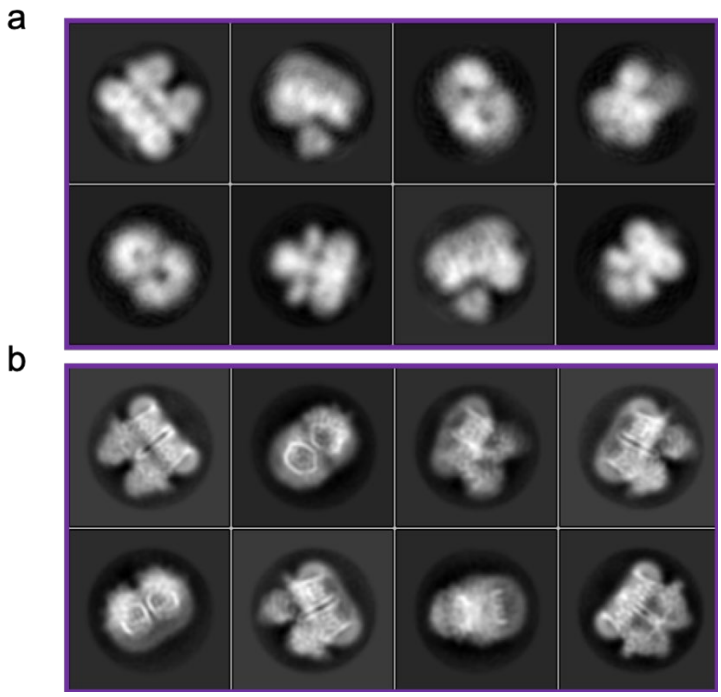


Figure 3.21 Results of the second round of 2D classification for RagAB incubated with P4 peptide (**a**) and RagAB subjected to extensive washing by anion exchange chromatography (**b**). Datasets contained ~170,000 and ~110,000 particles respectively. Despite this, the quality of the 2D classes derived from RagAB incubated with P4 is much poorer than those derived from the flushed sample

In contrast, incubation of RagAB with P21 (50-fold molar excess) resulted in complete loss of a clear OO state. 1100 micrographs provided 49,873 particles after 2D classification. After initial 3D classification, only the CC and OC states were apparent. The third most populated class, whilst resembling the global structure of RagAB, is ambiguous, with disjointed density in the region of the TM barrels (see **Figure 3.22**). Further dissection of this class did not provide convincing reconstructions of any conformational state. The presence of an ambiguous state was surprising since similar ambiguous classes were not observed in the control datasets. Even in the P4 experiment, where the quality of the data is comparatively poor, all RagAB-like classes (not junk) identified in the first round of 3D classification clearly resemble either the OO or OC state. Irrespective of this, the loss of a clear OO state and marked increase of the CC state of RagAB (> 2-fold) observed upon incubation with P21 suggests that the binding of this peptide is associated with closing of the RagB lid.

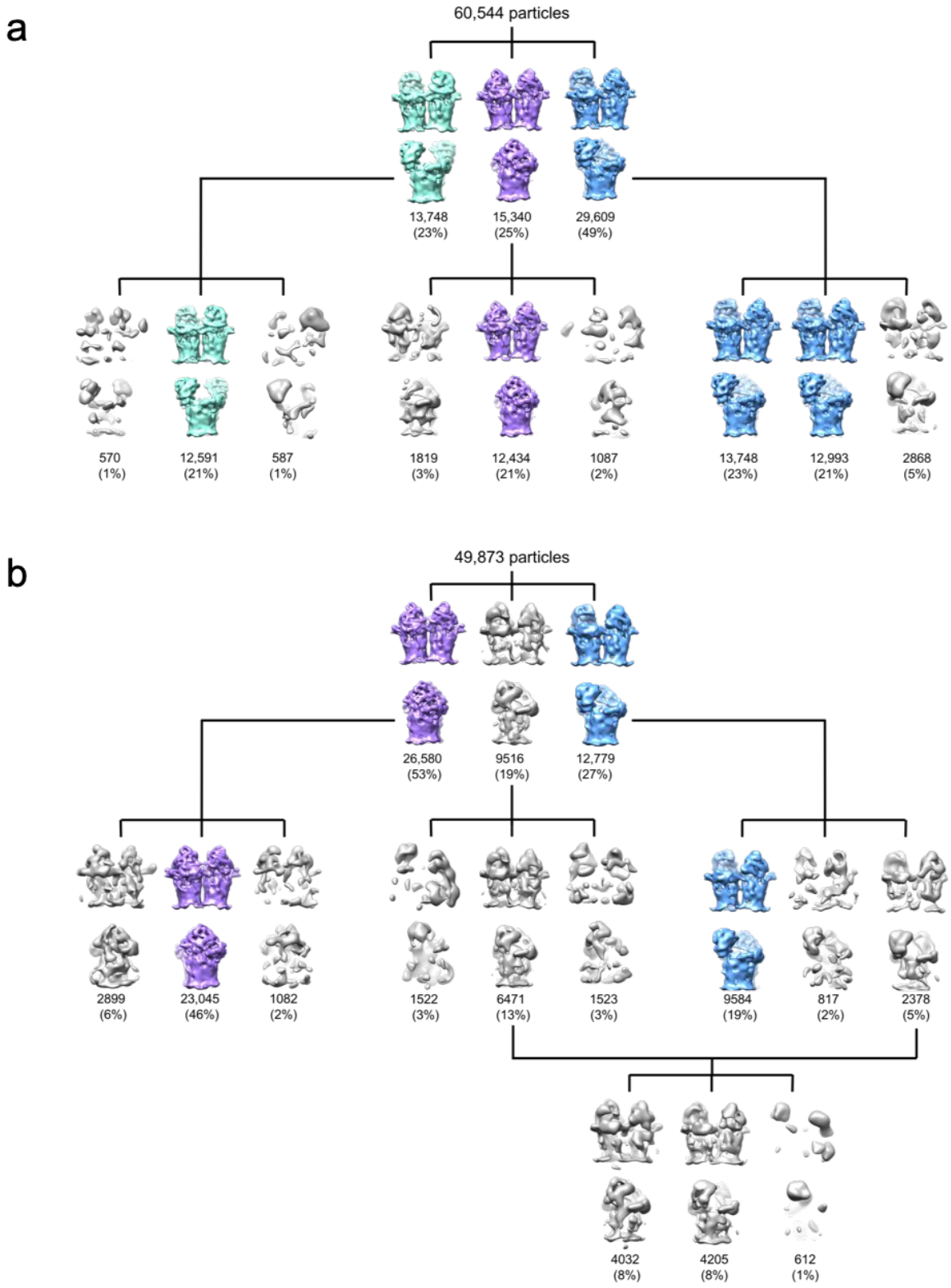


Figure 3.22 Perturbation of RagAB conformational states upon incubation with P21. a, Results of 3D classification for untreated RagAB. **b**, Results of 3D classification for RagAB in the presence of 50-fold molar excess P21. Classes corresponding to the CC, OC and OO states are coloured purple, blue and green respectively. Junk or ambiguous classes *e.g*. where RagA barrels are incomplete are coloured grey. In the presence of P21 there was no clear OO state whilst the proportion of the CC state increased, supporting the proposed mechanism of substrate capture.

3.10 Examining density at the binding site of P21-doped RagAB

MST experiments show that P21 binds RagAB, and that the apparent strength of this interaction (K_d = 0.4 μM) is comparable to classical TBDT-ligand interactions in *P. gingivalis* (Moynié et al., 2019). Moreover, cryoEM demonstrated that incubation of the RagAB complex with P21 results in a shift in population distribution to favour the CC state of the complex. Taken together, these results suggest that P21 is behaving as a 'productive' substrate, binding the open state of the transporter and causing the RagB lid to close. A reconstruction of the CC state from the P21-doped RagAB sample was solved to 3.8 Å, allowing substrate density at the binding site to be compared with that of the untreated sample (see **Figure 3.23**).

Both maps were lowpass filtered to 3.8 Å for consistent comparison. In both cases, the binding pose for the peptide is very similar. Some small differences in the prominence of sidechain densities are observed towards the arbitrarily labelled C-terminus of the peptide. More substantial differences can be seen at the N-terminus where density for the peptides begins to diverge. However, the visible density in the P21-doped sample is not extended and would accommodate no more than the 13-residues built into the density for the untreated sample. Densities for sidechains are no more distinct than for the untreated sample and, as such, it was not possible to unambiguously assign the position of the P21 peptide.

A comparison of the crystal structure of WT W83 RagAB with a co-crystal structure of this complex with 50-fold molar excess of P21 shows similar results (see **Figure 3.23**). The higher resolution of these structures provides more sidechain information but it was still not possible to model the P21 sequence. Moreover, the maps obtained by cryo-EM and crystallography are for the W83 KRAB and W83 wild-type strains respectively. These strains bind different ensembles of peptides, yet maps for the untreated samples are similar to each other, and both are similar to the P21-doped maps. It is likely that not all endogenous peptides are replaced by P21 when added in 50-fold molar excess. LC-MS/MS has since revealed that 100-fold molar excess is sufficient to completely displace co-purified peptides, therefore this experiment may be revisited in future work. It is also possible that P21 (and substrates in general) bind with register shifts and perhaps different chain directions, precluding unambiguous building of the sequence.

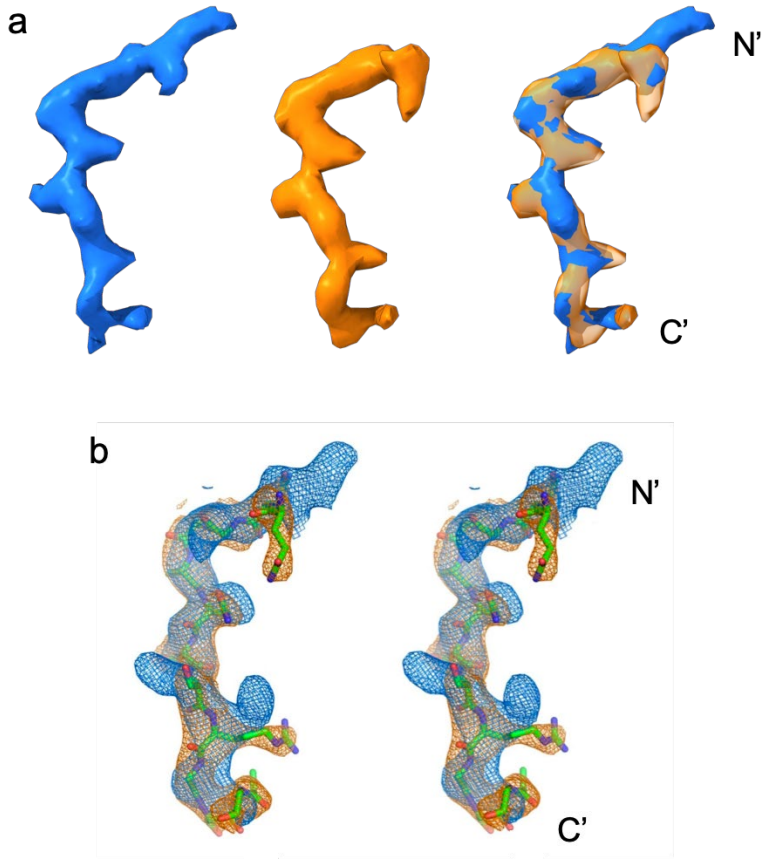


Figure 3.23 Cryo-EM and X-ray crystallography data showing peptide density for untreated and P21-doped samples. a, Cryo-EM peptide density for the untreated (blue) and P21-doped (orange) samples of W83 KRAB RagAB, and superposed (right). **b**, Stereo view of superposed 2Fo-FcElectron density maps (1.0 σ, carve = 2.0) for wild-type W83 RagAB in the absence (blue) and presence (orange) of the P21 peptide generated with the same high-resolution cut-off (3.0 Å). The peptide sequence of the P21 co-crystal structure is arbitrarily modelled as QNGGANTSRGSAG. X-ray crystallography was performed by the van dan Berg group, Newcastle University.

3.11 Discussion on RagAB

Crystal structures of SusCD-like transporters have recently provided unprecedented insight into their mechanism of action (Glenwright et al, 2017). The observed encapsulation of substrate in a solvent excluded cavity formed by both SusC and SusD suggested that separation of these subunits is required for substrate binding. MD simulations, set up to predict the nature of substrate capture, suggest a rigid-body, hinge-like movement of SusD away from SusC that exposes the substrate binding site. Interestingly this movement did not occur when substrate was modelled in, suggesting that substrate binding locks the SusD lid in a closed conformation. This is biologically sensible since it would prevent unproductive

substrate binding and release. However, structural support for the conformational changes predicted by these simulations was lacking, with available crystal structures of SusCD complexes showing SusD lids closed against their SusC counterparts whether substrate is present or not (Glenwright et al., 2017). The data presented here are complementary, providing confirmation of the dimeric complex observed by crystallography and conclusive structural evidence in support of the lid open conformations predicted by molecular dynamics.

Unfortunately, the structure of RagAB was not solved in the context of a nanodisc. Whilst negative stain-TEM data appeared to provide some evidence of partial reconstitution into nanodiscs, no such evidence was observed in cryoEM studies. This is difficult to reconcile given that grids for cryoEM were prepared from the same sample that was used for negative stain TEM and on the same day. Thus, we would expect the same species to be visible regardless of technique. Aggregation was more extensive for the vitrified samples, with concatemers observed that were not seen in negative stain TEM. One possibility is that the small population of RagAB in discs was lost to these aggregates and therefore not observed. Future work should consider simplifying the system by using a single lipid type such as DMPC rather than a lipid mixture as was used here. This would permit more accurate ratio calculations which may in turn improve reconstitution. Evidence suggests that insufficient lipid was used in the assembly mixture. In practice, ratios estimated from these calculations should be treated as starting points and a comprehensive screen of ratios will likely be required to identify optimal conditions for reconstitution into nanodiscs. In addition, size-exclusion chromatography should be used to separate nanodiscs containing RagAB from empty nanodiscs and aggregates. Following this, SDS-PAGE or native PAGE should be employed in addition to negative stain TEM to confirm successful reconstitution.

Single particle cryo-EM of detergent solubilised RagAB provided near-atomic resolution structures of three conformational states of the complex. One identical to the crystal structure in which both RagA barrels were tightly capped by RagB subunits, and two novel states where either one or both RagB subunits were flipped open, away from their respective barrels. These structures neatly substantiate the predictions from MD simulations. Interestingly, RagB was only ever observed to occupy either a closed position or an open

position (~45° swing away from RagA). This lack of intermediate states in the cryo-EM data is interesting to consider in the context of substrate capture. It would appear that the energy landscape for opening and closing of the lid is very rugged; with the open and closed states of the complex representing stable minima. Strong density corresponding to bound peptide was visible in the enclosed binding sites at the interface of RagA and RagB, while weaker peptide densities were observed in the open states of the transporter. Peptide binding is therefore expected to alter the minutiae of the interaction between RagA and RagB, shifting the equilibrium towards the closed state. The absence of intermediate states was unexpected given that electrophysiology experiments conducted on the apo levan transporting system BT1762-BT1763 from *B. theta* show rapid fluctuations in conductivity attributed to dynamic opening and closing of the SusD lid (Glenwright et al., 2017). However, the nature of these experiments requires that they be performed using a transporter from which the plug domain of the SusC component has been removed. This is a substantial change to the protein architecture and is perhaps responsible for the observed differences.

Avoiding unproductive transport cycles, i.e. expending energy to disrupt the plug in the absence of substrate, is an important aspect of TonB-dependent transport. To achieve this, formation of a transport-competent (substrate-loaded) complex must be signalled across the OM. For 'classical' TBDTs such as BtuB and FecA, evidence points to a signalling role for the TonB box. Substrate binding at the extracellular side of these transporters induces conformational changes in the TonB-box region at the periplasmic side. How these conformational changes are propagated through the plug domain remains an open question. Identifying allosteric pathways is not trivial even when structures are available, and few studies have addressed this question directly. Ferguson et al. (2007) identified a sparse network of residues responsible for linking distant functional sites (extracellular and periplasmic pockets) in the ferric citrate transporter (FecA). Such signal transduction networks may be conserved across all TBDTs. However, a consistent feature of classical TBDTs is that apical loops of the plug domain contribute to ligand binding and undergo conformational changes when they do so (Locher et al., 1998; Ferguson et al., 2002; Chimento et al., 2003). In these cases, it is logical to propose that these conformational changes could be allosterically propagated through the plug domain causing disordering and increased accessibility of the TonB box region. However, in RagAB, there are no direct interactions

between the substrate and the plug domain (minimum distance is ~8 Å). It is possible that the extremities of the peptide(s) are not observed due to mobility and that these may contact the plug. However, despite substrate density being observed in the open sites of the double open structure, the TonB box region (res. 102-108) was visible well within the barrel and looks to be inaccessible to TonB (**Figure 3.16**). Thus, substrate binding alone does not appear to determine whether the RagAB complex is transport-competent.

As described in section **3.4.3**, lid closure results in the disordering of residues 103-115 at the N-terminus of RagA. It is conceivable that this might result in a more extended, dynamic conformation of the TonB box region, allowing it to protrude from the mouth of the barrel. Thus, if we consider lid closing to be integral to substrate translocation, our data is consistent with the consensus hypothesis that a substrate-loaded, transport-competent state of the complex is communicated to the periplasm via disordering of the TonB box region, making it accessible to TonB. Whether this is a general property of the SusCD-like transporter family remains to be seen. The proposed mechanism of substrate capture and translocation by RagAB is shown in **Figure 3.24**.

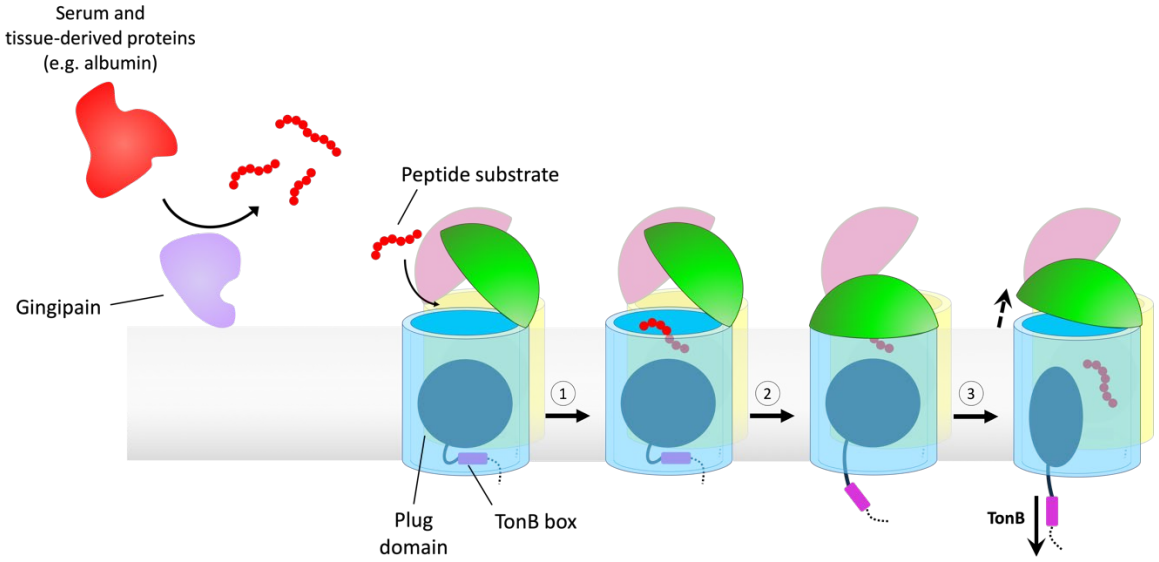


Figure 3.24 Schematic demonstrating the proposed mechanism of substrate capture and translocation by RagAB. Peptide ligands to be imported by the RagAB system are predominantly generated by the action of gingipains on serum and tissue-derived proteins. **1**. A lid-open state of RagAB permits peptide binding. **2**. Contributions from both RagA and RagB to peptide binding elicits closure of the lid, forming the transport-competent state of the complex. This is signalled across the OM by perturbation of the TonB box region on the periplasmic side of the plug domain, making it accessible to TonB. **3**. According to the literature consensus, TonB-mediated disruption of the plug permits substrate translocation and a return to the open state of RagAB.

The presence of substrate density in the open states of RagAB is intriguing since the working hypothesis predicts that substrate binding elicits lid closure and 'locks' the lid in the closed conformation. However, this hypothesis was originally based on the Bt2263-Bt2264 system where the relative contributions of the SusC and SusD-like subunits to substrate binding are more equal. The Bt2263-Bt2264 crystal structure shows the putative peptide(s) substrate(s) residing at the interface of the SusC- and SusD-like components with near-equal contributions from both in terms of interaction surface. Indeed, crystal structures of isolated SusD-like proteins in complex with various oligosaccharide ligands indicate a general and significant contribution of this subunit to substrate binding by SusCD-like transporters (Koropatkin et al., 2008; Tauzin et al., 2016). In contrast, the binding site for the putative peptide(s) in RagAB is predominantly formed by RagA, and a PISA analysis indicates a stable RagA-substrate complex in the absence of RagB. As such, it is perhaps unsurprising that we observed peptide density in the open state and it appears that stringent substrate-induced 'locking' of the RagB lid does not occur (at least for this artificially reconstituted system with co-purified peptide(s)). The more pronounced density for substrate in the closed states probably indicates a higher occupancy. Alternatively, it may be that, in the absence of RagB, the peptide is more mobile. The latter seems unlikely given that density for the 90 ° kink, albeit considerably weaker, is still observable for the peptide in the open-open structure. As mentioned previously, the difference in peptide density observed in the open sites of the OO structure and the open site of the OC structure is unexpected. Whilst this may indicate cooperativity between the transport hubs, it is unclear how this could be verified experimentally, especially given that attempts to generate a monomeric form of the protein have been unsuccessful (expression too low for further study). The functional relevance of the dimeric structure therefore remains an interesting enigma, unique to SusCD-like TonB-dependent transporters.

An alternative hypothesis for the observation of substrate bound open states, is that not all peptides are recognised as productive substrates and, as such, do not cause closing of the RagB lid. Support for this hypothesis comes from the cryo-EM experiments in which RagAB was incubated with either P4 or P21. Attempts to establish a cause and effect relationship between substrate binding and lid closure through the addition of P4 peptide were unsuccessful. Surprisingly, addition of P4 led to loss of the CC state of the complex and a

corresponding increase in the OO state. As described previously, these results indicate that closed complexes are periodically open and that, in this open state, P4 is able to displace bound peptide. It is possible that P4 is a non-productive substrate and therefore the RagB lids remain open.

Further examination of the MST data provides support for this. Experiments with labelled P4-FAM demonstrated binding of P4 to both ATCC RagAB and W83 RagAB, albeit 10-fold weaker for the latter. However, substrate-dependent quenching of the P4-FAM fluorescence signal was observed with the ATCC RagAB but not with W83 RagAB. It is proposed that the observed quenching is due to closing of the RagB lid over the labelled peptide, in which case these data indicate that P4 is a productive substrate for ATCC RagAB but not W83 RagAB. The apparent lack of binding observed for unlabelled P4 with fluorescently labelled W83 RagAB can be explained in a similar way. Here, the fluorescent tag is on the RagAB complex. If the binding of P4 does not elicit lid closure, the binding event may not alter the thermophoretic properties of the complex sufficiently enough to allow detection. In contrast, the cryo-EM experiment where W83 RagAB was incubated with P21 showed a pronounced shift towards the CC state of the complex and a corresponding reduction in the OO state. Accordingly, binding was detected in the MST experiment involving unlabelled P21 and fluorescently labelled RagAB, presumably because the peptide induces closing of the lid and thus causes a large change in the thermophoretic mobility of the complex.

It is interesting to speculate on what it is that determines whether a peptide engages in productive vs non-productive binding with RagAB; as judged by closing of the lid. The most striking feature of the peptides modelled in both X-ray and cryoEM structures is the pronounced 90° kink adopted. When considered in the context of the capture and signalling mechanism, this peptide topology has important ramifications. One section of the peptide runs parallel to the RagAB interface, contributing bridging interactions that plausibly hold the lid of the complex closed, while the presence of the sharp 90° turn means that the remainder of the peptide continues towards the plug domain. Assuming that closing of the lid and allosteric signalling to the Ton box region via the plug domain are critical steps in the translocation mechanism, then the ability of the peptide to adopt this conformation may be crucial. Consequently, peptide properties such as length and sequence may determine

whether they engage productively or not. Notably, the P4 peptide was only 11-residues in length c.f. 21 for the P21 peptide. Given that RagA is shown to contribute the majority of binding interactions, it is plausible that shorter peptides preferentially occupy a position closer to the plug, thus precluding the formation of bridging interactions with RagA and the RagB lid. Selection of longer peptides in this way would also be biologically advantageous to *P. gingivalis* by maximising the nutritional value of each translocation cycle.

Whilst broad selectivity for peptide properties such as length can be justified biologically as described above, the work demonstrating strain-dependent growth on BSA is less easily reconciled. The data presented in **Figure 3.16a** clearly demonstrate RagB-mediated substrate specificity. However, it is important to consider the context and limitations of these experiments. Chiefly, conclusions drawn from growth-assays with the relatively small substrate pool generated by gingipain-mediated proteolysis of BSA as the sole carbon source are likely not transferrable to *in vivo* behaviour of *P. gingivalis.* The ecological niche of the periodontal pocket offers a myriad of peptide sources including host cell surface and serum proteins as well as microbial peptide sources (Hočevar et al., 2020; Davies et al., 2021). Indeed, the pro-inflammatory nature of *P. gingivalis* and its ability to promote dysbiosis and unchecked growth of microbial communities in the periodontal pocket works to expand the pool of peptides present through tissue destruction and an increase in host pro-inflammatory mediators (Chopra et al., 2020). Under these conditions, subtle differences in substrate specificity of the RagAB transporter are unlikely to make significant contributions to the survival and pathogenicity of *P. gingivalis*. As such, the lid appears more likely to play a principal role in the translocation mechanism such as facilitating directional transport, or preserving the selective permeability of OM by preventing large pore formation upon putative disruption of the plug.

Interestingly, the open position of RagB observed in the cryo-EM structures is more extreme than that predicted by MD simulations (see **Figure 3.25**). Of three simulations set up with a generated apo closed state of the transporter, one resulted in opening of the RagB lid (Cα rmsd ~15 Å relative to the closed position of RagB from the X-ray crystal structure). Simulations starting from the open cryo-EM structure settled to a more closed state, again

with Cα rmsd ~15 Å. It therefore appears that the energy minimum for the open state differs between *in silico* modelling and 'in solution' structure determination for reasons that are unclear. In all simulations that started with a closed complex containing bound peptide, the complexes remain closed for the duration of the simulation, suggesting that lid opening is unfavourable in the presence of substrate. However, Cryo-EM data suggest that the binding sites of closed complexes may become transiently accessible. Longer simulation times might be required to reveal opening of the holo-complex (containing a modelled peptide) in the absence of an energy input.

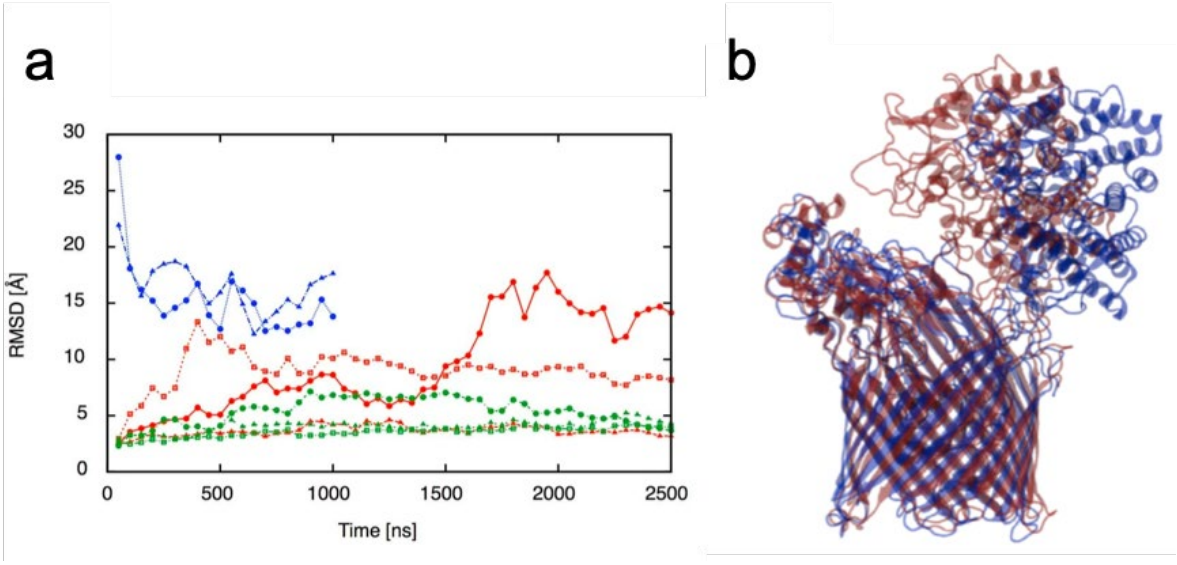


Figure 3.25 MD simulations of RagAB. **a**, Cα-rmsd values of RagB lids in apo-RagAB (red) and peptide bound RagAB (green) with reference to the starting crystal structure in the closed conformation. The Cα-rmsd values of the RagB lids in $RagA_2B_2$ with reference to the OO cryo-EM state are shown in blue. Three independent simulations were performed for apo-RagAB, of which one showed a clear opening of the RagB lid which shifted ~15Å compared to the closed starting structure. Considering the lids of the double open structure, Cα-rmsd values for both RagB subunits decrease from ~30 Å to ~15 Å (blue) relative to the closed structure, resulting in a state similar to that observed after opening of the apo-closed structure. Simulations of peptide-bound RagAB show that the complex remains closed for the duration of the simulation. Figure adapted from Madej and White et al., 2020. MD simulations were performed by Karunakar Pothula, Jacobs University, Bremen, Germany.

Finally, despite having observed the position of the NTE for the first time, its function remains enigmatic. Its proximity to the TonB box may indicate a role in stabilisation of the interaction with TonB. An interesting future experiment would be to include the TonB box-binding C-terminal domain of TonB in further structural studies in an attempt to visualise potential interactions with the NTE.

Chapter 4 Structural and functional insights into SusCD-mediated glycan import by a prominent gut symbiont

4.1 Introduction

The work presented in chapter 3 represents a significant leap in our understanding of SusCD-like transporters. However, analysis of the RagAB system was not straightforward and was complicated by the fact that the complex co-purified with its substrate. Thus, in order to explore the mechanism of substrate capture further, an apo system was required.

The SusCD system for levan transport (Bt1762-Bt1763) from *B. theta* was an attractive choice. The associated PUL is well characterised, and can be upregulated by growing the organism in the presence of fructose, thereby increasing expression of the complex (Sonnenburg et al., 2010). Given that levan is not required for growth, it can be omitted from the growth medium, and we therefore had a mechanism to express high levels of the complex of interest in an apo state. Moreover, the well-defined substrate of this system meant that substrate addition experiments were possible, allowing prospective structural comparison of apo and ligand-bound states.

Levan is a β2,6-linked poly-fructose derivative with occasional β2,1 branch points (see **Figure 4.1a**). It is predominantly found as a bacterial exopolysaccharide but is also present in some cereal plants e.g wheat (Öner et al., 2016; Benigar et al., 2014). A schematic of the levan PUL is shown in **Figure 4.1b**. The OM-associated components of this system are $Bt1763^{SusC}$, $Bt1762^{SusD}$ and two surface exposed lipoproteins: a SGBP, Bt1761, and an endo-acting levanase, Bt1760 (a glycoside hydrolase family 32 (GH32) member). These OM components have complementary functions: the SGBP, from here on referred to as the levan binding protein, is expected to bind levan polysaccharides at the cell surface and the levanase acts to cleave the levan chains into transport competent substrates (Sonnenburg et al., 2010; Mardo et al., 2017). These substrates are then bound by the SusCD complex and imported into the periplasm. Here, two exo-acting GH32 glycosidases, Bt1759 and Bt3082, further break down the imported fructo-oligosaccharides (FOS). It should be noted that Bt3082 is not physically associated with the levan PUL, but is co-regulated with the rest of the PUL components

(Sonnenburg et al., 2010). The resulting monomeric fructose acts as the signal for the IM hybrid two component system (HTCS) sensor regulator Bt1754, leading to upregulation of the PUL in the presence of its target substrate. The IM monosaccharide transporter Bt1758 permits influx of fructose into the cytoplasm where it is phosphorylated by the PUL-encoded fructokinase Bt1757 and thus shunted into the glycolytic pathway (Bolam and Sonnenburg, 2011).

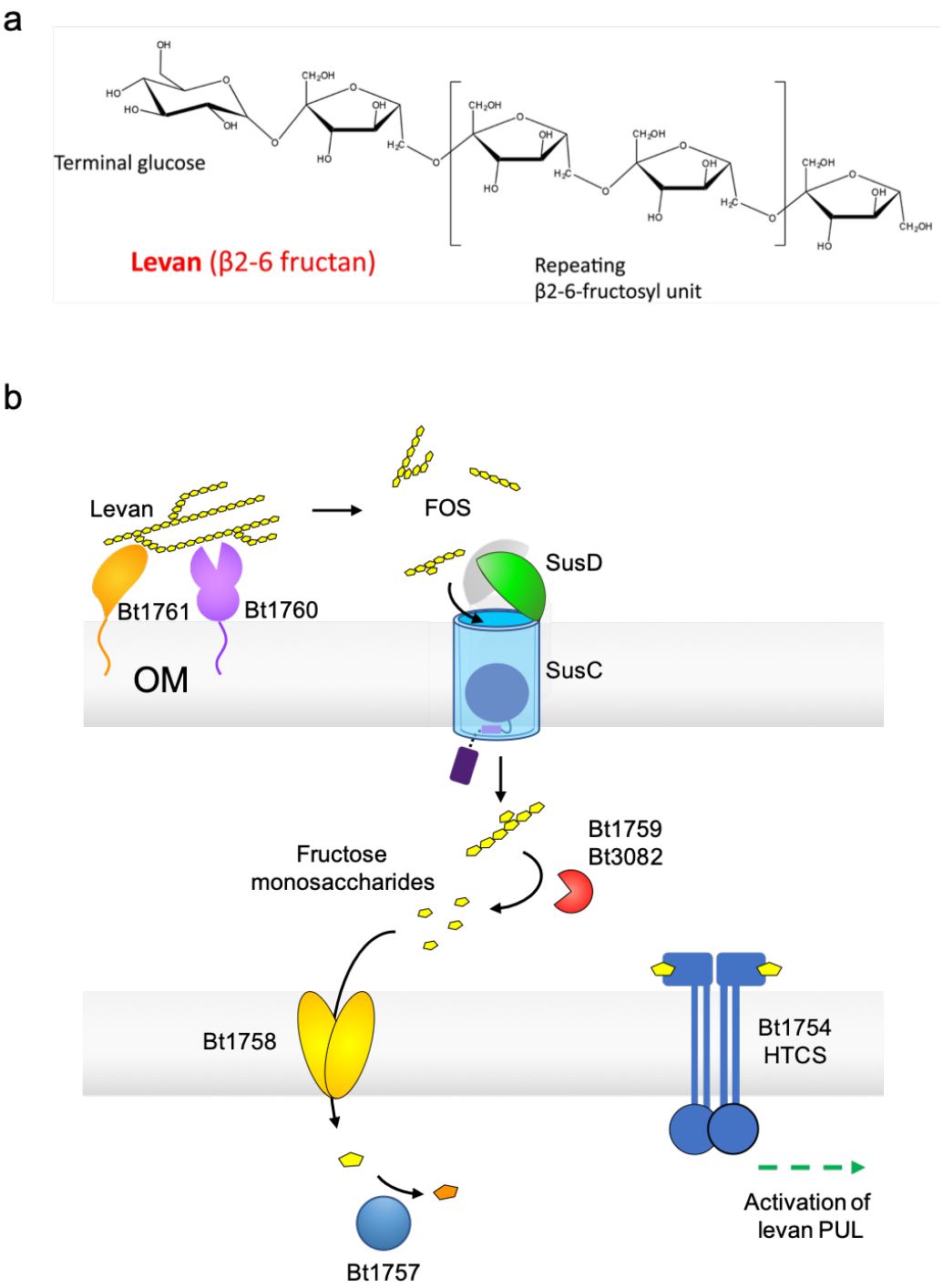


Figure 4.1 Levan structure and the levan polysaccharide utilisation locus. a, Structure of the levan polysaccharide. The chain is characterised by a terminal glucose residue and repeating β2,6 linked fructosyl units. **b**, Schematic demonstrating the function of PUL encoded components in levan utilisation. At the cell surface, levan is bound by Bt1761 and partially digested by Bt1760, a GH32 family endo-levanase. Breakdown products are FOS of various sizes that are captured and transported across the OM by the levan SusCD pair Bt1762-Bt1763. FOS molecules undergo further digestion in the periplasm through the action of exo-acting glycosidases Bt1759 and Bt3082. Resulting fructose monosaccharides, are imported into the cytoplasm via the MFS family member Bt1758. Periplasmic fructose also serves as a signal for the HTCS Bt1754, inducing activation of the levan PUL. Cytoplasmic fructose is phosphorylated by the fructokinase Bt1767 and enters the glycolytic pathway.

Not only did the levan SusCD system offer a valuable opportunity to study the effects of substrate binding on the conformation of the transporter, it is also more representative of this family of transporters. Like most SusCD systems it operates as part of a PUL and facilitates the transport of glycans rather than peptides. A glycan-bound structure of a SusCD complex was not yet available and whether the molecular details of substrate capture and binding are consistent between peptide and glycan transporters remained to be seen. As outlined in the introduction to this thesis, glycan utilisation mediated by PUL-encoded machinery is essential for the survival of important members of the human gut microbiome and therefore understanding the mechanisms that underpin this process is a key research goal. The aim of this chapter was to use cryoEM to characterise the behaviour of the levan transporting SusCD complex (Bt1762-Bt1763) in the absence of substrate, complementing the work done on the RagAB complex and developing the working 'pedal-bin' hypothesis. Successful structure determination would also set up future experiments investigating the behaviour of the complex in response to substrate addition. This project was part of a large international and interdisciplinary collaboration and, where necessary, results generated by collaborators have been included to contextualise the work carried out by the candidate. Work carried out by others is explicitly identified both in the text and in relevant figure legends.

4.2 Assessment of sample quality by negative stain TEM

Levan SusCD solubilised in LDAO (~3 mg/ml) was provided by collaborators at Newcastle University. The complex was purified from *B. theta* grown on minimal medium containing fructose as the sole carbon source and was therefore assumed to be in the apo state. Sample quality was assessed by negative-stain TEM and a representative micrograph is shown in **Figure 4.2.** A 600-fold dilution of the stock sample was found to give a useful distribution of particles on the grid. Particles were homogeneous and monodisperse, with the expected dimeric arrangement of β-barrels visible in the raw micrographs. The sample was therefore deemed suitable for downstream cryoEM studies.

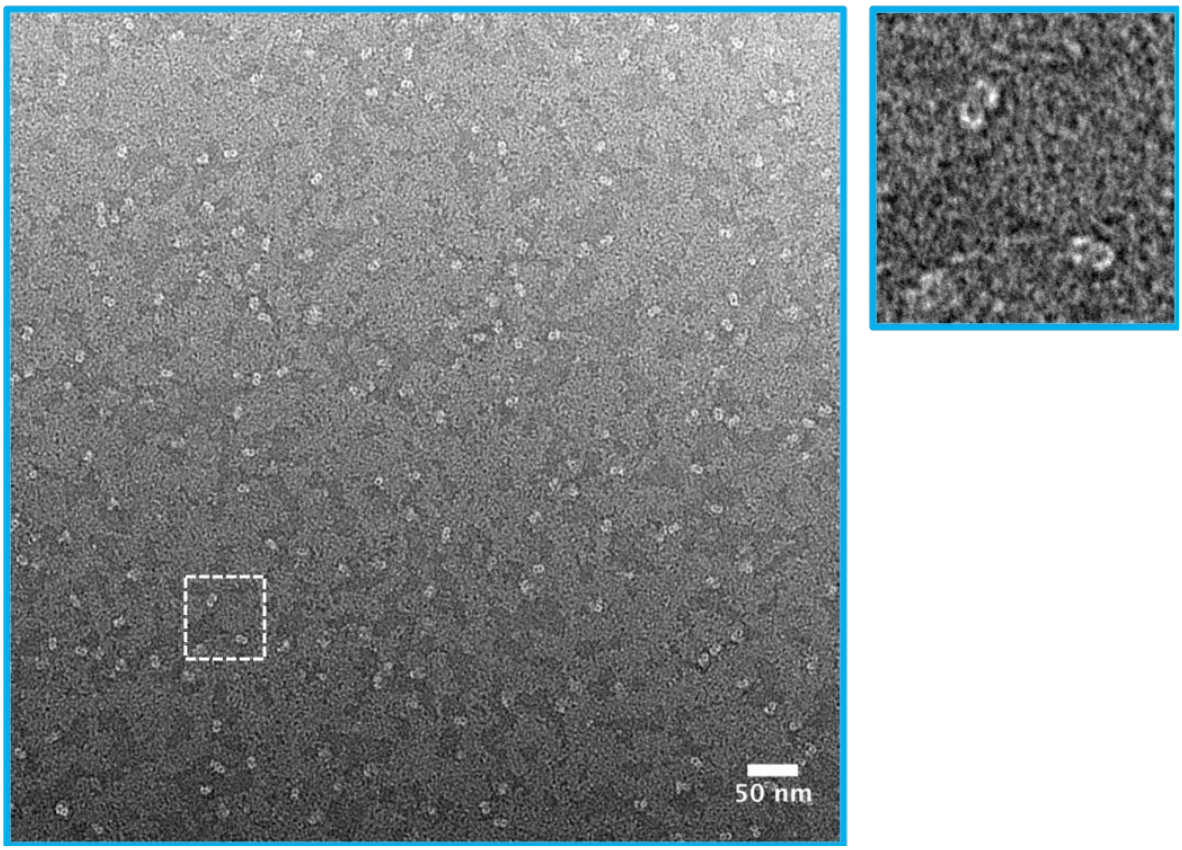


Figure 4.2 Uranyl acetate stained levan SusCD solubilised in LDAO at 50k x magnification. Particles are homogeneous and monodisperse. A zoomed inset of the area marked with a white dashed box is shown (right). The twin barrel structure is clearly visible in the zoomed inset.

4.3 Single particle cryo-electron microscopy of the levan transporting SusCD complex in the absence of substrate

Screening of the sample applied neat to quantifoil grids (1.2/1.3, 300 mesh) revealed wide-spread self-association of the particles. While some individual particles were visible, the majority were present in chains (see **Figure 4.3**), that hindered particle picking and resulted in a strong preferred orientation. Optimisation of grid preparation conditions was therefore necessary. The high dilution factor required for negative stain imaging indicated that the complex has a high affinity for the carbon support films and, since self-association of particles was not observed by negative stain, it was reasoned that the affinity of particles for the carbon may be greater than the affinity of the particles for each other. The sample was therefore applied to 200 mesh lacy carbon grids with an ultrathin carbon support; A 200-fold dilution of the sample (0.02 mg/ml) was required to achieve a suitable particle distribution. The ultrathin support film provided a uniformly monodisperse population of particles suitable for data collection (see **Figure 4.3**).

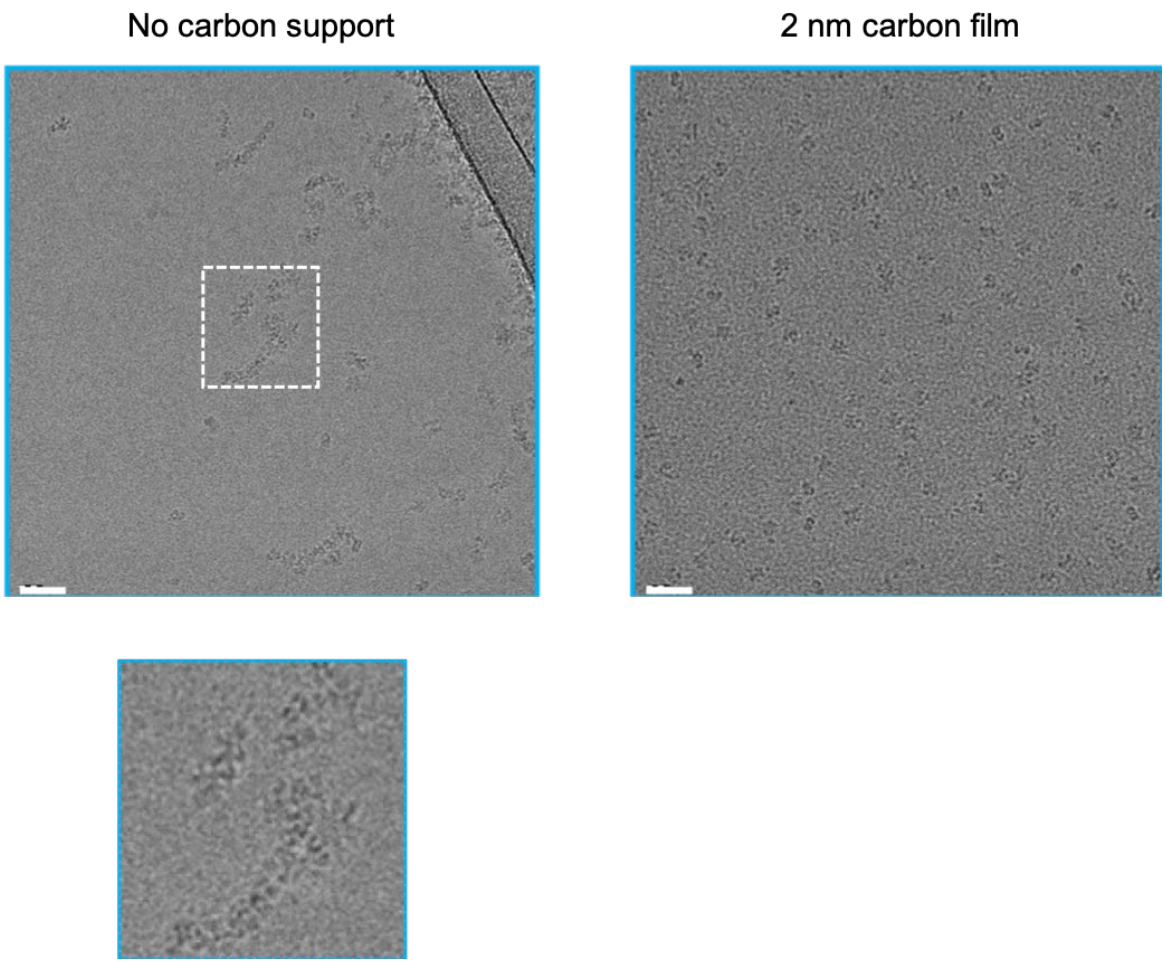


Figure 4.3 Particle distribution of the levan SusCD complex in the absence and presence of an ultrathin carbon support film. Self-association of particles can be seen in the absence of the support (left) while this behaviour is eliminated when the same sample (diluted 200-fold) is applied to a grid with an ultrathin carbon support (right). A zoomed inset of the area within the white dashed box is shown to highlight the extent of aggregation. Scale bars are 50 nm.

A dataset comprising 3276 micrographs was collected. After motion correction and CTF estimation, ~2000 particles were manually picked, extracted and classified in 2D to provide templates for automated particle picking. Using these templates, ~775,000 particles were picked and extracted. Unfortunately, auto-picking was not stringent and, despite attempts to optimise the picking parameters, images of carbon edges were prevalent in the extracted particle stack. In order to efficiently remove these 'false' particles, an initial classification in 2D was performed with restrictions on search space during image alignment (see **Methods 2.8.4**) to remove the majority of carbon edges resulting in a culled particle stack of ~575,000

particles. These were subjected to further rounds of more stringent 2D classification, after which ~177,000 particles contributed to classes containing high-resolution structural features (see **Figure 4.4a**). Two subsequent rounds of classification in 3D revealed considerable heterogeneity in the dataset (see **Figure 4.4b**).

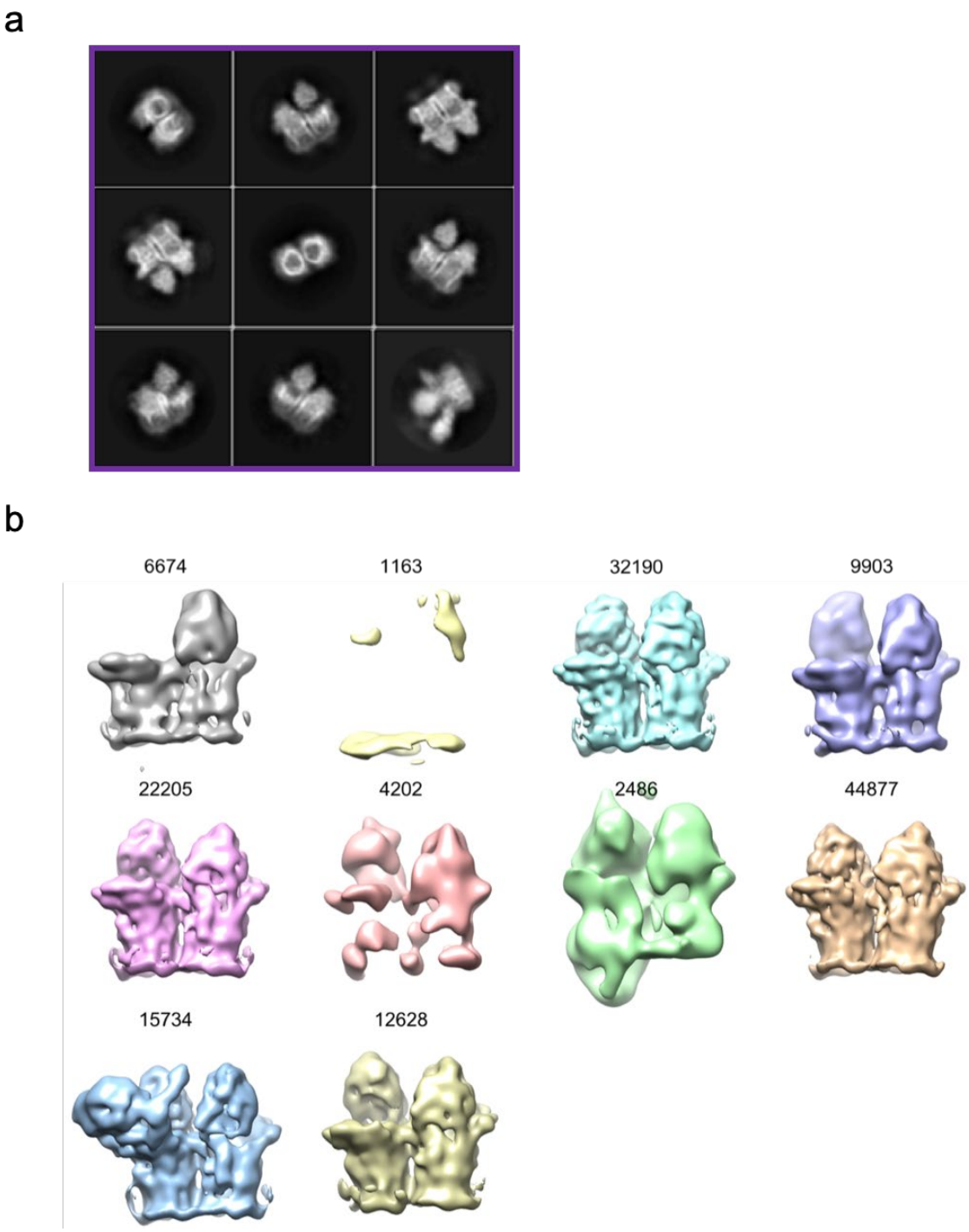


Figure 4.4 Overview of early outputs from image processing of the levan SusCD. a, Representative 2D class averages from the final round of 2D classification showing clear internal structural detail, indicative of good quality data. **b**, Output 3D class averages from the second round of 3D classification demonstrating the considerable amount of heterogeneity in the data. Particle numbers contributing to each class are shown.

4.3.1 Understanding conformational heterogeneity in the levan SusCD complex

Given the apo state of the transporter, the pedal-bin hypothesis would predict that all complexes should occupy an open state. However, as for RagAB, three principle conformational states were observed and taken through to refinement: closed-closed (CC), open-closed (OC) and open-open (OO) (see **Figure 4.5**). Final reconstructions of the OO, OC and CC states yielded global resolutions of 3.9 Å, 4.7 Å and 4.2 Å respectively, with contributing particle numbers of 32,190, 22,205 and 17,416. Only the reconstruction of the OO state was of a suitable quality for model building and refinement. The quality of reconstructions of the OC and CC states was limited by low particle number and a sub-optimal distribution of viewing angles respectively. In the open states, SusD has undergone a rigid-body hinge-like movement relative to its position in the closed state, although this movement is smaller than that observed for RagAB. Strikingly, where lids were closed, density corresponding to the plug domain was not observed within the SusC barrel (see **Figure 4.5**).

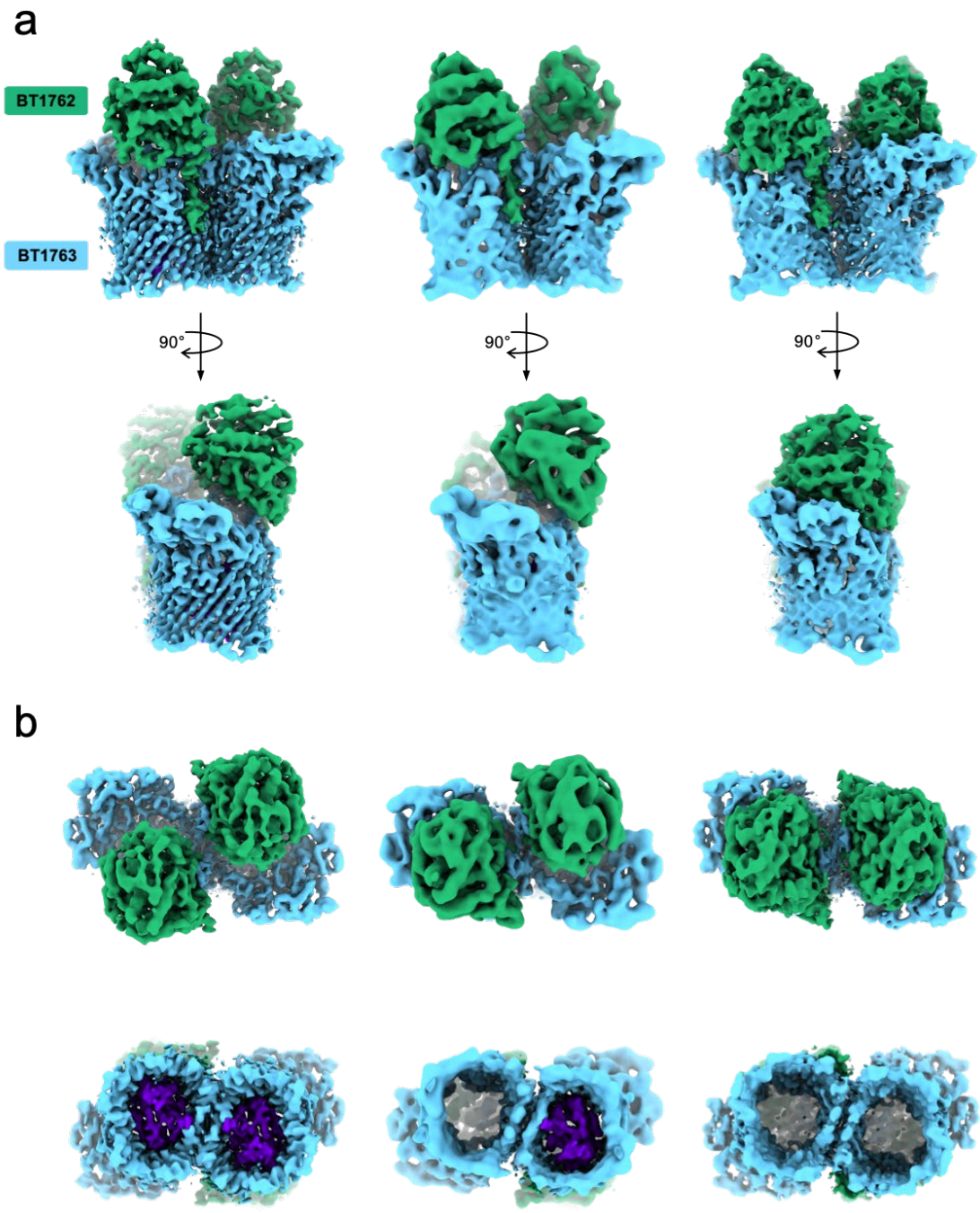


Figure 4.5 Three predominant conformational states of the levan SusCD complex observed by cryoEM. a, Local resolution filtered reconstructions of the SusCD complex in the open-open (left panels), closed-open (middle panels), and closed-closed states (right panels) viewed in the plane of the membrane. **b**, Local resolution filtered reconstructions of the complex viewed from the extracellular space (upper panels) and from the periplasm (lower panels). Clear density for the plug domain (purple) is invariably associated with the open state of the transporter. Where the lid is closed, no density is observed within the SusC barrel.

While working to obtain homogenous particle stacks for refinement, the double closed state that was identified in global 3D classification was found to possess weak density for the plug domain within one of the SusC barrels when the map was viewed at low threshold levels (see **Figure 4.6**). It was proposed that the relatively weak density for the plug was indicative of heterogeneity in the particle stack contributing to the reconstruction and, as such, a focussed classification approach was employed. Here, a mask encompassing only the barrel containing the weak plug density was constructed. This mask was then supplied in a 3D classification job which was run without alignment, allowing separation based on presence or absence of the plug domain. Classification was successful and separated out distinct populations of plugged and unplugged barrels. Independent refinement of the corresponding particle stacks revealed a 'true' double closed state and a novel partially open state (see **Figure 4.6**).

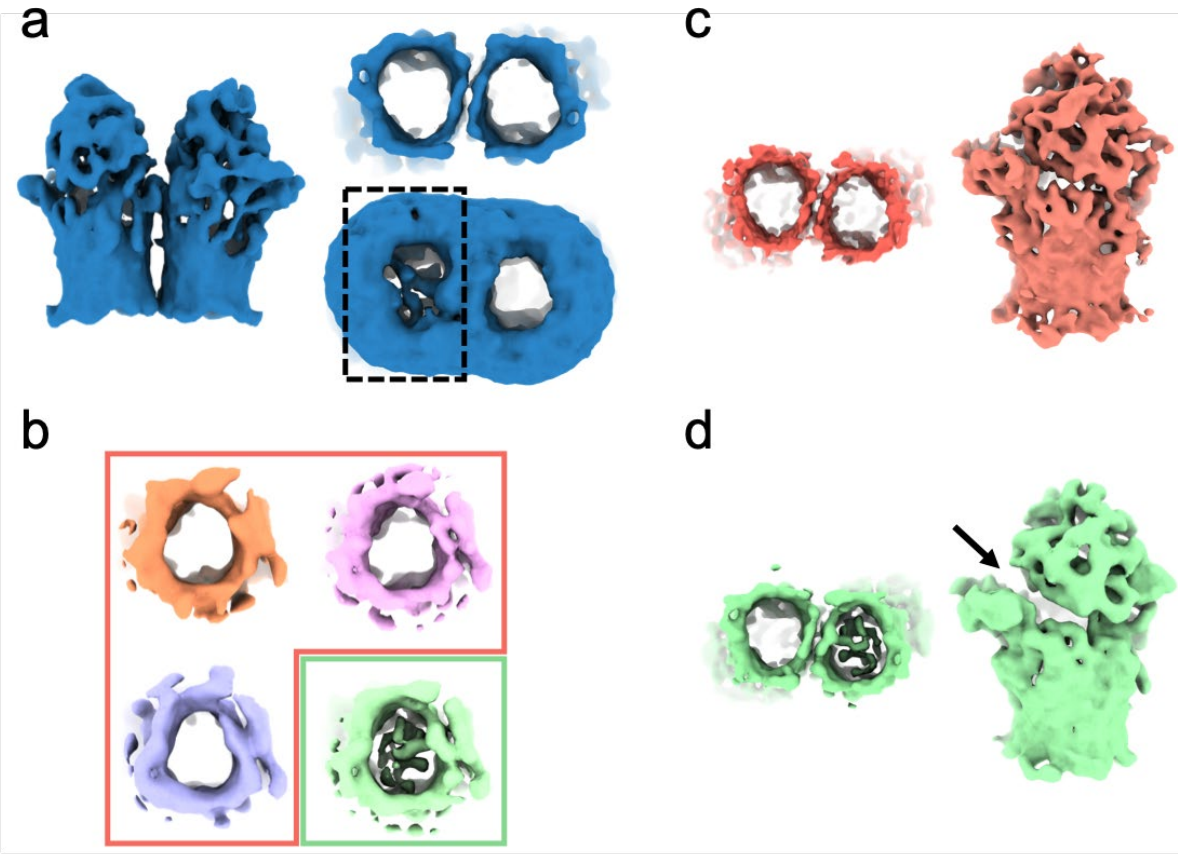


Figure 4.6 Focused classification of the levan SusCD closed-closed state. a, The closed-closed state of the transporter observed after global 3D classification viewed in the plane of the membrane (left) and from the periplasm (right). Note that at higher contour levels, density corresponding to the plug domain can be observed in just one of the two SusC barrels. A mask encompassing only the plugged barrel (area marked with black dashed box) was generated and applied in focused 3D classification without alignment. **b**, Results of the focused classification demonstrating clear separation of particles into plugless (orange box) and plugged (green box) classes. Empty and plugged particles were independently classified again in the absence of a mask. The particles contributing to the best-looking classes were refined and post-processed resulting in the reconstructions shown in **c** and **d**: A 'true' closed-closed state lacking density for the plug domain in either barrel (**c**), and a closed-barely open state where plug density is associated with a partially open state of the transporter (**d**).

Moreover, in addition to the predominant open position of SusD described above, a less populated wide-open state was observed. As for the other open states, density for the plug domain is clearly visible within the barrel of the corresponding SusC subunit. The position of SusD in this state is more akin to the position of RagB in the open state of the RagAB complex. Thus, under the conditions used, the energy landscape for the opening and closing of the levan-transporting SusCD complex appears to be rugged, with the resolved conformational states corresponding to local energy minima (see **Figure 4.7**). By extension, the greater

number of conformational states observed for the levan system relative to RagAB suggests a greater number of minima in the energy landscape of the former.

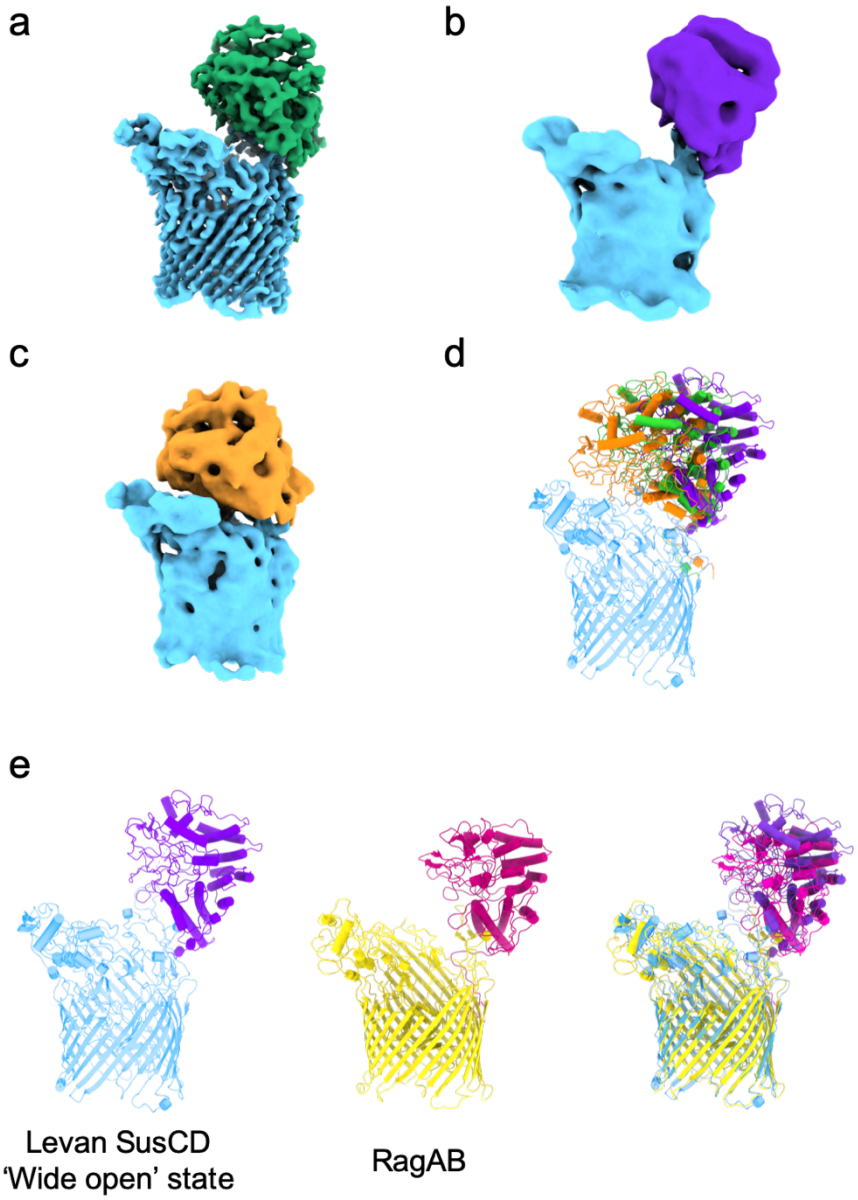


Figure 4.7 Variability in the position of Bt1762^{SusD} observed in the open states of the transporter. **a**, Predominant open position observed in both the principle open-open and open-closed states. Less populated classes with wide open (**b**) and barely open (**c**) lid positions were also observed. EM density is filtered by local resolution and monomers are shown for clarity. **d**, Overlaid rigid-body fits of SusD into the maps shown in **a**-**c**. **e**, Comparison of the 'wide open' state of levan SusCD (left) with the open state of RagAB (middle). The angles of opening are comparable and more obtuse than that of the principle open state of the levan SusCD transporter.

4.4 X-ray crystallography of the levan transporting SusCD complex in the absence of substrate

In contrast to the results observed by cryoEM, an X-ray crystal structure of the apo levan SusCD transporter determined by collaborators at Newcastle University was in the closed-closed state. This is not unexpected as open states for SusCD transporters have never been observed by X-ray crystallography; an observation attributed to the fact that crystallisation favours less hydrated and more compact conformations i.e. lids closed (Glenwright et al., 2017). Electron density for the plug domain in this apo closed state is poor but unambiguously present, indicating that this region is perhaps more mobile (see **Figure 4.8**). Given the cryoEM data, it is also possible that some of the plugs have been ejected from the barrels. The crystals have a high solvent content (~63%) and assessment of crystal packing shows large spaces beneath the SusC barrels that could plausibly accommodate ejected plug domains. However, it should be noted that evidence of the plug domain outside of the barrel was not observed by either X-ray crystallography or cryoEM.

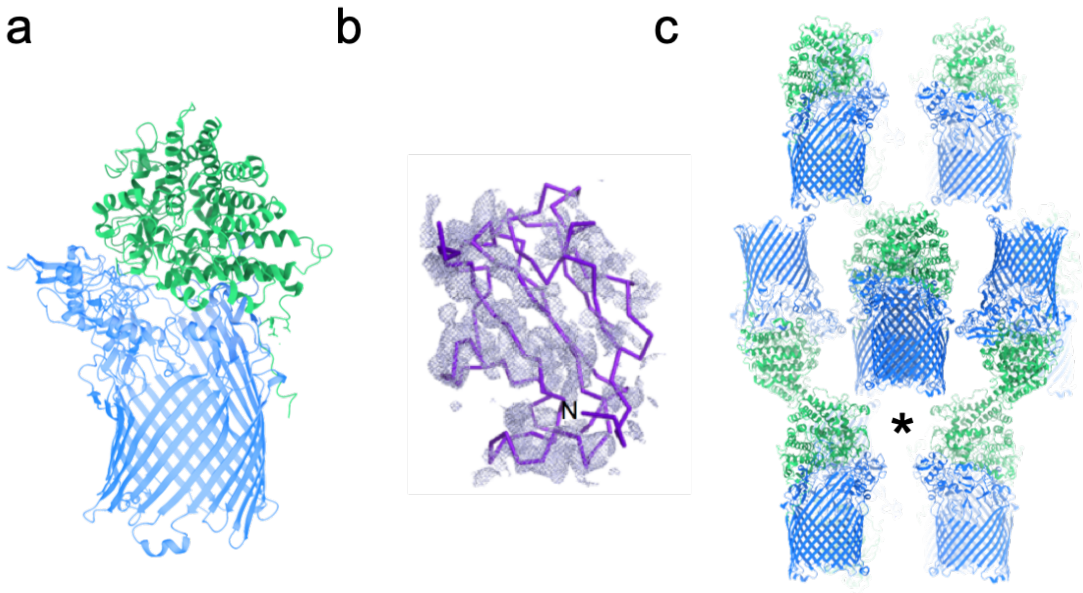


Figure 4.8 Crystal structure of the apo levan SusCD complex. a, Cartoon viewed in the plane of the membrane. A monomer is shown for clarity. Note that the plug domain is absent from the model owing to poor density in this region. **b**, Poor density observed for the plug domain of the apo transporter. A Cα ribbon of the plug domain from the cryoEM structure is superimposed (purple). **c**, Crystal packing arrangement of the apo SusCD transporter. An asterisk indicates the large spaces beneath the SusC barrels that could plausibly accommodate ejected plug domains. Figure adapted from Gray et al., 2021 Supplementary Figure 2. X-ray crystallography experiments were performed by the van den Berg group, Newcastle University.

4.5 Co-crystallisation with FOS reveals SusCD residues involved in levan binding

The levanase (Bt1760) had previously been shown to be essential for the growth of *B. theta* on levan, responsible for generating transport-competent FOS from long chain levan in the growth medium (Sonnenburg et al., 2010). Thus, collaborators at Newcastle University generated suitable substrates for co-crystallisation experiments by partial digestion of levan from *Erwinia herbicola* using the levanase. FOS of varying degrees of polymerisation (DP) were separated by SEC and characterised by thin-layer chromatography (TLC) and MS. Through co-crystallisation of the levan SusCD complex with β2,6-linked FOS (DP6-12) collaborators at Newcastle University were able to determine the structure of the substrate-loaded state of the transporter (see **Figure 4.9**).

Consistent with the position of peptide binding in RagAB and the Bt2261-64 complexes, an FOS is positioned at the top of the solvent excluded cavity, interacting with both the SusC and SusD components of the transporter (see **Figure 4.9**). A stretch of six β2,6-linked fructose units was assigned to the density, with the FOS adopting a twisted conformation similar to that observed for levantetraose bound to the levanase Bt1760 in a previously determined crystal structure (PDB ID: 6R3U). Residues from both SusC and SusD are involved in numerous polar contacts with substrate hydroxyls while stacking interactions are provided by W85 of SusD. Notably, the bound FOS possesses a β2,1 decoration on fructose 4. Whether this is evidence of preference for branched substrate molecules or a result of the source levan (from *Erwinia herbicola*) containing a high proportion of β2,1-linkages (~10%) is uncertain (Benigar et al., 2014). The branch point is accommodated by a non-polar surface provided by a vicinal disulphide between Cys298 and Cys299 of SusD.

Surprisingly, a second FOS molecule is also evident in the structure. Occupying a position at the bottom of the SusC binding cavity, this FOS contains 4 fructose units and makes contacts with the plug domain of SusC (see **Figure 4.9**). To best fit the density, this second FOS was modelled as a 3-mer with a terminal β2,1-branch rather than a stretch of 4 β2,6-linked fructose units.

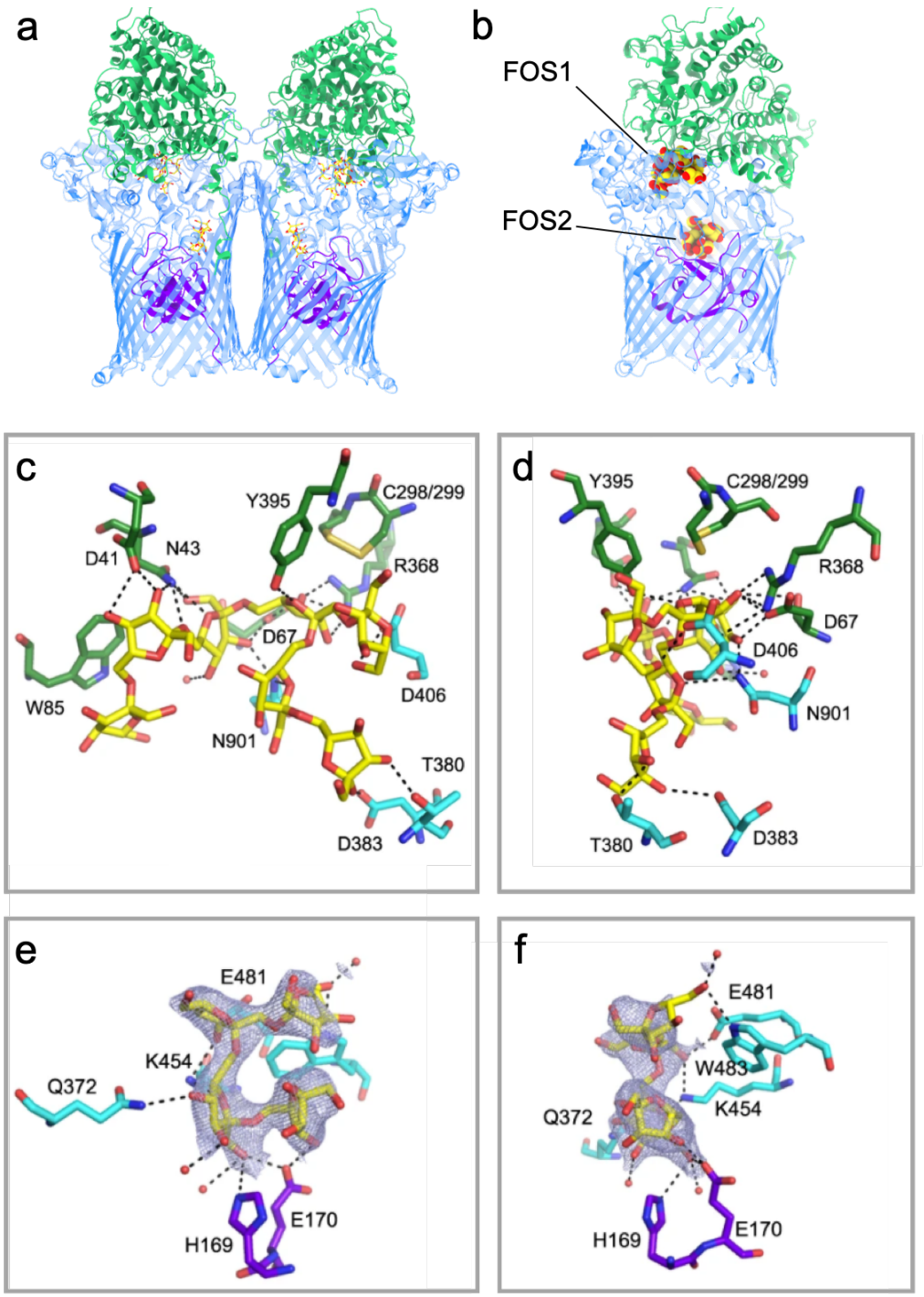


Figure 4.9 X-ray crystal structure of the levan SusCD transporter in the presence of β2,6-FOS. a, Cartoon representation of the $SusC_2D_2$ complex viewed in the plane of the membrane with bound FOS molecules shown as stick models. The SusC plug domains are shown in purple. **b**, Side-on view of a single SusCD unit with the two bound FOS molecules shown as spheres to more clearly demonstrate the two binding locations. **c**, Close-up of the FOS molecule bound in the upper site (designated FOS1) and rotated 90° (**d**). **e**, Close-up of the FOS molecule bound in the lower site (designated FOS2) and rotated 90° (**f**). Interacting residues are shown as stick models. Polar interactions are denoted by dashed lines and water molecules are shown as red spheres. Figure adapted from Gray et al., 2021 Figure 2. X-ray crystallography experiments and structure determination were performed by the van den Berg group, Newcastle University.

4.6 Investigation of the substrate size range of the levan-transporting SusCD complex by ITC

Substrate binding by SusCD transporters involves closing of the SusD lid and, consequently, encapsulation of the nutrient to be transported. Furthermore, surface-exposed endo-acting glycosidases are crucial for the generation of transport competent substrates in the levan and starch-utilising PUL systems of *B. theta* (Shipman et al., 1999; Sonnenburg et al., 2010) and in xylan utilisation by *B. ovatus* (Rogowski et al., 2015). Both lines of evidence indicate the presence of an upper size limit for substrate transport. To investigate the size range of compatible substrates, collaborators at Newcastle University assessed binding of the levan SusCD to FOS with differing degrees of polymerisation (DP) using ITC.

FOS of DP5 were the smallest substrates for which binding was observed and also possessed the lowest binding affinity (K_d ~ 30 μM) (see **Figure 4.10**). Affinity increased with substrate size from DP6 to DP8 (K_d 17 μM – K_d ~ 1 μM). Larger FOS, up to at least ~DP14, bind with similar affinity to DP8.

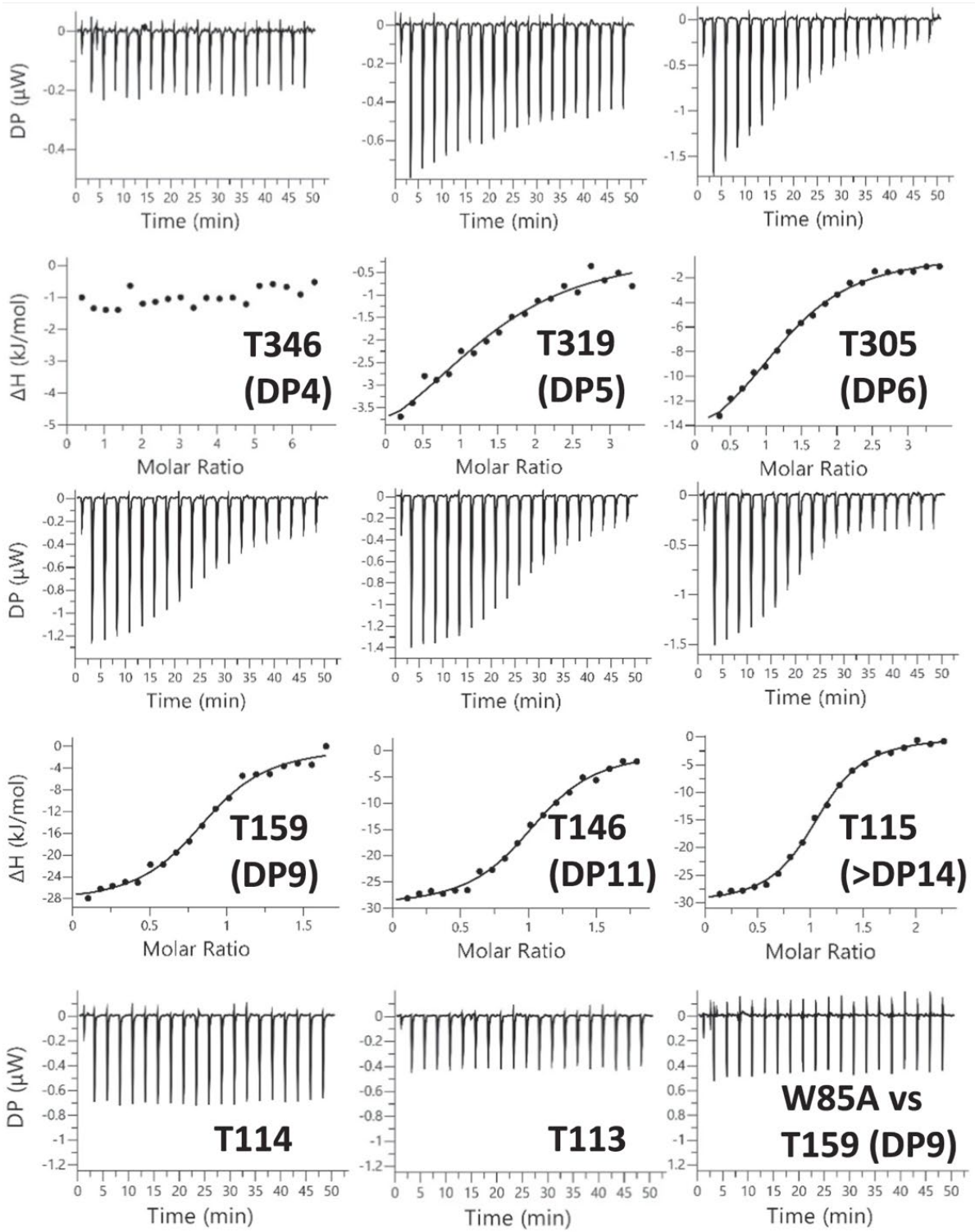


Figure 4.10 FOS binding to the levan SusCD complex analysed by ITC. Fractions from size exclusion chromatography of partially digested *Erwinia* levan were tested for binding affinity to the SusCD complex. The fraction identity is shown (e.g. T346 = Tube 346) above the approximate DP of the predominant FOS in the fraction, as determined by MS. W85A in the bottom right panel corresponds to a SusCD complex in which SusD carries the W85A mutation. The upper panels show raw heats of injection and lower panels show integrated heats fit to a one set of sites model if binding was observed. Figure taken from Gray et al., 2021 Figure 6. ITC experiments were carried out by Declan Gray, Newcastle University.

Investigation of substrates larger than this is hampered by the fact that they are difficult to characterise. FOS >DP14 do not move from the origin in TLC, and poor ionisation efficiency of long oligosaccharide molecules makes analysis of FOS >DP12 difficult by MS. What is clear is that SEC fractions corresponding to larger FOS species contain a broad range of oligosaccharide sizes. Fraction 115 from SEC contained FOS of at least ~DP14 that were shown to bind by ITC. Surprisingly, fractions 113 and 114 did not exhibit binding despite them both possessing similar MS profiles to fraction 115, with a broad range of FOS sizes present (see **Figure 4.11**). The average molecular weight of FOS in fraction 114 is larger than that of fraction 115 which may explain the lack of binding. However, it is also possible that the population of β2,1 branched substrate varies substantially between the fractions; a property that cannot be assessed by TLC or MS. The observation of β2,1 decorations on SusCD-bound substrates in the crystal structure indicates that this feature should not be overlooked.

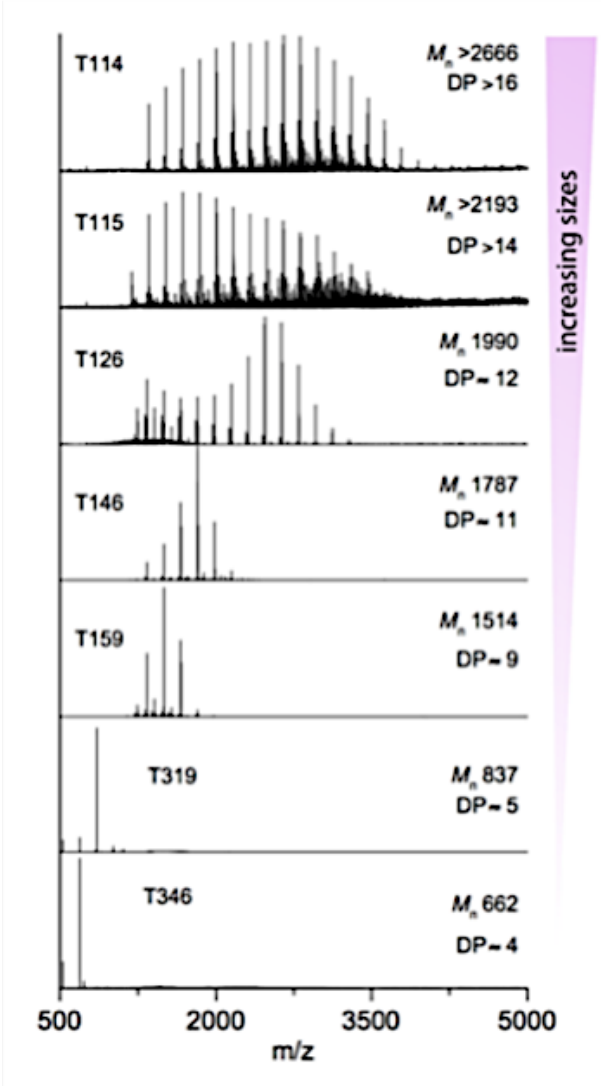


Figure 4.11 Characterisation of β2,6 FOS fractions by MS. Representative mass spectra of FOS from SEC fractions of partially digested *Erwinia* levan. The average molar mass (M_n), based on the intensity-weighted signals, is shown. Higher DP fractions show increased polydispersity relative to lower DP fractions. Figure adapted from Gray et al., 2021 Supplementary Figure 8. Mass spectrometry experiments were carried out by Abraham Oluwole, University of Oxford.

4.7 Investigation of substrate binding by the levan-transporting SusCD complex by native MS

FOS binding to the levan SusCD complex was also assessed by native MS; carried out by collaborators at the University of Oxford. For the apo state of the transporter, charge states corresponding to the expected $SusC_2D_2$ dimer were observed, along with low intensity peaks for monomeric SusCD. Notably, addition long FOS e.g. > ~DP14 resulted in an increase in intensity of monomer peaks, suggesting that interaction with the substrate results in

destabilisation of the dimer (see **Figure 4.12**). Despite this, low intensity adducts corresponding to substrate-bound $SusC_2D_2$ dimers were observed upon addition of either ~DP14 or ~DP16 FOS, corresponding to fractions 114 and 115 of SEC respectively. The broad range of FOS sizes in these fractions is reflected in the complexity of the binding pattern in the corresponding spectra. Binding of FOS from SEC fraction 114 observed by MS was not detected ITC studies. This discrepancy is attributed to differences in experimental setup, where the FOS:protein ratio used in MS experiments was 3-fold higher than in the ITC experiments.

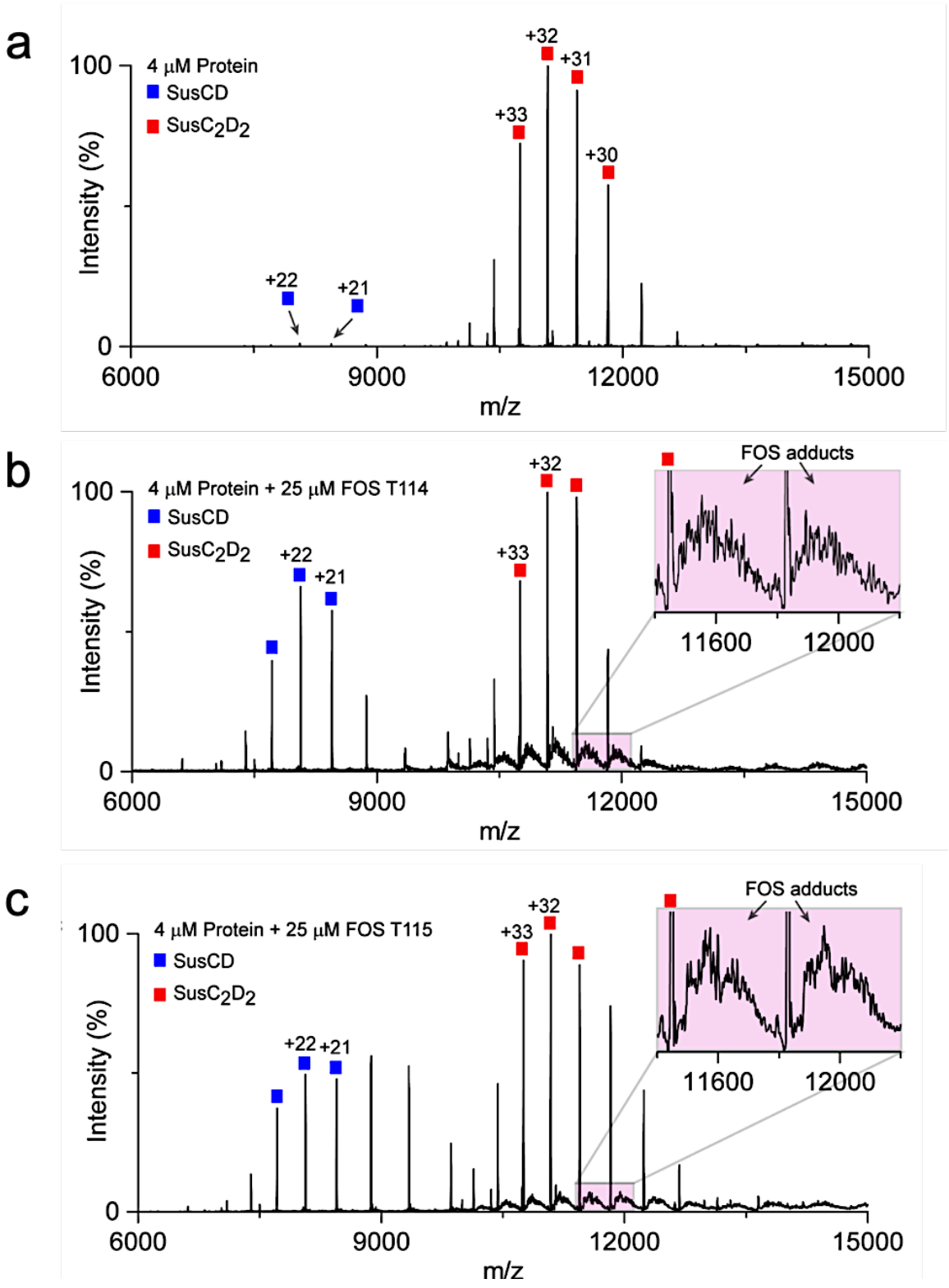


Figure 4.12 Native MS analysis of high DP FOS binding to levan SusCD. a, Native mass spectrum of the levan SusCD complex in the absence of FOS showing low intensity monomer peaks and higher intensity peaks corresponding to the $SusC_2D_2$ dimer. Addition of high DP FOS from either T114 (**b**) or T115 (**c**) results in an increase in monomer peak intensity as well as the emergence of FOS adducts. Thus both fractions tested exhibit some binding to the $SusC_2D_2$ complex. Figure reproduced from Gray et al., 2021 Supplementary Figure 10. Native mass spectrometry experiments were carried out by Abraham Oluwole, University of Oxford.

Repeating the experiment with shorter FOS ~DP9 (7-10 fructose units) showed preferential binding to the $SusC_2D_2$ dimer over the monomer such that no substrate-free dimer was observed (see **Figure 4.13**). Moreover, at higher FOS concentrations, charge states corresponding to $SusC_2D_2$ bound to 1-4 FOS molecules emerged, corroborating the structural data that show that more than one FOS molecule can occupy the binding cavity of SusCD.

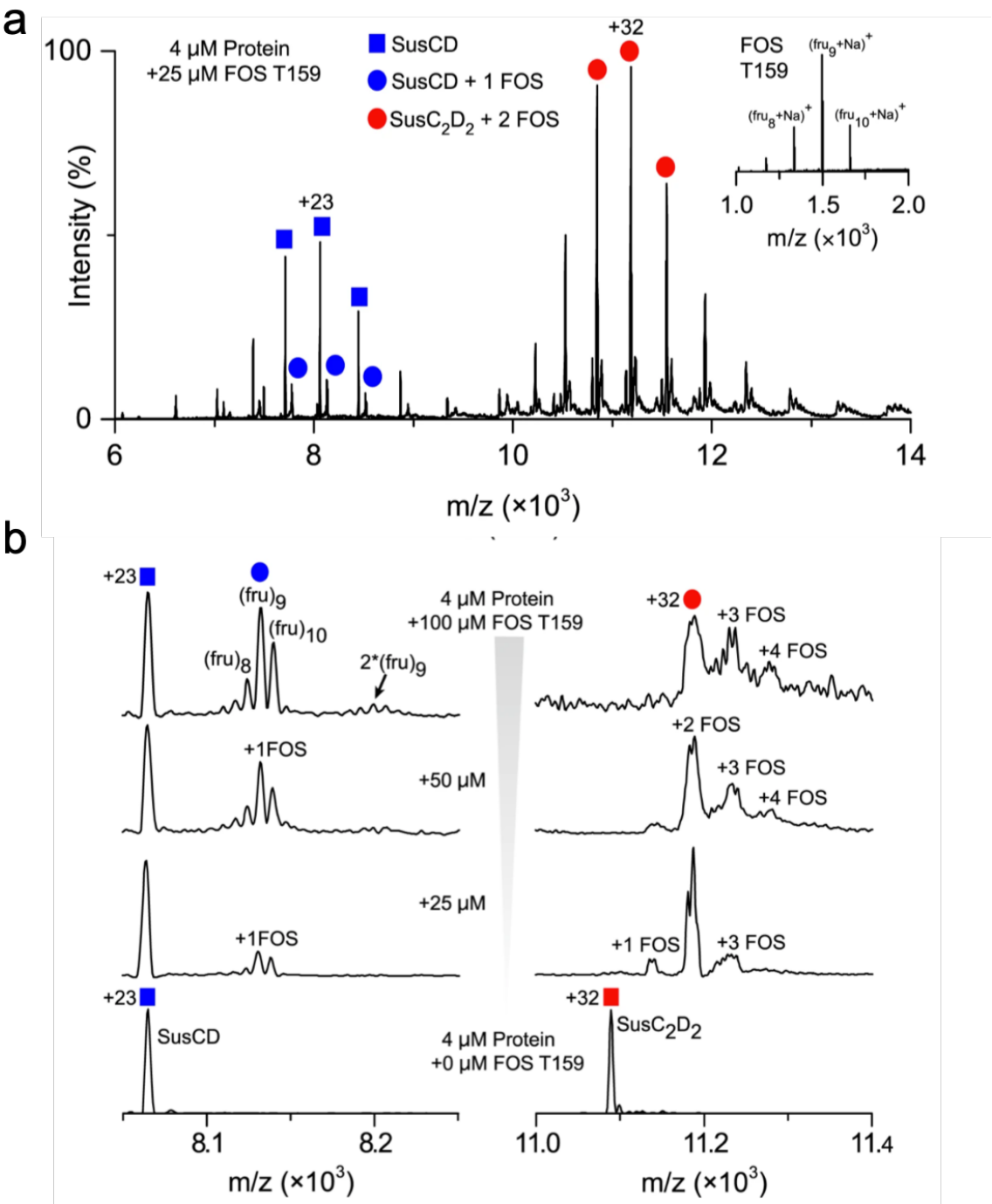


Figure 4.13 Native MS analysis of medium DP FOS binding to SusCD. a, Native mass spectrum of SusCD in the presence of FOS fraction T159 (~DP9) showing both bound and unbound SusCD monomers but only bound $SusC_2D_2$ dimers. The mass spectrum of the T159 fraction alone is shown as an inset (upper right). **b,** Incubation with different concentrations of T159 FOS. At higher concentrations, the intensity of peaks corresponding to dimers with 3 or more bound FOS molecules are increased. Figure reproduced from Gray et al., 2021 Figure 7. Native mass spectrometry experiments were carried out by Abraham Oluwole, University of Oxford.

4.8 Investigating the upper size limit for SusCD-mediated FOS transport

In order to ascertain the upper size limit of FOS transported by the levan SusCD system, collaborators at Newcastle University conducted growth assays on a mutant strain of *B. theta* that lacks the levanase component (Δ1760). Surprisingly, this strain possessed WT-like growth on all levan sources tested. This indicates that all levan sources contained enough low DP/transport competent FOS to sustain growth in the absence of the endo-levanase activity of Bt1760. The spent media from these growth assays was not able to support growth of the Δ1760 strain upon re-inoculation. Unfortunately, technical limitations precluded size characterisation of the remaining high DP FOS by MS or high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD); poor ionisation efficiency, and insufficient separation and detection are the respective limiting factors.

4.9 Structure-function studies on the levan-transporting SusCD complex

Previous work used ITC to examine the roles of individual SusD residues in FOS binding (Glenwright et al., 2017). Alanine substitutions of Trp85 and Cys298 independently abolished substrate binding, while a Tyr395 substitution resulted in a 6-fold reduction in affinity. These observations are concordant with the binding interactions observed for FOS in the substrate-bound crystal structure described above. Here, collaborators at Newcastle University assessed the effect of these mutations *in vivo* via growth assays in which levan was the sole carbon source. The W85A variant showed WT levels of expression and no growth defect was observed for this mutant strain. It therefore appears that this variant is fully functional in the context of the complete transporter. Simultaneous alanine substitution of all SusD residues implicated in substrate binding (D41, N43, W85, C298, R368 and Y395) resulted in a strain ($SusD^{BR}$) with a delayed growth phenotype. The prolonged lag phase (~8 hrs of this strain is comparable to the growth of a strain lacking SusD entirely (ΔSusD) (see **Figure 4.14**). However, this effect is likely due extremely low expression levels relative to WT.

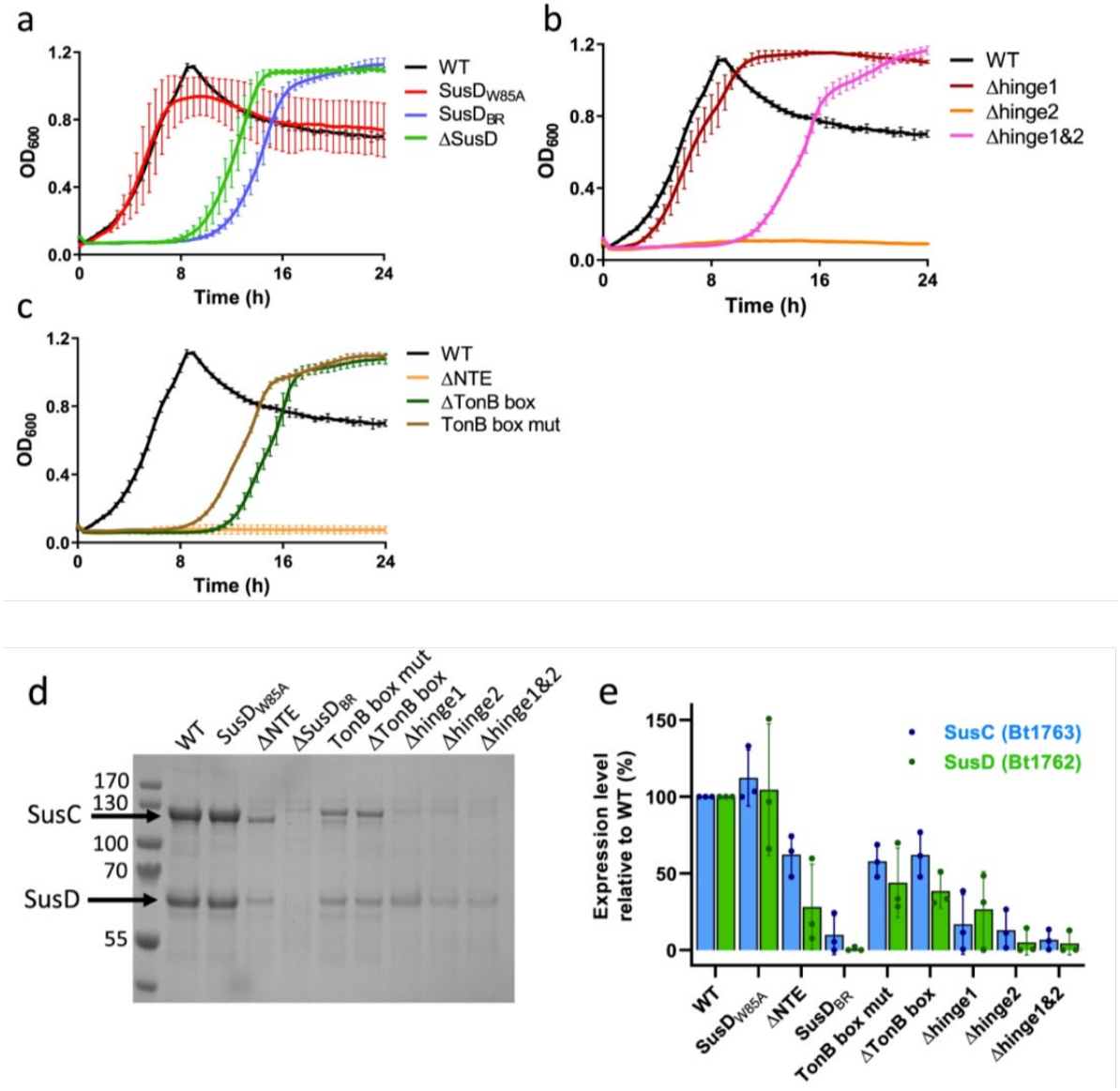


Figure 4.14 Probing the function of the levan SusCD complex by mutational analysis. a-c, Growth curves for SusD mutant strains (**a**), SusC hinge loop mutant strains (**b**), and ΔNTE and TonB box mutant strains (**c**). Growth curves show mean (solid line) ± SD of triplicate wells and are representative of three independent experiments. **d**, Representative SDS-PAGE gel of IMAC purified SusCD complexes from OM fractions of WT and mutant strains grown on minimal medium supplemented with fructose. The MW marker is in the first lane and sizes are shown in kDa. **e**, Relative expression levels from band intensities on SDS-PAGE gels using ImageJ (n = 3 independent experiments; means ± SD are shown, plus individual values). Band intensities of WT SusC and SusD were set to 100%. Figure from Gray et al., 2021 Figure 4. Growth assays and SDS-PAGE densitometry were carried out by Declan Gray, Newcastle University.

Comparison of the open structure determined by cryoEM and the substrate-loaded, closed crystal structure identifies loops 6 and 7 of SusC (L7 and L8) as hinge loops (**Figure 4.15**). Both undergo large conformational changes upon opening of the SusD lid, maintaining contact with it, and comprise the majority of SusC-SusD interactions in the open state. Deletions of one or

both hinges resulted in interesting growth phenotypes (see **Figure 4.14**). A strain in which L7 was deleted (ΔHinge1), showed near WT growth despite expression levels much lower than WT, perhaps indicating enhanced transport activity. An L8 deletion mutant (ΔHinge2) showed similar expression levels to ΔHinge1, but this mutant strain showed no growth over a 24 hr period. Simultaneous deletion of both hinges produced a growth phenotype comparable to ΔSusD i.e. a prolonged 8 hr lag phase. Expression levels of the double hinge mutant were comparable to that of single hinge deletions. On this basis, it is proposed that the ΔSusD-like growth observed for the double hinge mutant is due to uncoupling of SusC and SusD. Without the hinge loops, SusC-SusD interactions are limited to the ~15 N-terminal residues that follow the lipid anchor. Thus, the mobility of SusD relative to SusC is likely much greater, precluding cooperative binding of substrate by the two components. The complete lack of growth observed for the ΔHinge2 strain is attributed to a 'dominant negative' effect on an essential cellular process, independent of the PUL.

Considering SusC, the putative TonB box and the NTE domain were of particular interest. Although SusC-like proteins from *Bacteroides spp.* possess all of the structural hallmarks of TonB-dependent transporters, direct evidence demonstrating the importance of the putative TonB box is lacking for these systems. To demonstrate that the levan transporting SusC is a genuine TBDT, collaborators at Newcastle University constructed two mutant strains in which the putative TonB box (^{81}VDEVVV^{86}) had either been deleted (ΔTonB) or replaced by all-Ala (TonB box mut). Both SusC mutants were reasonably well expressed (~50 % of WT) and the mutant strains exhibit similar growth defects on levan, characterised by a prolonged (~8 hr) lag phase. These data show that Bt1763^{SusC} functions as a genuine TBDT and, importantly, indicate that growth phenotypes with a lag phase of ~8 hrs, like the one seen for ΔSusD, are the result of a non-functional transporter. Furthermore, structural data supports the literature consensus that disordering of the TonB box region acts as a signal for substrate loading of the transporter. In the cryoEM structure of the open state the TonB box (^{82}DEVVVTG^{88}) is visible, and is held within the SusC barrel, interacting with the main body of the plug domain. In comparison, the most N-terminal residue observed in the closed crystal structure is L97, indicating that the preceding section is mobile/disordered and, taken further, is likely to protrude from the barrel where it will be accessible to TonB (see **Figure 4.15**).

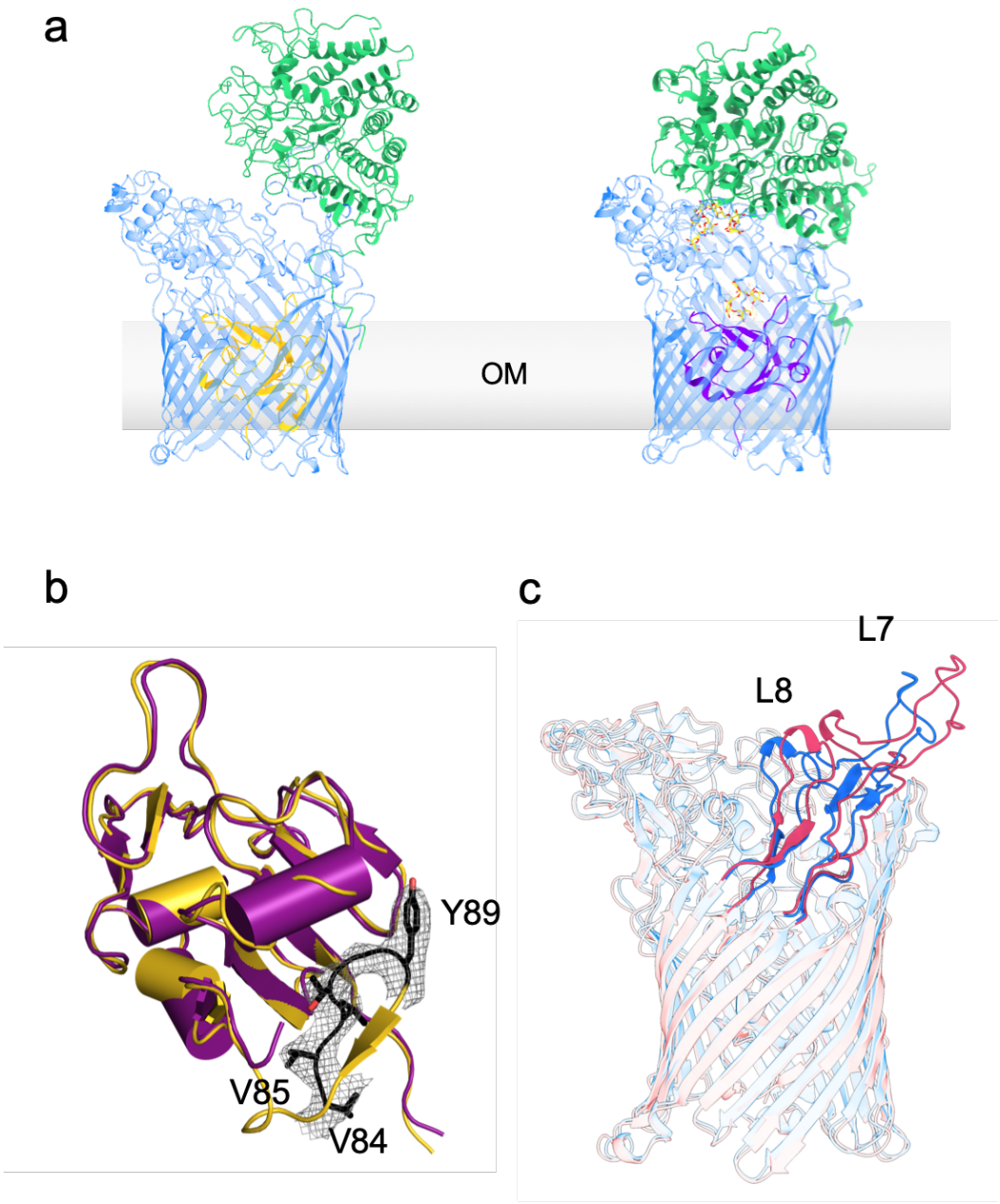


Figure 4.15 Comparison of the open and closed states of the levan SusCD transporter. a, Cartoon representations of the open (left) and substrate-bound, closed (right) SusCD transporter structures derived by cryoEM and X-ray crystallography respectively. For clarity, transporters are shown as monomers. **b**, Overlay of plug domains from the open (gold) and closed, substrate-bound (purple), forms of the transporter. The TonB box (^{82}DEVVVTG^{88}) is visible only in the open state and is coloured black with sidechains displayed as sticks. The equivalent region in the closed structure is not visible and is therefore assumed to be disordered and likely protrudes from the barrel, leaving the TonB box accessible to TonB. Visible density for the N-terminus starts at residue 96 of the 'closed' plug and at residue 84 for the 'open' plug. **c**, Superposition of SusC subunits from the open (red) and closed (blue) states of the transporter. Major conformational changes are localised to the hinge loops which are highlighted and labelled L7 and L8. The X-ray crystal structure presented here was determined by collaborators at Newcastle University.

B. theta contains 121 predicted TBDTs. Of these, only 12 contain just the plug and barrel domains. The rest possess an additional N-terminal extension (NTE) domain located just N-terminal of the TonB box. This domain is annotated as a carboxypeptidase D regulatory-like domain and is ~9 kDa in size but its function is enigmatic. In the case of the levan SusC NTE, deletion mutants (ΔNTE) show no detectable growth on levan within 24 hrs, despite expression levels ~20-50 % of WT. This dominant-negative effect is intriguing but, unfortunately, the NTE domain is not visible in the crystal structures and is present only as diffuse density when viewing cryoEM maps at high threshold levels. Thus, these techniques provide no clues on function from structure. Instead, collaborators at the University of Basel solved the structure of the NTE domain in isolation using high-resolution nuclear magnetic resonance (NMR) spectroscopy of uniformly ^{15}N, ^{13}C-labelled protein expressed in *E. coli*. Complete sequence-specific assignments for backbone resonances were obtained and distance constraints were determined from nuclear Overhauser effect spectroscopy (NOESY) spectra, permitting structure calculations. The NTE domain possesses an Ig-like fold with 7 defined β-strands (see **Figure 4.16**). N and C-terminal regions including the His tag and TonB box respectively are unstructured with no long-range NOEs, and chemical shifts corresponding to random coils. A DALI analysis of this structure returns several hits with Z-scores >9, corresponding to transthyretin-like domains of carboxy peptidase D from various species. The highest Z-score was 10.2 and corresponded to the transthyretin-like domain of human carboxy peptidase D (PDB ID 5aq0). However, this and the other DALI hits offered no insight into the potential function of the NTE domain.

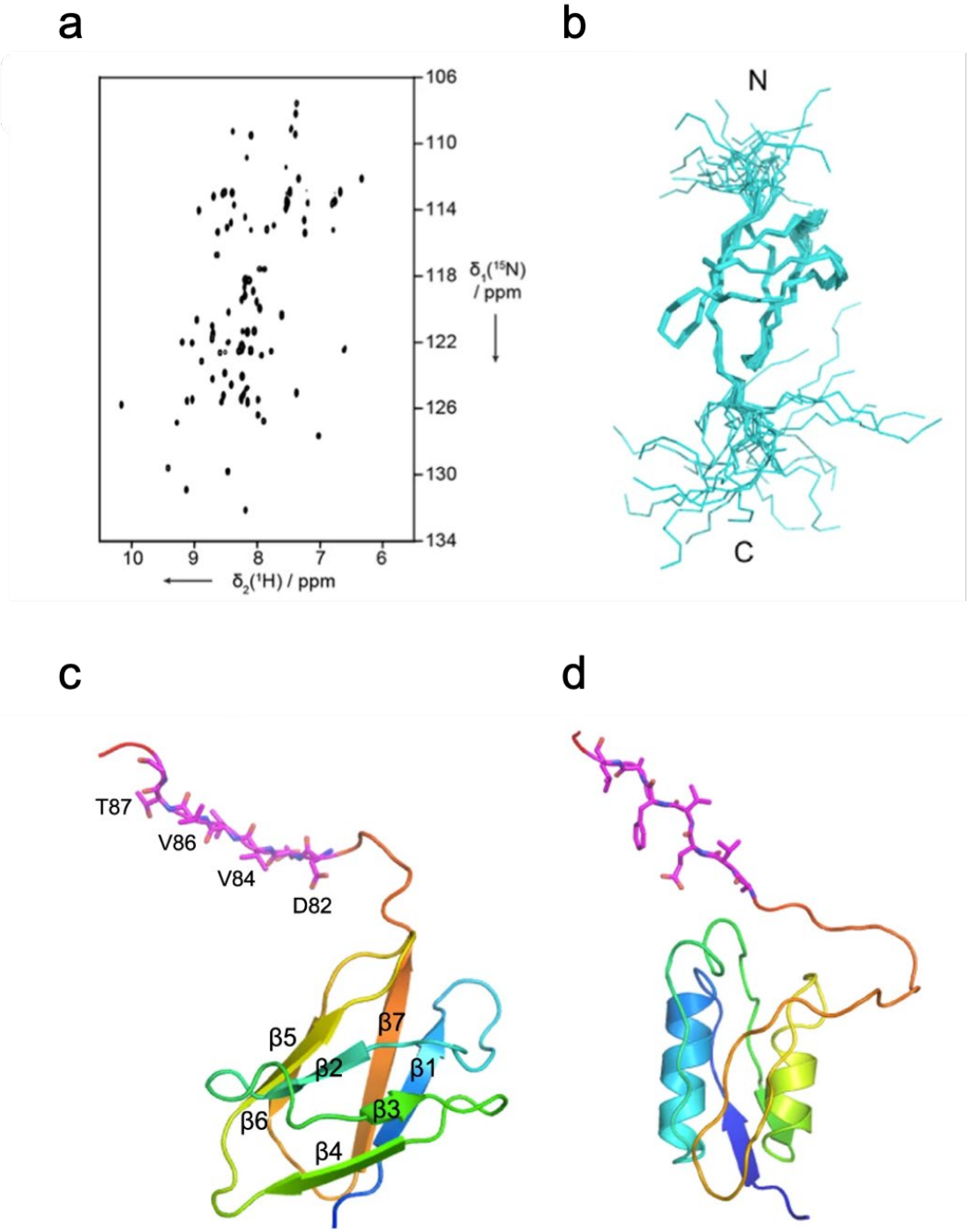


Figure 4.16 Solution NMR structure of the levan SusC NTE. a, 2D [^{15}N, ^{1}H]-HSQC spectrum of the NTE. **b**, Ensemble of the 20 lowest-energy structures of the NTE presented as main-chain stick view. N and C termini are labelled. **c**, Lowest energy NMR structure from the ensemble in panel (**b**). **d**, STN domain of *Pseudomonas aeruginosa* TBDT FoxA (PDB ID 6I97) shown to highlight the difference in structure of NTE and STN domains. The TonB boxes for the NTE and STN domains in **c** and **d** are coloured magenta with sidechains shown as sticks. NMR experiments and data analysis were performed by Adam Mazur and Michael Zahn, University of Basel, Switzerland.

4.10 Investigating the compositional heterogeneity of the levan-transporting SusCD complex observed by cryoEM

Returning to the 3D classes shown in **Figure 4.4**, while the majority of particles correspond to the $SusC_2D_2$ complex, there is some compositional heterogeneity in the data. For example, a small population of particles (<5 %) have an appearance that is consistent with them missing a single SusD subunit. More interestingly, a class representing ~10% of the particles contained density in addition to that expected for the SusC and SusD components. Refinement of this class produced a reconstruction at 7.2 Å and visual assessment of this map suggested that the unassigned density could not be explained by a higher order assembly of either SusC or SusD (see **Figure 4.17**). The sample used for cryoEM grid preparation was analysed by SDS-PAGE. Clear bands were visible for the SusC and SusD proteins, along with two weaker bands running at ~58 kDa and ~45 kDa (see **Figure 4.17**). These molecular weights were reasonably close to those of the lipoprotein components Bt1760 and Bt1761 (respective MW of 57.1 kDa and 48.5 kDa with signal sequences omitted) co-expressed as part of the levan PUL. Excision of these bands and downstream protein ID mass spectrometry confirmed the predicted identity, indicating co-purification of sub-stoichiometric amounts of the levanase and levan-binding proteins with the core SusCD-like transporter. To confirm whether the additional density observed corresponded to one or both of these components, an existing crystal structure of Bt1760 (PDB ID: 6R3R) was docked into the map (see **Figure 4.17**). The model was found to occupy only part of the unassigned density, leading us to tentatively assign the remaining density to Bt1761, of which there was no available structure. Thus, it appears that the surface exposed lipoproteins of the levan PUL are able to assemble on the core SusCD transporter in the absence of substrate.

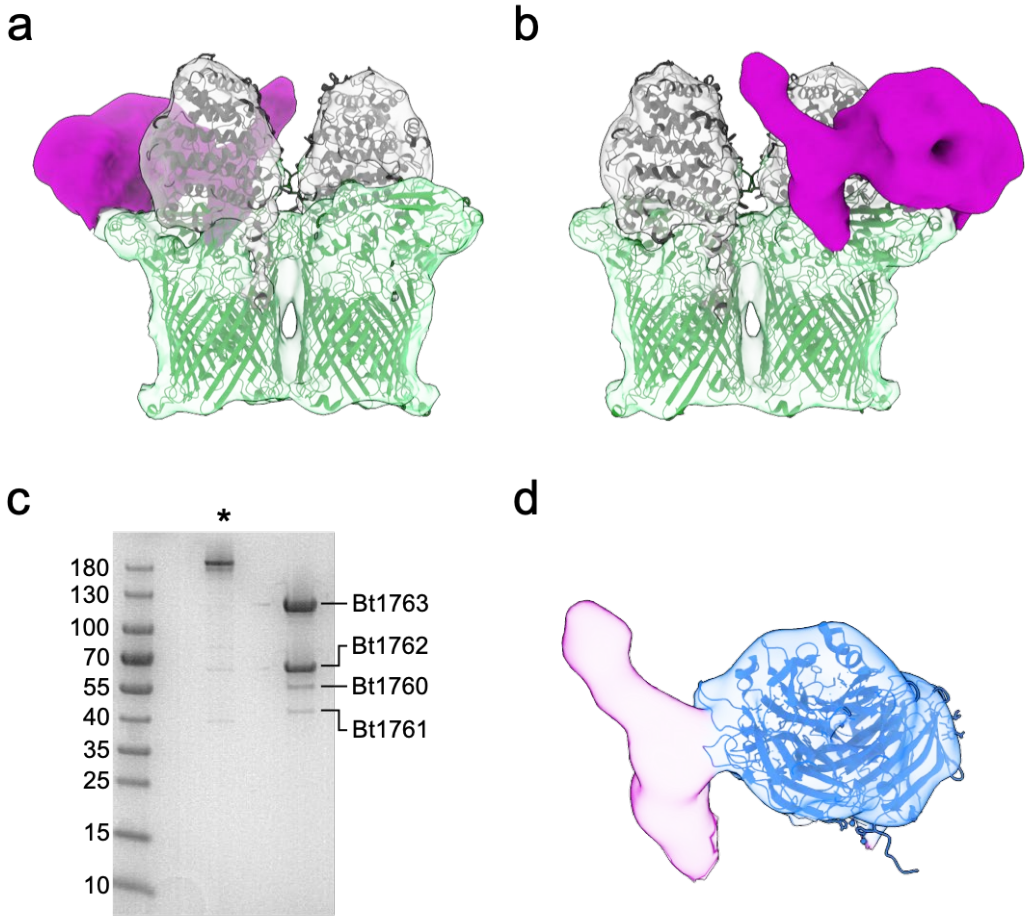


Figure 4.17 Additional density identified in 3D classification of the levan SusCD transporter. a, A class average obtained during 3D classification of the levan $SusC_2D_2$ transport complex. The SusC and SusD components (now coloured green and grey respectively) are docked into the corresponding density. A large region of density remains unassigned (magenta). **b**, as for **a** rotated 180 degrees. **c**, SDS-PAGE of the sample that was applied to cryoEM grids before (asterisk) and after boiling. The boiled sample shows two clear bands in addition to those for Bt1762 (SusD) and Bt1763 (SusC). The additional bands were identified as Bt1760 and Bt1761 by mass spectrometry. **d**, Isolated view of the previously unassigned density with the crystal structure of Bt1760 (63RU) rigid body fit. The fit is reasonable for only part of the unoccupied density, now coloured blue. The remaining density was therefore attributed to Bt1761 and remains magenta.

4.11 Discussion on the levan transporting SusCD complex

Work on the RagAB complex demonstrated that single particle cryoEM is a viable technique for the structural study of SusCD-like transporters. However, the presence of co-purified substrate limited our ability to probe the mechanism of substrate capture in that system. To overcome this limitation, the SusCD-like transporter of the levan utilisation system from *B. theta* was selected for further study. The levan PUL is well characterised and can be upregulated by growing *B. theta* on fructose in the absence of levan. Our primary aim was

therefore to determine the structure of the apo transporter by cryoEM, complementing the work done on the RagAB complex and furthering our mechanistic understanding of this transporter family. This would also lay the foundation for future structural studies on the levan transporter in the presence of substrate.

Based on the pedal-bin hypothesis, it was anticipated that in the absence of levan, the SusD lids of the transporter should uniformly occupy an open state. However, this was not the case with open-closed and closed-closed states also observed for the apo complex. Considering the open states of the transporter, the levan SusCD exhibits greater conformational variability than was observed for RagAB. The reason for this is unclear and it may simply be that these intermediate states are intrinsically more stable in this system. Whether this is of mechanistic importance remains to be seen. Interestingly, the unexpected closed states were invariably associated with a lack of density for the plug domain within the corresponding SusC barrel. An SDS-PAGE analysis of the sample used for cryoEM studies is shown in **Figure 4.17**, and shows a single band for $Bt1763^{SusC}$ that runs at the expected molecular weight for the full-length protein, ruling out proteolytic cleavage of the plug domain that has been reported for this protein previously (Glenwright et al., 2017). It therefore appears that in some instances the plug has exited the barrel and its mobility is such that density is averaged out in the reconstruction. This is, to our knowledge, the most convincing evidence that the plug domain of at least some TBDTs, exits the barrel completely. However, expulsion of the plug is likely non-physiological in our system. LDAO was used as the solubilising detergent in this work and is harsher than the DDM used for structural studies on RagAB owing to its zwitterionic head group. Given that the plug domains of TBDTs are highly solvated and proposed to be conducive to conformational change, it is possible that small, detergent-induced changes to protein structure and/or dynamics result in loss of the plug. Despite this, the fact that expulsion of the plug is invariably associated with a closed conformation of the SusD lid indicates that conformational changes in these domains are intimately linked. Perhaps closing of the lid induces small conformational changes in the barrel that partially disrupt plug-barrel interfaces, facilitating removal of the entire plug domain by TonB, rather than localised unfolding and channel formation as has been proposed for classical TBDTs (Hickman et al., 2017; Gumbart et al., 2007). Studies have shown that TBDT plug domains are stable when expressed alone (Oke et al., 2004) and that for the classical TBDT FhuA, separately expressed

plug and barrel domains are able to form functional units *in vivo* (Braun et al., 2003). It is therefore plausible that re-insertion of the complete plug domain following TonB-mediated removal is a spontaneous event.

It is interesting to consider the role of the SusD subunit in light of the above discussion. Growth assays show that the ability of this subunit to bind substrate is dispensable *in vivo* with the $SusD_{W85A}$ strain exhibiting WT-like growth kinetics. In contrast, a SusD deletion mutant exhibits a clear growth defect consistent with a non-functional transporter. This evidence makes a strong case that the presence of SusD and its ability to function as a lid are integral to the transport mechanism. In the event of complete unplugging, a closed state of the lid would prevent the formation of a large (~20-25 Å diameter) channel, thus preserving the selective permeability of the OM.

Co-crystal structures of the levan SusCD transporter with FOS permitted the first characterisation of glycan binding by members of this transporter family. The longer FOS binds at the top of the solvent excluded cavity, at the interface between SusC and SusD. This position is consistent with the position of bound peptides in both the RagAB and Bt2263-2264 systems (see **Figure 4.18**). The unique observation of a second substrate molecule at the bottom of the levan SusCD binding cavity may provide important insight into the way in which substrate loading of the transporter is signalled to the periplasm. Considering the $Bt2263^{SusD}$-2264^{SusC} system, the SusC plug domain possesses a long loop that contacts the putative peptide in the binding site, providing a route for allosteric communication between the binding site and the TonB box region at the periplasmic face of the plug. Similar extended plug loops are absent in the levan SusC and RagA proteins, raising the question of how substrate-loading induces exposure of the TonB box. In the case of the levan transporter, loading of a second substrate molecule at a site proximal to the plug seems like a plausible solution. However, questions remain about the nuances of the mechanism: Is the binding of one FOS proximal to the plug sufficient to initiate signalling across the OM? Or, is the double bound structure observed here a pre-requisite for initiation of transport? A schematic depicting the working hypothesis is shown in **Figure 4.19**.

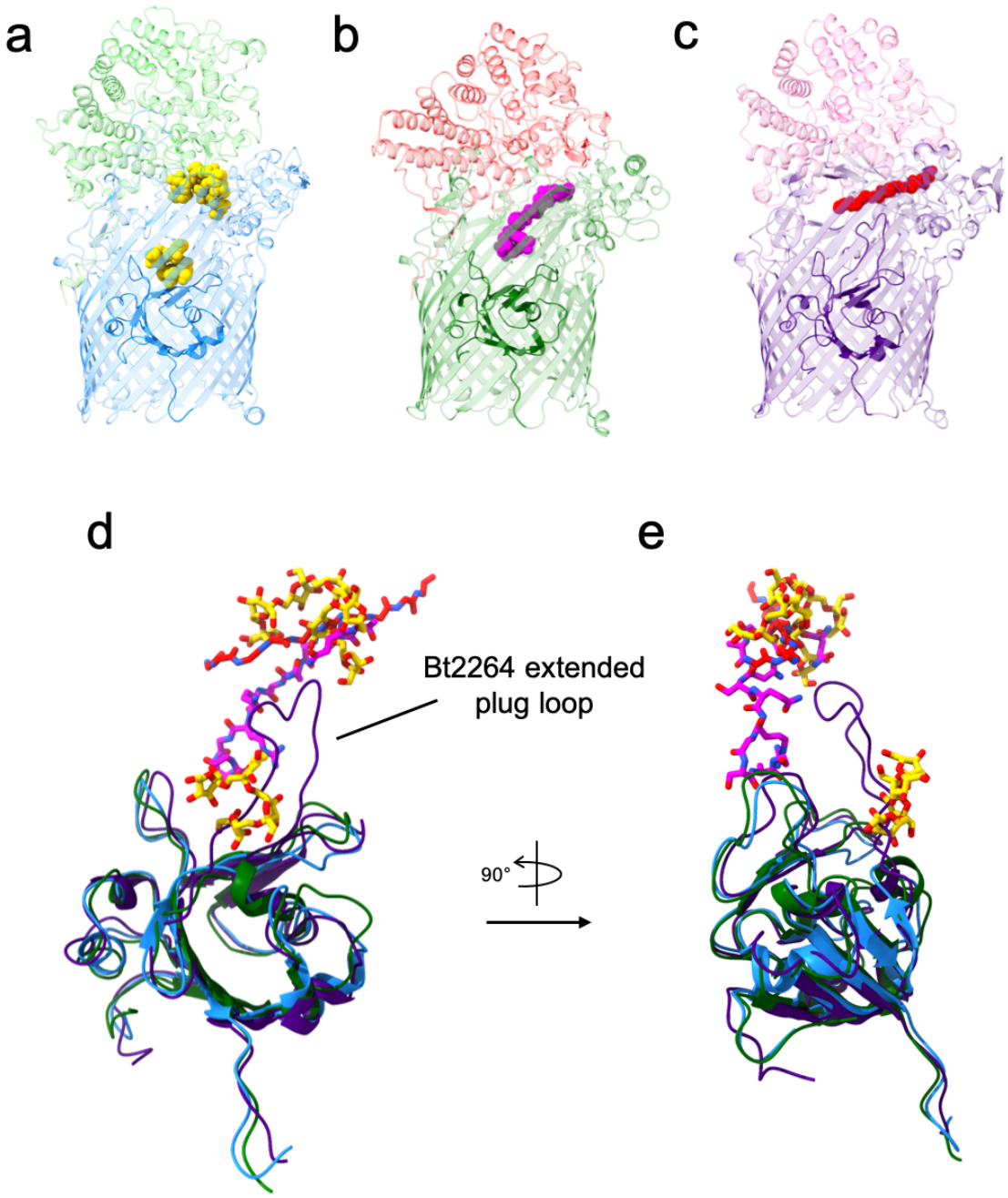


Figure 4.18 Comparison of ligand binding positions in substrate loaded SusCD structures. a-c, Cartoon representations of (**a**) Levan SusCD with bound FOS (yellow), (**b**) *Porphyromonas gingivalis* RagAB with bound peptide (magenta; PDB ID 6SLN), and (**c**) Bt2263-64 with bound peptide (red; PDB ID 5FQ8). **d**, Isolated views of overlaid plug domains and bound substrate molecules from structures shown in (**a**-**c**). The extended plug loop of Bt2264 (absent in the other two transporters) is highlighted. **e**, As for (**d**) but rotated 90° counter clockwise.

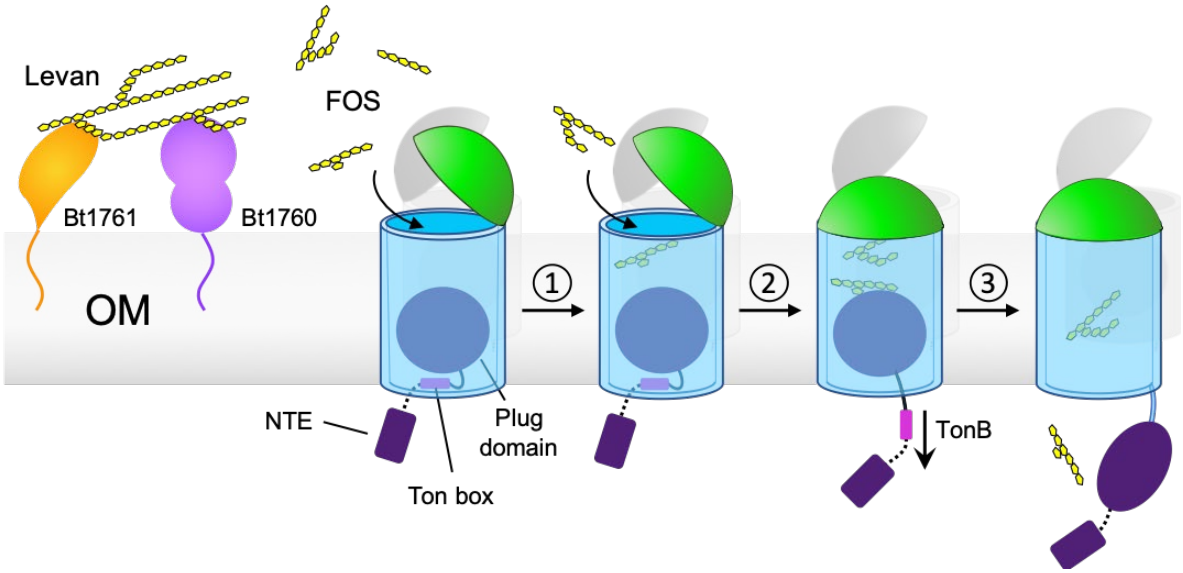


Figure 4.19 Schematic illustrating the working hypothesis for substrate capture and translocation by the levan transporting SusCD complex. Levan polysaccharide is initially bound and hydrolysed at the cell surface by the levan binding SGBP Bt1761 and the GH32 endo-levanase Bt1760 respectively, although the precise role of the SGBP is unclear. ①Contributions from both SusC and SusD to FOS binding elicits closure of the lid. ② Multiple cycles of lid opening and closing occur until the SusC 'bin' is fully loaded with substrate, forming the transport competent state of the complex. Substrate loading is communicated across the outer membrane by an allosteric mechanism likely dependent on a closed lid and direct contact of FOS with the SusC plug domain, inducing perturbation of the TonB box region on the periplasmic side of the plug, rendering it accessible to TonB. ③ TonB-mediated disruption/extraction of the plug permits substrate translocation. Details of the 'reset' mechanism are unclear, but re-insertion of the plug is likely a prerequisite to restoring the open state of the transporter.

Despite discerning its structure, the function of the NTE domain remains an enigma. Domains located just N-terminal of the TonB box are not unique to SusC-like TBDTs. Indeed, so-called STN domains (Secretin and TonB N-terminus domain, Pfam 07660) are present in a number of TBDTs and have a well-defined role in ECF-sigma/anti-sigma regulation, interacting directly with anti-sigma factors at the IM to promote TBDT expression in the presence of their cognate substrates (Noinaj et al., 2010; Garcia-Herrero & Vogel, 2005; Malki et al., 2014; Jensen et al., 2020). Structures of the STN domains of FoxA, FecA, FpvA and HasR reveal a globular topology comprising two α-helices and 5 β-strands, markedly distinct from the NTE domain structure of the levan SusC described above (see **Figure 4.16**) (Brillet et al., 2007; Josts et al., 2019). Moreover, at least 19 TBDTs in *B. theta* possess both an NTE and a preceding STN domain, indicating that these domains are functionally distinct (Pollet et al., 2021).

Finally, the region of unexpected density tentatively assigned to the additional lipoproteins Bt1760 and Bt1761 is important to consider. This data is the first indication that the glycosidase and SGBP lipoproteins of the levan PUL can assemble onto the core SusCD transporter to form a stable complex in the absence of substrate. Given that the combined action of these lipoproteins generates transport-competent substrates, their spatial association with each other as well as with the transporter has important implications on the size of the substrates generated and the efficiency with which they are captured by the SusCD complex. The putative lipoprotein density is associated with just one of the two available SusC barrels, resulting in an asymmetric complex. Given the dimeric nature of the $SusC_2D_2$ core complex, it is likely that the other barrel could accommodate an identical arrangement of lipoproteins and thus the hexameric complex observed is plausibly 'incomplete'. Switching to a milder detergent as well as moving affinity tags to the additional lipoprotein components may aid purification of these higher order complexes. The follow-up work is presented in Chapter 5.

Chapter 5 Outer membrane utilisomes mediate oligosaccharide uptake in gut Bacteroides

5.1 Introduction and build-up work

Data presented at the end of Chapter 4 suggest that the additional lipoproteins Bt1760 and Bt1761 are able associate with the core SusCD transporter of the levan PUL in the absence of substrate. This is an important observation as the organisation of OM-associated PUL components remains one of the major open questions in the field.

Some information is available for the archetypal starch utilisation system. This PUL encodes two SGBPs (SusE and SusF) in addition to the SusCD transporter and surface glycosidase SusG. Seminal studies utilising cross-linking, native PAGE and proteolysis experiments provide evidence for interactions between SusC and SusD, as well as the SusE and SusF components (Cho and Salyers, 2001). Later work involving super-resolution microscopy of live *B. theta* cells reports that SusG is highly mobile, interacting transiently with other Sus complex components, and that colocalization of components is increased in the presence of amylopectin (Karunatilaka et al., 2014). Building on this work, Tuson and colleagues describe a role for SusE and SusF as immobile starch-binding centres, around which the SusCD transporter and the SusG glycosidase can assemble (Tuson et al., 2018). Together, these observations paint a dynamic picture of the OM-associated starch PUL components, characterised by weak or transient interactions of lipoproteins with SusC. This idea is supported by co-immunoprecipitation and downstream quantitative MS experiments on SusE in which co-elution with SusC and SusD components requires cross-linking by formaldehyde treatment (Tuson et al., 2018).

In contrast, the Bt2263^{SusD}-Bt2264^{SusC} complex from an uncharacterised PUL in *B. theta* co-purifies with additional lipoproteins Bt2261 and Bt2262 (Glenwright et al., 2017). Structure determination by X-ray crystallography reveals an octameric complex where the dimeric $SusC_2D_2$ transporter is decorated by the additional lipoproteins which are positioned on the lips of the SusC barrels. The putative substrate for this complex is a peptide or peptide ensemble for which density is visible in the binding cavity created by the lid closed

conformation of the SusC and SusD components. Successful purification of this detergent-solubilised four-component complex strongly suggests that it is stable in the OM. Thus, two opposing models have emerged to describe the organisation of OM PUL components: 1. A dynamic assembly model in which free lipoproteins transiently associate with the core transporter, possibly in response to substrate, and 2. The stable complex model, in which the additional lipoprotein components assemble on the core SusCD transporter and no free pools of additional lipoproteins exist at the OM (see **Figure 5.1**).

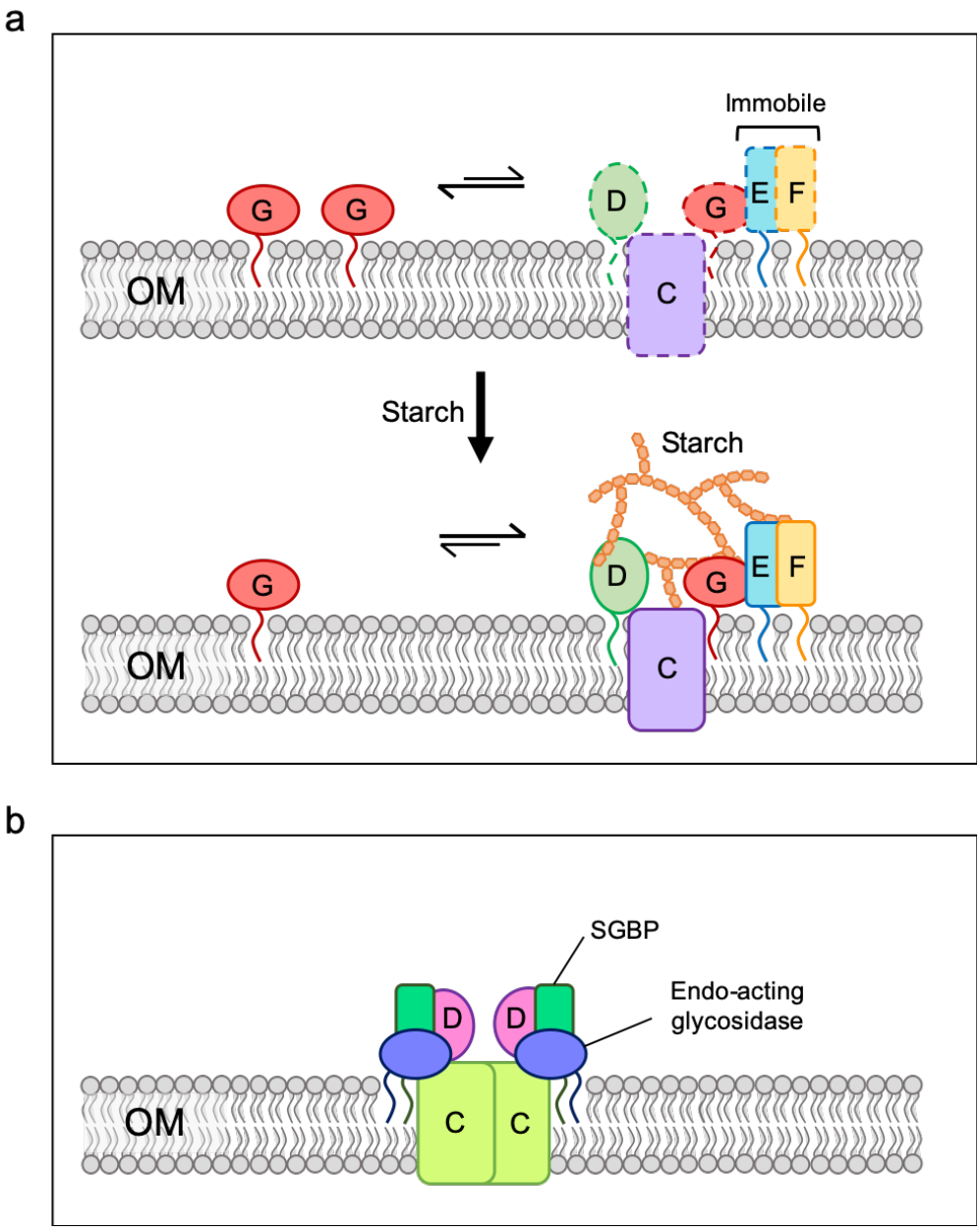


Figure 5.1 Two alternative models for the arrangement of OM-associated PUL components. a, The dynamic assembly model proposed for the starch utilisation system. Here, interactions between OM-associated PUL components are proposed to be transient in the absence of starch substrates (illustrated by dashed borders). Live-cell super-resolution microscopy studies show that, under these conditions, SusG exists predominantly as a highly mobile population, with a smaller proportion exhibiting slow movement that is attributed to complexation with other Sus components. In contrast SusE and SusF have been shown to be largely immobile. The addition of starch results in an increased population of slow-moving SusG suggesting starch-induced complex formation. This complex moves faster in ΔSusD and ΔSusEF cells indicating that these components are involved in this starch-induced complex. **b**, The stable assembly model. Here, the additional lipoprotein components stably assemble on the core SusCD transporter, and no free pools of additional lipoproteins exist at the OM. Evidence in support of this hypothesis is currently limited to the structure of the functionally uncharacterised Bt2261-Bt2264 complex (PDB ID: 5FQ7).

The work in this chapter aimed to investigate the aforementioned hypotheses in the context of well-characterised PULs. More specifically, single particle cryoEM was used to study higher-order assemblies of OM components of the levan PUL and characterise their behaviour in response to substrate addition. The primary aim was to provide a molecular level understanding of the arrangement of components and the way in which they cooperate in the capture, processing and transport of the target glycan across the OM. The dextran utilisation system from *B. theta* was also studied, providing an opportunity for comparison, and with the intention of establishing a general model for glycan utilisation systems in the Bacteroidetes.

Our previous work on the levan transporting $SusC_2D_2$ complex revealed that a proportion of the particles (~10%) gave rise to reconstructions with additional unexpected density that was later assigned to the two additional lipoproteins of the levan utilisation system: the endo-acting levanase Bt1760, and the SGBP Bt1761 hereafter referred to as the levan binding protein (LBP). The levanase is well characterised, with X-ray crystal structures of the enzyme in isolation and bound to levan tetraose available (Ernits et al., 2019). In contrast, no experimentally-derived structure is available for the LBP and no structurally-characterised homologues are present in the PDB. Thus, to provide initial structural insights on the LBP, a predicted model was generated using AlphaFold2. The model possesses a tripartite domain structure comprising two N-terminal Ig-like domains, and a globular C-terminal domain that was hypothesised to be the site of levan binding (see **Figure 5.2**). Density for the additional lipoproteins in our cryoEM maps was associated with just one of the SusC barrels, and this seemed unlikely to represent a 'complete' complex given the dimeric nature of the $SusC_2D_2$ core. Collaborators at Newcastle University therefore adapted the purification protocol to use milder detergents (a mixture of decyl maltoside (DM) and dodecyl maltoside (DDM) rather than the harsher LDAO)) to extract the complex from membranes, and the rest of the purification was also carried out in DDM rather than LDAO. In addition, the His-tag was moved from the C-terminus of SusD to the C-terminus of the LBP in order to allow the purification of higher order complexes. A qualitative assessment of the newly purified sample by SDS-PAGE indicated the presence of all four components in approximately stoichiometric amounts, within a complex that ran as a single peak in analytical size exclusion chromatography (see **Figure 5.2**).

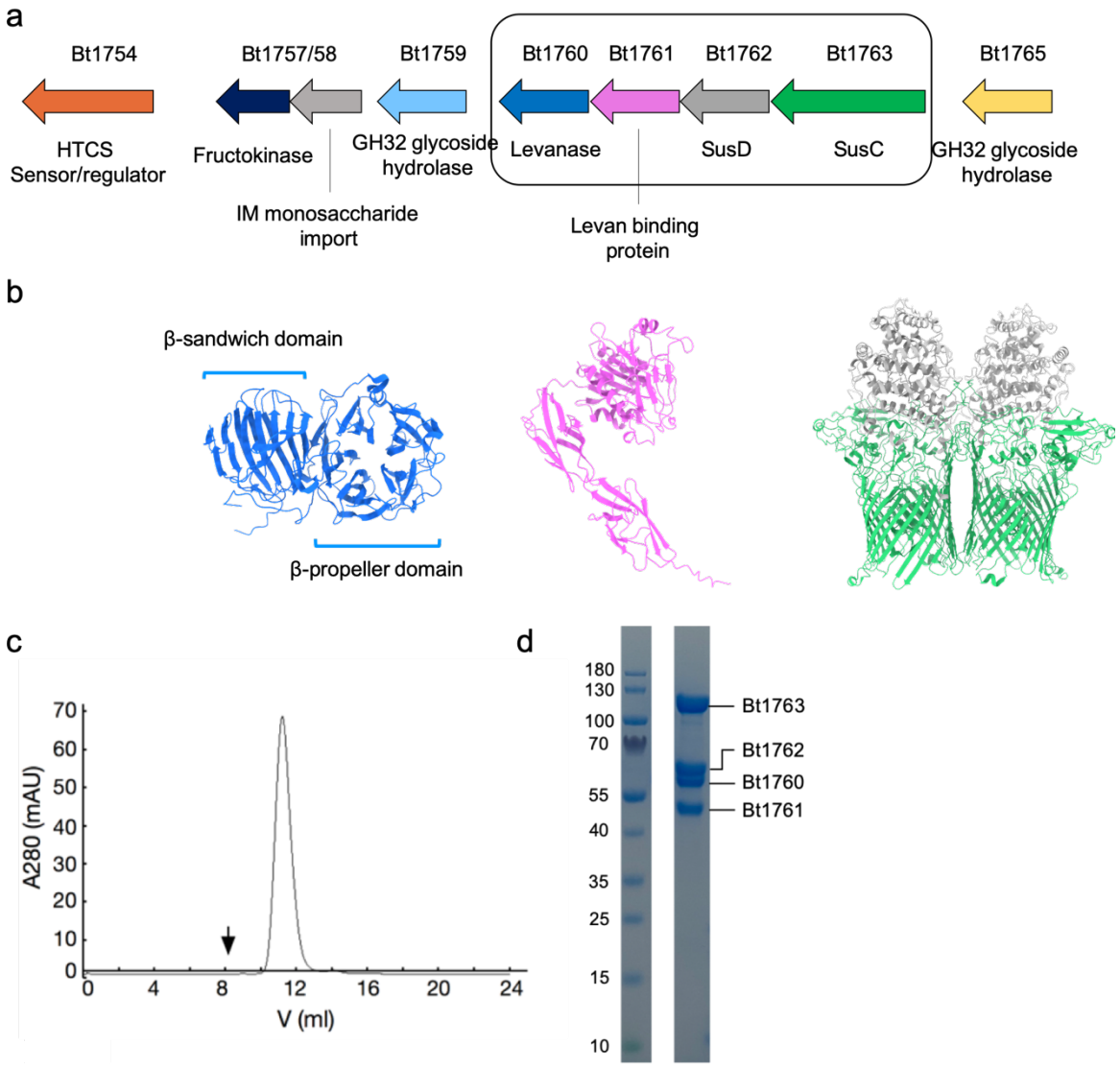


Figure 5.2 OM-associated proteins of the levan PUL. a, Organisation of the levan PUL showing gene positions within the locus with functions labelled. OM-associated PUL components are within the black box. **b**, Experimental and predicted structures of the OM-associated levan components. An X-ray crystal structure of the endo-levanase Bt1760 is shown (blue) with its β-sandwich and β-propeller domains clearly identifiable (PDB ID: 6R3R). The AlphaFold2 predicted model for Bt1761 is shown (pink) oriented such that the N-terminus is at the bottom and the globular C-terminal domain hypothesised to bind levan is at the top. The dimeric SusCD complex structure solved in Chapter 4 is shown with SusC components in green and SusD components in grey. **c**, Analytical SEC elution profile for the Bt1760-36 levan four-component complex run on a Superdex-200 column. The void volume of the column is indicated by an arrow. **d**, SDS-PAGE of the complex purified in (**c**) with the identity of the bands indicated. Analytical SEC and SDS-PAGE were carried out by Bert van den Berg, Newcastle University.

5.2 Negative stain TEM of the newly purified levan four-component complex

The newly purified four-component complex solubilised in DDM (~3 mg/ml) was provided by collaborators at Newcastle University. The complex was purified from *B. theta* grown on

fructose and is therefore assumed to be in the apo state. Sample quality was assessed by negative-stain TEM and a representative micrograph is shown in **Figure 5.3**. Particles were homogeneous in size and monodisperse. Notably, various particle views are observable in contrast to negative stain images of the core levan SusCD complex in which top-down views showing the twin-barrel structure of the SusC dimer were predominant. In light of these results, the sample was deemed suitable for downstream cryoEM studies.

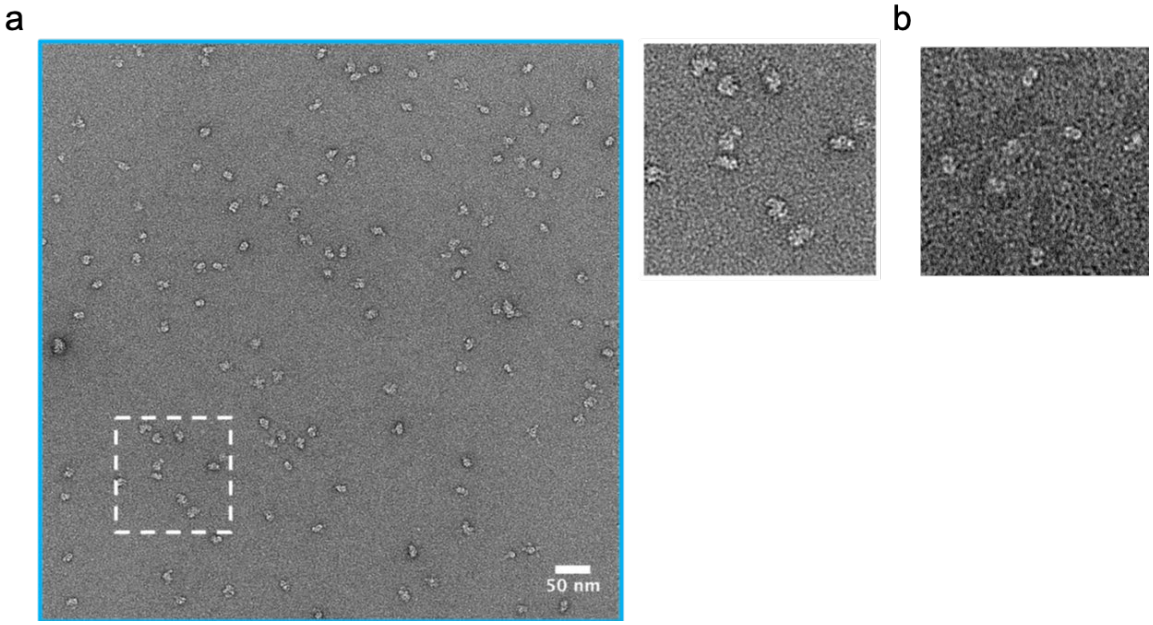


Figure 5.3 Uranyl acetate stained putative four-component complex solubilised in DDM at 50k x magnification. **a**, Representative micrograph showing a monodisperse distribution of particles. A zoomed inset corresponding to the area marked by the dashed white box shows a variety of particle orientations. **b**, An inset from a micrograph of the levan $SusC_2D_2$ core complex for comparison. Here particles show preferential orientation with the viewing angle along the axis of the SusC barrels, resulting in a characteristic 'figure eight' shape for the particles. Scale bar is 50 nm

5.3 Single-particle cryoEM of the levan four-component complex in the absence of substrate

Following the promising results obtained by negative stain TEM, grids were prepared for screening by cryoEM. Quantifoil and lacy carbon grids were trialled with the best particle distribution observed for the latter (see **Figure 5.4**). Notably, the extensive self-association previously observed for the core $SusC_2D_2$ complex was not observed for the putative four-component complex, meaning that micrographs suitable for single particle image processing were obtainable in the absence of a continuous carbon support. This difference in behaviour also provided an initial indication that the particles present likely corresponded to higher

order four-component complexes, where the surfaces that led to self-association of the $SusC_2D_2$ complex are 'quenched' by the presence of the additional lipoproteins.

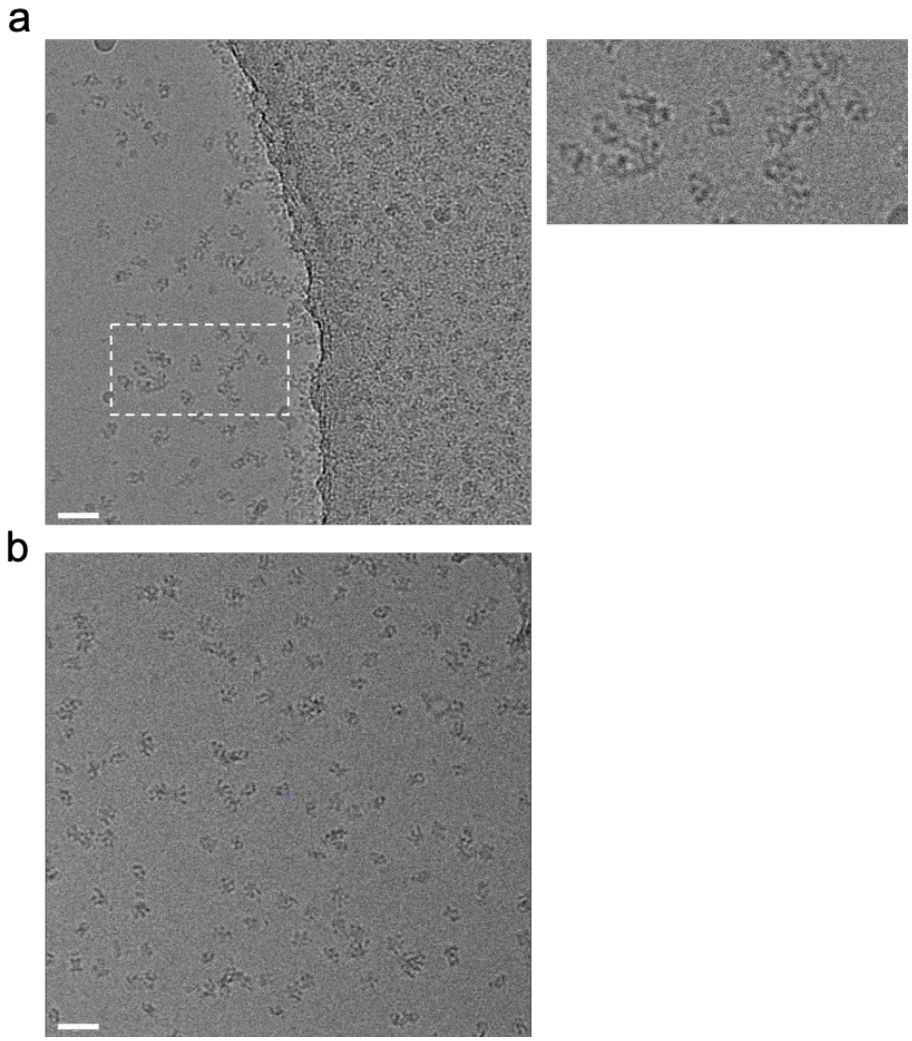


Figure 5.4 CryoEM screening images of the putative four-component complex. a, Representative micrograph of the sample loaded onto Quantifoil 1.2/1.3 grids. Some small aggregates are visible as shown in the zoomed inset. **b**, Representative micrograph of the sample loaded onto a lacy carbon grid demonstrating improved particle distribution. Scale bars are 40 nm.

An initial data set of 2057 micrographs was collected. At the time of first processing, the presence of large amounts of carbon in the images was found to severely hinder auto-picking, despite attempts to optimise picking parameters. In order to circumvent this problem, micrographs in which carbon accounted for ~50 % of the image were manually excluded, such that 1093 micrographs remained. The dataset was later revisited with the neural network-based picker crYOLO which does not suffer from the same performance issues, but unfortunately the originally discarded micrographs had already been deleted in the interest

of storage space. Using crYOLO, 96,639 particles were identified in the remaining micrographs and were extracted, aligned, and classified in 2D. After removal of junk, 89,305 particles remained and contributed to 2D class averages with high-resolution structural features. These classes are markedly distinct from those of the core complex, with additional density clearly observable in a number of projections (see **Figure 5.5**).

An initial model was generated *de* novo from the data and used as a reference for 3D classification, allowing characterisation of the assemblies present (see **Figure 5.5**). Two classes (grey and green) correspond to the hexameric complex previously observed in the dataset of the core levan $SusC_2D_2$ transporter and comprise ~42% of the data. The remaining non-junk classes (yellow, purple and blue) correspond to the hypothesised octameric four-component complex and together account for >50% of the data. Here, putative density for the additional lipoprotein subunits is associated with the lip of each SusC barrel in the $SusC_2D_2$ dimer, maintaining the symmetry of the molecule. As before, the levanase subunits were easily identifiable owing to the availability of a crystal structure of a soluble form of the protein lacking the lipid anchor (PDB ID: 6R3R). The characteristic β-propeller fold of this enzyme allowed it to be assigned to the more compact of the two additional densities associated with each SusC subunit. In contrast, no structure is available for the LBP Bt1761 and there are no homologous structures in the PDB. The density corresponding to this protein in our EM maps is located adjacent to that of the levanase and stretches out in the direction of the extracellular space. The most N-terminal domains of the protein appear stable, while mobility increases towards the C-terminus (see **Figure 5.6**). Indeed, the diffuse density at the visible C-terminus indicates that the remainder of the protein is too mobile to resolve, having been averaged out in the reconstructions. Importantly, the positions of the additional lipoproteins are arranged such that they do not obstruct opening and closing of the SusD lids.

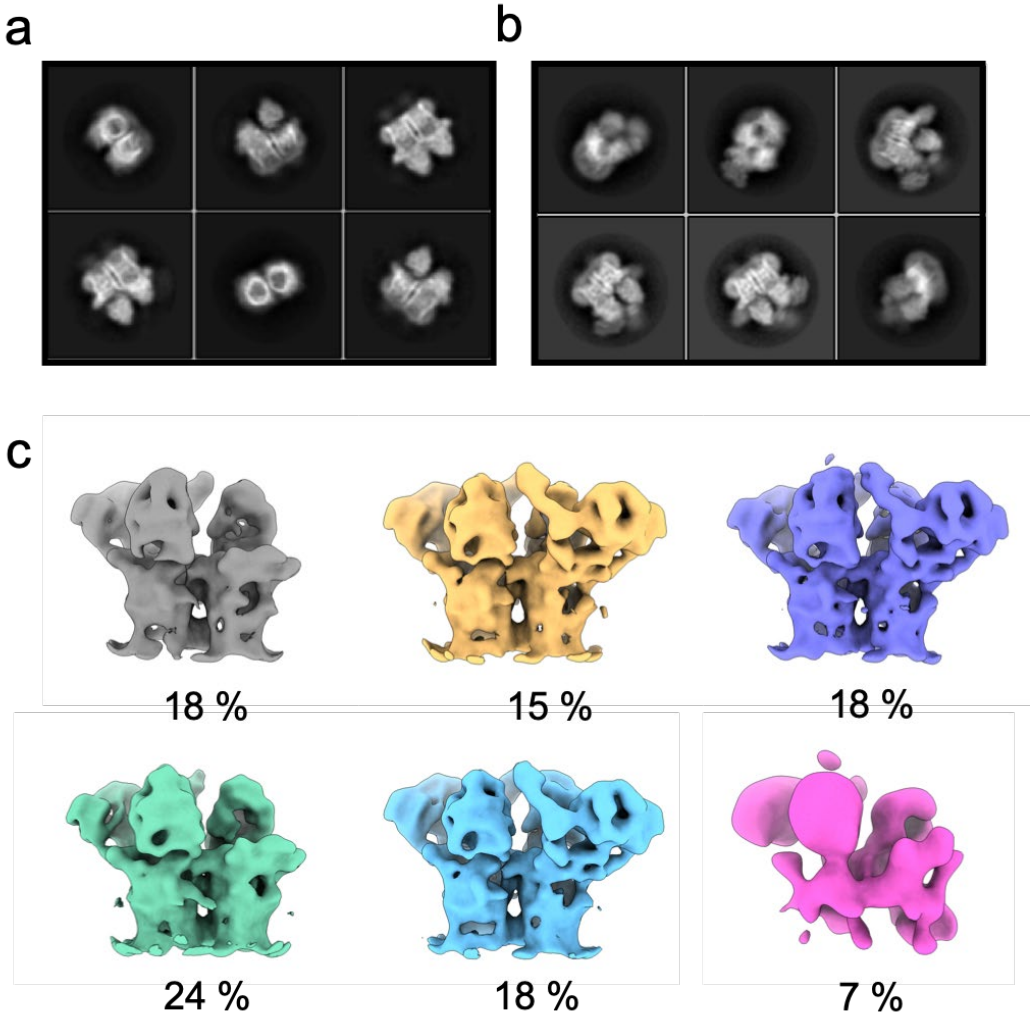


Figure 5.5 Early processing outputs of the levan utilising four-component complex. a, Example 2D class averages of the LDAO-solubilised levan core complex from our previous work. **b**, Example 2D class averages of the levan four-component complex solubilised in DDM. Density in addition to that of the core complex (**a**) can be seen in all views. **c**, Output of the first round of 3D classification. Yellow, purple and blue classes represent the octameric four-component complex. The grey and green classes show additional lipoproteins associated with just one SusC unit. The pink class is considered 'junk'.

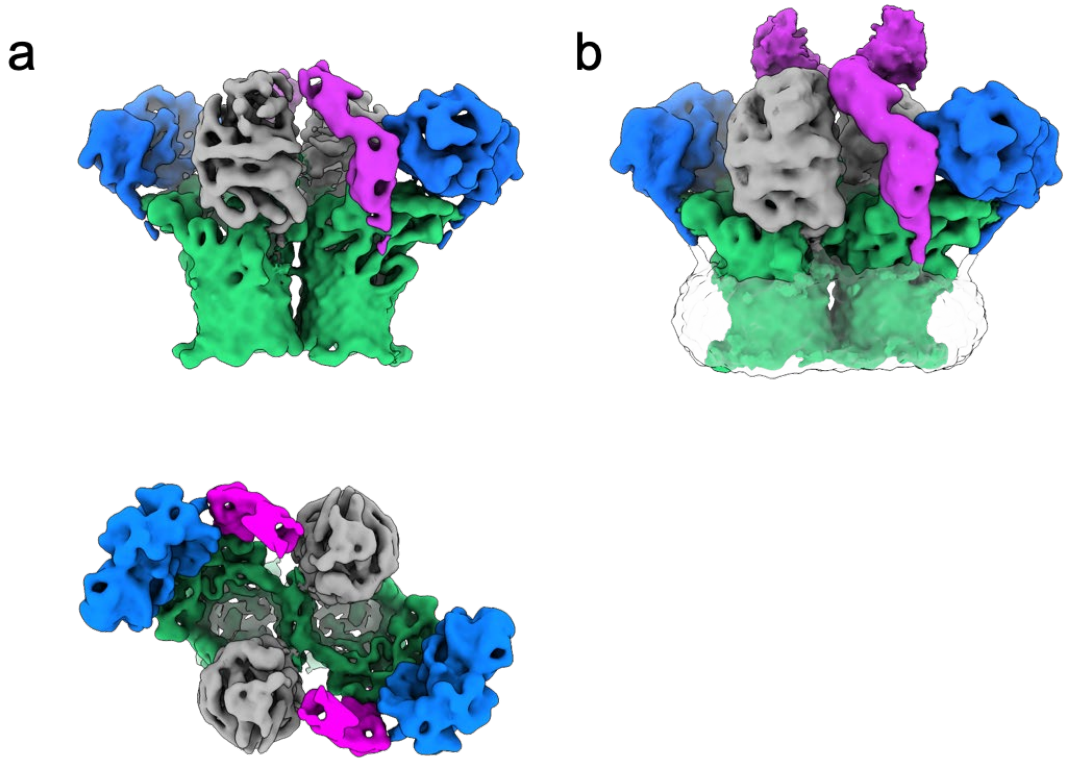


Figure 5.6 Overview of the octameric four-component complex coloured by subunit. a, Map generated by 3D auto-refinement of particles contributing to the purple class in **Figure 5.4**. SusC and SusD subunits are coloured green and grey respectively. Compact density corresponding to the levanase is coloured blue and the extended snake-like density of the LBP is shown in pink. The upper panel is viewed in the plane of the membrane and the lower panel is viewed from the extracellular space. **b**, Same map as in (**a**) but viewed at a higher threshold level to reveal diffuse density in the C-terminal region of the LBP. The translucent density corresponds to the micelle belt.

All particles assigned to classes containing a full complement of components, that is, the $SusC_2D_2$ core as well as two copies of both the levanase and the LBP, were pooled, resulting in a compositionally homogenous particle stack for further processing. We propose that this four-component, octameric complex (~570 kDa) represents the complete OM levan utilisation machine, which we term the levan utilisome.

5.3.1 Investigating conformational heterogeneity of the levan utilisome

To assess the conformational heterogeneity present in the levan utilisome, additional rounds of 3D classification were employed. Major conformational differences between the resulting classes were predominantly localised to the SusD subunit, which displayed variation in its extent of opening relative to SusC (see **Figure 5.7**), as observed for the core complex in isolation. The lids within a complex appear to operate independently of one another such that

differing lid positions on either side of the complex can break the C2 symmetry of the molecule (see **Figure 5.7**). Extents of lid opening vary from an open state comparable to the predominant lid position observed in the study of the core $SusC_2D_2$ complex, to a wide-open state also described in the previous chapter (**see Figure 5.7**). Thus, three principle open classes emerged: 1. An open-open state where both lids occupy what we describe as a 'narrow' open position, consistent with the predominant open position observed for the core complex and abbreviated to NN, 2. An open-open state where one SusD subunit adopts a wider open position while the other is narrow open, which we abbreviate to WN, 3. An open-open state where both SusD subunits are in the wide position, abbreviated to WW. Importantly, there is no evidence of a closed lid position and, in contrast to our findings for the LDAO-solubilised core complex, all SusC barrels contain unambiguous density for the plug domain that occludes the barrel. The first of these observations is consistent with the working 'pedal-bin' hypothesis that would predict an invariantly open transporter population in the absence of substrate, as is the case here. The absence of transporters that have lost their plugs in this DDM-solubilised sample corroborates our previous proposal that ejection of the plug domain observed for the LDAO-solubilised core levan transporting complex is non-physiological. Rather, it appears to be a result of the harsh detergent environment, or perhaps the distinct grid preparation conditions, used. Aside from the predominant open and wide-open states displayed in **Figure 5.7**, a small population of particles (~4%) are assigned to a class with weaker density for SusD on one side of the transporter (see **Figure 5.8**). Given the small particle number, further classification was not possible and these particles were excluded from downstream processing steps. Finally, while the major conformational differences between classes involve the SusD subunit, mobility of the C-terminal region of the LBP Bt1761 is also evident in the output of 3D classification (see **Figure 5.7**).

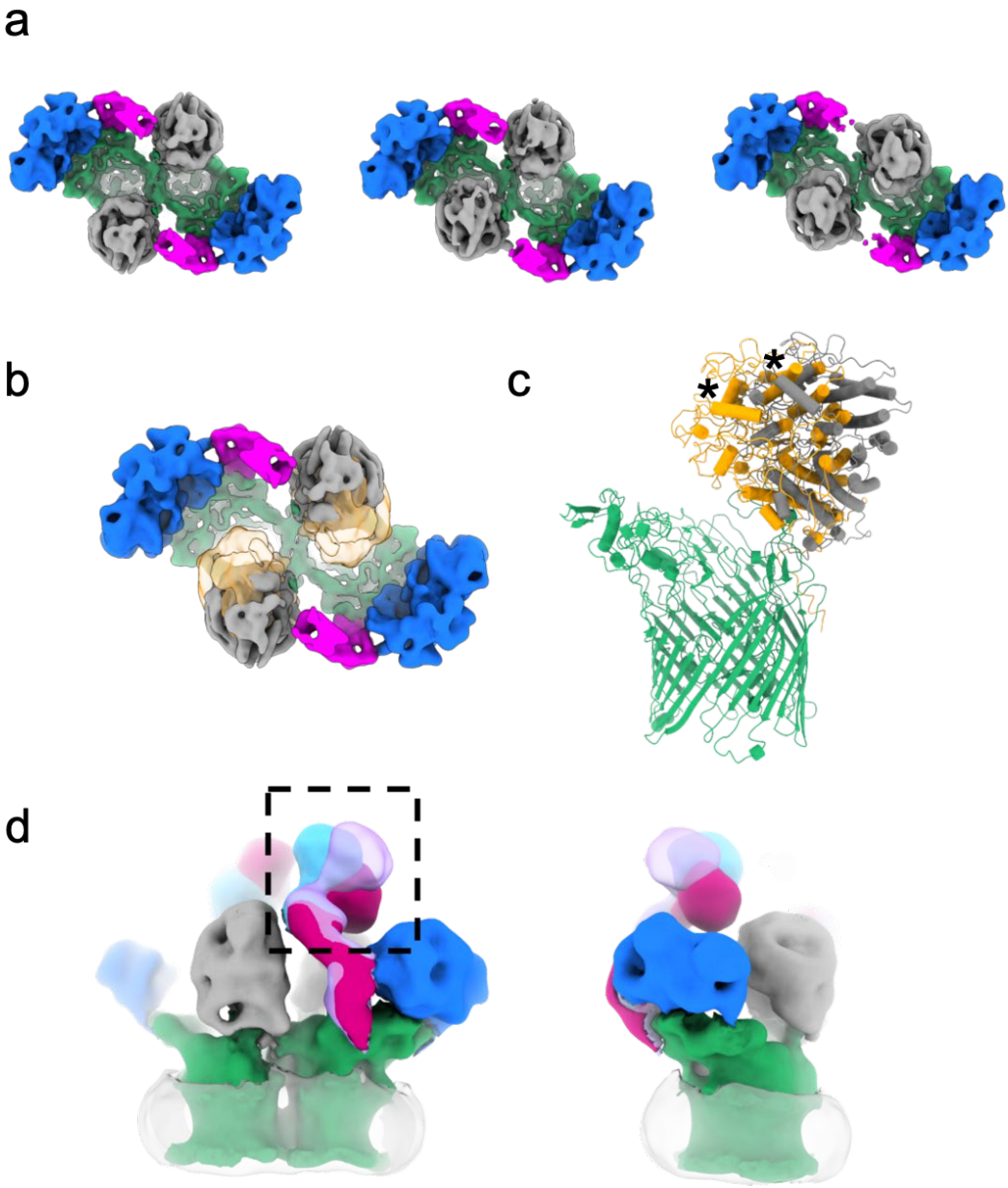


Figure 5.7 Conformational variability of the substrate free levan utilisome. a, Refined outputs from 3D classification of the octameric utilisome viewed from the extracellular space. Classes separated on the basis of lid position. Wide-wide, narrow-wide, and narrow-narrow open states are presented from left to right. **b**, Overlay of the wide-wide (grey) and narrow-narrow (orange) open states of the complex. **c**, Overlay of models for the narrow open versus wide open state of the transporter generated by a rigid-body fit of SusD into the density of the respective maps. A monomer is shown for clarity and an asterisk marks the same SusD helix in both models. **d**, A view of the utilisome shown at high threshold levels in the plane of the membrane (left). Different conformations of the levan binding protein observed in 3D classification are overlaid to demonstrate the flexibility of this subunit (boxed region). The same view rotated 90° is shown (right). Disordered micelle density is shown as translucent grey.

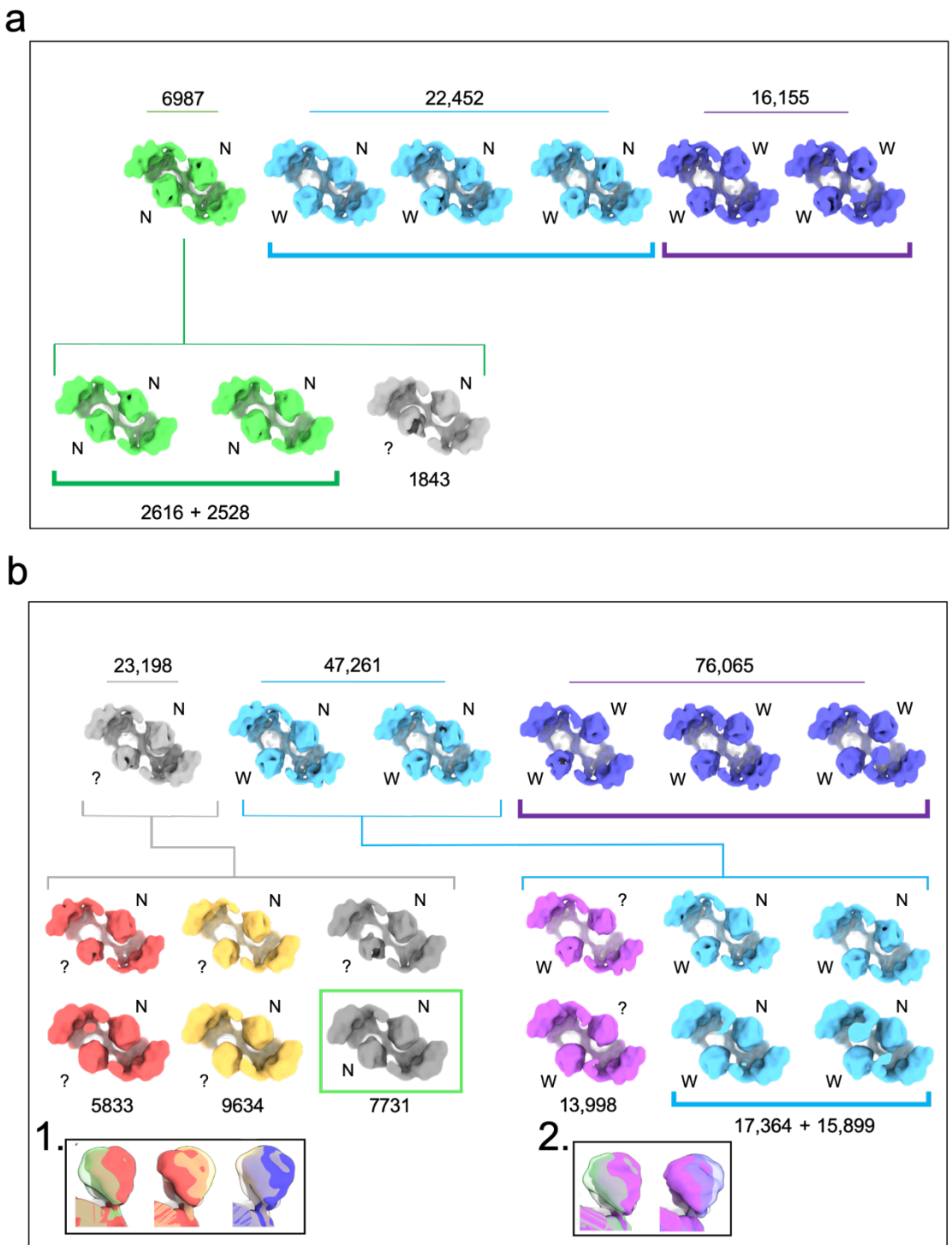


Figure 5.8 Results of 3D classification for two independent datasets of the substrate-free levan utilisome. a, Results of classification starting from the compositionally homogeneous utilisome particle stack from the first dataset. Classes are coloured according to their conformational state where NN is green, WN is blue and WW is purple. The conformation of each SusD subunit within a complex is also labelled. Lid positions that do not correspond to W or N are marked '?'. **b**, Results of classification starting from the compositionally homogeneous utilisome particle stack from the second dataset. In contrast to the first dataset no clear NN conformation was observed in the early stages of classification. Instead more particles possessed weaker density for SusD on one side (grey class). The next branch of the classification tree shows classes at low (upper panels) and high (lower panels) threshold levels to enable comparison of weak lid density positions. The class boxed in green

appears comparable to the NN state at high thresholds. Box **1.** Overlay comparisons of ambiguous lid positions from grey branch. Left: comparison of red ambiguous with the narrow open state. Middle: comparison of the red and yellow ambiguous lid positions. Right: comparison of yellow ambiguous with wide-open state. Box **2.** Overlay comparisons of ambiguous lid positions from blue branch. Left: comparison of pink ambiguous with the narrow open state. Right, comparison of pink ambiguous with the wide-open state. Classification trees terminate where further rounds of classification revealed no additional heterogeneity.

In an attempt to improve the resolution of reconstructions for the apo levan utilisome, a second dataset comprising 3142 micrographs was collected. After cleaning through several rounds of 2D classification, 280,696 particles remained and were taken forward to 3D classification. As for the first dataset, classes displaying the octameric utilisome complex were pooled and subjected to further rounds of classification in 3D (a classification tree is shown in **Figure 5.8**). However, unlike the first dataset where NN, WN and WW conformations emerged in the first utilisome classification, here only the WW and WN conformations were unambiguously observed. The remaining class has weak density for SusD on one side of the complex while the remaining SusD subunit occupied a narrow open position. This state comprises ~16 % of the particles and was therefore investigated further. An additional round of classification yielded three classes, all with a weaker SusD subunit on one side of the complex relative to the other. Viewing these maps at increased threshold levels compensated for the weaker density allowing comparison of lid positions. Two of the ambiguous SusD lid positions observed in the red and yellow classes from **Figure 5.8** are distinct from one another and occupy open positions between that of the narrow-open and wide-open states. The remaining class (grey) is again distinct from the other classes and looks equivalent to the NN conformation at high threshold levels. The cause of the weaker SusD density is unclear, however its increased prevalence in the second dataset indicates that it is perhaps the result of an additional freeze-thaw cycle that occurred between the two grid preparation sessions. Further classification of the WN state revealed another novel class, again characterised by weaker lid density on one side, that when viewed at high threshold levels shows a wide lid position for one SusCD subunit and an intermediate lid position comparable to that observed in the red class described above for the other. The WW classes identified in the first round of 3D classification were found to be homogeneous.

The WW and WN particles from the preliminary dataset were combined with the unambiguous WW and WN particle stacks from the second dataset (denoted by bold brackets

in the corresponding classification trees). Particles were subjected to per-particle CTF refinement and Bayesian polishing with distinct optics groups used for each dataset. Sharpened reconstructions of WW and WN states had resolutions of 3.7 Å and 4.0 Å respectively, although large variations in local resolution are present in the maps (see **Figure 5.9**). Despite the stringent classification, density for the SusD subunits is markedly weaker than that of the SusC barrels and precluded model building and refinement. However, while the SusD lids occupy several conformations, SusC, the levanase and the N-terminal domain of the LBP appeared to constitute a relatively rigid unit with C2 symmetry. Thus, to provide higher resolution structural information for these regions, a mask that excluded density for the SusD lids was created and supplied in a focused refinement. Including all octameric complexes irrespective of lid position gave rise to a 3.5 Å reconstruction after sharpening. At this resolution, it was possible to dock and refine models for SusC and the levanase, as well as to build the N-terminal domain of the LBP (see **Figure 5.10**).

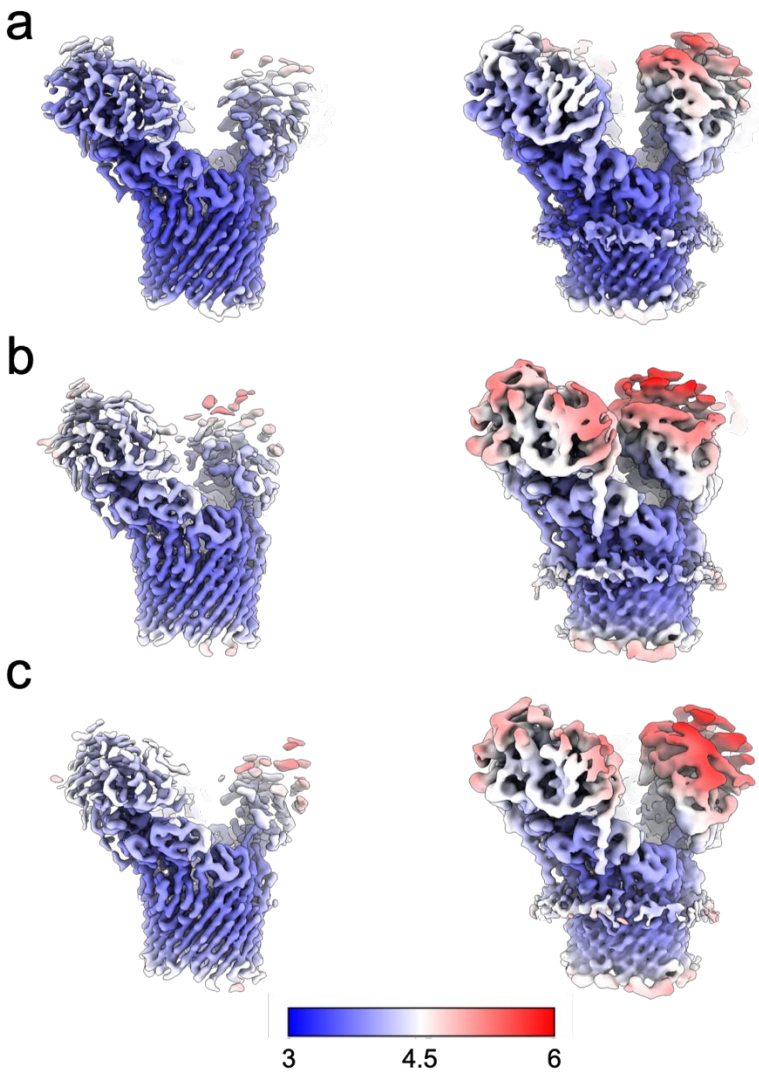


Figure 5.9 Local resolution maps of the wide-wide and wide-narrow open states using particles from both datasets. a, wide-wide apo utilisome conformation (one side shown as C2 symmetry was employed in the reconstruction. **b**, narrow open side of the narrow-wide utilisome conformation. **c**, wide open side of the narrow-wide utilisome conformation. Maps are displayed at low (left) and high (right) threshold levels and are filtered and coloured by local resolution. Colour key is resolution in Angstroms.

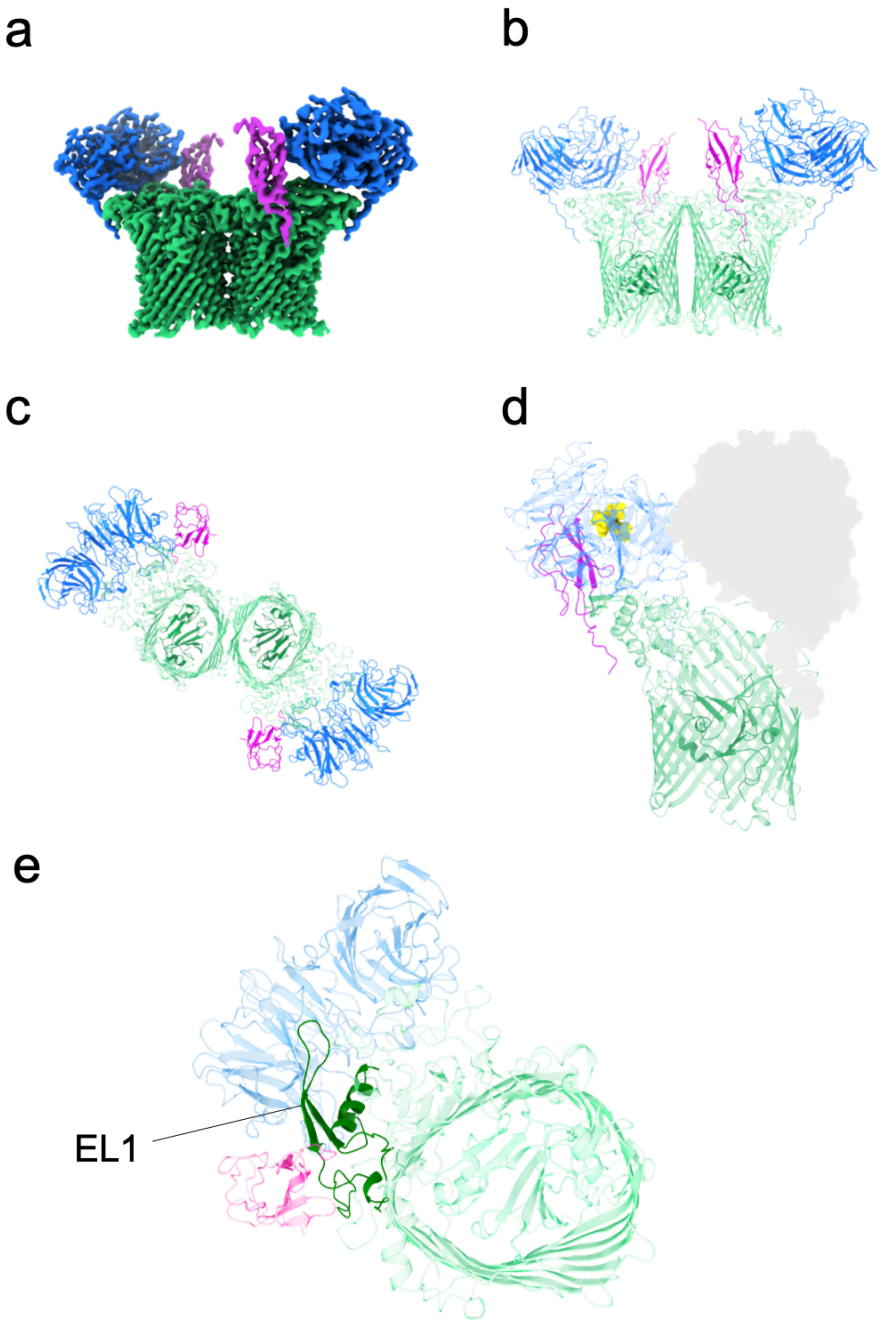


Figure 5.10 The organisation of the levan utilisome revealed by cryoEM a, C2 symmetric reconstruction at 3.5 Å from a focused refinement excluding the variable SusD subunits, viewed in the plane of the membrane. **b**, Atomic model built from the density shown in **a**. **c**, the same model but viewed top down from the extracellular side. **d**, One half of the complex shown in the context of the SusD lid (grey shadow). Levan tetraose (yellow) has been artificially displayed at the active site of the endo-levanase highlighting proximity to the open mouth of the barrel. **e**, One half of the complex viewed bottom up from the periplasmic space. The extended extracellular loop 1 that constitutes the majority of interfacing interactions with the levanase and LBP is shown in dark green and labelled. Note, all particles possess SusD components, and their absence here is a result of focused classification.

5.3.2 The arrangement of the levanase and levan binding proteins indicates concerted function

The levanase is mounted on the lip of the SusC barrel, at the opposite edge from the location of the hinge contacts between SusC and SusD. Extracellular loops 1 and 9 contribute to the binding interface which buries a surface area of ~820Å. This arrangement means that the opening and closing of the SusD lid would not be hindered by the presence of the levanase. Further to this, the levanase is oriented in a way that is consistent with concerted function with the core transporter, with its active site close to the binding cavity in the SusC transporter such that the minimum distance between FOS bound to the respective subunits is ~30 Å. The levan binding protein is mounted adjacent to the levanase, and in this substrate-free utilisome structure, only its N-terminal Ig-like domain is unambiguously resolved at high resolution. Interactions between this domain and SusC again occur at extracellular loop 1 (see **Figure 5.10**). The lip of SusC therefore appears to serve as a platform for the association of the additional lipoproteins, potentially explaining why SusC-like TBDTs are ~40 % larger than their classical non-complex forming counterparts. This structure, alongside the observation that the LBP subunits appear to be flexible, especially towards its C-terminal region, provides important insight into how the utilisome architecture contributes to efficient levan utilisation. The flexible LBP arms of the complex are capable of binding lengths of levan chain and positioning them for processing by the adjacent levanase units. Resulting cleavage products would then be released close to their binding site within the SusC barrel, promoting efficient loading of the transporter.

5.4 Addition of levan FOS DP8-12 gives rise to concerted conformational changes in the utilisome

To probe the mechanism of substrate capture and processing by the levan utilisome, 2.5 mM levan fructo-oligosaccharide (β2,6 FOS) with a degree of polymerisation (DP) of ~8-12 (DP8-12) was added to the detergent-solubilised utilisome complex (3 mg/ml). Screening of alternative grid preparation conditions led to the discovery that glow-discharging in the presence of amylamine vapour yielded good particle distributions on Quantifoil grids which had previously not been suitable for data collection. Ice quality under these conditions was far superior to that seen with lacy carbon grids. Moreover, the use of Quantifoil grids allowed

for a more tailored data acquisition set-up, where holes were stringently selected on each grid square such that only ideal ice conditions were imaged. Micrographs for Quantifoil grids glow-discharged in air and in the presence of amylamine vapour are shown side-by-side in **Figure 5.11**.

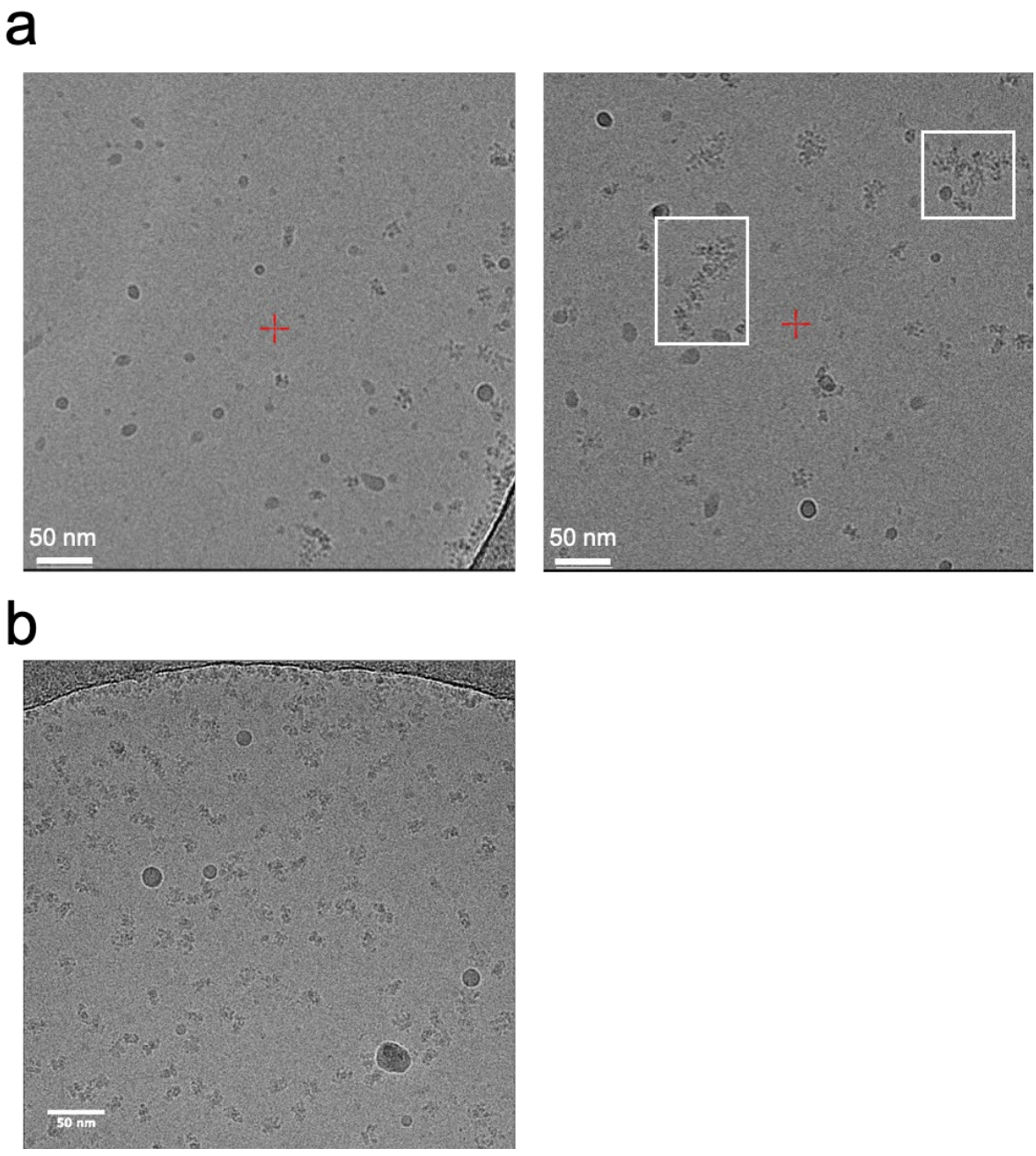


Figure 5.11 Glow discharging in the presence of amylamine vapour improves utilisome on-grid particle distributions. a, Screening micrographs showing the behaviour of levan utilisome particles on quantifoil grids glow-discharged in air. In thin ice (left), particles are found exclusively near the edge of the hole. In thicker ice (right), particles were distributed further into the holes but small aggregates were prevalent (white boxes). **b**, Representative micrograph of levan utilisome particles applied to a quantifoil grid glow-discharged in the presence of amylamine vapour. Particles are well distributed and monodisperse. Note that the concentration of the utilisome sample used was not changed between the two preparation conditions. Scale bars are 50 nm.

A small dataset of 974 micrographs was collected from which 72,373 particles were extracted, aligned and classified in 2D to remove junk. The remaining 63,789 particles were classified in 3D to assess the level of heterogeneity present in the data. In accordance with previous data collections on the levan utilisome, compositional heterogeneity was evident with classes corresponding to the naked core $SusC_2D_2$ transporter as well as the previously described hexameric complex and the complete utilisome assembly. Strikingly however, the SusD subunits invariably occupy a closed position, tightly capping the extracellular lumen of their corresponding SusC barrel (see **Figure 5.12**). Thus, the addition of FOS substrates gives rise to a major concerted conformational change in the complex.

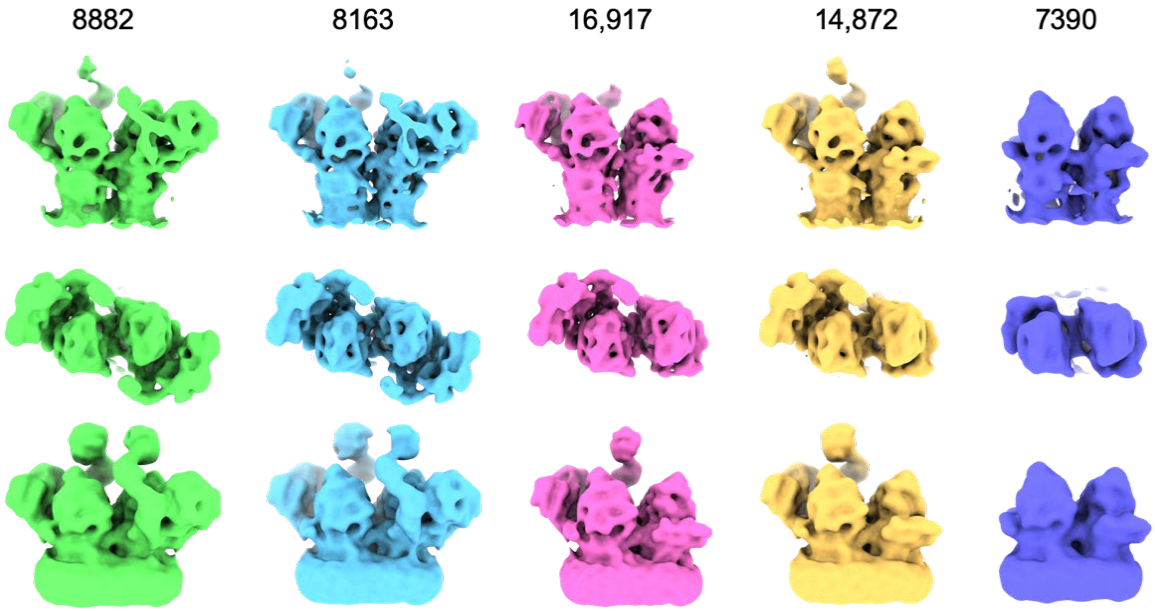


Figure 5.12 Results of 3D classification for the active utilisome complex in the presence of short FOS. Class averages are shown in the plane of the membrane (top) and from the extracellular space (middle). The bottom row shows the classes at a higher threshold level to allow visualisation of the mobile C-terminal regions of the levan binding proteins. The particle number associated with each class is provided.

5.4.1 FOS substrates are visible within the SusCD binding cavity

To confirm that FOS had indeed been captured by the transporter complex, we first targeted refinement of the $SusC_2D_2$ core. To achieve this, we used particle subtraction to remove signal for the levanase and LBP from the images, and were thus able to include all particles in the refinement, regardless of their lipoprotein complement. A final reconstruction of the $SusC_2D_2$ core was obtained at 2.9 Å resolution, which contains unambiguous density for FOS substrate

at two distinct sites within the SusC barrel (see **Figure 5.13**). These densities are comparable to those observed in the crystal structure of the substrate-bound $SusC_2D_2$ complex (PDB ID 6ZAZ; Gray et al., 2021). The oligosaccharide bound at the interface of SusC and SusD consists of 6 β2,6-linked fructose units with a β2,1 decoration on Frc-4 (numbered from non-reducing end). The FOS adopts a compact, twisted topology similar that observed for the levantetraose in the levanase complex (PDB 6R3U). The second FOS molecule occupies a site at the bottom of the SusC binding cavity, where it is in contact with the plug domain. Here, the density corresponds to three β2,6 linked fructose units with a β2,1 decoration on the terminal fructose (non-reducing end).

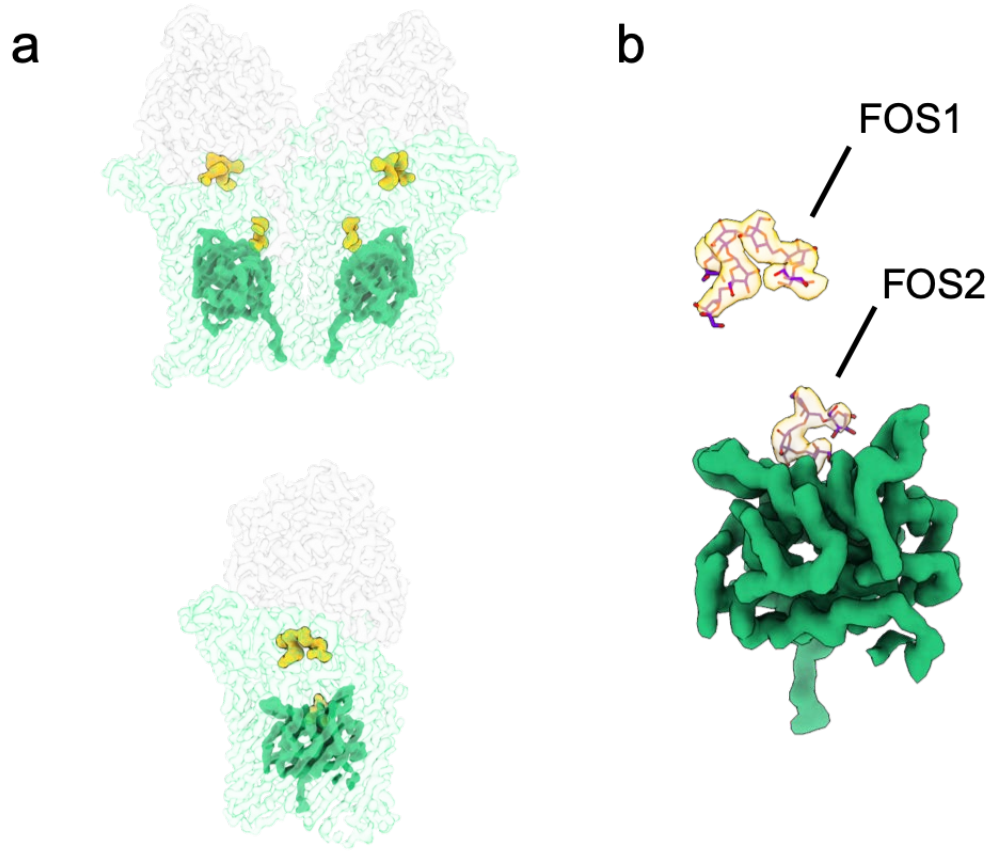


Figure 5.13 A concerted conformational change of the levan utilisome upon addition of short FOS. **a,** Sharpened reconstruction at 2.9 Å of the $SusC_2D_2$ core complex generated from a focused C2 refinement after particle subtraction of additional lipoprotein components. Densities for bound FOS (yellow) and the plug domain of SusC are highlighted. **b**, Isolated view of plug and substrate density. The FOS modelled are identical to those in the previously determined crystal structure of $SusC_2D_2$ (PDB: 6ZAZ).

5.4.2 Substrate binding induces conformational changes at the periplasmic face of the SusC plug domain

Given the biological role of the utilisome, it is expected that substrate loading of the transporter should be signalled across the OM via disordering of the SusC TonB box region at the periplasmic face of the protein. Indeed, the TonB box and the region preceding it are not visible in the substrate-bound structure (see **Figure 5.14**). This is in stark contrast to the apo state of the transporter in which the region preceding the TonB box is resolved, and adopts a sharp turn towards the plug domain, retaining the TonB box within the barrel. Our data therefore suggest that substrate binding is allosterically signalled through the transporter (and thus across the OM) in a manner consistent with the prevailing models for TonB-dependent transport. Although determining components of allosteric pathways from structures is difficult, some important insights are gained by comparing the substrate-free and substrate bound-states of the levan transporter. Most notably, in the apo state of the transporter, Tyr191 is involved in putative stacking interactions with Tyr89 in the region preceding the TonB box that is tucked up against the plug domain (see **Figure 5.14**). Upon substrate binding and lid closure, Tyr191 undergoes a shift of ~4 Å, disrupting the stacking interactions and adopting a position that is incompatible with the tucked conformation of the Ton region owing to steric hinderance. A similar shift was observed for Tyr215 of RagA in the substrate-bound, lid closed structure when compared with the lid open structure. It is possible that this movement is the culmination of the allosteric signalling pathway and, if true, the data suggest that at least part of this pathway is conserved between the peptide transporting RagAB complex and the levan transporting SusCD complex.

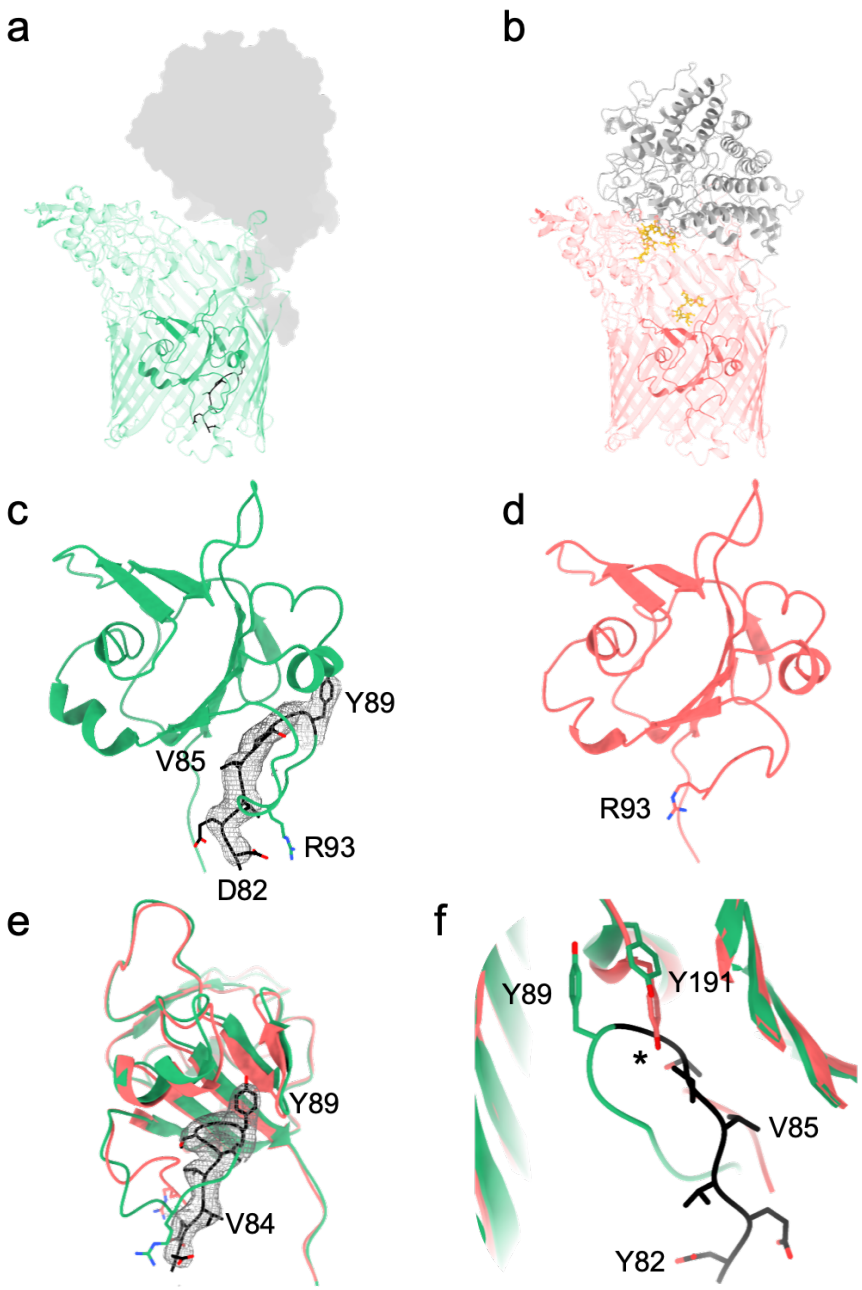


Figure 5.14 Analysis of the TonB box region in cryoEM structures of the apo and substrate-bound transporter structures. **a**, Structure of SusC of the levan transporter (Bt1763) in the absence of substrate. The TonB box is coloured black. The corresponding SusD component is shown as a shaded zone. **b**, Structure of SusCD complex in the presence of short FOS. Isolated plug domains from each state can be seen in (**c**) and (**d**). Density is displayed for the TonB box region visible only in the apo state of the transporter. **e**, Superposition of plug domains for the apo (green) and substrate-bound (yellow) structures of the Bt1763; rotated 90° relative to (**c**). The visible N-terminus is D82 for the substrate-free transporter compared with R93 for the substrate-bound transporter. This corresponds to a disordering of 11 residues upon substrate binding. **f**, Overlay showing the TonB box region tucked against the plug domain in the apo state with Tyr89 participating in stacking interactions with Tyr191 of the plug. Tyr191 undergoes a shift of ~4 Å in the substrate-loaded transporter structure, disrupting the stacking interactions and adopting a position that is incompatible with the tucked conformation of the TonB box region due to steric hinderance; an asterisk indicates the site of clashing.

5.4.3 Substrate-dependent conformational changes were not observed for the lipoprotein components in the presence of levan FOS DP8-12

To evaluate the effect of substrate binding on the additional lipoprotein components, a subset of particles corresponding to the complete levan utilisome was selected. Refinement of these particles resulted in a 3.2 Å reconstruction after sharpening (see **Figure 5.15**). The addition of short FOS was shown to have no influence on the conformations of the additional lipoproteins. As expected, no substrate density was observed within the active site of the levanase, since the levanase included in this experiment was active and therefore capable of hydrolysing the levan chains, and unfortunately the flexibility of the levan binding protein still prevented further structural analysis. Notably, in both the apo and substrate-bound datasets, the levanase and the LBP were invariantly associated with one another, i.e. there is no evidence of pentameric or heptameric complexes where one SusCD unit is associated with a single lipoprotein.

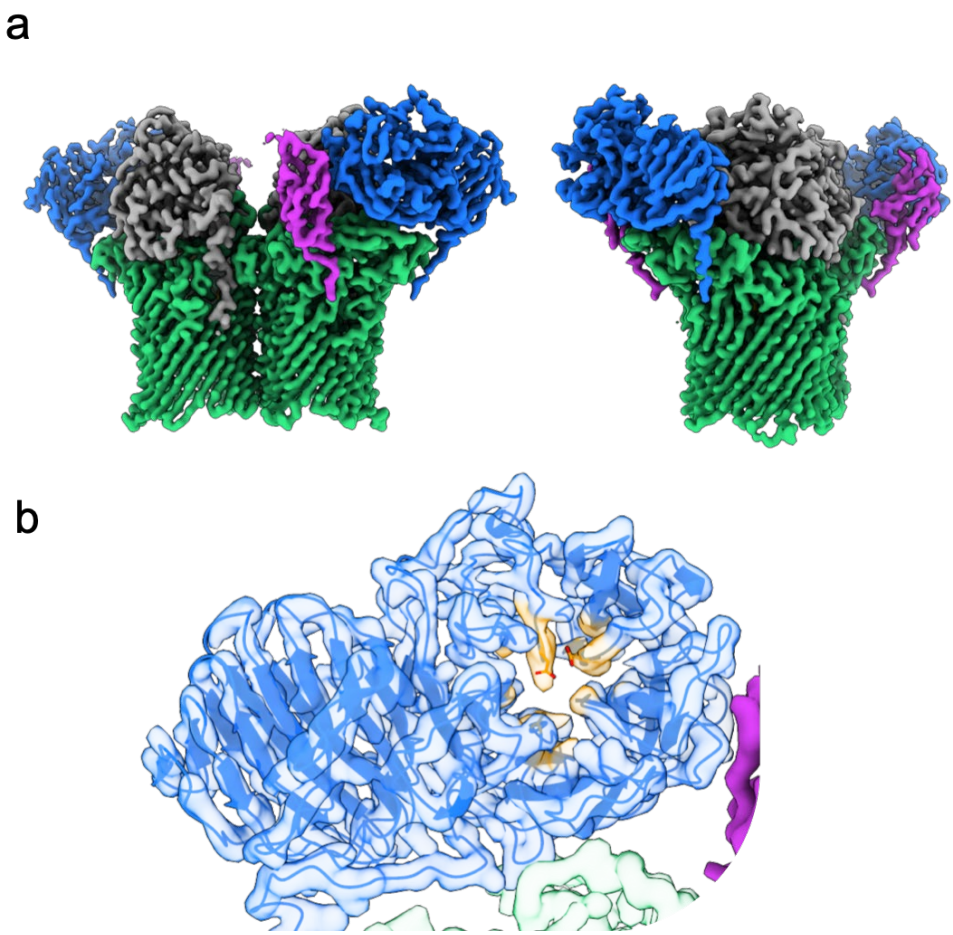


Figure 5.15 Levan utilisome architecture in the presence of short FOS. a, Sharpened 3.2 Å reconstruction of the closed, substrate-bound transporter viewed in the plane of the membrane (left) and rotated 90° (right). **b**, Transparent view of the levanase density containing the refined model. The pocket of the active site within the β-propeller domain is coloured orange with no density corresponding to FOS observed.

5.5 A utilisome containing an inactive levanase permits structure determination of the LBP in the presence of longer FOS substrates

While the cooperative behaviour of the levan utilisome components is implied by their arrangement, the structure of the complete levan binding protein remained enigmatic. We reasoned that by using a catalytically inactive levanase (bearing the active site D41A mutation) and longer FOS chains (DP ~15-25), cooperative binding of a single levan chain by the LBP and the levanase might restrict the mobility of the LBP. CryoEM screening revealed that the grid preparation conditions used for the previous substrate-bound utilisome also yielded micrographs suitable for data collection with the new sample. A dataset of 1388 micrographs was collected and, using crYOLO, 157,953 particles were identified and

extracted. Junk particles were removed through 2D classification and the remaining 146,056 particles were taken forward to classification in 3D.

5.5.1 Classification in 3D reveals a 'docked' conformation of the levan binding protein

As was the case for the active utilisome complex, all 3D classes corresponded to closed states of the $SusC_2D_2$ transporter, in accordance with the proposed pedal-bin hypothesis (see **Figure 5.16**). Compositional heterogeneity is also present in the data with both hexameric and octameric complexes observed; there was no significant population of naked $SusC_2D_2$ complexes in this dataset. Notably, a novel conformation of the LBP was present in several of the classes. Here, the LBP adopts a 'docked' state, in which its C-terminal domain is held proximal to both SusD and the levanase. Viewing these classes at high contour levels revealed more density for the C-terminal domain of the LBP than had been seen in any of the previous reconstructions across all datasets. With the aim of providing high resolution structural information on the LBP, all classes showing evidence of the docked state of LBP were pooled and used in 3D auto-refinement. Unfortunately, this consensus refinement produced a map in which the C-terminal domain of Bt1761 was still poorly resolved (see **Figure 5.16**), indicating that local conformational heterogeneity was still present in the particle stack.

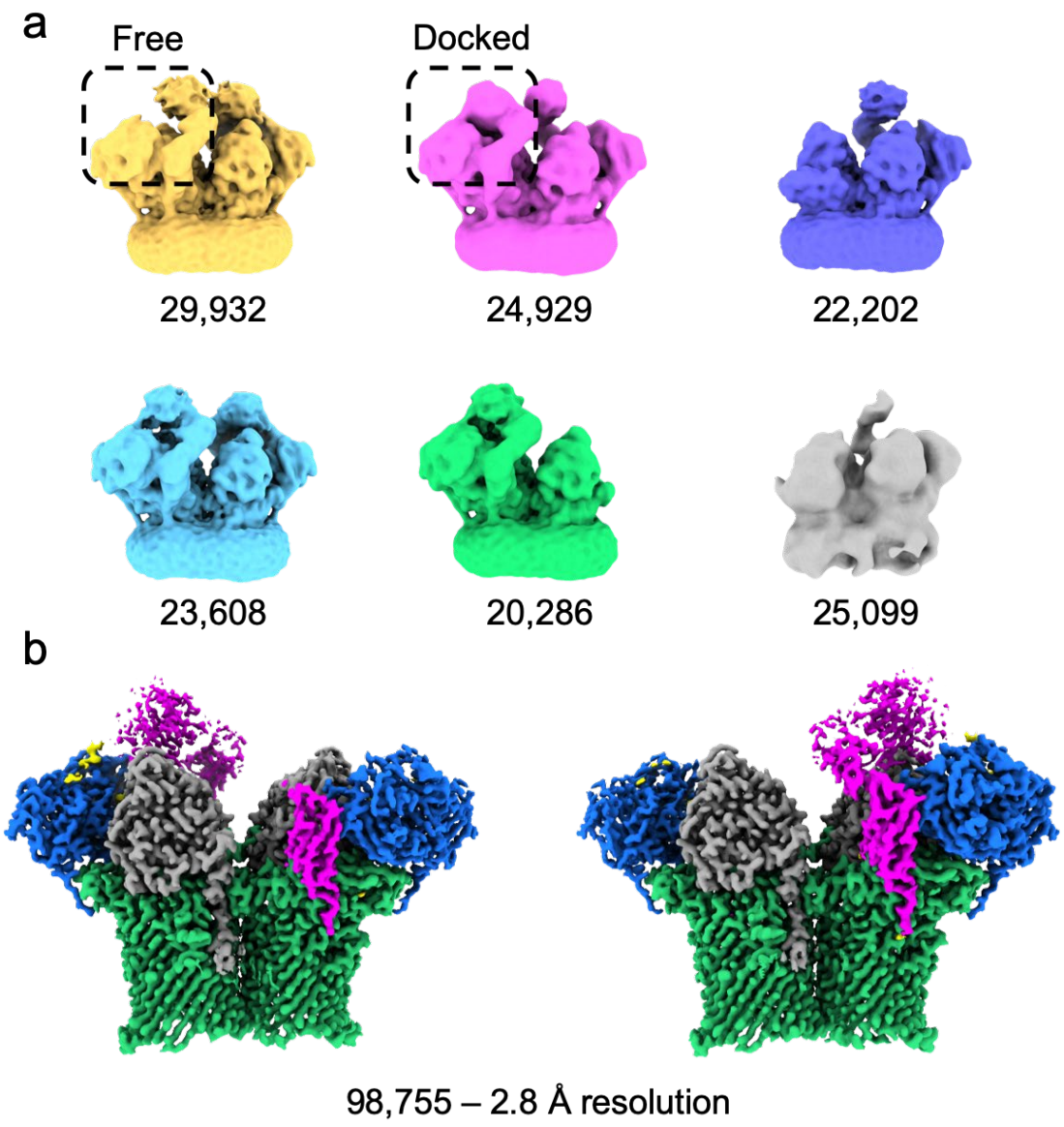


Figure 5.16 Addition of FOS DP ~15-25 to levan utilisomes containing an inactive glycosidase results in a docked conformation of the LBP. a, Outputs of 3D classification showing that the levan binding protein can adopt a 'docked' conformation in some instances. Examples of docked and free conformations are labelled. **b**, A consensus refinement of all classes containing docked LBP (yellow, pink, blue and green) to a global resolution of 2.8 Å. Note that density for the C-terminal region of LBP is poor compared with the rest of the map. Note that the handedness of maps in (**a**) is incorrect. Maps in (**b**) are displayed with the correct handedness.

In an attempt to collate a more homogeneous subset of particles, a focused classification approach was employed in the absence of particle alignment, using a mask that included only the docked SGBP position and some surrounding density (see **Figure 5.17**). Through this approach, the angles assigned in the original refinement are maintained, overcoming the low signal-to-noise problem that would, for example, be present in a particle subtraction experiment. Additional details on this approach can be found in **Methods 2.9.7**. Focused

classification yielded a class with high-resolution information for the C-terminal domain of the LBP (see **Figure 5.17**). Particles belonging to this class were used to generate independent half maps of the complete utilisome complex for post-processing, resulting in a final reconstruction at 3.0 Å with well resolved density for the complete LBP.

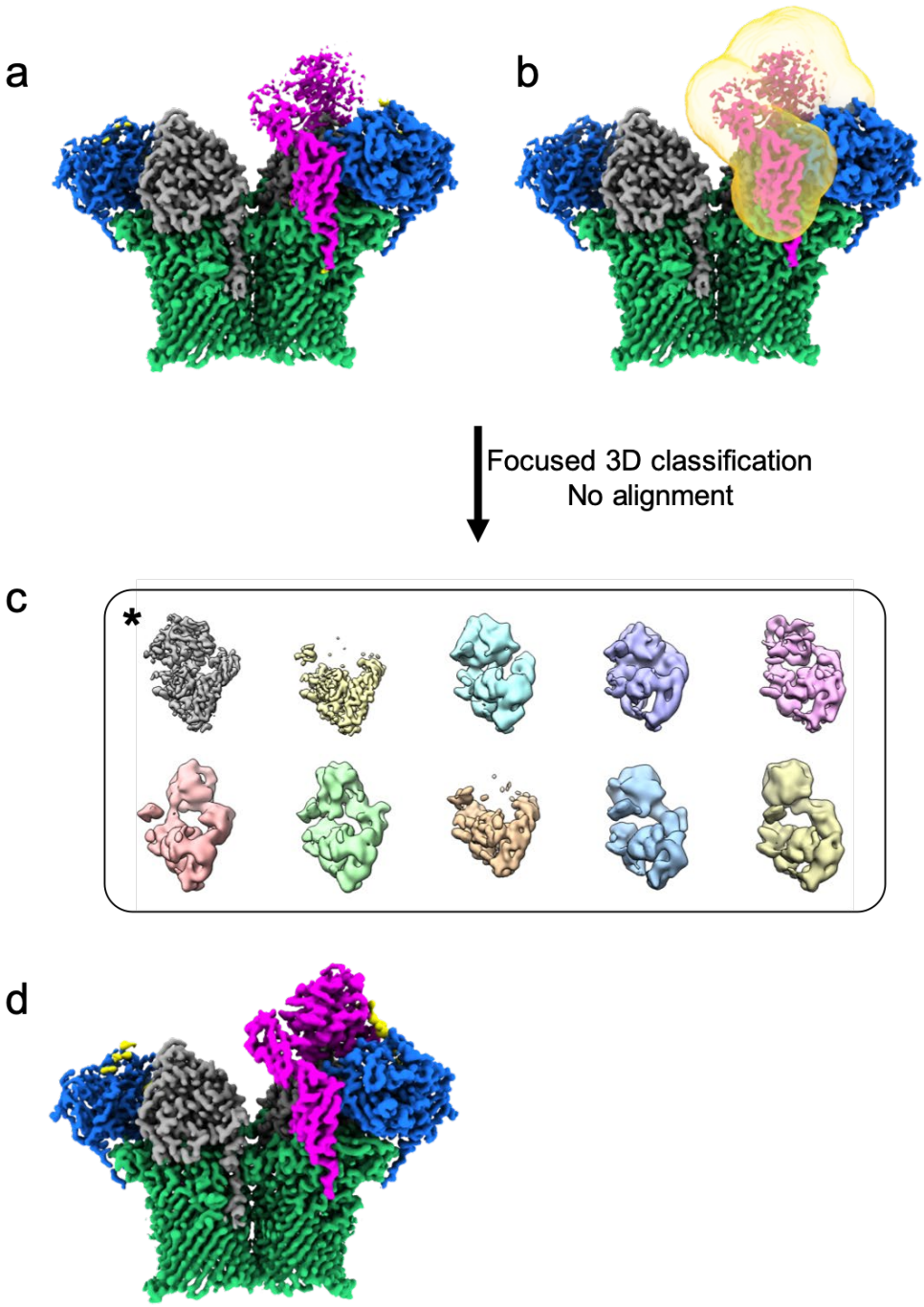


Figure 5.17 A focused classification approach improved map density for the C-terminal domain of the LBP. a, 2.8 Å reconstruction obtained prior to focused classification. **b**, The mask to be applied in focused classification shown on the same map as (**a**). **c**, Outputs of focused classification on the masked region without alignment. A class displaying high resolution features in the region of interest is marked with an asterisk. Independent half maps were reconstructed using the 27,310 particles belonging to this class. **d**, sharpened reconstruction generated from the aforementioned half maps showing improved density for the levan binding protein. The global resolution of the new map is 3.0 Å.

Using a model generated by AlphaFold2 (Jumper et al., 2021) as a starting point, the complete LBP was built and refined into the cryoEM density (see **Figure 5.18**). In agreement with the alphafold predicted structure, the LBP is organised into three domains, with the lipidated N-terminal and central domains each possessing an Ig-like fold. The C-terminal levan binding domain (res. 180-438) is globular, with a central β-sheet decorated with six α-helices. A distance-matrix alignment (DALI) analysis of the C-terminal domain returned a number of structures with significant similarity. A ThuA-like protein (PDB ID 1T0B) with a putative role in the conversion of disaccharides to their 3-keto derivatives, and a homoserine O-succinyl transferase (PDB ID 7CBE) possessed Z-scores of ~15 and ~14 respectively.

5.5.2 Longer substrate chains are responsible for tethering the levan binding protein

After model building of the LBP was complete, additional density was observed to bridge the levanase and the LBP. This novel density was attributed to a length of levan chain, bound by the C-terminal domain of the LBP and the inactive levanase. The density is compatible with a stretch of 12 β2,6-linked fructose units, with a putative β2,1 decoration observed on Frc7 (see **Figure 5.18**). Interestingly, the bound levan chain adopts a relatively extended conformation, markedly distinct from the twisted conformations of bound levan FOS seen previously. While the resolution of this linking density is insufficient for a detailed description of binding interactions, several tryptophan residues are clearly implicated in binding. These include Trp297 and Trp359 from the LBP, and Trp 195 from the levanase, which appear to cradle the FOS chain via stacking interactions. Notably, the levan chain binds across the top of the levanase β-propeller domain rather than at the active site of the enzyme (in the centre of the β-propeller). The levan density then proceeds towards the C-terminal β-sandwich domain of the levanase (see **Figure 5.18**). Evidence of levan binding at this secondary 'tethering' site on the levanase is also seen in the absence of bridging interactions with the LBP (see **Figure 5.18**). Levan binding in this region was not predicted, and was not observed in a recent crystal structure of the levanase with levantetraose. It was also not observed in the cryoEM structure of the utilisome containing the active levanase and short FOS, suggesting that this site is perhaps important in the processing of longer FOS. Indeed, the presence of a secondary binding site is relatively common in endo-acting glycoside hydrolases and is proposed to be involved in enhancing enzyme-substrate targeting and processivity (Cuyvers et al., 2012).

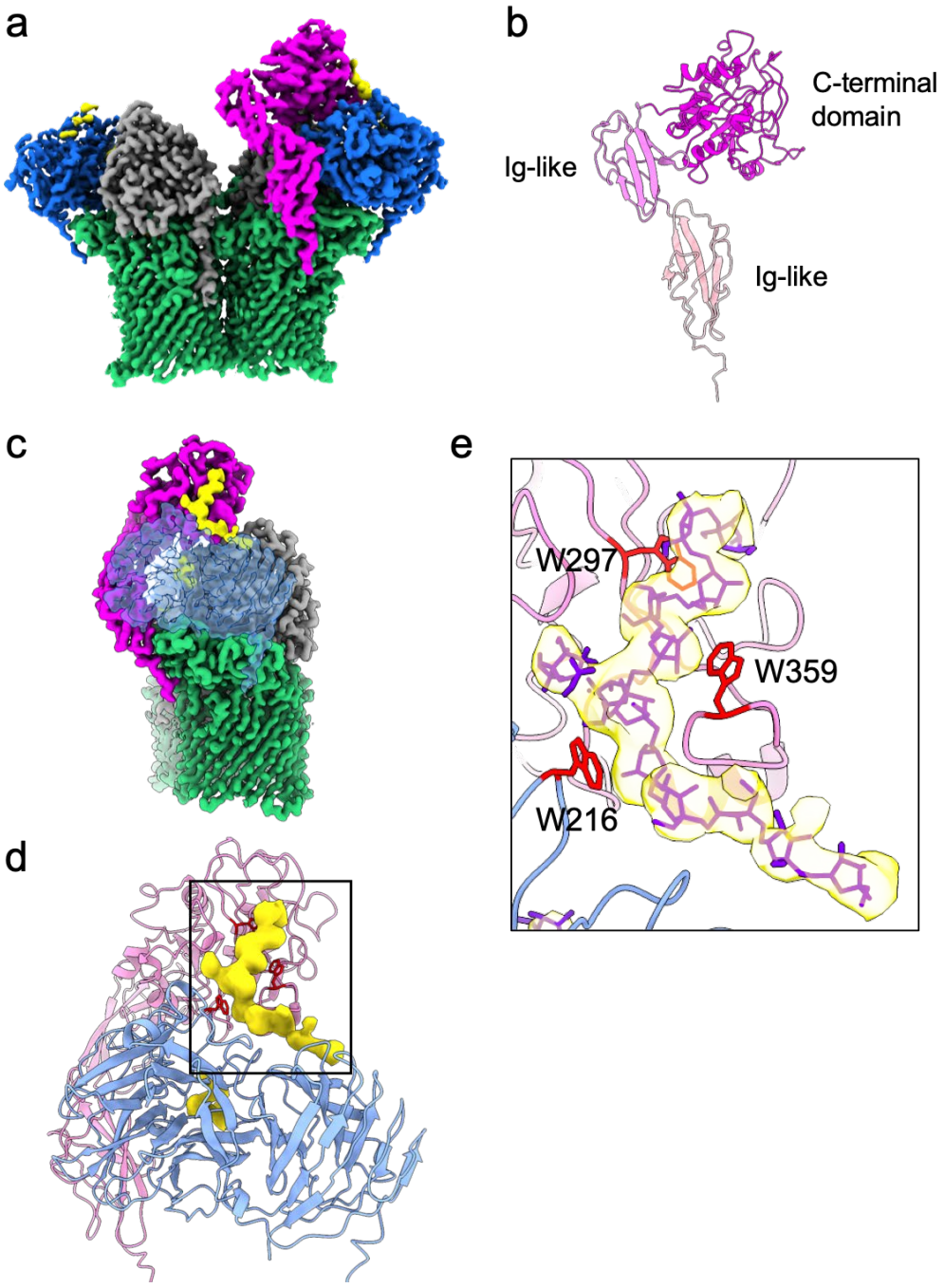


Figure 5.18 Structure of the levan binding protein tethered to the levanase via cooperative binding of a levan substrate. a, Reconstruction of the levan utilisome after focused classification viewed in the plane of the membrane. One levan binding protein is now fully resolved. **b**, Atomic model of the LBP showing the tripartite domain structure. Two Ig-like domains project the C-terminal levan binding domain away from the cell surface and towards the extracellular space. **c**, Side-on view of the complex with transparent density for Bt1760, highlighting the position of the cooperatively bound levan chain in the context of the complete utilisome. **d**, Structures of the levan binding protein and endo-levanase with additional density attributed to the cooperatively bound FOS shown in yellow (boxed). **e**, Zoomed inset of the boxed region in (**d**). The modelled levan chain is shown in purple. Trp297 and Trp359 from the LBP and Trp216 from the levanase are proposed to contribute to levan binding and are shown in red.

Finally, bound FOS is also observed within the inactivated site of this levanase variant, with density for five β2-6 linked fructose units that adopt the same twisted conformation as levan tetraose in the crystal structure of Bt1760 (PDB ID: 6R3U) (see **Figure 5.19**). No density linking levan at the tethering and the active site is present.

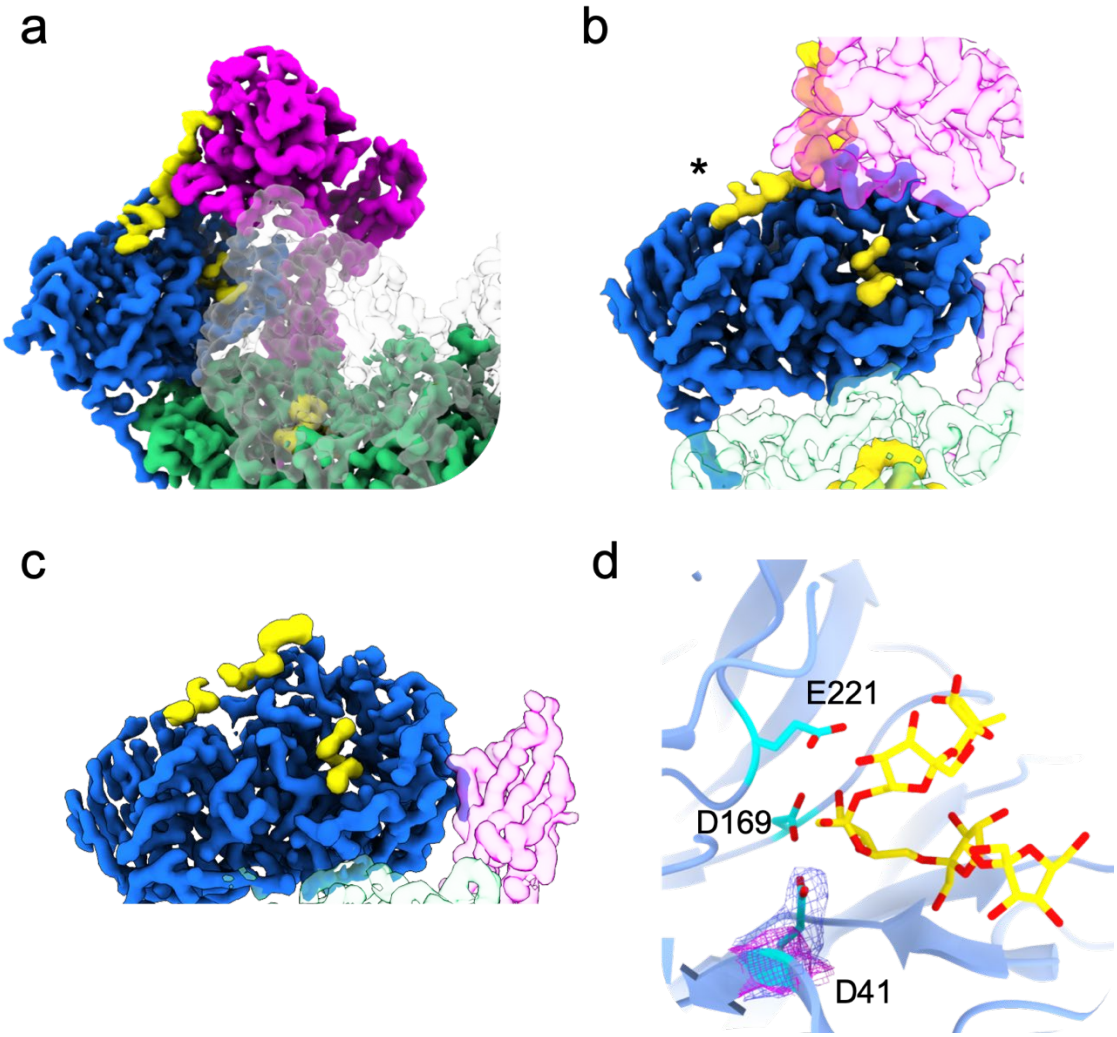


Figure 5.19 FOS are observed to bind two distinct sites on the levanase. a, Zoomed view of one half of the levan utilisome highlighting the tethered position of the LBP relative to the levanase, the SusD subunit (translucent density) and the top of the SusC barrel. **b**, Face on view of the levanase showing the distinct densities for levan bound at the secondary site (marked with an asterisk) and at the active site. No density linking these two sites was observed. In both (**a**) and (**b**) FOS bound at the upper site in SusC is observed. The minimum distance between FOS at the active site of the levanase and in the binding cavity of SusC is ~30 Å. **c**, View of the levanase associated with the alternate SusC unit. Here, the adjacent LBP is not tethered but levan is still observed to bind across the top of the levanase β-propeller and β-sandwich domains. **d**, Zoomed view showing levan FOS modelled into the active site of the inactive levanase. Residues comprising the catalytic triad are shown as stick models. The inactive levanase possesses a D41A mutation which is clearly resolved when overlaying density from the active (blue) and inactivated (pink) levanase reconstructions.

Why SusCD transporters exist as dimers is one of the key unanswered questions in the field. While the functional relevance of dimersation is not clear when thinking about the core complex in isolation, considering the question in the context of the utilisome provides some important insights. During 3D classification of both the active and inactive substrate-bound utilisome datasets, the inherent flexibility of the LBP was evident. However, in the inactive utilisome dataset, a novel conformation was observed in which an untethered LBP appears to reach across and contact a tethered LBP associated with the alternate SusC subunit (see **Figure 5.20**). The fact this 'reach over' conformation was observed exclusively in the inactive utilisome dataset, perhaps indicates that this conformation is the result of both LBPs interacting with the same stretch of levan chain; an event likely requiring longer FOS than those found in the active utilisome sample. Such cooperative binding would result in higher avidity for the substrate, offering a potential explanation as to how dimerisation contributes to function.

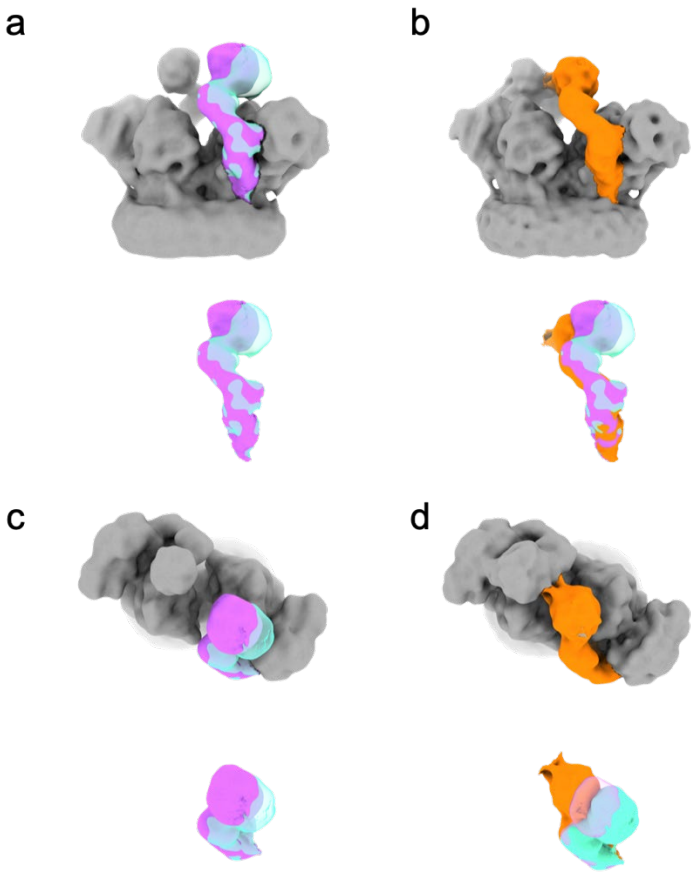


Figure 5.20 Evidence for cooperative substrate binding by both levan binding proteins of the levan utilisome. a, Variability in the conformation of the levan binding protein observed for the substrate bound utilisome with active glycosidase and FOS DP8-12. **b**, A novel state observed in 3D classification of the substrate bound utilisome with inactive glycosidase and FOS DP15-25. Here, the LBP (orange) appears to reach across and contact the LBP associated with the other SusC subunit that is present in a docked state. Given that this state was exclusively observed in the presence of long FOS and inactive glycosidase, it is postulated that this conformation is the result of both levan binding proteins interacting with the same stretch of substrate. **c** and **d** show the same maps as **a** and **b** rotated 90° such that they are viewed top-down from the extracellular space.

5.6 FOS can bind within the core $SusC_2D_2$ transporter in different ways

Throughout 3D classification, the core $SusC_2D_2$ transporter components were shown to uniformly occupy the closed state, and reconstructions of the complete utilisome contained FOS density enclosed with the SusCD units. To obtain the best resolution for SusCD-bound FOS, and to compare this substrate density with that observed in the active utilisome, a particle subtraction approach was again used so that all particles could be included in the refinement procedure, regardless of their lipoprotein complement. A final reconstruction of the core complex was resolved to 2.7 Å. Interestingly, while substrate density is present in the two previously defined sites within the SusC barrel, this density differs from that observed

in the active utilisome complex with short FOS. At the upper site located at the SusC-SusD interface, density previously assigned to a putative β2-1 decoration of Frc4 is missing (see **Figure 5.21**). Conversely, contiguous density can be seen at the previously defined terminus (reducing end) of the FOS molecule, with a novel β2,1 decoration on Frc5. At the lower FOS binding site close to the SusC plug domain, a similar trend is observed whereby the β2,1-linked fructose decoration on the non-reducing terminal fructose modelled previously is much weaker in the presence of longer FOS, and additional density attributed to another β2,6-linked monomer extends the reducing end of the chain in the direction of the upper FOS binding site. FOS molecules built into the cryoEM density map are depicted in **Figure 5.22**. At the upper site, the FOS molecule makes numerous polar contacts with side chains from both SusC and SusD. For SusD, these include D41, N43, D67, R368 and Y395, and for SusC, D383, D406 and N901. Notably, the ring of Frc2 engages in stacking interactactions with W395 of SusD. In contrast, the FOS molecule at the lower site interacts exclusively with SusC. Barrel wall residues Q372, K454, E481 and W483 engage in polar interactions with FOS, as do residues H169 and E170 of the plug domain. It should be noted that at the resolution of the data, density predominantly corresponds to the fructose rings, with only limited information available for hydroxyl groups. As such the hydrogen bonding networks depicted in **Figure 5.22** should not be over-interpreted. Nonetheless, residues engaging in multiple hydrogen-bonding interactions such as R368 of SusD as well as E170 and Q372 of SusC are likely to make important contributions to substrate binding.

Strikingly, when viewing the map at higher threshold levels, diffuse density is visible that connects both FOS binding sites, indicating that longer FOS (~DP15) can occupy both sites simultaneously. The density between the two FOS binding sites is relatively weak and indicative of multiple conformations, consistent with the absence of any contacts from SusC in this bridging segment. Attempts to characterise this heterogeneity and shed light on the diversity of FOS bound using focused 3D classification were unsuccessful, likely owing to the extremely low signal-to-noise ratio associated with this experiment. It is important to interpret these results with some caution given that the sample did not contain an active levanase component. In this experiment, the FOS added were heterogeneous in size and likely in branching pattern and it is not known whether or not such a heterogenous substrate pool

is biologically relevant. Indeed, the close association of the levanase, the LBP and the SusC subunits in the utilisome assembly may favour the production and capture of FOS that are smaller and more uniform in size. Structure determination of the levan utilisome containing an active levanase purified from *B. theta* grown on *Erwinia* levan would be valuable in addressing this question.

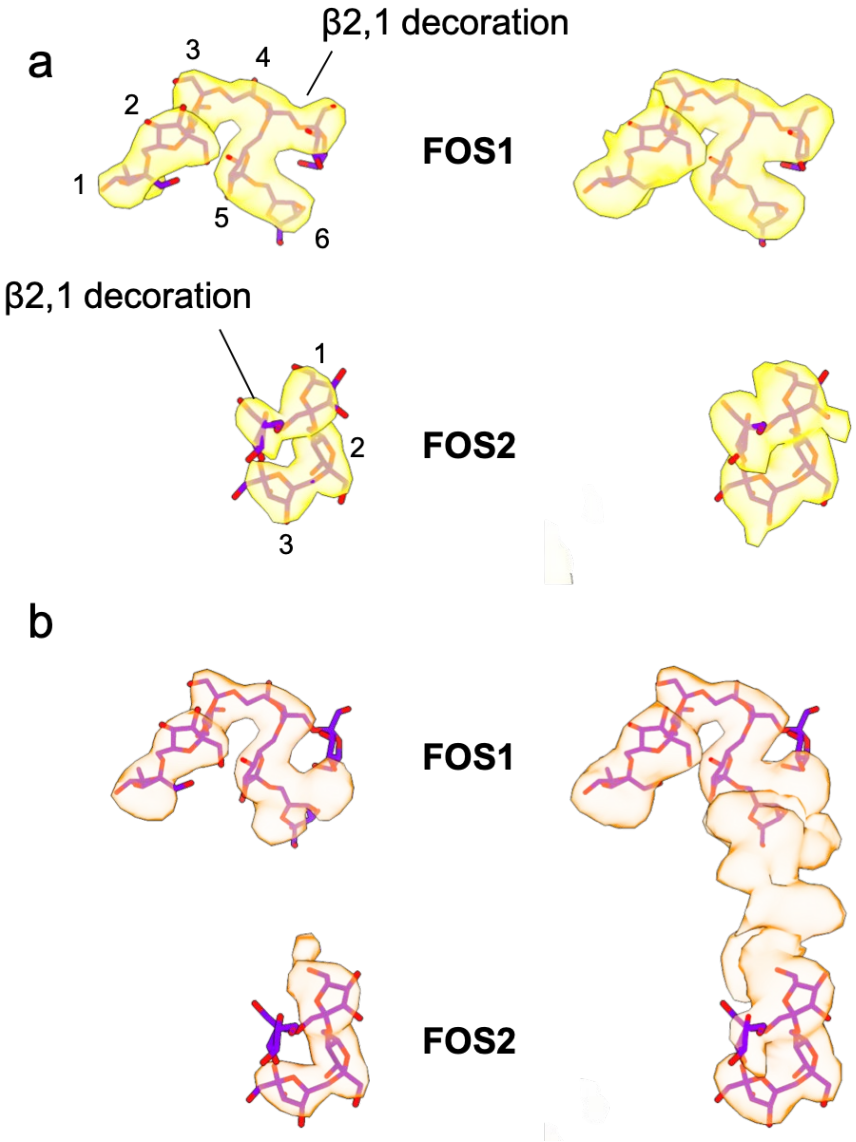


Figure 5.21 Assessing differences in substrate density for short and long FOS bound in the cavity of SusC. a, Isolated FOS density obtained from the dataset with active levanase and short FOS. Density for substrate (yellow) is shown at low (left) and high (right) thresholds. **b**, Isolated FOS density obtained from the dataset with inactive Bt1760 and long FOS. Density for substrate (orange) is shown at low (left) and high (right) thresholds. Note the presence of unmodeled substrate density bridging the binding sites visible at high thresholds. FOS models shown are from the original X-ray crystal structure.

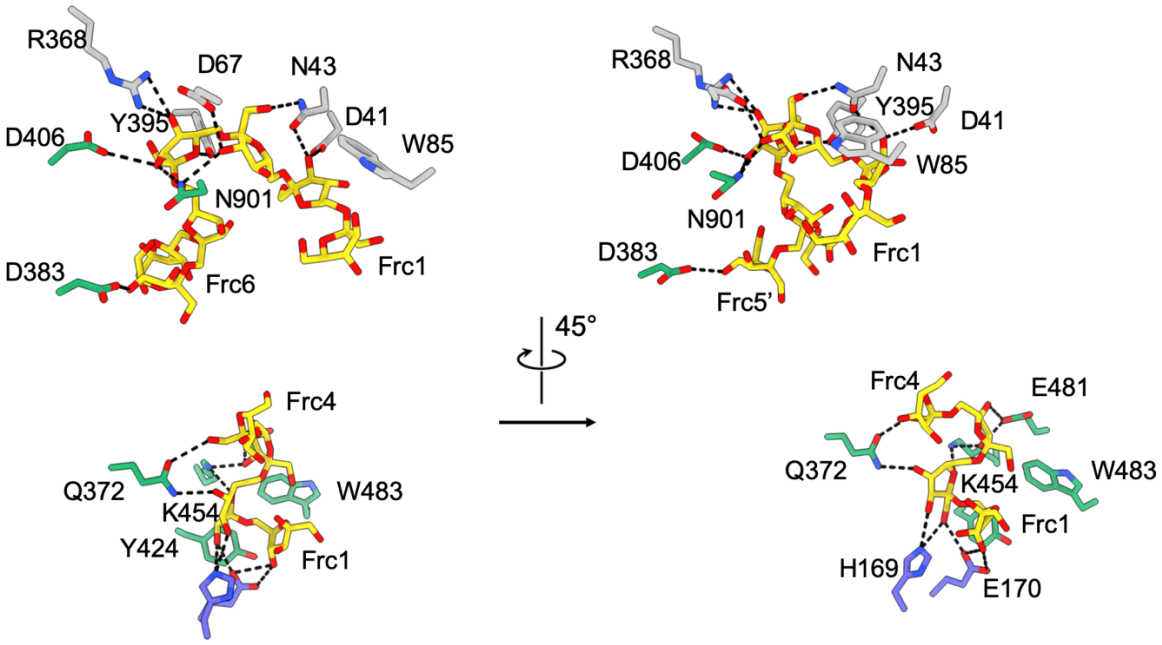


Figure 5.22 Interactions of FOS with the core SusCD unit of the transporter. FOS molecules observed within the binding cavity of a single SusCD monomer from the inactive utilisome sample in the presence of FOS ~DP15-25. FOS molecules are shown in yellow. Side chains of residues involved in polar interactions with the substrate are displayed as stick models. Hydrogen bonding interactions are shown as dashed black lines. Residues from SusD are coloured grey, while residues from the SusC barrel are coloured green, and residues from the SusC plug domain are coloured purple.

5.7 Stable utilisome complexes are not unique to the levan system

As outlined in the introduction to this chapter, two models have emerged to describe the association of additional lipoprotein components with SusCD transporters: the stable complex model and the dynamic assembly model. To investigate whether stable utilisome architecture observed for the levan system is a general feature of PUL encoded systems, a second example was selected for further study; the dextran utilisation system of *B. theta*. Dextran is an α1-6 linked glucose polymer with occasional α1-4 glucose branches and, like the levan utilisation system, the dextran PUL encodes four OM components: the SusC transporter and its SusD lid (Bt3090 and Bt3089), a GH66 endo-dextranase (Bt3087) and a putative surface dextran binding protein (Bt3088) (see **Figure 5.23**).

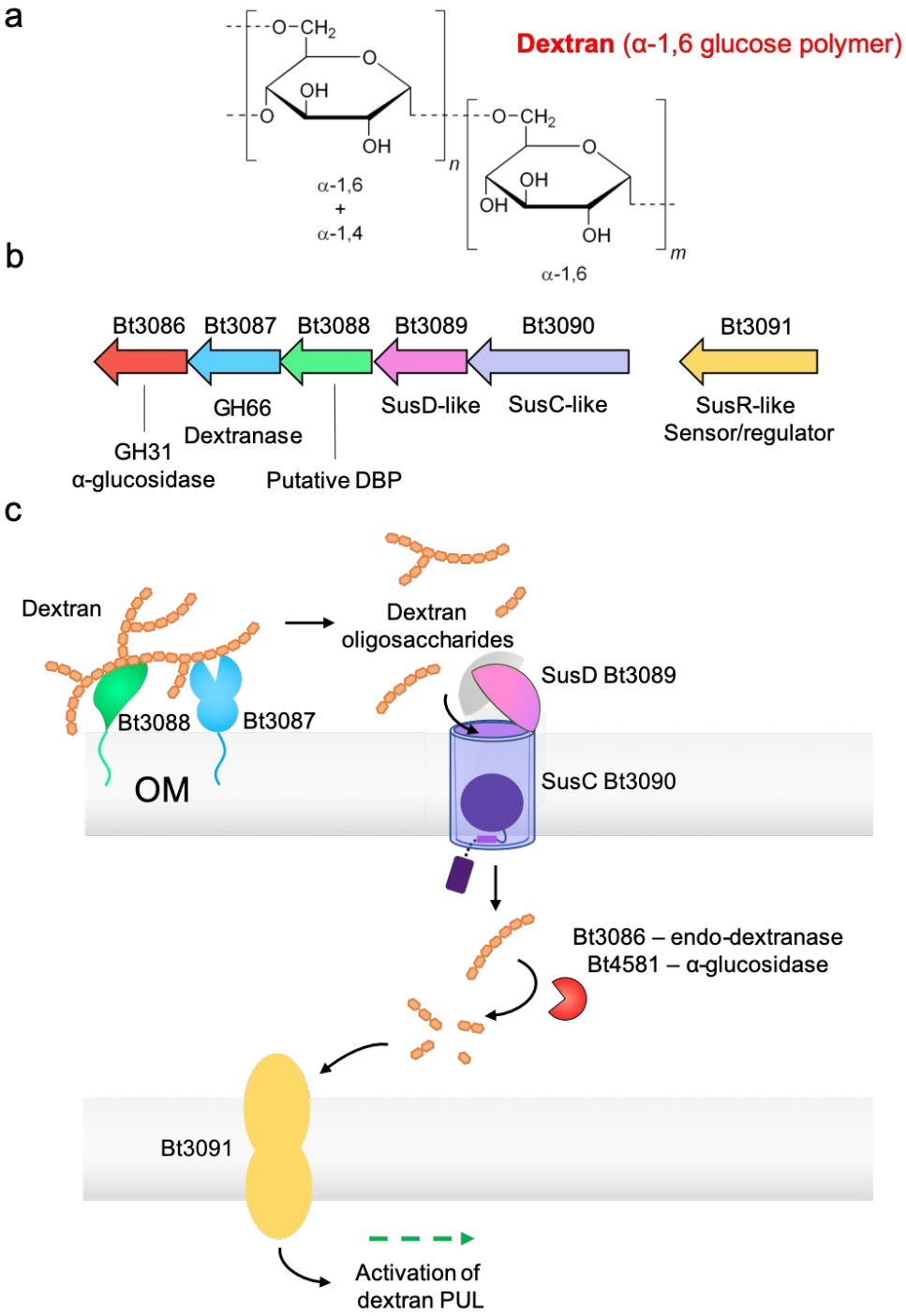


Figure 5.23 Overview of the dextran utilisation system. a, Structure of the dextran polysaccharide. The sugar is composed of repeating α1,6-linked glucose units with occasional α1,4 branch points. **b**, Schematic of the dextran PUL. Bt3090 and Bt3089 correspond to the SusC and SusD components respectively. Bt3088 is a SusE-positioned gene and thus has a putative role as a dextran binding protein. Bt3087 is a surface exposed endo-dextranase while Bt3086 is a periplasmic endo-dextranase. The PUL is regulated by Bt3091, a SusR-like sensor/regulator (one component system) located upstream of the PUL. Bt4581 is a periplasmic α-glucosidase not physically associated with the dextran PUL, but it is co-regulated with the other PUL components. **C**, Schematic demonstrating the arrangement and function of PUL-encoded components.

Collaborators at Newcastle University purified the putative dextran utilisome complex from a dextran-grown *B. theta* strain expressing a His-tagged SusD protein. The putative complex was extracted in a mix of DM and DDM and the remainder of the purification protocol was carried out in DDM. As was observed for the levan utilisome, the putative dextran utilisome ran as a single, monodisperse peak in analytical size exclusion chromatography which corresponded to a complex containing all four OM-associated PUL components as judged by SDS-PAGE (see **Figure 5.24**).

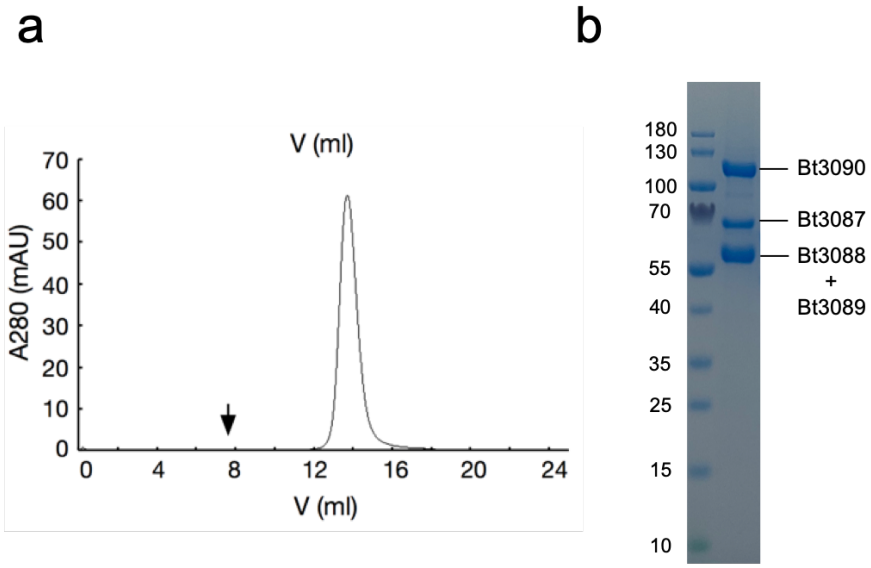


Figure 5.24 Purification of a stable dextran utilisome complex. a, Analytical SEC elution profile for the Bt3087-90 dextran four-component complex run on a Superose-6 column. The void volume of the column is indicated by an arrow. **b**, SDS-PAGE of the complex purified in (**a**) with the identity of the bands indicated. Note that Bt3088 and Bt3089 are almost the same MW (56 kDa and 55.7 kDa, respectively) and so appear as a single band on the gel. Analytical SEC and SDS-PAGE were carried out by Bert van den Berg, Newcastle University

5.8 Negative stain TEM of the dextran utilisome complex solubilised in DDM

Using negative stain TEM, the quality of the DDM-solubilised dextran four-component complex was assessed. The sample was provided at a stock concentration of 3 mg/ml, and screening revealed that a 200-fold dilution was necessary and sufficient to achieve useful particle distributions on continuous carbon coated grids. The well-defined symmetrical architecture observed for the levan utilisome, as well as the considerable size difference between this higher order assembly and the core $SusC_2D_2$ transporter, suggested that it might be possible to investigate complexation of the dextran PUL components by negative stain. Samples were stained using uranyl formate since this stain was found to provide better stain

depth and greater particle definition than uranyl acetate (see **Figure 5.25**). A small negative stain dataset of 59 micrographs was collected and processed through to 2D classification (see **Figure 5.25**). Manual inspection of the resulting classes shows strong preferred particle orientation where the predominant view is down the axes of the SusC barrels. Nonetheless, comparison of the 2D classes indicates that particles are compositionally heterogeneous, with some classes possessing a clean 'figure eight' view characteristic of the core $SusC_2D_2$ complex, and others displaying decorated variations that likely indicate the association of additional lipoprotein components. Unfortunately, the preferred orientation precluded generation of a 3D reconstruction from this data. In an attempt to circumvent this problem by improving the distribution of viewing angles, two additional datasets of 63 and 55 micrographs were collected with stage tilts of 35° and 50° respectively. In the case of the 35° tilted collection, 2D class averages show some additional views relative to the data collected at 0° of tilt. However, even when combining the 0° and 35° tilt data, attempts at a 3D reconstruction were still unsuccessful. Results of 2D classification on particles collected at a 50° tilt showed no further improvement in the angular distribution, instead the class averages contain less detail than observed for previous datasets. This is likely explained by the increase in specimen thickness at higher tilts. In this case, the thicker specimen makes multiple scattering events more likely, resulting in noisier images and a lower data quality.

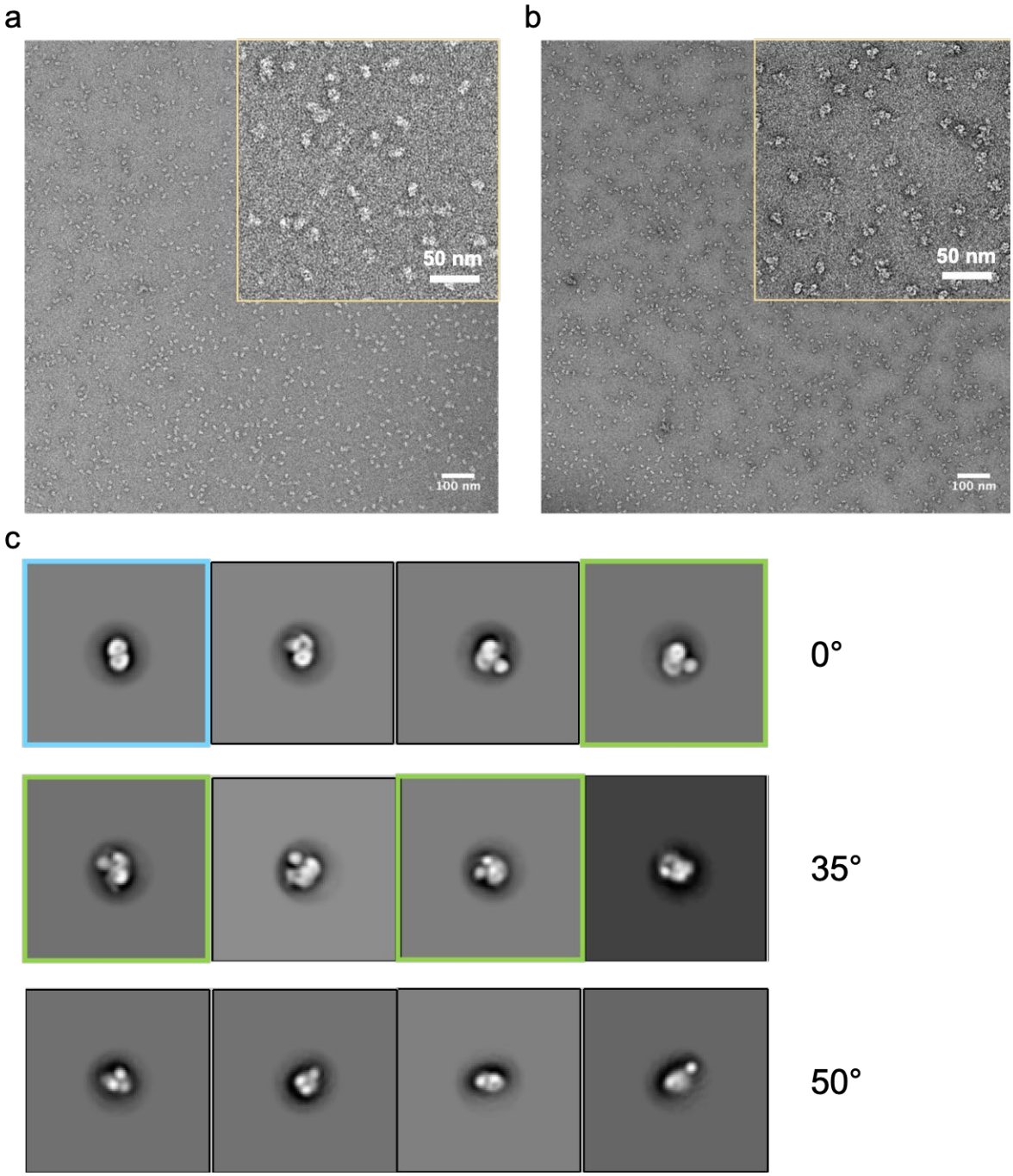


Figure 5.25 Negative stain TEM on the putative dextran four-component complex. a, Representative micrograph of the putative four-component complex stained with uranyl acetate. **B**, Same sample as (**a**) but stained with uranyl formate which yielded improved stain depth and better-defined particle features. Sample was loaded at 0.015 mg/ml and images were acquired at 39,886 x magnification. **c**, Example 2D class averages from negative stain data sets at 0, 35 and 50 degrees of stage tilt. The blue box highlights a class that corresponds to a naked $SusC_2D_2$ transporter while green boxes identify some example classes that show additional density, causing deviation from the characteristic figure eight shape of the core complex and thus indicating association of additional lipoproteins.

5.9 Single particle cryo-electron microscopy of the dextran utilisome

The 2D class averages obtained from the negative stain TEM datasets provided an early indication that the dextran four-component sample may contain some higher order assemblies. However, without 3D reconstructions, it was not possible to assess whether the additional components interact with the core transporter in a manner analogous to the levan utilisome. To probe this question further, single particle cryoEM was required.

5.9.1 Grid optimisation and data collection on the dextran four-component complex

The on-grid behaviour of the four-component dextran utilising system was markedly distinct from that of the levan utilisome. Under the same preparation conditions, namely amylamine glow-discharged quantifoil grids and a sample concentration of 3 mg/ml, the dextran system exhibited extensive aggregation with no areas suitable for data acquisition on the grids (see **Figure 5.26**). Alternative solubilising detergents were trialled, including LMng and DMng, but in both cases particle behaviour was comparable to the DDM-solubilised sample. Higher sample concentrations of 5mg/ml and 10 mg/ml were trialled on both carbon and gold quantifoil grids in an attempt to induce formation of a protein monolayer, an approach that has proven successful for G-protein coupled receptors (Danev et al., 2021). Unfortunately, neither approach was observed to alleviate aggregation (data not shown).

Having been unable, to optimise conditions in the absence of a continuous carbon support layer, carbon-coated lacy grids were trialled and a DDM-solubilised sample concentration of 0.06 mg/ml was found to yield particle distributions suitable for data collection (see **Figure 5.26**). A dataset of 6331 micrographs was collected from which 820,184 particles were picked using crYOLO. After cleaning the particle stack over several rounds of 2D classification, 477,707 particles remained. Representative 2D classes are shown in **Figure 5.26**.

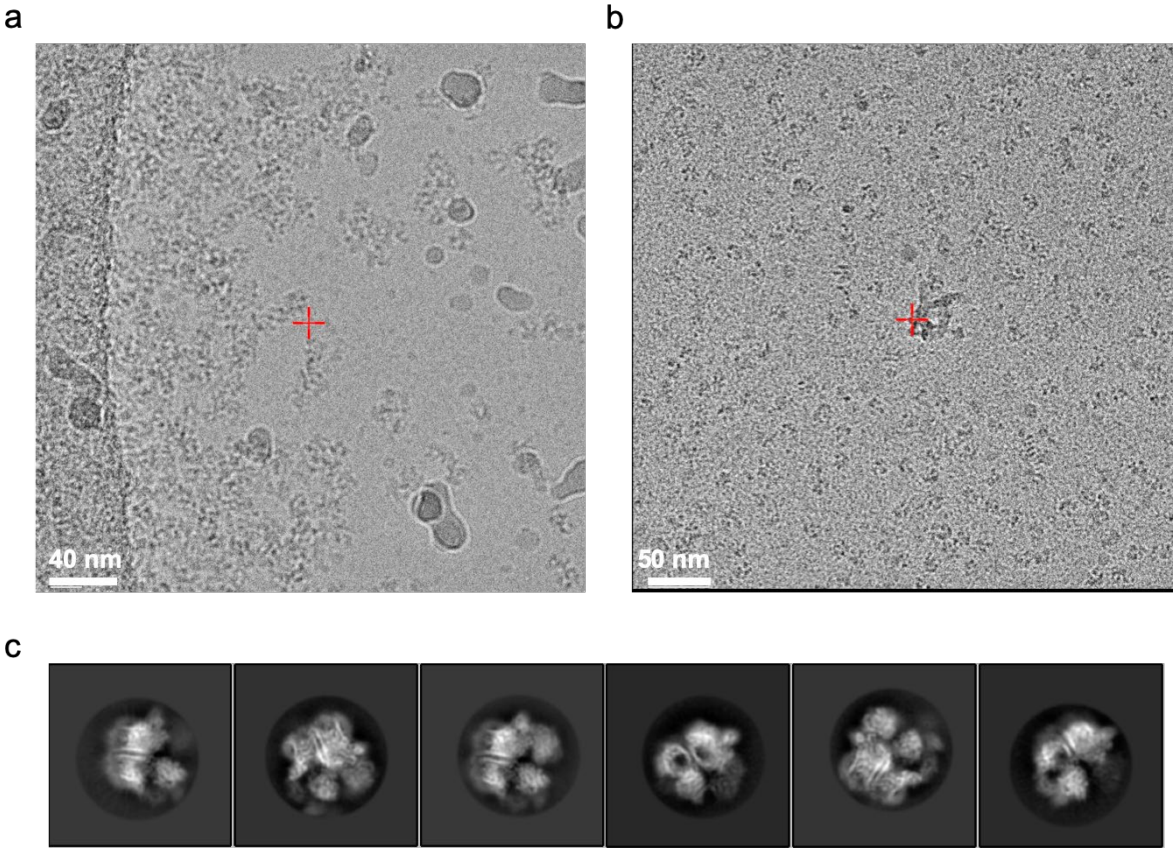


Figure 5.26 Grid optimisation and representative 2D class averages of the dextran four-components complex. a, Micrograph of 3 mg/ml dextran four-component complex on holey grids in the absence of a carbon support. Extensive aggregation is observed. **b**, Micrograph of 0.06 mg/ml dextran four-component complex on lacy carbon grids with an ultrathin carbon support. There are a good number of particles in the image and the particles present are monodisperse.

An initial model was generated *de novo* from the data and used as a reference for 3D classification. Three of the six classes generated show good density for the core $SusC_2D_2$ transporter and, in each case, there is evidence of the association of additional lipoproteins (see **Figure 5.27**). As for the levan utilisome complex, two distinct density morphologies were observed: an extended snake-like density, and a compact globular density. Structure predictions of the dextranase and putative dextran binding (DBP) protein using AlphaFold2 allowed assignment of the globular density to the dextranase and of the extended density to the first two domains of the DBP (see **Figure 5.27**).

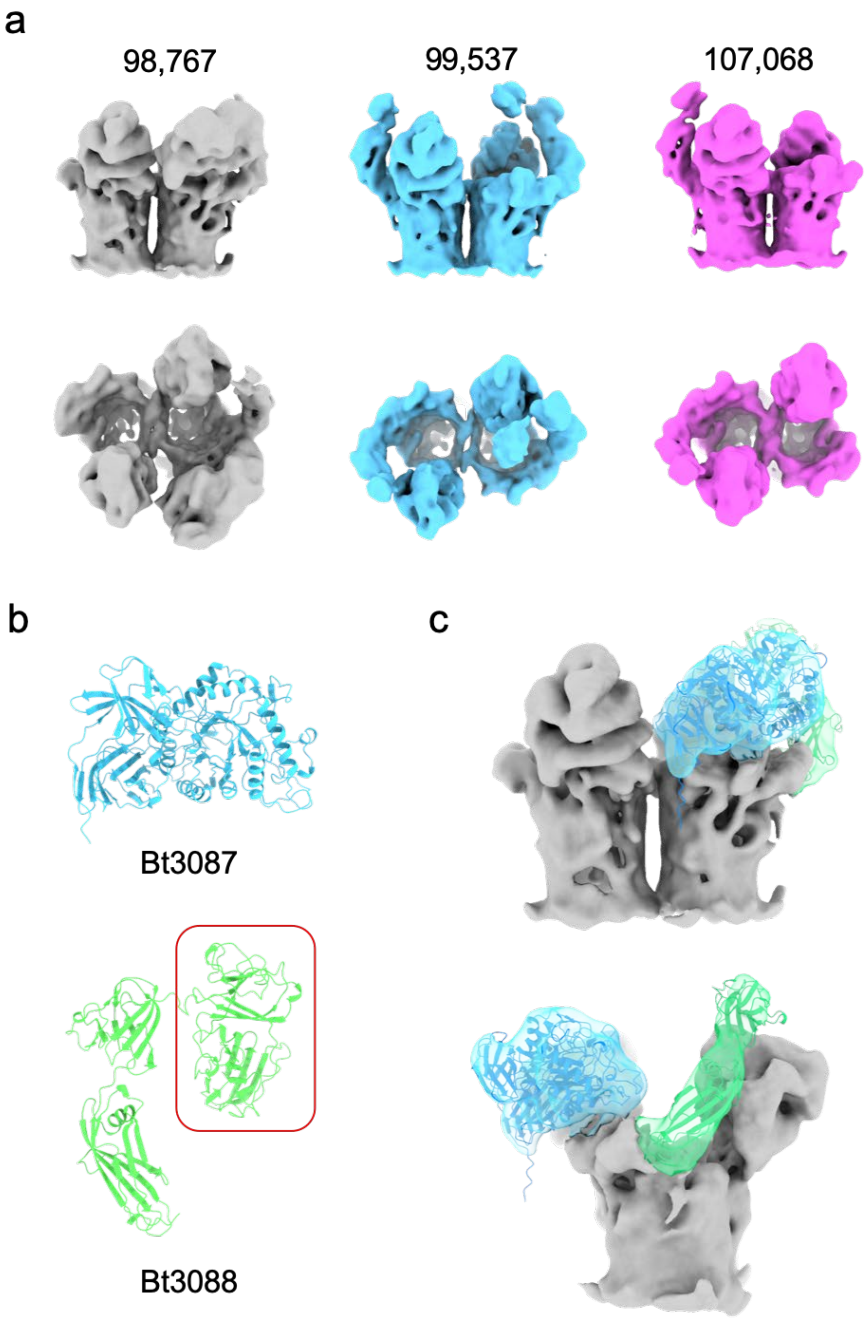


Figure 5.27 Initial outputs from 3D classification of the dextran utilisome and assignment of lipoprotein density. a, Good 3D classes with clear density for the core transporter viewed in the plane of the membrane (top) and from the extracellular space (bottom). Unassigned density is also present in all classes. Particle numbers contributing to each class are displayed. **b**, AlphaFold2 predictions of the dextranase (blue) and putative DBP (green) of the dextran utilisation system. **c**, Grey class shown in (**a**) with additional densities assigned to the dextranase and putative DBP on the basis of predicted structures. Note that the C-terminal domain of the DBP (red box) is not visible in the EM map and the docked model for the DBP has been truncated accordingly.

5.9.2 Assessment of compositional heterogeneity of the dextran utilisome

Further classification of particles revealed considerable compositional heterogeneity in the data. However, unlike for the levan system, there were no classes corresponding to an octameric dextran utilisome, which we attribute to a higher propensity of the complex to disassemble during grid preparation, perhaps as a result of the continuous carbon support required for data collection on this sample. The highest order complex was instead heptameric, containing two copies of the dextran binding protein and a single dextranase. Other compositions include hexameric and pentameric complexes comprised of all possible combinations of additional lipoproteins (see **Figure 5.28**). In all cases, SusD subunits were found to occupy an open position despite the complex having been purified from dextran grown *B. theta*.

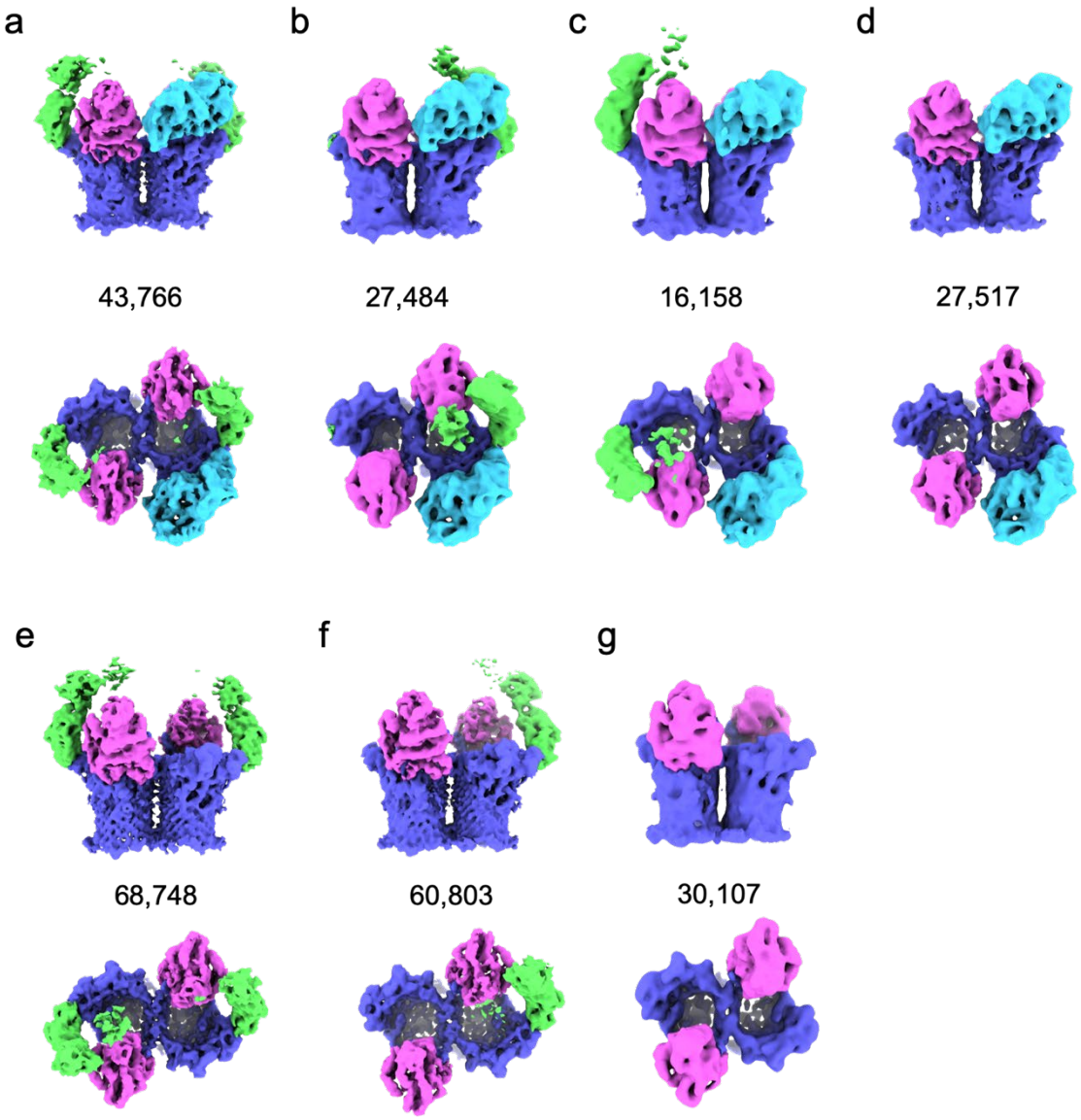


Figure 5.28: Compositional heterogeneity of the dextran utilisome observed by cryoEM. a-g, Refined outputs of 3D classification viewed in the plain of the membrane (top) and from the extracellular space (bottom), where each map corresponds to a unique complement or arrangement of auxiliary components. SusC (Bt3090) is purple, SusD (Bt3089) is pink, the dextran binding protein (Bt3088) is green and the dextran glycosidase (Bt3087) is blue. Particle numbers contributing to each class are shown.

5.9.3 High-resolution structure determination of the dextran SusC component by cryoEM

Alongside these structural studies, collaborators at Newcastle University used X-ray crystallography to solve the structures of the dextranase and SusD components in isolation. To complement this work, we targeted a high resolution cryoEM reconstruction that included the SusC component of the complex. A mask was generated that encompassed only the core $SusC_2D_2$ transporter, since this complex was present in all good particles. This mask was then applied in a consensus focused refinement which yielded a 3.1 Å map after sharpening. Using an AlphaFold2 predicted model for the SusC component as a starting point, an atomic model was refined into the EM density. In accordance with the uniformly open state of the transporter and the apparent absence of dextran substrate, the TonB box region of SusC was visible and held with the barrel. As for Bt1763, the region preceding the TonB box is observed to take a sharp turn towards the plug with Tyr88 engaging in stacking interactions with Tyr190 of the plug domain (see **Figure 5.29**). Density for the SusD component was of insufficient quality to permit model building and refinement, and this was not pursued given the availability of the crystal structure of this subunit.

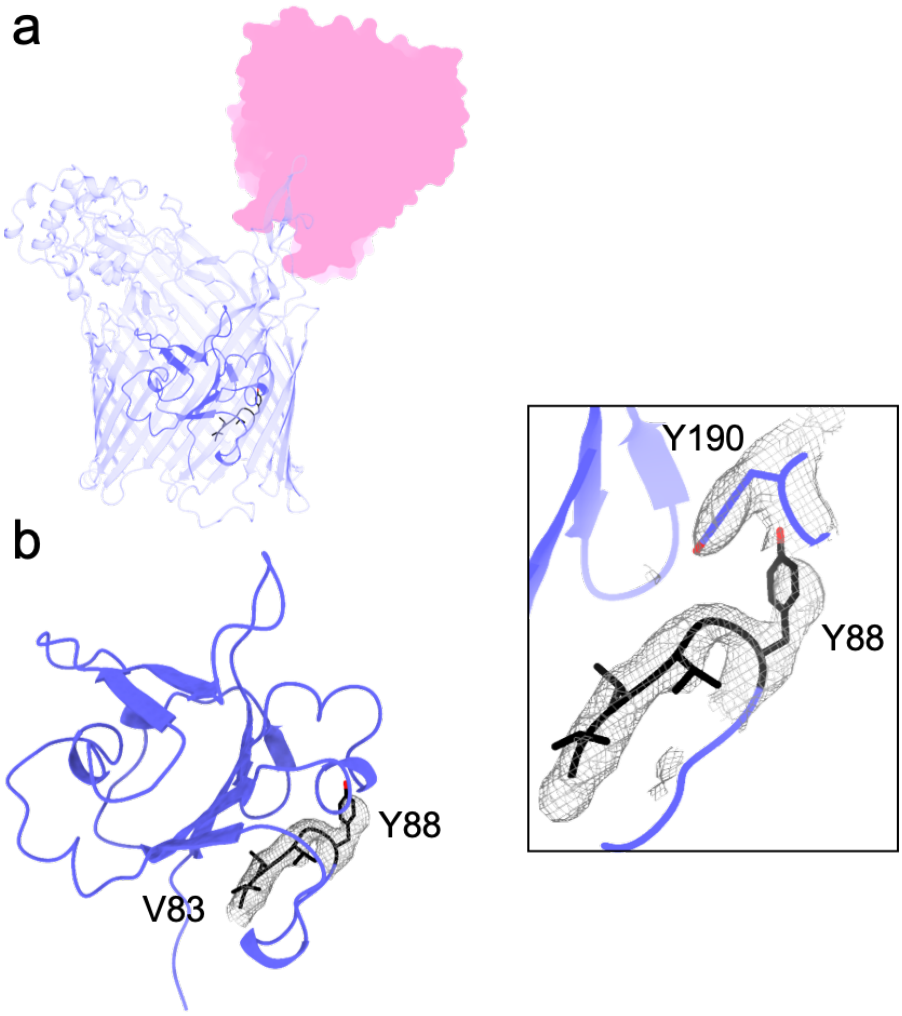


Figure 5.29 Analysis of the Ton box region in the dextran SusC subunit. a, Refined model of Bt3090. The corresponding SusD component is shown as a shaded zone. **b**, Zoomed view of the SusC plug domain in isolation showing the TonB box region (density shown) held close to the plug domain. The zoomed inset (boxed) displays a putative stacking interaction between Y88 which precedes the TonB box and Y190 of the plug domain.

5.10 Levan and dextran utilisomes display different arrangements of components

Returning to the heptameric form of dextran utilisome, the relative positions of the additional lipoproteins on the core complex differ between the levan and dextran utilising systems (see **Figure 5.30**). For the former, the levanase is closest to the hinge region of SusD whereas for the latter, the DBP is closest. Regardless, the arrangement of components in the dextran utilisome is still compatible with movement of the SusD lid.

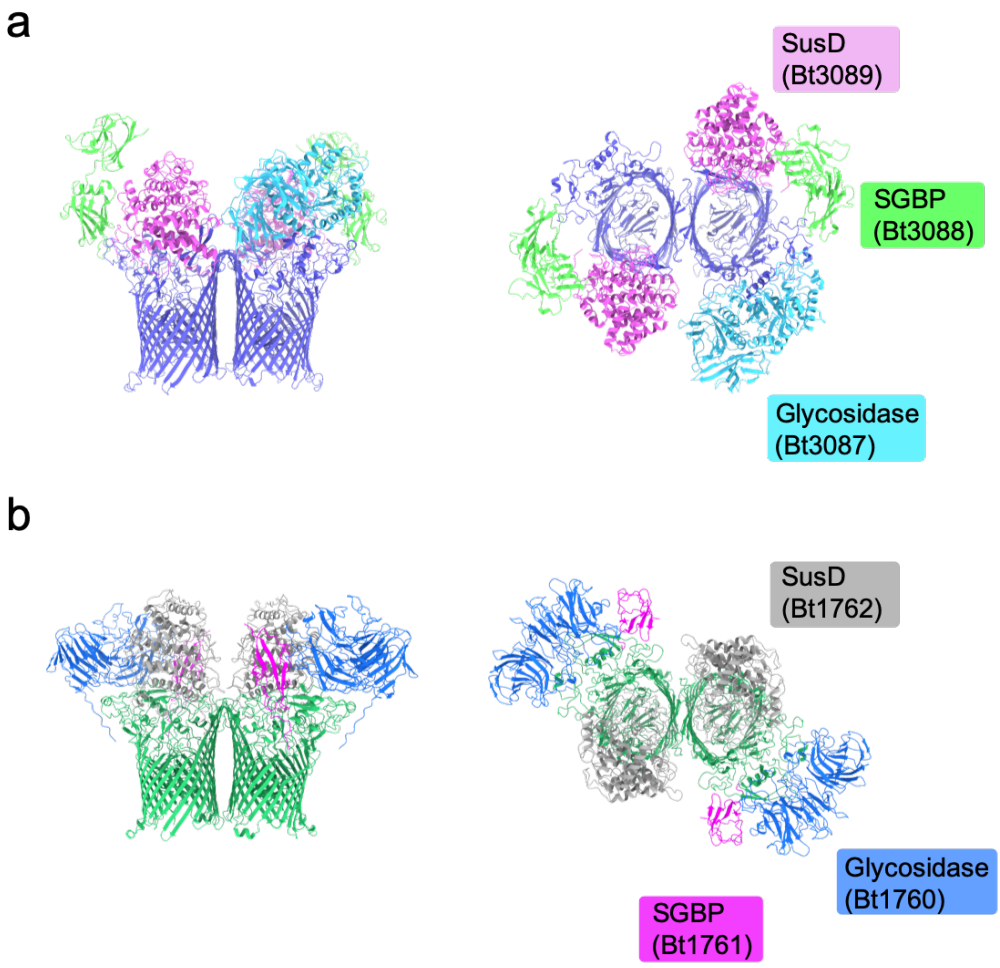


Figure 5.30 Arrangement of auxiliary proteins in glycan utilisomes. a, A composite model demonstrating the arrangement of the dextran utilisome. CryoEM data permitted refinement of the SusC components. Crystal structures were determined for SusD (Bt3089) and the dextran glycosidase (Bt3087) and docked into the cryoEM map for the heptameric complex. An alphafold2 structure prediction for the dextran binding protein (Bt3088) was also rigid fit to the cryoEM density. Unambiguous density is visible only for the first two domains of the dextran binding protein and the predicted model was therefore truncated prior to the C-terminal domain. **b**, Structure of the levan utilisome. SusD components (Bt1762) are positioned based on a rigid body fit to a low-resolution map in which unambiguous density for the lids was observed. Note the different arrangement of the glycosidase and SGBP components relative to SusD in the levan and dextran systems.

5.11 Proteomics experiments support the stable assembly hypothesis

The data presented thus far strongly support the hypothesis that the levan and dextran utilisation machineries assemble as stable utilisome complexes at the OM. Both complexes can be purified from membranes by detergent extraction and both run as a single peak in analytical SEC. However, to further support this notion and to shed light on the relative abundance of the components at the OM, collaborators performed semi-quantitative proteomics using intensity-based absolute quantification (iBAQ) on the total membrane fractions from *B. theta* grown on either fructose or dextran. In the presence of these cognate substrates as the sole carbon source, the OM-associated components of the corresponding PUL are some of the most highly expressed membrane proteins in *B. theta.* Accounting for the semi-quantitative nature of iBAQ analyses of abundance, the individual OM-associated components were found to be present in similar amounts i.e. there is no excess of additional lipoproteins relative to SusC and SusD (see **Figure 5.31**).

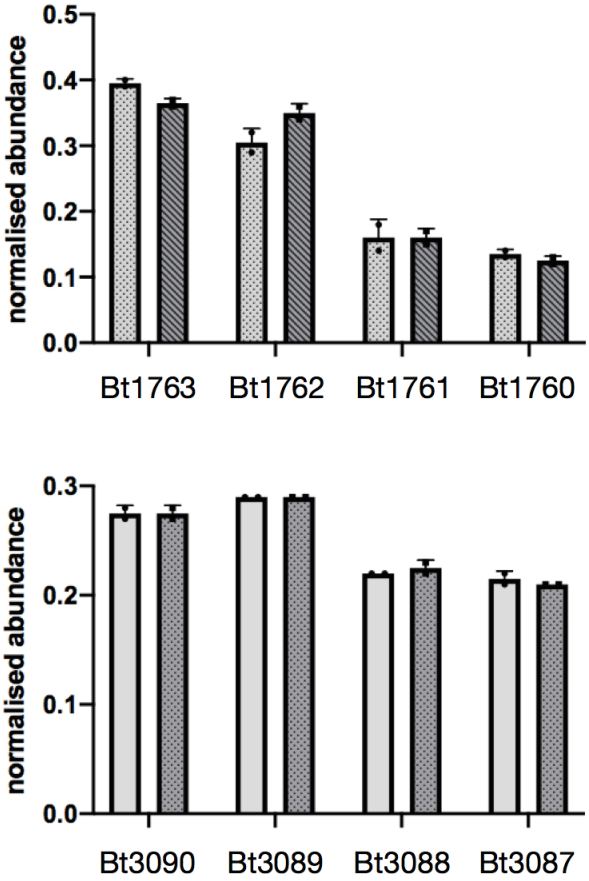


Figure 5.31 Proteomic analysis of levan and dextran PUL components present in *B. theta* total membrane fractions and purified utilisome complexes. a, Normalised abundance of the four components of the levan (top panel) and the dextran utilisomes (bottom panel) obtained from fructose- and dextran-grown cells respectively (light bars). The dark bars show the normalised abundance of detergent-purified complexes spiked into the proteome samples of dextran- or fructose-grown cells (top and bottom panels, respectively). Proteomics experiments were carried out in the group of Matthias Trost, Newcastle University.

Importantly, expression of the levan PUL is not activated when *B. theta* is grown on dextran, with the reverse true for expression of the dextran PUL. As a result, in dextran-grown cells, components of the levan PUL are present at just ~1% of the level found when cells are grown in fructose, and in fructose-grown cells, dextran PUL components were not detectable. This provided the opportunity to spike membrane proteome samples from cells grown on fructose or dextran with detergent-purified dextran or levan four-component complexes respectively. These spiking experiments allow comparison of the relative abundance of components in purified complexes versus endogenous membrane samples. Relative abundances in purified and endogenous sample were found to be very similar for both the levan and dextran complexes (see **Figure 5.31**), suggesting that there is no excess of any component in the OM. Collectively these data support the hypothesis that the additional lipoproteins form stable,

equimolar, four-component utilisome complexes with their cognate SusCD transporters in the outer membrane of Bacteriodetes.

5.12 Discussion on levan and dextran utilisomes

Our cryoEM structures of the levan utilisome provide remarkable insight into glycan acquisition by *Bacteroides spp*.. Furthermore, as all sequenced Bacteroidetes to date contain SusCD homologues it can be assumed that these findings apply more generally to all glycan utilisomes and potentially SusCD apparatus involved in transport of non-glycan macromolecules e.g the Bt2261-Bt2264 complex, the substrate of which is tentatively assigned as a peptide (Glenwright et al., 2017). As Bacteroidetes are found in a diverse range of terrestrial and marine niches these findings also expand our understanding of glycan import outside the animal gut. We demonstrate that for two distinct PULs, all of the gene products that localise to the bacterial outer membrane form a stable utilisome complex with a defined architecture. The position of each component within this complex appears to be optimised for their respective functions. Firstly, the SusD lids are able to open and close without clashes with the other components. Secondly, the observed mobility of the levan binding protein within the apo utilisome structure aligns with its putative levan binding activity, and its presumed role in grappling long lengths of levan chain, positioning them for cleavage by the proximal levanase. The active site of the levanase is then, in turn, proximal to the substrate binding site within the barrel of the transporter.

Substrate addition experiments on the levan utilisome provided a direct demonstration of the pedal bin mechanism of substrate capture for the first time. The SusD lids were invariably open in the substrate free levan utilisome, allowing access to the substrate binding site within the SusC transporter from the extracellular space. Upon addition of levan substrate, we observed complete closure of the lids, such that bound FOS were encapsulated within the sealed transporter, as well as evidence of allosteric signalling of substrate binding; as judged by a conformational change at the periplasmic face of the SusC plug which results in disordering of the TonB box region. Thus, our findings align with the literature consensus for the mechanism of TonB-dependent transport whereby substrate binding renders the TonB

box region available for interaction with TonB, facilitating plug disruption/ejection and translocation of the encapsulated levan oligosaccharide into the periplasm.

Outside of the core transporter, long levan chains (DP ~15-25) form a bridge between the levan binding domain of the SGBP and the levanase, and this raises several interesting questions. Firstly, the long FOS chain is bound in an extended conformation, making it distinct from FOS bound elsewhere in this and other structures. This may arise because the chain is held under tension, and indeed in the absence of substrate, the preferred position of the LBP is in an extended conformation which reaches out towards the extracellular space. The substrate-bound utilisome structure captured here with long FOS may therefore represent a state that is 'strained', with a bound levan chain that is pulled taut. It is important to note that this conformation has been selected for during image processing owing to the fact that it represents the least mobile and therefore most resolvable state. The binding of this linking levan by the levanase is independent of its catalytic site in our structure and as such it is not immediately clear how the LBP-levanase interaction contributes to concerted utilisome function. However, given the inactive state of the levanase it is possible that short FOS preferentially bind the inactivated active site, precluding the observation of longer levan chains that bridge the secondary and active sites. Alternatively, if longer levan chains are indeed bound in our structures, flexibility away from the interaction sites may preclude resolution of contiguous density. Nevertheless, the linking levan interaction likely facilitates recruitment of the C-terminal domain of the LBP to the levanase (see **Figure 5.32**). Longer FOS, or chains possessing branches of significant length may be able to bind at both sites and have enough 'free' chain left to feed into the active site of the levanase. We anticipate that, for active utilisomes in the OM, these interactions are dynamic and as such the presence of the secondary binding site on the levanase likely increases the probability of retaining levan close to the active site, promoting efficient capture and cleavage. However, the functional relevance of this site remains to be demonstrated by further experiments.

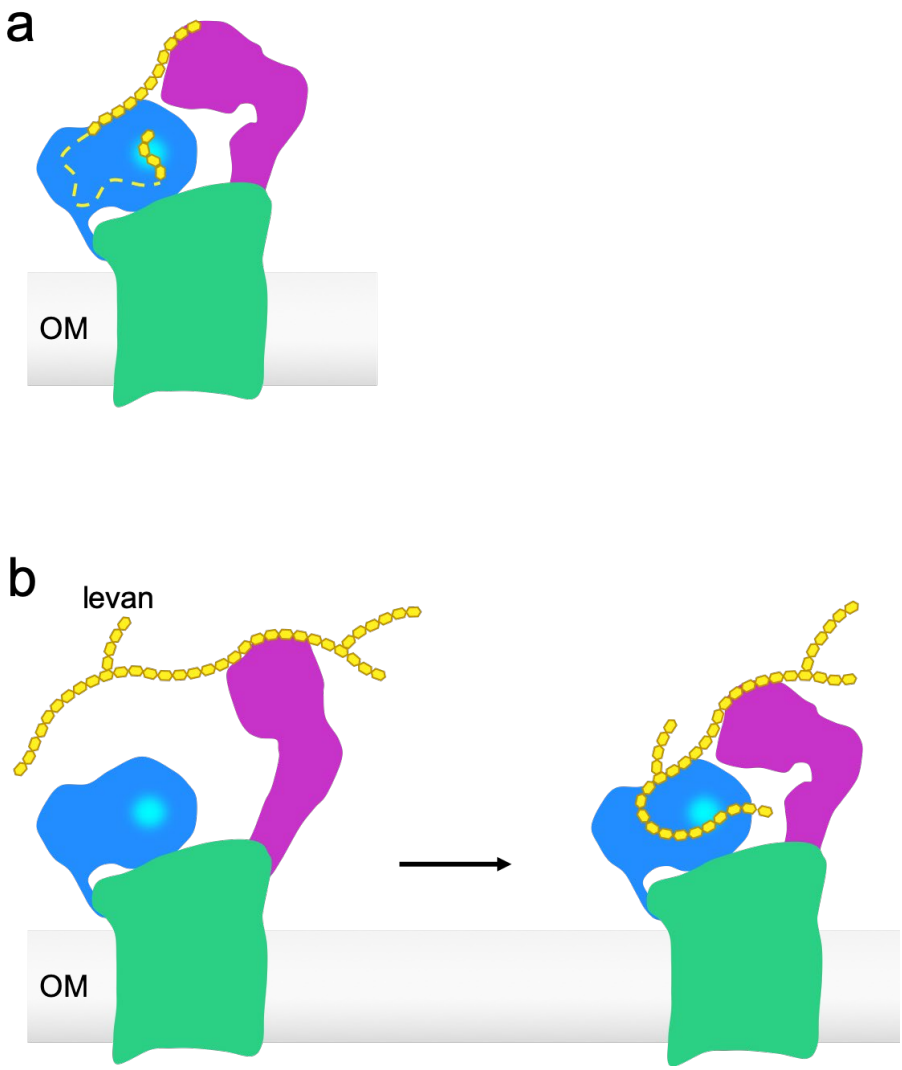


Figure 5.32: Schematic demonstrating the proposed concerted function of the levan binding protein and levanase subunits of the levan utilisome. a, Schematic representation of the levan utilisome with tethered levan binding protein observed by cryoEM. A dotted yellow line indicates the possibility of contiguous and dynamic levan chain not resolved in the reconstruction. **b**, Proposed model for role of the secondary binding site in recruitment of the levan binding protein C-terminal domain. The levan binding protein captures a levan chain from the surrounding medium and the inherent flexibility of this lipoprotein subunit allows it to sample different conformations. The C-terminal domain is recruited to the levanase as a result of simultaneous interactions of the bound levan chain with the secondary binding site on the levanase. Stretches of chain not involved in binding interactions are likely dynamic and can therefore access the levanase active site (marked in cyan) where cleavage results in the generation of transport competent FOS substrates. Levan chains are shown in yellow, SusC in green, the levanase (Bt1760) in blue (active site cyan) and the levan binding protein in pink. SusD is omitted from the schematic and one half of the utilisome is shown for clarity.

Considering the data on the dextran utilisome, the universally open states observed for its $SusC_2D_2$ core are, on first sight, difficult to reconcile with our model, given that the complex was isolated from cells grown on oligomeric dextran (~DP8-10). As such, one might expect to

see closed states of the complex containing the dextran that was present in the growth medium. However, assuming that the transporters continue to function until cell lysis, the lack of added dextran in the resuspension buffer after cell pelleting would result in the rapid depletion of dextran, and generate the observed all-open transporters.

The successful purification and structure determination of the complete levan utilisome demonstrates that the complex is highly stable. By extension, the absence of a complete dextran utilisome in our cryoEM data may indicate that this complex is less stable. However, given that we observe heptameric complexes, and that naked $SusC_2D_2$ pairs represent only ~10 % of the dataset, we conclude that complete utilisome complexes are likely predominant in the OM. The dextran utilisome was more difficult to work with, owing to its tendency to aggregate, and the difference in experimental conditions required for data collection may be responsible for disrupting octameric complexes. Importantly, it is clear for the levan system that utilisome formation is independent of the presence of substrate.

Whether the organisation of OM PUL components into stable utilisome complexes is a general feature of glycan transport in the Bacteroidetes is an important question. Utilisomes offer a clear advantage in environments where complex polymer substrates are the predominant nutrient source since they are capable of concomitant binding, cleavage, capture and transport of their relatively large target molecules across the OM. While our data argue in favour of the presence of these assemblies at the OM, recent studies on the archetypal starch PUL propose a much more dynamic model, with immobile starch binding centres comprised of the SGBPs SusE and SusF, around which transporter components and the endo-amylase (SusG) can transiently assemble (Tuson et al., 2018). Proteomics data for the starch utilisation system reveal similar amounts of SusC and SusD at the OM whilst SusEFG are much less abundant, providing support for the assembly of complexes with differing stoichiometries. Intriguingly, co-immunoprecipitation with SusD antibodies reportedly captures twice as much SusD as SusC, suggesting that even the core SusCD transporter may not form a stable complex in the starch system (Tuson et al., 2018). This is in clear conflict with the observations for SusCD complexes from other PULs where separate SusC or SusD components have not been observed upon purification. Moreover, while it is not clear how much the *E. coli* OM resembles that of *B. theta*, recent work has shown that *E. coli* OMP mobility is limited, casting

doubts on the efficiency of dynamically (dis)assembling transport complexes in a crowded OM (Sun et al., 2021).

It is not immediately clear why the starch system appears to operate so differently to the levan and dextran systems presented here. One consideration is that these models are based on live-cell fluorescence studies that require C-terminal fusion of OM PUL components with relatively large fluorescent tags. Considering the proximity of components in both the levan and dextran utilisomes (see **Figure 5.30**), it is possible that the fluorescent tags hinder complex formation, and therefore contribute to the dynamic behaviour observed. The influence of tagging on protein function was assessed by cell growth assays. However, this approach may be limited. The starch PUL encodes two SGBPs (SusE and SusF), and the corresponding fusion proteins were assessed independently. An alphafold2 structure prediction of SusC from the starch system shows that the transporter offers no extra interaction surface for the association of auxiliary components relative to the levan and dextran SusC units (see **Figure 5.33**). As such, it is unlikely that both SGBPs and SusG can bind the same SusCD transporter. Thus, functional redundancy of the SusE and SusF components may preclude an observable growth phenotype. Furthermore, the loss of SusG from a hypothetical starch utilisome may not be detectable by growth assay. SusG will generate transport competent substrates and release them into the growth medium regardless of whether or not it is bound to SusCD. In an environment where starch is in excess, the advantage conferred by complexation with the transporter may be less evident.

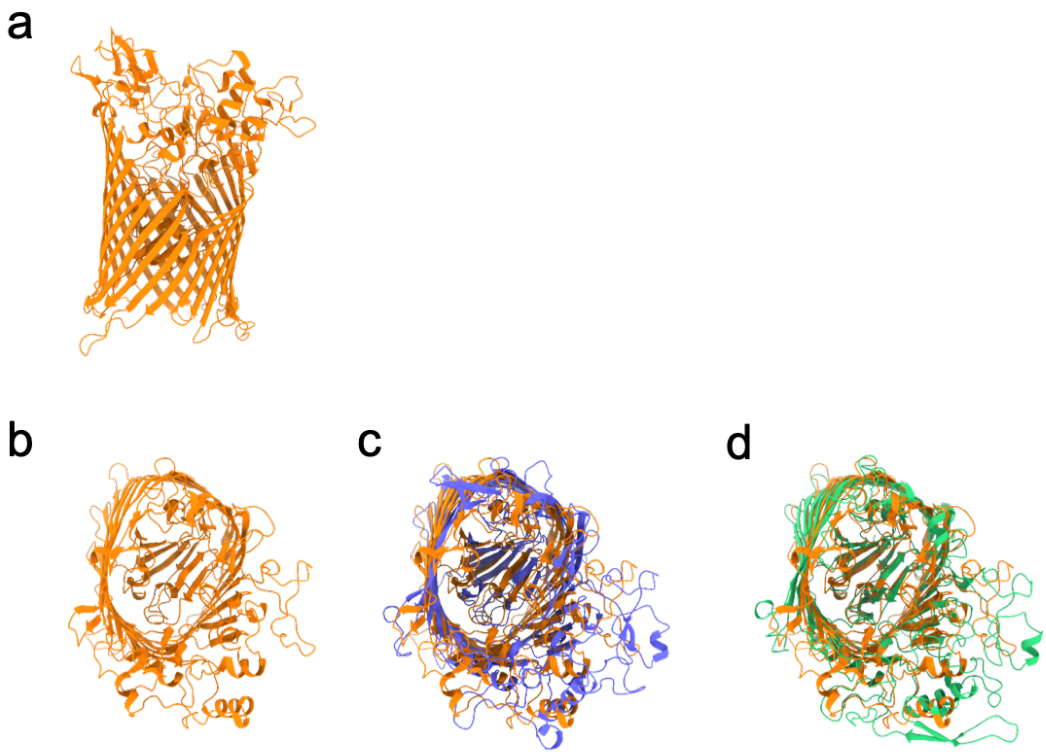


Figure 5.33 SusC from the starch utilisation system provides no additional interaction surface for the assembly of additional lipoproteins. a, Alphafold2 structure prediction of SusC from the starch utilisation system viewed in the plane of the membrane and (**b**) from the extracellular space. Alignments of the predicted structure with experimentally determined structures for SusC components of the dextran and levan systems are shown in (**c**) and (**d**) respectively.

Considering the bigger picture, it is also plausible that the starch utilisation system is fundamentally different and indeed, the presence of a second SGBP sets it apart from the systems investigated here. Transient interactions of SusG with the putative starch utilisome complex might indicate diversity in the function of this amylase. Proteomic analysis shows that SusG is enriched in outer membrane vesicles (OMVs) produced by *B. theta*, while SusC remains mostly in the OM (Valguarnera et al., 2018). Dissemination of such enzymatically active OMVs may have an important role in promoting the growth of other members of the gut microbiome with different metabolic capabilities, and thus in establishing mutualistic relationships. Whether specific PULs contribute more heavily than others to the lipoprotein complement of OMVs remains to be seen. It is plausible that, once trafficked to the cell surface, the additional lipoproteins follow one of two routes: utilisome assembly vs packaging into OMVs, and that the balance of these routes is reflected in the stability of the corresponding utilisomes.

5.13 Future work

The work presented here represents a significant leap in our understanding of glycan utilisation by prominent members of our gut microbiota. Utilisome assemblies offer a neat solution to the challenge of capturing large and complex substrates at the cell surface and transporting them across the OM. However, determining whether these stable architectures are a hallmark feature of PULs in general requires the study of more systems not only from *B. theta*, but from other gut *Bacteroides spp.*, as well as from more distant Bacteroidetes occupying niches in terrestrial and marine environments. It is important to note that the levan and dextran utilisation systems are relatively simple. The complexity of PULs i.e. the number of encoded binding proteins, enzymes and even SusCD pairs increases in-line with the complexity of their target substrate. For example, the characterised PULs targeting arabinogalactan, complex N-glycans and rhamnogalacturonan II all contain more than one SusCD pair (Briliute et al., 2019; Martens et al., 2011; Ndeh et al., 2017). This raises the question of whether SusC and SusD components can undergo mix-and-match style complex assembly with other SusC and SusD components from the same PUL. A similar question can be asked in the context of utilisomes, with the additional variable of which additional lipoproteins might assemble on the core transporter.

Despite having observed utilisome complexes, the way in which they are assembled remains enigmatic. The lipoprotein components, including SusD, must be flipped from the inner leaflet to the outer leaflet of the OM for presentation on the cell surface. The size and complexity of these subunits suggests that dedicated cellular machinery is required, although no candidates have yet been identified. It is plausible that lipoprotein flipping is coordinated with the folding of SusC components into the OM, in which case the β-barrel assembly machinery would be closely associated with any putative lipoprotein flippase.

Regarding the SusCD transporters themselves, although we have shown how these systems capture substrate and signal the resulting substrate-bound state to the periplasmic space, the fine details of the allosteric pathway that connects these events remains to be explored. Substrate binding, lid closure or both will result in large-scale re-organisation of residue contacts, shifting non-covalent interaction networks and consequently propagating

conformational changes from the extracellular face of the transporter to the Ton box at the periplasmic face of the plug. Identifying and dissecting these networks is non-trivial and few studies have addressed this challenge for TBDTs. However, effective methodologies for the elucidation of these networks are well established in the study of G-protein coupled receptors (GPCR) (Flock et al., 2015; Venkatakrishnan et al., 2016; Zhou et al., 2019; Hauser et al., 2021; Heydenreich et al., 2021). Similarly to TBDTs, these systems must transduce a signal in the form of ligand binding at the extracellular side, across the membrane, to the intracellular face of the receptor in order to facilitate interactions with the partner G-protein and initiate downstream signalling events. Considering the largest subfamily of GPCRs in humans (class A), the availability of hundreds of structures, including both active and inactive states, has permitted dissection of the mechanism of activation (Zhou et al., 2019). Residue-residue contact scores (RRCS) based on interatomic distances were calculated which allowed a ΔRRCS to be determined when comparing the RRCS of a residue pair between two conformational states of the receptor e.g. active versus inactive. This power of this approach is that it is capable of capturing a global picture of conformational changes in receptor structure upon activation. Major conformational changes such as helix-level movements result in contact switching between states, i.e. contacts present in the inactive state that are lost in the active state and vice versa. More subtle changes in side chain positions are also detectable and are identified as repacking contacts where ΔRRCS for a residue pair is non-zero. By carrying out this analysis for hundreds of members of the class A GPCRs, it was possible to establish that a common and conserved receptor activation pathway exists within this subfamily, directly linking the ligand binding pocket and G protein-coupling regions (Zhou et al., 2019). Site-directed mutagenesis of the implicated residues was used to validate the pathway. These findings allow rationalisation of disease-causing mutations on the basis of whether they stabilise or destabilise contacts in the common activation pathway and are therefore extremely valuable in the context of pharmacology research and drug discovery. Notably, the usefulness of this approach depends heavily on the quality of modelled side chain conformations and thus on the resolution of the structural data. Equally important for a complete picture of an allosteric mechanism, yet not considered in this approach, are interactions contributed by non-protein components such as water molecules, the observation of which again requires high resolution. The role of water molecules may be especially significant in the context of TBDTs given that the plug domain is highly solvated,

with interactions between the barrel and the plug predominantly mediated via bridging waters (Chimento et al., 2005), it is therefore important that these potential components of allosteric pathways not be overlooked. With this in mind, future work should target improved resolution of the maps presented in this thesis. The availability of next-generation detectors such as the Falcon 4, will facilitate the improvement in resolution that is required to identify ordered water molecules present in the structure.

At present, the limited number of SusCD-like transporter structures preclude the analysis described above. However, methods that rely on sequence data as opposed to atomic structures offer a valuable starting point. Statistical coupling analysis (SCA) is one such sequence-based approach that facilitates identification of evolutionary interactions between residues in multiple sequence alignments (Lockless and Ranganathan, 1999). The underlying principle is that conserved residue interactions that confer function should drive co-evolution of the residues involved. Ferguson and colleagues applied this approach to classical TBDTs and mapped the co-evolving residues onto the crystal structure of FecA (Ferguson et al., 2007). In this way, they identified a bifurcating network of sparse but structurally connected residues from both the internal barrel wall and the plug domain. Together, these residues link the extracellular binding pocket to the periplasmic surface of the transporter and are therefore implicated in the allosteric signal transduction mechanism. Support for this conclusion comes from site directed mutagenesis where co-evolving residues distal to both the ligand binding site and the periplasmic face of the transporter were still shown to impair FecA function. It would be interesting to compute similar coupling analyses for SusCD-like systems, not only with the intention of providing an initial overview of the putative allosteric pathway, but also to investigate whether systems transporting different substrate classes, such as glycans or peptides, utilise different allosteric pathways.

Bringing together the available structural information for different SusCD-like transport systems provides an early indication that while there may be a general allosteric framework that elicits a common outcome (unfolding of the Ton box region), mechanistic nuances are present. In RagAB, the bound peptide adopts a kinked structure that allows one end of the peptide to run parallel to, and interact with, the RagA-RagB interface, while the other end is directed down towards the plug domain. Despite the absence of a direct interaction of the

peptide with the plug in the data presented here, the well-defined topology of the former has led us to hypothesise that native peptides captured by the *in vivo* system may form a physical/mechanical link between the lid and the plug that is plausibly crucial for allosteric signalling and the translocation process itself. Similarly, in the cryoEM map of the levan SusCD system obtained in the presence of levan FOS DP15-25, long FOS molecules are shown to be capable of spanning the binding cavity to contact both the plug domain and the SusC-SusD interface. In contrast, the putative peptide in the crystal structure of the functionally uncharacterised Bt2263-Bt2264 system runs parallel to the SusC-SusD interface with no evidence of this substrate extending in the direction of the plug. Instead, Bt2263 possesses an extended plug loop that reaches up and contacts the putative peptide. Thus, the means of initiating the allosteric signal transduction pathway may differ substantially between members of the SusCD family of transporters. Future structural studies should therefore target a diverse range of SusCD transporters from Bacteroidetes occupying terrestrial and marine environments, as well as those constituting part of the human microbiota so as to gain a wider understanding of these systems.

Just as importantly, questions still remain about the function of the SusC NTE domain. Our data offers little insight into a potential role and it is possible that the right questions have not yet been asked. Future work should target the pairing of SusC TBDTs with the TonB/ExbB/ExbD inner membrane complex. *B. theta* possesses around 10 TonB homologues and it is not known whether these exhibit specificity for particular SusC transporters or vice versa. The proximal location of the NTE domain relative to the TonB box region makes a role in governing TonB binding interactions an interesting possibility. Finally, the pedal-bin model for SusCD-mediated nutrient transport is not yet complete. While we have shown how these systems capture substrate and signal this capturing event to the periplasmic space, the translocation step itself must be investigated. The precise details of plug disruption in TonB-dependent transport are still unclear, and, by extension, it is not known whether SusC-like TBDTs operate in the same way as their classical counterparts. It would be interesting to attempt single molecule unfolding experiments similar to those performed by Hickman et al. (2017) on SusCD complexes. Work in chapter 4 indicated that spontaneous expulsion of the entire plug domain is possible in SusCD complexes, albeit when solubilised in detergent. This observation does not agree with the literature consensus for classical TBDTs, where partial

plug unfolding is proposed to result in the formation of a channel whose size is tailored to its specific substrate. Ultimately, a complete understanding of TonB-dependent transport, be it in classical or SusCD-like systems, will likely require new methodologies that permit reconstitution of the complete trans-envelope TBDT-TonB subcomplex in the presence of an energised bilayer.

Chapter 6 Concluding remarks

This thesis describes high-resolution cryoEM structures of SusCD transport systems from two important members of the human microbiota. Novel conformational states revealed by these studies provide unprecedented insight into the molecular mechanisms by which these organisms capture large, complex nutrients.

Chapter 3 describes structures of the peptide-transporting RagAB (SusCD equivalent) complex from *P. gingivalis* in three conformational states. In all cases, reconstructions possessed the twin β-barrel architecture previously observed for X-ray crystal structures of SusCD systems, confirming that this dimeric arrangement was not an artefact of crystal packing. In contrast to all structural data available at the time, the principle finding of this work was the observation of novel conformational states in which one or both RagB subunits occupied an open position, flipped up and away from the extracellular mouth of their associated RagA barrel. These structures corroborated predictions from MD simulations that indicated SusD-like components may function as lids, opening and closing via a rigid-body movement about an N-terminal hinge region to alternately expose and occlude the substrate binding cavity within SusC. Prior to this study, the working hypothesis for SusCD-mediated nutrient capture described a 'pedal bin' mechanism where, in the absence of substrate, SusCD transport units possess an open conformation until substrate binding elicits closure of the SusD lid and the substrate-loaded state of the transporter is signalled to the periplasm via disordering of the TonB box region. However, evidence of substrate density in open states of the RagAB transporter demonstrated that this system is more complex than previously appreciated. Indeed, substrate addition experiments indicate that some peptides may interact with the transporter in a non-productive manner, occupying the binding site without inducing closure of the lid. Nonetheless, strong substrate density was invariably associated with lid-closed conformations and structural analysis of plug domains from open and closed states of the RagAB complex showed ordered and disordered conformations of the TonB box regions, respectively. Thus, while this signal transduction mechanism could not be directly linked to the absence or presence of substrate, it was shown to be associated with closing of the lid, highlighting the substrate-bound lid closed conformation as the transport competent state of

the complex. Ultimately, the inability to remove co-purified peptides from the RagAB complex precluded definitive investigation of conformational changes in response to substrate addition.

Chapter 4 describes the structures of the levan transporting SusCD complex from *B. theta* observed in the absence of substrate. Similarly to RagAB, a mixed population of open and closed SusCD units were present. This was unexpected as the pedal bin hypothesis would predict a uniformly open population of transporters when no substrate is present. Further structural analysis revealed that where SusD lids were closed, the corresponding SusC barrels lacked density for the plug domain, indicating spontaneous expulsion of the plug under the experimental conditions used. As discussed in Chapter 4, this observation is likely an artefact of the detergent-solubilised system. Notably, the literature consensus for the mechanism of TonB-mediated plug disruption involves a partial unfolding event that culminates in the formation of a channel through the transporter, the size of which is tailored to the size of the transported substrate. Whether the expulsion of the plug domain observed in this work provides insights into a novel mechanism of plug disruption unique to SusCD transporters remains to be seen. Closed conformations of the SusD lid were exclusively associated with the absence of the plug domain in the corresponding SusC barrel, indicating that conformational changes in these domains are intimately linked. It is plausible that, in the event of complete unplugging, SusD acts to preserve selective permeability of the outer membrane by capping the extracellular face of the SusC barrel and thus preventing large channel formation. Steered MD simulations and complementary single molecule unfolding experiments that investigate the behaviour of the SusC plug domain in response to mechanical force may be sufficient to identify possible differences in the mechanisms of plug disruption for SusC-like and classical TBDTs. The second notable finding of this work was that the levanase and levan binding protein, co-expressed from the levan PUL, co-purified with SusCD and, remarkably, were observed to assemble on the lip of a single SusC barrel, giving rise to an asymmetric, hexameric complex. This provided the first structural evidence that OM-associated PUL components involved in glycan utilisation can form stable, higher order assemblies at the OM.

Chapter 5 first describes the structural characterisation of a levan utilising macromolecular machine comprising two copies of each of the outer-membrane components of the levan PUL, namely SusC, SusD, the levanase, and the levan binding protein. This stable octameric complex, or utilisome, possesses all of the activities required for levan utilisation, with components arranged in a way that appears to promote efficient levan biding, cleavage and capture. Structure determination in the absence of substrate showed a uniformly open population of SusCD transporters within utilisome complexes as predicted by the pedal bin hypothesis. Furthermore, the addition of substrate was shown to result in large, concerted conformational changes of the utilisome, including closure of all SusD lids, which revealed substrate-bound states of SusCD, the levanase and levan binding protein. These structures therefore provide a direct demonstration of the mechanism of nutrient capture for the first time and offer insights into how utilisome components function in concert. Subsequent structural studies on the dextran utilisation system indicated that utilisome architecture is a general feature of outer membrane PUL components, with the dextranase and dextran binding proteins of this system also shown to assemble on the core $SusC_2D_2$ transporter.

Taken together, the work presented in this thesis represents a major advancement in our understanding of nutrient acquisition by SusCD transport systems of Bacteroidetes. This is especially valuable for Bacteroidetes occupying the distal gut where the ability to access complex nutrients is a crucial factor in determining the community composition of the colonic microbiome. Community dynamics are central to microbiota-host symbiosis/dysbiosis and, in turn, human health. As such, understanding the molecular mechanisms that facilitate nutrient accessibility is an important step towards informed manipulation of the gut microbiome to benefit the health of the host. However, a great deal more work is required to establish whether or not utilisome assembly is a general hallmark of polysaccharide utilisation systems. As discussed in Chapter 5, the levan and dextran utilisation systems are relatively simple, and the complexity of PULs increases in-line with the complexity of their target substrate. Where PULs encode multiple SusCD homologues, key questions arise including whether SusC and SusD components can undergo mix-and-match style complex assembly with other SusC and SusD components from the same PUL, or whether discrete SusCD complexes target different constituent oligosaccharides from the target glycan. Where PULs, such as the starch utilisation system, contain multiple SGBPs it is not clear what utilisome organisation would

look like. The number of additional components that can assemble on SusC is limited, suggesting either that some SGBPs do not localise to utilisomes, or that utilisomes possessing heterogeneous complements of additional lipoprotein components exist at the OM. Structural characterisation of these more complex systems is therefore an important target of future work.

Bibliography

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010) 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallographica Section D,* 66(2), pp. 213-221.

Arnosti, C., Wietz, M., Brinkhoff, T., Hehemann, J. H., Probandt, D., Zeugner, L. and Amann, R. (2021) 'The Biogeochemistry of Marine Polysaccharides: Sources, Inventories, and Bacterial Drivers of the Carbohydrate Cycle', *Annual Review of Marine Science,* 13(1), pp. 81-108.

Arruda, S., Bomfim, G., Knights, R., Huima-Byron, T. and Riley, L. W. (1993) 'Cloning of an M. tuberculosis DNA fragment associated with entry and survival inside cells', *Science,* 261(5127), pp. 1454-1457.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Doré, J., Antolín, M., Artiguenave, F., Blottiere, H. M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K. U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Kristiansen, K., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Mérieux, A., Melo Minardi, R., M'Rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S. D., Bork, P. and Meta, H. I. T. C. (2011) 'Enterotypes of the human gut microbiome', *Nature,* 473(7346), pp. 174-180.

Asmar, A. T., Ferreira, J. L., Cohen, E. J., Cho, S.-H., Beeby, M., Hughes, K. T. and Collet, J.-F. (2017) 'Communication across the bacterial cell envelope depends on the size of the periplasm', *PLoS Biology,* 15(12), pp. e2004303.

Auclair, S. M., Bhanu, M. K. and Kendall, D. A. (2012) 'Signal peptidase I: cleaving the way to mature proteins', *Protein Science,* 21(1), pp. 13-25.

Balusek, C. and Gumbart, J. C. (2016) 'Role of the native outer-membrane environment on the transporter BtuB', *Biophysical journal,* 111(7), pp. 1409-1417.

Bammes, B. E., Jakana, J., Schmid, M. F. and Chiu, W. (2010) 'Radiation damage effects at four specimen temperatures from 4 to 100 K', *Journal of structural biology,* 169(3), pp. 331-341.

Bayburt, T. H. and Sligar, S. G. (2010) 'Membrane protein assembly into Nanodiscs', *FEBS letters,* 584(9), pp. 1721-1727.

Benigar, E., Dogsa, I., Stopar, D., Jamnik, A., Cigić, I. K. and Tomšič, M. (2014) 'Structure and dynamics of a polysaccharide matrix: aqueous solutions of bacterial levan', *Langmuir,* 30(14), pp. 4172-4182.

Bennion, D., Charlson, E. S., Coon, E. and Misra, R. (2010) 'Dissection of β-barrel outer membrane protein assembly pathways through characterizing BamA POTRA 1 mutants of Escherichia coli', *Molecular microbiology,* 77(5), pp. 1153-1171.

Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. and Berger, B. (2019) 'Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs', *Nature methods,* 16(11), pp. 1153-1160.

Berg, B. V. D., Clemons, W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. and Rapoport, T. A. (2004) 'X-ray structure of a protein-conducting channel', *nature,* 427(6969), pp. 36-44.

Bhella, D. (2019) 'Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility', *Biophysical Reviews,* 11(4), pp. 515-519.

Bishop, C. M., Walkenhorst, W. F. and Wimley, W. C. (2001) 'Folding of β-sheets in membranes: specificity and promiscuity in peptide model systems', *Journal of molecular biology,* 309(4), pp. 975-988.

Bogomolnaya, L. M., Santiviago, C. A., Yang, H. J., Baumler, A. J. and Andrews-Polymenis, H. L. (2008) ''Form variation'of the O12 antigen is critical for persistence of Salmonella Typhimurium in the murine intestine', *Molecular microbiology,* 70(5), pp. 1105-1119.

Bolam, D. N. and Koropatkin, N. M. (2012) 'Glycan recognition by the Bacteroidetes Sus-like systems', *Current opinion in structural biology,* 22(5), pp. 563-569.

Bolam, D. N. and Sonnenburg, J. L. (2011) 'Mechanistic insight into polysaccharide use within the intestinal microbiota', *Gut microbes,* 2(2), pp. 86-90.

Bolam, D. N. and van den Berg, B. (2018) 'TonB-dependent transport by the gut microbiota: novel aspects of an old problem', *Current opinion in structural biology,* 51, pp. 35-43.

Botos, I., Majdalani, N., Mayclin, S. J., McCarthy, J. G., Lundquist, K., Wojtowicz, D., Barnard, T. J., Gumbart, J. C. and Buchanan, S. K. (2016) 'Structural and functional characterization of the LPS transporter LptDE from Gram-negative pathogens', *Structure,* 24(6), pp. 965-976.

Botte, M., Ni, D., Schenck, S., Zimmermann, I., Chami, M., Bocquet, N., Egloff, P., Bucher, D., Trabuco, M., Cheng, R. K. Y., Brunner, J. D., Seeger, M. A., Stahlberg, H. and Hennig, M. (2021) 'Insight into Lipopolysaccharide Translocation by Cryo-EM structures of a LptDE Transporter in Complex with Pro-Macrobodies', *bioRxiv*, pp. 2021.03.23.436624.

Braun, M., Endriss, F., Killmann, H. and Braun, V. (2003) 'In vivo reconstitution of the FhuA transport protein of Escherichia coli K-12', *Journal of bacteriology,* 185(18), pp. 5508-5518.

Braun, V. and Hantke, K. (2019) 'Lipoproteins: structure, function, biosynthesis', *Bacterial Cell Walls and Membranes*, pp. 39-77.

Braun, V. and Wu, H. (1994) 'Lipoproteins, structure, function, biosynthesis and model for protein export', *New comprehensive biochemistry*: Elsevier, pp. 319-341.

Briliūtė, J., Urbanowicz, P. A., Luis, A. S., Baslé, A., Paterson, N., Rebello, O., Hendel, J., Ndeh, D. A., Lowe, E. C., Martens, E. C., Spencer, D. I. R., Bolam, D. N. and Crouch, L. I. (2019) 'Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci', *Nature Microbiology,* 4(9), pp. 1571-1581.

Brillet, K., Journet, L., Célia, H., Paulus, L., Stahl, A., Pattus, F. and Cobessi, D. (2007) 'A β strand lock exchange for signal transduction in TonB-dependent transducers on the basis of a common structural motif', *Structure,* 15(11), pp. 1383-1391.

Brockwell, D. J., Paci, E., Zinober, R. C., Beddard, G. S., Olmsted, P. D., Smith, D. A., Perham, R. N. and Radford, S. E. (2003) 'Pulling geometry defines the mechanical resistance of a β-sheet protein', *Nature Structural & Molecular Biology,* 10(9), pp. 731-737.

Bryant, C. E., Spring, D. R., Gangloff, M. and Gay, N. J. (2010) 'The molecular basis of the host response to lipopolysaccharide', *Nature Reviews Microbiology,* 8(1), pp. 8-14.

Bui, F. Q., Almeida-da-Silva, C. L. C., Huynh, B., Trinh, A., Liu, J., Woodward, J., Asadi, H. and Ojcius, D. M. (2019) 'Association between periodontal pathogens and systemic disease', *Biomedical journal,* 42(1), pp. 27-35.

Burgess, S., Walker, M., Sakakibara, H., Oiwa, K. and Knight, P. (2004) 'The structure of dynein-c by negative stain electron microscopy', *Journal of structural biology,* 146(1-2), pp. 205-216.

Cadieux, N. and Kadner, R. J. (1999) 'Site-directed disulfide bonding reveals an interaction site between energy-coupling protein TonB and BtuB, the outer membrane cobalamin transporter', *Proceedings of the National Academy of Sciences,* 96(19), pp. 10673-10678.

Calabrese, A. N., Schiffrin, B., Watson, M., Karamanos, T. K., Walko, M., Humes, J. R., Horne, J. E., White, P., Wilson, A. J., Kalli, A. C., Tuma, R., Ashcroft, A. E., Brockwell, D. J. and Radford, S. E. (2020) 'Inter-domain dynamics in the chaperone SurA and multi-site binding to its outer membrane protein clients', *Nature Communications,* 11(1), pp. 2155.

Cameron, E. A., Maynard, M. A., Smith, C. J., Smith, T. J., Koropatkin, N. M. and Martens, E. C. (2012) 'Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism', *J Biol Chem,* 287(41), pp. 34614-25.

Carpenter, T. S., Parkin, J. and Khalid, S. (2016) 'The free energy of small solute permeation through the Escherichia coli outer membrane has a distinctly asymmetric profile', *The journal of physical chemistry letters,* 7(17), pp. 3446-3451.

Cartmell, A., Lowe, E. C., Baslé, A., Firbank, S. J., Ndeh, D. A., Murray, H., Terrapon, N., Lombard, V., Henrissat, B., Turnbull, J. E., Czjzek, M., Gilbert, H. J. and Bolam, D. N. (2017) 'How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans', *Proceedings of the National Academy of Sciences,* 114(27), pp. 7037-7042.

Celia, H., Botos, I., Ni, X., Fox, T., De Val, N., Lloubes, R., Jiang, J. and Buchanan, S. K. (2019) 'Cryo-EM structure of the bacterial Ton motor subcomplex ExbB–ExbD provides information on structure and stoichiometry', *Communications biology,* 2(1), pp. 1-6.

Chaturvedi, D. and Mahalakshmi, R. (2017) 'Transmembrane β-barrels: Evolution, folding and energetics', *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1859(12), pp. 2467-2482.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity: all-atom structure validation for macromolecular crystallography', *Acta Crystallographica Section D: Biological Crystallography,* 66(1), pp. 12-21.

Cheng, Y. (2015) 'Single-particle cryo-EM at crystallographic resolution', *Cell,* 161(3), pp. 450-457.

Cheng, Y., Grigorieff, N., Penczek, P. A. and Walz, T. (2015) 'A primer to single-particle cryo-electron microscopy', *Cell,* 161(3), pp. 438-449.

Chimento, D. P., Kadner, R. J. and Wiener, M. C. (2005) 'Comparative structural analysis of TonB-dependent outer membrane transporters: implications for the transport cycle', *Proteins: Structure, Function, and Bioinformatics,* 59(2), pp. 240-251.

Chimento, D. P., Mohanty, A. K., Kadner, R. J. and Wiener, M. C. (2003) 'Substrate-induced transmembrane signaling in the cobalamin transporter BtuB', *Nature Structural & Molecular Biology,* 10(5), pp. 394-401.

Cho, K. H. and Salyers, A. A. (2001) 'Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron', *Journal of bacteriology,* 183(24), pp. 7224-7230.

Chong, Z. S., Woo, W. F. and Chng, S. S. (2015) 'Osmoporin OmpC forms a complex with MlaA to maintain outer membrane lipid asymmetry in E scherichia coli', *Molecular microbiology,* 98(6), pp. 1133-1146.

Chopra, A., Bhat, S. G. and Sivaraman, K. (2020) 'Porphyromonas gingivalis adopts intricate and unique molecular mechanisms to survive and persist within the host: a critical update', *Journal of Oral Microbiology,* 12(1), pp. 1801090.

Cohen, E. J., Ferreira, J. L., Ladinsky, M. S., Beeby, M. and Hughes, K. T. (2017) 'Nanoscale-length control of the flagellar driveshaft requires hitting the tethered outer membrane', *Science,* 356(6334), pp. 197-200.

Cuskin, F., Lowe, E. C., Temple, M. J., Zhu, Y., Cameron, E. A., Pudlo, N. A., Porter, N. T., Urs, K., Thompson, A. J., Cartmell, A., Rogowski, A., Hamilton, B. S., Chen, R., Tolbert, T. J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munōz-Munōz, J. L., Day, A., Peña, M. J., McLean, R., Suits, M. D., Boraston, A. B., Atherly, T., Ziemer, C. J., Williams, S. J., Davies, G. J., Abbott, D. W., Martens, E. C. and Gilbert, H. J. (2015) 'Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism', *Nature,* 517(7533), pp. 165-169.

Dai, X. and Zhou, Z. H. (2018) 'Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes', *Science,* 360(6384), pp. eaao7298.

Danev, R., Belousoff, M., Liang, Y.-L., Zhang, X., Eisenstein, F., Wootten, D. and Sexton, P. M. (2021) 'Routine sub-2.5 Å cryo-EM structure determination of GPCRs', *Nature Communications,* 12(1), pp. 1-10.

Dashper, S., Brownfield, L., Slakeski, N., Zilm, P., Rogers, A. and Reynolds, E. (2001) 'Sodium ion-driven serine/threonine transport in Porphyromonas gingivalis', *Journal of Bacteriology,* 183(14), pp. 4142-4148.

Davies, J. R., Kad, T., Neilands, J., Kinnby, B., Prgomet, Z., Bengtsson, T., Khalaf, H. and Svensäter, G. (2021) 'Polymicrobial synergy stimulates Porphyromonas gingivalis survival and gingipain expression in a multi-species subgingival community', *BMC Oral Health,* 21(1), pp. 639.

Dawaliby, R., Trubbia, C., Delporte, C., Masureel, M., Van Antwerpen, P., Kobilka, B. K. and Govaerts, C. (2016) 'Allosteric regulation of G protein–coupled receptor activity by phospholipids', *Nature chemical biology,* 12(1), pp. 35-39.

De Carlo, S. and Harris, J. R. (2011) 'Negative staining and cryo-negative staining of macromolecules and viruses for TEM', *Micron,* 42(2), pp. 117-131.

Death, A., Notley, L. and Ferenci, T. (1993) 'Derepression of LamB protein facilitates outer membrane permeation of carbohydrates into Escherichia coli under conditions of nutrient stress', *Journal of bacteriology,* 175(5), pp. 1475-1483.

Deme, J. C., Johnson, S., Vickery, O., Aron, A., Monkhouse, H., Griffiths, T., James, R. H., Berks, B. C., Coulton, J. W., Stansfeld, P. J. and Lea, S. M. (2020) 'Structures of the stator complex that drives rotation of the bacterial flagellum', *Nature Microbiology,* 5(12), pp. 1553-1564.

Denisov, I. G. and Sligar, S. G. (2016) 'Nanodiscs for structural and functional studies of membrane proteins', *Nature structural & molecular biology,* 23(6), pp. 481-486.

Dobro, M. J., Melanson, L. A., Jensen, G. J. and McDowall, A. W. (2010) 'Plunge freezing for electron cryomicroscopy', *Methods in enzymology,* 481, pp. 63-82.

Dominguez-Martin, M. A., Sauer, P. V., Sutter, M., Kirst, H., Bina, D., Greber, B. J., Nogales, E., Polívka, T. and Kerfeld, C. A. (2021) 'Structure of the quenched cyanobacterial OCP-phycobilisome complex', *bioRxiv*.

Dominy Stephen, S., Lynch, C., Ermini, F., Benedyk, M., Marczyk, A., Konradi, A., Nguyen, M., Haditsch, U., Raha, D., Griffin, C., Holsinger Leslie, J., Arastu-Kapur, S., Kaba, S., Lee, A., Ryder Mark, I., Potempa, B., Mydel, P., Hellvard, A., Adamowicz, K., Hasturk, H., Walker Glenn, D., Reynolds Eric, C., Faull Richard, L. M., Curtis Maurice, A., Dragunow, M. and Potempa, J. 'Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors', *Science Advances,* 5(1), pp. eaau3333.

Dong, H., Xiang, Q., Gu, Y., Wang, Z., Paterson, N. G., Stansfeld, P. J., He, C., Zhang, Y., Wang, W. and Dong, C. (2014) 'Structural basis for outer membrane lipopolysaccharide insertion', *Nature,* 511(7507), pp. 52-56.

Doyle, R. J. and Marquis, R. E. (1994) 'Elastic, flexible peptidoglycan and bacterial cell wall properties', *Trends in microbiology,* 2(2), pp. 57-60.

Drulyte, I., Johnson, R. M., Hesketh, E. L., Hurdiss, D. L., Scarff, C. A., Porav, S. A., Ranson, N. A., Muench, S. P. and Thompson, R. F. (2018) 'Approaches to altering particle distributions in cryo-electron microscopy sample preparation', *Acta Crystallographica Section D: Structural Biology,* 74(6), pp. 560-571.

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowall, A. W. and Schultz, P. (1988) 'Cryo-electron microscopy of vitrified specimens', *Quarterly reviews of biophysics,* 21(2), pp. 129-228.

Dubochet, J., Ducommun, M., Zollinger, M. and Kellenberger, E. (1971) 'A new preparation method for dark-field electron microscopy of biomacromolecules', *Journal of ultrastructure research,* 35(1-2), pp. 147-167.

Dumas, F., Koebnik, R., Winterhalter, M. and Van Gelder, P. (2000) 'Sugar transport through maltoporin of Escherichia coli: role of polar tracks', *Journal of Biological Chemistry,* 275(26), pp. 19747-19751.

Dutzler, R., Wang, Y., Rizkallah, P., Rosenbusch, J. and Schirmer, T. (1996) 'Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway', *Structure,* 4(2), pp. 127-134.

Ekiert, D. C., Bhabha, G., Isom, G. L., Greenan, G., Ovchinnikov, S., Henderson, I. R., Cox, J. S. and Vale, R. D. (2017) 'Architectures of lipid transport systems for the bacterial outer membrane', *Cell,* 169(2), pp. 273-285. e17.

Emsley, P. and Cowtan, K. (2004) 'Coot: model-building tools for molecular graphics', *Acta crystallographica section D: biological crystallography,* 60(12), pp. 2126-2132.

Erickson, H. P. and Klug, A. (1970) 'The Fourier transform of an electron micrograph: effects of defocussing and aberrations, and implications for the use of underfocus contrast enhancement', *Berichte der Bunsengesellschaft für physikalische Chemie,* 74(11), pp. 1129-1137.

Ernits, K., Eek, P., Lukk, T., Visnapuu, T. and Alamäe, T. (2019) 'First crystal structure of an endo-levanase–the BT1760 from a human gut commensal Bacteroides thetaiotaomicron', *Scientific reports,* 9(1), pp. 1-13.

Fanucci, G. E., Cadieux, N., Kadner, R. J. and Cafiso, D. S. (2003) 'Competing ligands stabilize alternate conformations of the energy coupling motif of a TonB-dependent outer membrane transporter', *Proceedings of the National Academy of Sciences,* 100(20), pp. 11382-11387.

Feathers, J. R., Spoth, K. A. and Fromme, J. C. (2021) 'Experimental evaluation of super-resolution imaging and magnification choice in single-particle cryo-EM', *Journal of structural biology: X,* 5, pp. 100047.

Feng, X., Fu, Z., Kaledhonkar, S., Jia, Y., Shah, B., Jin, A., Liu, Z., Sun, M., Chen, B., Grassucci, R. A., Ren, Y., Jiang, H., Frank, J. and Lin, Q. (2017) 'A Fast and Effective Microfluidic Spraying-Plunging Method for High-Resolution Single-Particle Cryo-EM', *Structure,* 25(4), pp. 663-670.e3.

Ferguson, A. D., Amezcua, C. A., Halabi, N. M., Chelliah, Y., Rosen, M. K., Ranganathan, R. and Deisenhofer, J. (2007) 'Signal transduction pathway of TonB-dependent transporters', *Proceedings of the National Academy of Sciences,* 104(2), pp. 513-518.

Ferguson, A. D., Chakraborty, R., Smith, B. S., Esser, L., Van Der Helm, D. and Deisenhofer, J. (2002) 'Structural basis of gating by the outer membrane transporter FecA', *Science,* 295(5560), pp. 1715-1719.

Ferguson, A. D., Hofmann, E., Coulton, J. W., Diederichs, K. and Welte, W. (1998) 'Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide', *science,* 282(5397), pp. 2215-2220.

Flock, T., Ravarani, C. N. J., Sun, D., Venkatakrishnan, A. J., Kayikci, M., Tate, C. G., Veprintsev, D. B. and Babu, M. M. (2015) 'Universal allosteric mechanism for Gα activation by GPCRs', *Nature,* 524(7564), pp. 173-179.

Foley, M. H., Cockburn, D. W. and Koropatkin, N. M. (2016) 'The Sus operon: a model system for starch uptake by the human gut Bacteroidetes', *Cellular and Molecular Life Sciences,* 73(14), pp. 2603-2617.

Freed, D. M., Horanyi, P. S., Wiener, M. C. and Cafiso, D. S. (2010) 'Conformational exchange in a membrane transport protein is altered in protein crystals', *Biophysical journal,* 99(5), pp. 1604-1610.

Freed, D. M., Lukasik, S. M., Sikora, A., Mokdad, A. and Cafiso, D. S. (2013) 'Monomeric TonB and the ton box are required for the formation of a high-affinity transporter–TonB Complex', *Biochemistry,* 52(15), pp. 2638-2648.

Freinkman, E., Okuda, S., Ruiz, N. and Kahne, D. (2012) 'Regulated assembly of the transenvelope protein complex required for lipopolysaccharide export', *Biochemistry,* 51(24), pp. 4800-4806.

Frey, L., Lakomek, N. A., Riek, R. and Bibow, S. (2017) 'Micelles, bicelles, and nanodiscs: comparing the impact of membrane mimetics on membrane protein backbone dynamics', *Angewandte Chemie International Edition,* 56(1), pp. 380-383.

Furse, S. and Scott, D. J. (2016) 'Three-dimensional distribution of phospholipids in gram negative bacteria', *Biochemistry,* 55(34), pp. 4742-4747.

Gao, Y., Cao, E., Julius, D. and Cheng, Y. (2016) 'TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action', *Nature,* 534(7607), pp. 347-351.

Garcia-Herrero, A. and Vogel, H. J. (2005) 'Nuclear magnetic resonance solution structure of the periplasmic signalling domain of the TonB-dependent outer membrane transporter FecA from Escherichia coli', *Molecular microbiology,* 58(5), pp. 1226-1237.

Glauert, A. M. and Thornley, M. J. (1969) 'The topography of the bacterial cell wall', *Annual review of microbiology,* 23(1), pp. 159-198.

Glenwright, A. J., Pothula, K. R., Bhamidimarri, S. P., Chorev, D. S., Baslé, A., Firbank, S. J., Zheng, H., Robinson, C. V., Winterhalter, M., Kleinekathöfer, U., Bolam, D. N. and van den Berg, B. (2017) 'Structural basis for nutrient acquisition by dominant members of the human gut microbiota', *Nature,* 541(7637), pp. 407-411.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. and Ferrin, T. E. (2018) 'UCSF ChimeraX: Meeting modern challenges in visualization and analysis', *Protein Science,* 27(1), pp. 14-25.

Goemans, C., Denoncin, K. and Collet, J.-F. (2014) 'Folding mechanisms of periplasmic proteins', *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research,* 1843(8), pp. 1517-1528.

Goulas, T., Garcia-Ferrer, I., Hutcherson, J. A., Potempa, B. A., Potempa, J., Scott, D. A. and Xavier Gomis-Rüth, F. (2016) 'Structure of RagB, a major immunodominant outer-membrane surface receptor antigen of Porphyromonas gingivalis', *Molecular oral microbiology,* 31(6), pp. 472-485.

Grabowicz, M. and Silhavy, T. J. (2017) 'Redefining the essential trafficking pathway for outer membrane lipoproteins', *Proceedings of the National Academy of Sciences,* 114(18), pp. 4769-4774.

Gray, D. A., White, J. B. R., Oluwole, A. O., Rath, P., Glenwright, A. J., Mazur, A., Zahn, M., Baslé, A., Morland, C., Evans, S. L., Cartmell, A., Robinson, C. V., Hiller, S., Ranson, N. A., Bolam, D. N. and van den Berg, B. (2021) 'Insights into SusCD-mediated glycan import by a prominent gut symbiont', *Nature Communications,* 12(1), pp. 44.

Grenier, D., Imbeault, S., Plamondon, P., Grenier, G., Nakayama, K. and Mayrand, D. (2001) 'Role of gingipains in growth of Porphyromonas gingivalis in the presence of human serum albumin', *Infection and immunity,* 69(8), pp. 5166-5172.

Grigorieff, N. and Harrison, S. C. (2011) 'Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy', *Current opinion in structural biology,* 21(2), pp. 265-273.

Gu, Y., Li, H., Dong, H., Zeng, Y., Zhang, Z., Paterson, N. G., Stansfeld, P. J., Wang, Z., Zhang, Y., Wang, W. and Dong, C. (2016) 'Structural basis of outer membrane protein insertion by the BAM complex', *Nature,* 531(7592), pp. 64-69.

Gu, Y., Stansfeld, P. J., Zeng, Y., Dong, H., Wang, W. and Dong, C. (2015) 'Lipopolysaccharide is inserted into the outer membrane through an intramembrane hole, a lumen gate, and the lateral opening of LptD', *Structure,* 23(3), pp. 496-504.

Gudmundsdottir, A., Bell, P., Lundrigan, M., Bradbeer, C. and Kadner, R. J. (1989) 'Point mutations in a conserved region (TonB box) of Escherichia coli outer membrane protein BtuB affect vitamin B12 transport', *Journal of bacteriology,* 171(12), pp. 6526-6533.

Gumbart, J., Wiener, M. C. and Tajkhorshid, E. (2007) 'Mechanics of force propagation in TonB-dependent outer membrane transport', *Biophysical journal,* 93(2), pp. 496-504.

Gupta, K., Donlan, J. A. C., Hopper, J. T. S., Uzdavinys, P., Landreh, M., Struwe, W. B., Drew, D., Baldwin, A. J., Stansfeld, P. J. and Robinson, C. V. (2017) 'The role of interfacial lipids in stabilizing membrane protein oligomers', *Nature,* 541(7637), pp. 421-424.

Hagan, C. L., Kim, S. and Kahne, D. (2010) 'Reconstitution of outer membrane protein assembly from purified components', *Science,* 328(5980), pp. 890-892.

Hagan, C. L., Silhavy, T. J. and Kahne, D. (2011) 'β-Barrel membrane protein assembly by the Bam complex', *Annual review of biochemistry,* 80, pp. 189-210.

Hajishengallis, G., Darveau, R. P. and Curtis, M. A. (2012) 'The keystone-pathogen hypothesis', *Nature Reviews Microbiology,* 10(10), pp. 717-725.

Hajishengallis, G. and Lamont, R. J. (2012) 'Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology', *Molecular oral microbiology,* 27(6), pp. 409-419.

Hall, L. M., Fawell, S. C., Shi, X., Faray-Kele, M.-C., Aduse-Opoku, J., Whiley, R. A. and Curtis, M. A. (2005) 'Sequence diversity and antigenic variation at the rag locus of Porphyromonas gingivalis', *Infection and immunity,* 73(7), pp. 4253-4262.

Han, L., Zheng, J., Wang, Y., Yang, X., Liu, Y., Sun, C., Cao, B., Zhou, H., Ni, D., Lou, J., Zhao, Y. and Huang, Y. (2016) 'Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins', *Nature Structural & Molecular Biology,* 23(3), pp. 192-196.

Hauser, A. S., Kooistra, A. J., Munk, C., Heydenreich, F. M., Veprintsev, D. B., Bouvier, M., Babu, M. M. and Gloriam, D. E. (2021) 'GPCR activation mechanisms across classes and macro/microscales', *Nature Structural & Molecular Biology,* 28(11), pp. 879-888.

Hayashi, S. and Wu, H. C. (1990) 'Lipoproteins in bacteria', *Journal of bioenergetics and biomembranes,* 22(3), pp. 451-471.

Henderson, R. (1995) 'The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules', *Quarterly reviews of biophysics,* 28(2), pp. 171-193.

Heydenreich, F. M., Marti-Solano, M., Sandhu, M., Kobilka, B. K., Bouvier, M. and Babu, M. M. (2021) 'Dissecting the allosteric networks governing agonist efficacy and potency in G protein-coupled receptors', *bioRxiv*, pp. 2021.09.14.460253.

Hickman, S. J., Cooper, R. E., Bellucci, L., Paci, E. and Brockwell, D. J. (2017) 'Gating of TonB-dependent transporters by substrate-specific forced remodelling', *Nature communications,* 8(1), pp. 1-12.

Ho, M.-H., Lamont, R. J. and Xie, H. (2017) 'A novel peptidic inhibitor derived from Streptococcus cristatus ArcA attenuates virulence potential of Porphyromonas gingivalis', *Scientific Reports,* 7(1), pp. 16217.

Hočevar, K., Vizovišek, M., Wong, A., Kozieł, J., Fonović, M., Potempa, B., Lamont, R. J., Potempa, J. and Turk, B. (2020) 'Proteolysis of Gingival Keratinocyte Cell Surface Proteins by Gingipains Secreted From Porphyromonas gingivalis - Proteomic Insights Into Mechanisms Behind Tissue Damage in the Diseased Gingiva', *Front Microbiol,* 11, pp. 722.

Holt, S. C. and Ebersole, J. L. (2005) 'Porphyromonas gingivalis, Treponema denticola, and Tannerella forsythia: the 'red complex', a prototype polybacterial pathogenic consortium in periodontitis', *Periodontology 2000,* 38(1), pp. 72-122.

Iadanza, M. G., Higgins, A. J., Schiffrin, B., Calabrese, A. N., Brockwell, D. J., Ashcroft, A. E., Radford, S. E. and Ranson, N. A. (2016) 'Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM', *Nature communications,* 7(1), pp. 1-12.

Iadanza, M. G., Schiffrin, B., White, P., Watson, M. A., Horne, J. E., Higgins, A. J., Calabrese, A. N., Brockwell, D. J., Tuma, R., Kalli, A. C., Radford, S. E. and Ranson, N. A. (2020) 'Distortion of the bilayer and dynamics of the BAM complex in lipid nanodiscs', *Communications Biology,* 3(1), pp. 766.

Isom, G. L., Coudray, N., MacRae, M. R., McManus, C. T., Ekiert, D. C. and Bhabha, G. (2020) 'LetB structure reveals a tunnel for lipid transport across the bacterial envelope', *Cell,* 181(3), pp. 653-664. e19.

Jacobson, A. N., Choudhury, B. P. and Fischbach, M. A. (2018) 'The biosynthesis of lipooligosaccharide from Bacteroides thetaiotaomicron', *MBio,* 9(2), pp. e02289-17.

Jakobi, A. J., Wilmanns, M. and Sachse, C. (2017) 'Model-based local density sharpening of cryo-EM maps', *Elife,* 6, pp. e27131.

Jensen, J. L., Jernberg, B. D., Sinha, S. C. and Colbert, C. L. (2020) 'Structural basis of cell-surface signaling by a conserved sigma regulator in Gram-negative bacteria', *Journal of Biological Chemistry,* 295(17), pp. 5795-5806.

Jordan, L. D., Zhou, Y., Smallwood, C. R., Lill, Y., Ritchie, K., Yip, W. T., Newton, S. M. and Klebba, P. E. (2013) 'Energy-dependent motion of TonB in the Gram-negative bacterial inner membrane', *Proceedings of the National Academy of Sciences,* 110(28), pp. 11553-11558.

Josts, I., Veith, K. and Tidow, H. (2019) 'Ternary structure of the outer membrane transporter FoxA with resolved signalling domain provides insights into TonB-mediated siderophore uptake', *Elife,* 8, pp. e48528.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature,* 596(7873), pp. 583-589.

Kamio, Y. and Nikaido, H. (1976) 'Outer membrane of Salmonella typhimurium: accessibility of phospholipid head groups to phospholipase c and cyanogen bromide activated dextran in the external medium', *Biochemistry,* 15(12), pp. 2561-2570.

Karshikoff, A., Spassov, V., Cowan, S., Ladenstein, R. and Schirmer, T. (1994) 'Electrostatic properties of two porin channels from Escherichia coli', *Journal of molecular biology,* 240(4), pp. 372-384.

Karunatilaka, K. S., Cameron, E. A., Martens, E. C., Koropatkin, N. M. and Biteen, J. S. (2014) 'Superresolution imaging captures carbohydrate utilization dynamics in human gut symbionts', *MBio,* 5(6), pp. e02172-14.

Kaur, H., Jakob, R. P., Marzinek, J. K., Green, R., Imai, Y., Bolla, J. R., Agustoni, E., Robinson, C. V., Bond, P. J., Lewis, K., Maier, T. and Hiller, S. (2021) 'The antibiotic darobactin mimics a β-strand to inhibit outer membrane insertase', *Nature,* 593(7857), pp. 125-129.

Killian, J. A. and von Heijne, G. (2000) 'How proteins adapt to a membrane–water interface', *Trends in biochemical sciences,* 25(9), pp. 429-434.

Kim, M., Fanucci, G. E. and Cafiso, D. S. (2007) 'Substrate-dependent transmembrane signaling in TonB-dependent transporters is not conserved', *Proceedings of the National Academy of Sciences,* 104(29), pp. 11975-11980.

Kim, M. L. and Slauch, J. M. (1999) 'Effect of acetylation (O-factor 5) on the polyclonal antibody response to Salmonella typhimurium O-antigen', *FEMS Immunology & Medical Microbiology,* 26(1), pp. 83-92.

Kim, M., Xu, Q., Fanucci, G. E. and Cafiso, D. S. (2006) 'Solutes Modify a Conformational Transition in a Membrane Transport Protein', *Biophysical Journal,* 90(8), pp. 2922-2929.

Klebba, P. E. (2016) 'ROSET model of TonB action in Gram-negative bacterial iron acquisition', *Journal of bacteriology,* 198(7), pp. 1013-1021.

Klebl, D. P., Gravett, M. S. C., Kontziampasis, D., Wright, D. J., Bon, R. S., Monteiro, D. C. F., Trebbin, M., Sobott, F., White, H. D., Darrow, M. C., Thompson, R. F. and Muench, S. P. (2020) 'Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales', *Structure,* 28(11), pp. 1238-1248.e4.

Koebnik, R., Locher, K. P. and Van Gelder, P. (2000) 'Structure and function of bacterial outer membrane proteins: barrels in a nutshell', *Molecular microbiology,* 37(2), pp. 239-253.

Koh, A., De Vadder, F., Kovatcheva-Datchary, P. and Bäckhed, F. (2016) 'From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites', *Cell,* 165(6), pp. 1332-1345.

Konovalova, A., Grabowicz, M., Balibar, C. J., Malinverni, J. C., Painter, R. E., Riley, D., Mann, P. A., Wang, H., Garlisi, C. G., Sherborne, B., Rigel, N. W., Ricci, D. P., Black, T. A., Roemer, T., Silhavy, T. J. and Walker, S. S. (2018) 'Inhibitor of intramembrane protease RseP blocks the σ^E response causing lethal accumulation of unfolded outer membrane proteins', *Proceedings of the National Academy of Sciences,* 115(28), pp. E6614-E6621.

Konovalova, A., Kahne, D. E. and Silhavy, T. J. (2017) 'Outer membrane biogenesis', *Annual review of microbiology,* 71, pp. 539-556.

Konovalova, A. and Silhavy, T. J. (2015) 'Outer membrane lipoprotein biogenesis: Lol is not the end', *Philosophical Transactions of the Royal Society B: Biological Sciences,* 370(1679), pp. 20150030.

Korndörfer, I. P., Dommel, M. K. and Skerra, A. (2004) 'Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture', *Nature structural & molecular biology,* 11(10), pp. 1015-1020.

Koropatkin, N. M., Martens, E. C., Gordon, J. I. and Smith, T. J. (2008) 'Starch Catabolism by a Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices', *Structure,* 16(7), pp. 1105-1115.

Krissinel, E. and Henrick, K. (2007) 'Inference of macromolecular assemblies from crystalline state', *Journal of molecular biology,* 372(3), pp. 774-797.

Krojer, T., Sawa, J., Schäfer, E., Saibil, H. R., Ehrmann, M. and Clausen, T. (2008) 'Structural basis for the regulated protease and chaperone function of DegP', *Nature,* 453(7197), pp. 885-890.

Kucukelbir, A., Sigworth, F. J. and Tagare, H. D. (2014) 'Quantifying the local resolution of cryo-EM density maps', *Nature methods,* 11(1), pp. 63-65.

Kullman, L., Winterhalter, M. and Bezrukov, S. M. (2002) 'Transport of maltodextrins through maltoporin: a single-channel study', *Biophysical journal,* 82(2), pp. 803-812.

Laganowsky, A., Reading, E., Allison, T. M., Ulmschneider, M. B., Degiacomi, M. T., Baldwin, A. J. and Robinson, C. V. (2014) 'Membrane proteins bind lipids selectively to modulate their structure and function', *Nature,* 510(7503), pp. 172-175.

Larsbrink, J., Rogers, T. E., Hemsworth, G. R., McKee, L. S., Tauzin, A. S., Spadiut, O., Klinter, S., Pudlo, N. A., Urs, K., Koropatkin, N. M., Creagh, A. L., Haynes, C. A., Kelly, A. G., Cederholm, S. N., Davies, G. J., Martens, E. C. and Brumer, H. (2014) 'A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes', *Nature,* 506(7489), pp. 498-502.

Lauber, F., Deme, J. C., Lea, S. M. and Berks, B. C. (2018) 'Type 9 secretion system structures reveal a new protein transport mechanism', *Nature,* 564(7734), pp. 77-82.

Lebedev, A. A., Young, P., Isupov, M. N., Moroz, O. V., Vagin, A. A. and Murshudov, G. N. (2012) 'JLigand: a graphical tool for the CCP4 template-restraint library', *Acta Crystallographica Section D: Biological Crystallography,* 68(4), pp. 431-440.

Lidbury, I. D. E. A., Borsetto, C., Murphy, A. R. J., Bottrill, A., Jones, A. M. E., Bending, G. D., Hammond, J. P., Chen, Y., Wellington, E. M. H. and Scanlan, D. J. (2021) 'Niche-adaptation in plant-associated Bacteroidetes favours specialisation in organic phosphorus mineralisation', *The ISME Journal,* 15(4), pp. 1040-1055.

Lin, T.-L., Shu, C.-C., Chen, Y.-M., Lu, J.-J., Wu, T.-S., Lai, W.-F., Tzeng, C.-M., Lai, H.-C. and Lu, C.-C. (2020) 'Like cures like: pharmacological activity of anti-inflammatory lipopolysaccharides from gut microbiome', *Frontiers in Pharmacology,* 11, pp. 554.

Locher, K. P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J. P. and Moras, D. (1998) 'Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes', *Cell,* 95(6), pp. 771-778.

Lockless, S. W. and Ranganathan, R. (1999) 'Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families', *Science,* 286(5438), pp. 295-299.

Luckey, M. and Nikaido, H. (1980) 'Specificity of diffusion channels produced by lambda phage receptor protein of Escherichia coli', *Proceedings of the National Academy of Sciences,* 77(1), pp. 167-171.

Lugtenberg, E. and Peters, R. (1976) 'Distribution of lipids in cytoplasmic and outer membranes of Escherichia coli K12', *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism,* 441(1), pp. 38-47.

Madej, M., White, J. B. R., Nowakowska, Z., Rawson, S., Scavenius, C., Enghild, J. J., Bereta, G. P., Pothula, K., Kleinekathoefer, U., Baslé, A., Ranson, N. A., Potempa, J. and van den Berg, B. (2020) 'Structural and functional insights into oligopeptide acquisition by the RagAB transporter from Porphyromonas gingivalis', *Nature Microbiology,* 5(8), pp. 1016-1025.

Malinverni, J. C. and Silhavy, T. J. (2009) 'An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane', *Proceedings of the National Academy of Sciences,* 106(19), pp. 8009-8014.

Malinverni, J. C., Werner, J., Kim, S., Sklar, J. G., Kahne, D., Misra, R. and Silhavy, T. J. (2006) 'YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli', *Molecular microbiology,* 61(1), pp. 151-164.

Malki, I., Simenel, C., Wojtowicz, H., Cardoso de Amorim, G., Prochnicka-Chalufour, A., Hoos, S., Raynal, B., England, P., Chaffotte, A., Delepierre, M., Delepelaire, P. and Izadi-Pruneyre, N. (2014) 'Interaction of a Partially Disordered Antisigma Factor with Its Partner, the Signaling Domain of the TonB-Dependent Transporter HasR', *PLOS ONE,* 9(4), pp. e89502.

Marchesi, J. R., Adams, D. H., Fava, F., Hermes, G. D. A., Hirschfield, G. M., Hold, G., Quraishi, M. N., Kinross, J., Smidt, H., Tuohy, K. M., Thomas, L. V., Zoetendal, E. G. and Hart, A. (2016) 'The gut microbiota and host health: a new clinical frontier', *Gut,* 65(2), pp. 330-339.

Mardo, K., Visnapuu, T., Vija, H., Aasamets, A., Viigand, K. and Alamäe, T. (2017) 'A highly active endo-levanase BT1760 of a dominant mammalian gut commensal Bacteroides thetaiotaomicron cleaves not only various bacterial levans, but also levan of timothy grass', *PloS one,* 12(1), pp. e0169989.

Martens, C., Shekhar, M., Borysik, A. J., Lau, A. M., Reading, E., Tajkhorshid, E., Booth, P. J. and Politis, A. (2018) 'Direct protein-lipid interactions shape the conformational landscape of secondary transporters', *Nature communications,* 9(1), pp. 1-12.

Martens, C., Stein, R. A., Masureel, M., Roth, A., Mishra, S., Dawaliby, R., Konijnenberg, A., Sobott, F., Govaerts, C. and McHaourab, H. S. (2016) 'Lipids modulate the conformational dynamics of a secondary multidrug transporter', *Nature Structural & Molecular Biology,* 23(8), pp. 744-751.

Martens, E. C., Lowe, E. C., Chiang, H., Pudlo, N. A., Wu, M., McNulty, N. P., Abbott, D. W., Henrissat, B., Gilbert, H. J., Bolam, D. N. and Gordon, J. I. (2011) 'Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts', *PLOS Biology,* 9(12), pp. e1001221.

Mayrand, D. and Holt, S. C. (1988) 'Biology of asaccharolytic black-pigmented Bacteroides species', *Microbiological reviews,* 52(1), pp. 134-152.

McMullan, G., Chen, S., Henderson, R. and Faruqi, A. (2009) 'Detective quantum efficiency of electron area detectors in electron microscopy', *Ultramicroscopy,* 109(9), pp. 1126-1143.

McMullan, G., Faruqi, A. and Henderson, R. (2016) 'Direct electron detectors', *Methods in enzymology,* 579, pp. 1-17.

McNeil, N. (1984) 'The contribution of the large intestine to energy supplies in man', *The American journal of clinical nutrition,* 39(2), pp. 338-342.

Melchior, D. L. and Steim, J. M. (1976) 'Thermotropic transitions in biomembranes', *Annual review of biophysics and bioengineering,* 5(1), pp. 205-238.

Meredith, T. C., Mamat, U., Kaczynski, Z., Lindner, B., Holst, O. and Woodard, R. W. (2007) 'Modification of lipopolysaccharide with colanic acid (M-antigen) repeats in Escherichia coli', *Journal of Biological Chemistry,* 282(11), pp. 7790-7798.

Mi, W., Li, Y., Yoon, S. H., Ernst, R. K., Walz, T. and Liao, M. (2017) 'Structural basis of MsbA-mediated lipopolysaccharide transport', *Nature,* 549(7671), pp. 233-237.

Miller, S. I. and Salama, N. R. (2018) 'The gram-negative bacterial periplasm: Size matters', *PLoS biology,* 16(1), pp. e2004935.

Milne, J. L. S., Borgnia, M. J., Bartesaghi, A., Tran, E. E. H., Earl, L. A., Schauder, D. M., Lengyel, J., Pierson, J., Patwardhan, A. and Subramaniam, S. (2013) 'Cryo-electron microscopy – a primer for the non-microscopist', *The FEBS Journal,* 280(1), pp. 28-45.

Mohn, W. W., Van Der Geize, R., Stewart, G. R., Okamoto, S., Liu, J., Dijkhuizen, L. and Eltis, L. D. (2008) 'The actinobacterial mce4 locus encodes a steroid transporter', *Journal of Biological Chemistry,* 283(51), pp. 35368-35374.

Moon, C. P., Zaccai, N. R., Fleming, P. J., Gessmann, D. and Fleming, K. G. (2013) 'Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm', *Proceedings of the National Academy of Sciences,* 110(11), pp. 4285-4290.

Moran, A. P. (2010) 'Molecular mimicry of host glycosylated structures by bacteria', *Microbial glycobiology*: Elsevier, pp. 847-870.

Morrison, D. J. and Preston, T. (2016) 'Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism', *Gut Microbes,* 7(3), pp. 189-200.

Moynié, L., Milenkovic, S., Mislin, G. L. A., Gasser, V., Malloci, G., Baco, E., McCaughan, R. P., Page, M. G. P., Schalk, I. J., Ceccarelli, M. and Naismith, J. H. (2019) 'The complex of ferric-enterobactin with its transporter from Pseudomonas aeruginosa suggests a two-site model', *Nature Communications,* 10(1), pp. 3673.

Murata, K. and Wolf, M. (2018) 'Cryo-electron microscopy for structural analysis of dynamic biological macromolecules', *Biochimica et Biophysica Acta (BBA)-General Subjects,* 1862(2), pp. 324-334.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011) 'REFMAC5 for the refinement of macromolecular crystal structures', *Acta Crystallographica Section D: Biological Crystallography,* 67(4), pp. 355-367.

Nagano, K., Murakami, Y., Nishikawa, K., Sakakibara, J., Shimozato, K. and Yoshimura, F. (2007) 'Characterization of RagA and RagB in Porphyromonas gingivalis: study using gene-deletion mutants', *Journal of medical microbiology,* 56(11), pp. 1536-1548.

Narita, S.-i. and Tokuda, H. (2009) 'Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides', *FEBS letters,* 583(13), pp. 2160-2164.

Narita, S.-i. and Tokuda, H. (2017) 'Bacterial lipoproteins; biogenesis, sorting and quality control', *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids,* 1862(11), pp. 1414-1423.

Naydenova, K., Jia, P. and Russo, C. J. (2020) 'Cryo-EM with sub – 1 Å specimen movement', *Science,* 370(6513), pp. 223-226.

Ndeh, D., Rogowski, A., Cartmell, A., Luis, A. S., Baslé, A., Gray, J., Venditto, I., Briggs, J., Zhang, X., Labourel, A., Terrapon, N., Buffetto, F., Nepogodiev, S., Xiao, Y., Field, R. A., Zhu, Y., O'Neill, M. A., Urbanowicz, B. R., York, W. S., Davies, G. J., Abbott, D. W., Ralet, M.-C., Martens, E. C., Henrissat, B. and Gilbert, H. J. (2017) 'Complex pectin metabolism by gut bacteria reveals novel catalytic functions', *Nature,* 544(7648), pp. 65-70.

Nelson Karen, E., Fleischmann Robert, D., DeBoy Robert, T., Paulsen Ian, T., Fouts Derrick, E., Eisen Jonathan, A., Daugherty Sean, C., Dodson Robert, J., Durkin, A. S., Gwinn, M., Haft Daniel, H., Kolonay James, F., Nelson William, C., Mason, T., Tallon, L., Gray, J., Granger, D., Tettelin, H., Dong, H., Galvin Jamie, L., Duncan Margaret, J., Dewhirst Floyd, E. and Fraser Claire, M. (2003) 'Complete Genome Sequence of the Oral Pathogenic Bacterium Porphyromonas gingivalis Strain W83', *Journal of Bacteriology,* 185(18), pp. 5591-5601.

Nemoto, T. K. and Ohara-Nemoto, Y. (2016) 'Exopeptidases and gingipains in Porphyromonas gingivalis as prerequisites for its amino acid metabolism', *Japanese Dental Science Review,* 52(1), pp. 22-29.

Nikaido, H. (1976) 'Outer membrane of Salmonella typhimurium: transmembrane diffusion of some hydrophobic substances', *Biochimica Et Biophysica Acta (BBA)-Biomembranes,* 433(1), pp. 118-132.

Nikaido, H. (2003) 'Molecular basis of bacterial outer membrane permeability revisited', *Microbiology and molecular biology reviews,* 67(4), pp. 593-656.

Nilaweera, T. D., Nyenhuis, D. A. and Cafiso, D. S. (2021) 'Structural intermediates observed only in intact Escherichia coli indicate a mechanism for TonB-dependent transport', *Elife,* 10, pp. e68548.

Nishiyama, K., Hanada, M. and Tokuda, H. (1994) 'Disruption of the gene encoding p12 (SecG) reveals the direct involvement and important function of SecG in the protein translocation of Escherichia coli at low temperature', *The EMBO Journal,* 13(14), pp. 3272-3277.

Noble, A. J., Wei, H., Dandey, V. P., Zhang, Z., Tan, Y. Z., Potter, C. S. and Carragher, B. (2018) 'Reducing effects of particle adsorption to the air–water interface in cryo-EM', *Nature methods,* 15(10), pp. 793-795.

Nogales, E. and Scheres, S. H. (2015) 'Cryo-EM: a unique tool for the visualization of macromolecular complexity', *Molecular cell,* 58(4), pp. 677-689.

Noinaj, N., Guillier, M., Barnard, T. J. and Buchanan, S. K. (2010) 'TonB-dependent transporters: regulation, structure, and function', *Annual review of microbiology,* 64, pp. 43-60.

Noinaj, N., Kuszak, A. J., Balusek, C., Gumbart, J. C. and Buchanan, S. K. (2014) 'Lateral opening and exit pore formation are required for BamA function', *Structure,* 22(7), pp. 1055-1062.

Ohi, M., Li, Y., Cheng, Y. and Walz, T. (2004) 'Negative staining and image classification—powerful tools in modern electron microscopy', *Biological procedures online,* 6(1), pp. 23-34.

Oke, M., Sarra, R., Ghirlando, R., Farnaud, S., Gorringe, A., Evans, R. W. and Buchanan, S. (2004) 'The plug domain of a neisserial TonB-dependent transporter retains structural integrity in the absence of its transmembrane β-barrel', *FEBS letters,* 564(3), pp. 294-300.

Okuda, S., Freinkman, E. and Kahne, D. (2012) 'Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in E. coli', *Science,* 338(6111), pp. 1214-1217.

Okuda, S., Sherman, D. J., Silhavy, T. J., Ruiz, N. and Kahne, D. (2016) 'Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model', *Nature Reviews Microbiology,* 14(6), pp. 337-345.

Öner, E. T., Hernández, L. and Combie, J. (2016) 'Review of levan polysaccharide: from a century of past experiences to future prospects', *Biotechnology advances,* 34(5), pp. 827-844.

Orlova, E. and Saibil, H. R. (2011) 'Structural analysis of macromolecular assemblies by electron microscopy', *Chemical reviews,* 111(12), pp. 7710-7748.

Pandey, A. K. and Sassetti, C. M. (2008) 'Mycobacterial persistence requires the utilization of host cholesterol', *Proceedings of the National Academy of Sciences,* 105(11), pp. 4376-4380.

Passmore, L. A. and Russo, C. J. (2016) 'Specimen preparation for high-resolution cryo-EM', *Methods in enzymology,* 579, pp. 51-86.

Pautsch, A. and Schulz, G. E. (1998) 'Structure of the outer membrane protein A transmembrane domain', *Nature structural biology,* 5(11), pp. 1013-1017.

Pawelek, P. D., Croteau, N., Ng-Thow-Hing, C., Khursigara, C. M., Moiseeva, N., Allaire, M. and Coulton, J. W. (2006) 'Structure of TonB in complex with FhuA, E. coli outer membrane receptor', *Science,* 312(5778), pp. 1399-1402.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) 'UCSF Chimera—a visualization system for exploratory research and analysis', *Journal of computational chemistry,* 25(13), pp. 1605-1612.

Plant, L., Sundqvist, J., Zughaier, S., Lövkvist, L., Stephens, D. S. and Jonsson, A.-B. (2006) 'Lipooligosaccharide structure contributes to multiple steps in the virulence of Neisseria meningitidis', *Infection and immunity,* 74(2), pp. 1360-1367.

Pollet, R. M., Martin, L. M. and Koropatkin, N. M. (2021) 'TonB-dependent transporters in the Bacteroidetes: Unique domain structures and potential functions', *Molecular Microbiology,* 115(3), pp. 490-501.

Potempa, J., Banbula, A. and Travis, J. (2000) 'Role of bacterial proteinases in matrix destruction and modulation of host responses', *Periodontology 2000,* 24, pp. 153-192.

Potempa, J., Mydel, P. and Koziel, J. (2017) 'The case for periodontitis in the pathogenesis of rheumatoid arthritis', *Nature Reviews Rheumatology,* 13(10), pp. 606-620.

Preshaw, P., Alba, A., Herrera, D., Jepsen, S., Konstantinidis, A., Makrilakis, K. and Taylor, R. (2012) 'Periodontitis and diabetes: a two-way relationship', *Diabetologia,* 55(1), pp. 21-31.

Punjani, A., Rubinstein, J. L., Fleet, D. J. and Brubaker, M. A. (2017) 'cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination', *Nature methods,* 14(3), pp. 290-296.

Qiao, S., Luo, Q., Zhao, Y., Zhang, X. C. and Huang, Y. (2014) 'Structural basis for lipopolysaccharide insertion in the bacterial outer membrane', *Nature,* 511(7507), pp. 108-111.

Raetz, C. R. and Whitfield, C. (2002) 'Lipopolysaccharide endotoxins', *Annual review of biochemistry,* 71(1), pp. 635-700.

Ratliff, A. C., Buchanan, S. K. and Celia, H. (2021) 'Ton motor complexes', *Current Opinion in Structural Biology,* 67, pp. 95-100.

Razinkov, I., Dandey, V. P., Wei, H., Zhang, Z., Melnekoff, D., Rice, W. J., Wigge, C., Potter, C. S. and Carragher, B. (2016) 'A new method for vitrifying samples for cryoEM', *Journal of structural biology,* 195(2), pp. 190-198.

Rizzitello, A. E., Harper, J. R. and Silhavy, T. J. (2001) 'Genetic evidence for parallel pathways of chaperone activity in the periplasm of Escherichia coli', *Journal of bacteriology,* 183(23), pp. 6794-6800.

Rogowski, A., Briggs, J. A., Mortimer, J. C., Tryfona, T., Terrapon, N., Lowe, E. C., Baslé, A., Morland, C., Day, A. M., Zheng, H., Rogers, T. E., Thompson, P., Hawkins, A. R., Yadav, M. P., Henrissat, B., Martens, E. C., Dupree, P., Gilbert, H. J. and Bolam, D. N. (2015) 'Glycan complexity dictates microbial resource allocation in the large intestine', *Nature Communications,* 6(1), pp. 7481.

Rohou, A. and Grigorieff, N. (2015) 'CTFFIND4: Fast and accurate defocus estimation from electron micrographs', *Journal of structural biology,* 192(2), pp. 216-221.

Rollauer, S. E., Sooreshjani, M. A., Noinaj, N. and Buchanan, S. K. (2015) 'Outer membrane protein biogenesis in Gram-negative bacteria', *Philosophical Transactions of the Royal Society B: Biological Sciences,* 370(1679), pp. 20150023.

Rosenthal, P. B. and Henderson, R. (2003) 'Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy', *Journal of molecular biology,* 333(4), pp. 721-745.

Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J. M., Sorzano, C. O. S. and Vargas, J. (2021) 'DeepEMhancer: a deep learning solution for cryo-EM volume post-processing', *Communications biology,* 4(1), pp. 1-8.

Santiveri, M., Roa-Eguiara, A., Kühne, C., Wadhwa, N., Hu, H., Berg, H. C., Erhardt, M. and Taylor, N. M. (2020) 'Structure and function of stator units of the bacterial flagellar motor', *Cell,* 183(1), pp. 244-257. e16.

Scarff, C. A., Fuller, M. J., Thompson, R. F. and Iadanza, M. G. (2018) 'Variations on negative stain electron microscopy methods: tools for tackling challenging systems', *JoVE (Journal of Visualized Experiments),* (132), pp. e57199.

Scheres, S. H. (2012) 'RELION: implementation of a Bayesian approach to cryo-EM structure determination', *Journal of structural biology,* 180(3), pp. 519-530.

Scheres, S. H. (2016) 'Processing of structurally heterogeneous cryo-EM data in RELION', *Methods in enzymology,* 579, pp. 125-157.

Scheres, S. H. and Chen, S. (2012) 'Prevention of overfitting in cryo-EM structure determination', *Nature methods,* 9(9), pp. 853-854.

Schiffrin, B., Calabrese, A. N., Devine, P. W., Harris, S. A., Ashcroft, A. E., Brockwell, D. J. and Radford, S. E. (2016) 'Skp is a multivalent chaperone of outer-membrane proteins', *Nature structural & molecular biology,* 23(9), pp. 786-793.

Schirmer, T., Keller, T. A., Wang, Y.-F. and Rosenbusch, J. P. (1995) 'Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution', *Science,* 267(5197), pp. 512-514.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of image analysis', *Nature methods,* 9(7), pp. 671-675.

Schramm, E., Mende, J., Braun, V. and Kamp, R. (1987) 'Nucleotide sequence of the colicin B activity gene cba: consensus pentapeptide among TonB-dependent colicins and receptors', *Journal of bacteriology,* 169(7), pp. 3350-3357.

Schulz, G. E. (1992) 'Structure-function relationships in the membrane channel porin as based on a 1.8 Å resolution crystal structure', *Membrane Proteins: Structures, Interactions and Models*: Springer, pp. 403-412.

Schulz, G. E. (2002) 'The structure of bacterial outer membrane proteins', *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1565(2), pp. 308-317.

Sharma, S., Zhou, R., Wan, L., Feng, S., Song, K., Xu, C., Li, Y. and Liao, M. (2021) 'Mechanism of LolCDE as a molecular extruder of bacterial triacylated lipoproteins', *Nature Communications,* 12(1), pp. 1-11.

Sherman, D. J., Xie, R., Taylor, R. J., George, A. H., Okuda, S., Foster, P. J., Needleman, D. J. and Kahne, D. (2018) 'Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge', *Science,* 359(6377), pp. 798-801.

Shi, X., Hanley, S. A., Faray-Kele, M.-C., Fawell, S. C., Aduse-Opoku, J., Whiley, R. A., Curtis, M. A. and Hall, L. M. (2007) 'The rag locus of Porphyromonas gingivalis contributes to virulence in a murine model of soft tissue destruction', *Infection and immunity,* 75(4), pp. 2071-2074.

Shipman, J. A., Cho, K. H., Siegel, H. A. and Salyers, A. A. (1999) 'Physiological characterization of SusG, an outer membrane protein essential for starch utilization by Bacteroides thetaiotaomicron', *Journal of bacteriology,* 181(23), pp. 7206-7211.

Shultis, D. D., Purdy, M. D., Banchs, C. N. and Wiener, M. C. (2006) 'Outer membrane active transport: structure of the BtuB: TonB complex', *Science,* 312(5778), pp. 1396-1399.

Sigworth, F. J. (2016) 'Principles of cryo-EM single-particle image processing', *Microscopy,* 65(1), pp. 57-67.

Silhavy, T. J., Kahne, D. and Walker, S. (2010) 'The bacterial cell envelope', *Cold Spring Harbor perspectives in biology,* 2(5), pp. a000414.

Sklar, J. G., Wu, T., Kahne, D. and Silhavy, T. J. (2007) 'Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli', *Genes & development,* 21(19), pp. 2473-2484.

Socransky, S., Haffajee, A., Cugini, M., Smith, C. and Kent Jr, R. (1998) 'Microbial complexes in subgingival plaque', *Journal of clinical periodontology,* 25(2), pp. 134-144.

Some, D., Amartely, H., Tsadok, A. and Lebendiker, M. (2019) 'Characterization of proteins by size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS)', *JoVE (Journal of Visualized Experiments),* (148), pp. e59615.

Sonn-Segev, A., Belacic, K., Bodrug, T., Young, G., VanderLinden, R. T., Schulman, B. A., Schimpf, J., Friedrich, T., Dip, P. V., Schwartz, T. U., Bauer, B., Peters, J.-M., Struwe, W. B., Benesch, J. L. P., Brown, N. G., Haselbach, D. and Kukura, P. (2020) 'Quantifying the heterogeneity of macromolecular machines by mass photometry', *Nature Communications,* 11(1), pp. 1772.

Sonnenburg, E. D., Zheng, H., Joglekar, P., Higginbottom, S. K., Firbank, S. J., Bolam, D. N. and Sonnenburg, J. L. (2010) 'Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations', *Cell,* 141(7), pp. 1241-1252.

Sperandeo, P., Martorana, A. M. and Polissi, A. (2019) 'Lipopolysaccharide biosynthesis and transport to the outer membrane of Gram-negative bacteria', *Bacterial cell walls and membranes*, pp. 9-37.

Strauch, K. L., Johnson, K. and Beckwith, J. (1989) 'Characterization of degP, a gene required for proteolysis in the cell envelope and essential for growth of Escherichia coli at high temperature', *Journal of Bacteriology,* 171(5), pp. 2689-2696.

Sun, J., Rutherford, S. T., Silhavy, T. J. and Huang, K. C. (2021) 'Physical properties of the bacterial outer membrane', *Nature Reviews Microbiology*, pp. 1-13.

Tamura, K. and Brumer, H. (2021) 'Glycan utilization systems in the human gut microbiota: a gold mine for structural discoveries', *Current Opinion in Structural Biology,* 68, pp. 26-40.

Tang, X., Chang, S., Luo, Q., Zhang, Z., Qiao, W., Xu, C., Zhang, C., Niu, Y., Yang, W., Wang, T., Zhang, Z., Zhu, X., Wei, X., Dong, C., Zhang, X. and Dong, H. (2019) 'Cryo-EM structures of lipopolysaccharide transporter LptB2FGC in lipopolysaccharide or AMP-PNP-bound states reveal its transport mechanism', *Nature Communications,* 10(1), pp. 4175.

Tang, X., Chang, S., Zhang, K., Luo, Q., Zhang, Z., Wang, T., Qiao, W., Wang, C., Shen, C., Zhang, Z., Zhu, X., Wei, X., Dong, C., Zhang, X. and Dong, H. (2021) 'Structural basis for bacterial lipoprotein relocation by the transporter LolCDE', *Nature Structural & Molecular Biology,* 28(4), pp. 347-355.

Taniguchi, N., Matsuyama, S.-i. and Tokuda, H. (2005) 'Mechanisms underlying energy-independent transfer of lipoproteins from LolA to LolB, which have similar unclosed β-barrel structures', *Journal of Biological Chemistry,* 280(41), pp. 34481-34488.

Tauzin Alexandra, S., Kwiatkowski Kurt, J., Orlovsky Nicole, I., Smith Christopher, J., Creagh, A. L., Haynes Charles, A., Wawrzak, Z., Brumer, H., Koropatkin Nicole, M., Sonnenburg, J. and McFall-Ngai Margaret, J. 'Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont', *mBio,* 7(2), pp. e02134-15.

Thompson, R. F., Iadanza, M. G., Hesketh, E. L., Rawson, S. and Ranson, N. A. (2019) 'Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy', *Nature protocols,* 14(1), pp. 100-118.

Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. and Ranson, N. A. (2016) 'An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology', *Methods,* 100, pp. 3-15.

Thong, S., Ercan, B., Torta, F., Fong, Z. Y., Wong, H. Y. A., Wenk, M. R. and Chng, S.-S. (2016) 'Defining key roles for auxiliary proteins in an ABC transporter that maintains bacterial outer membrane lipid asymmetry', *Elife,* 5, pp. e19042.

Tomasek, D., Rawson, S., Lee, J., Wzorek, J. S., Harrison, S. C., Li, Z. and Kahne, D. (2020) 'Structure of a nascent membrane protein as it folds on the BAM complex', *Nature,* 583(7816), pp. 473-478.

Tonetti, M. S., D'Aiuto, F., Nibali, L., Donald, A., Storry, C., Parkar, M., Suvan, J., Hingorani, A. D., Vallance, P. and Deanfield, J. (2007) 'Treatment of periodontitis and endothelial function', *New England Journal of Medicine,* 356(9), pp. 911-920.

Tuson, H. H., Foley, M. H., Koropatkin, N. M. and Biteen, J. S. (2018) 'The starch utilization system assembles around stationary starch-binding proteins', *Biophysical journal,* 115(2), pp. 242-250.

Usher, A. K. and Stockley, R. A. (2013) 'The link between chronic periodontitis and COPD: a common role for the neutrophil?', *BMC medicine,* 11(1), pp. 1-11.

Valguarnera, E., Scott, N. E., Azimzadeh, P. and Feldman, M. F. (2018) 'Surface exposure and packing of lipoproteins into outer membrane vesicles are coupled processes in Bacteroides', *MSphere,* 3(6), pp. e00559-18.

Van Gelder, P., Saint, N., Phale, P., Eppens, E. F., Prilipov, A., van Boxtel, R., Rosenbusch, J. P. and Tommassen, J. 1997. Voltage sensing in the PhoE and OmpF outer membrane porins of Escherichia coli: role of charged residues. Elsevier.

Venkatakrishnan, A. J., Deupi, X., Lebon, G., Heydenreich, F. M., Flock, T., Miljus, T., Balaji, S., Bouvier, M., Veprintsev, D. B., Tate, C. G., Schertler, G. F. X. and Babu, M. M. (2016) 'Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region', *Nature,* 536(7617), pp. 484-487.

Vergalli, J., Bodrenko, I. V., Masi, M., Moynié, L., Acosta-Gutiérrez, S., Naismith, J. H., Davin-Regli, A., Ceccarelli, M., van den Berg, B., Winterhalter, M. and Pagès, J.-M. (2020) 'Porins and small-molecule translocation across the outer membrane of Gram-negative bacteria', *Nature Reviews Microbiology,* 18(3), pp. 164-176.

Vogt, J. and Schulz, G. E. (1999) 'The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence', *Structure,* 7(10), pp. 1301-1309.

Vollmer, W., Blanot, D. and De Pedro, M. A. (2008) 'Peptidoglycan structure and architecture', *FEMS microbiology reviews,* 32(2), pp. 149-167.

Voulhoux, R., Bos, M. P., Geurtsen, J., Mols, M. and Tommassen, J. (2003) 'Role of a highly conserved bacterial protein in outer membrane protein assembly', *Science,* 299(5604), pp. 262-265.

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T. R., Lill, P., Gatsogiannis, C. and Raunser, S. (2019) 'SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM', *Communications Biology,* 2(1), pp. 218.

Walton, T. A., Sandoval, C. M., Fowler, C. A., Pardi, A. and Sousa, M. C. (2009) 'The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains', *Proceedings of the National Academy of Sciences,* 106(6), pp. 1772-1777.

Walton, T. A. and Sousa, M. C. (2004) 'Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation', *Molecular cell,* 15(3), pp. 367-374.

Wang, M., Krauss Jennifer, L., Domon, H., Hosur Kavita, B., Liang, S., Magotti, P., Triantafilou, M., Triantafilou, K., Lambris John, D. and Hajishengallis, G. (2010) 'Microbial Hijacking of Complement–Toll-Like Receptor Crosstalk', *Science Signaling,* 3(109), pp. ra11-ra11.

Wang, Y.-F., Dutzler, R., Rizkallah, P. J., Rosenbusch, J. P. and Schirmer, T. (1997) 'Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin', *Journal of molecular biology,* 272(1), pp. 56-63.

Ward, A., Reyes, C. L., Yu, J., Roth, C. B. and Chang, G. (2007) 'Flexibility in the ABC transporter MsbA: Alternating access with a twist', *Proceedings of the National Academy of Sciences,* 104(48), pp. 19005-19010.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A P., Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T. (2018) 'SWISS-MODEL: homology modelling of protein structures and complexes', *Nucleic Acids Research,* 46(W1), pp. W296-W303.

Weiss, M., Abele, U., Weckesser, J., Welte, W., Schiltz, E. and Schulz, G. (1991) 'Molecular architecture and electrostatic properties of a bacterial porin', *Science,* 254(5038), pp. 1627-1630.

Whitfield, C. and Trent, M. S. (2014) 'Biosynthesis and export of bacterial lipopolysaccharides', *Annual review of biochemistry,* 83, pp. 99-128.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall Iii, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. and Richardson, D. C. (2018) 'MolProbity: More and better reference data for improved all-atom structure validation', *Protein Science,* 27(1), pp. 293-315.

Wilson, M. M., Anderson, D. E. and Bernstein, H. D. (2015) 'Analysis of the outer membrane proteome and secretome of Bacteroides fragilis reveals a multiplicity of secretion mechanisms', *PloS one,* 10(2), pp. e0117732.

Wilson, M. M. and Bernstein, H. D. (2016) 'Surface-exposed lipoproteins: an emerging secretion phenomenon in Gram-negative bacteria', *Trends in microbiology,* 24(3), pp. 198-208.

Wimley, W. C. (2003) 'The versatile β-barrel membrane protein', *Current opinion in structural biology,* 13(4), pp. 404-411.

Wu, M. and Lander, G. C. (2020) 'How low can we go? Structure determination of small biological complexes using single-particle cryo-EM', *Current opinion in structural biology,* 64, pp. 9-16.

Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T. J. and Kahne, D. (2005) 'Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli', *Cell,* 121(2), pp. 235-245.

Yamaguchi, K., Yu, F. and Inouye, M. (1988) 'A single amino acid determinant of the membrane localization of lipoproteins in E. coli', *Cell,* 53(3), pp. 423-432.

Yan, Z., Yin, M., Xu, D., Zhu, Y. and Li, X. (2017) 'Structural insights into the secretin translocation channel in the type II secretion system', *Nature structural & molecular biology,* 24(2), pp. 177-183.

Yau, W.-M., Wimley, W. C., Gawrisch, K. and White, S. H. (1998) 'The preference of tryptophan for membrane interfaces', *Biochemistry,* 37(42), pp. 14713-14718.

Zhai, Y. F., Heijne, W. and Saier Jr, M. H. (2003) 'Molecular modeling of the bacterial outer membrane receptor energizer, ExbBD/TonB, based on homology with the flagellar motor, MotAB', *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1614(2), pp. 201-210.

Zhang, K. (2016) 'Gctf: Real-time CTF determination and correction', *Journal of structural biology,* 193(1), pp. 1-12.

Zhang, X., Jin, L., Fang, Q., Hui, W. H. and Zhou, Z. H. (2010) '3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry', *Cell,* 141(3), pp. 472-482.

Zheng, R., Cai, R., Liu, R., Liu, G. and Sun, C. (2021) 'Maribellus comscasis sp. nov., a novel deep-sea Bacteroidetes bacterium, possessing a prominent capability of degrading cellulose', *Environ Microbiol,* 23(8), pp. 4561-4575.

Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y. and Agard, D. A. (2017) 'MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy', *Nature methods,* 14(4), pp. 331-332.

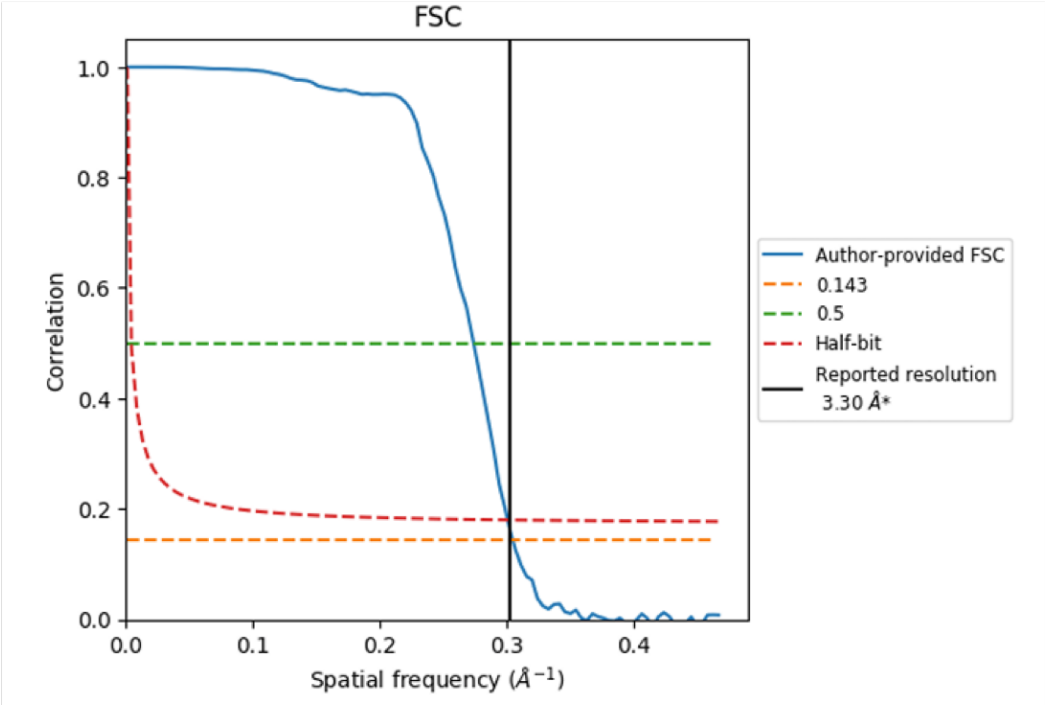
Zhou, Q., Yang, D., Wu, M., Guo, Y., Guo, W., Zhong, L., Cai, X., Dai, A., Jang, W., Shakhnovich, E. I., Liu, Z.-J., Stevens, R. C., Lambert, N. A., Babu, M. M., Wang, M.-W. and Zhao, S. (2019) 'Common activation mechanism of class A GPCRs', *eLife,* 8, pp. e50279.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E. and Scheres, S. H. (2018) 'New tools for automated high-resolution cryo-EM structure determination in RELION-3', *elife,* 7, pp. e42166.

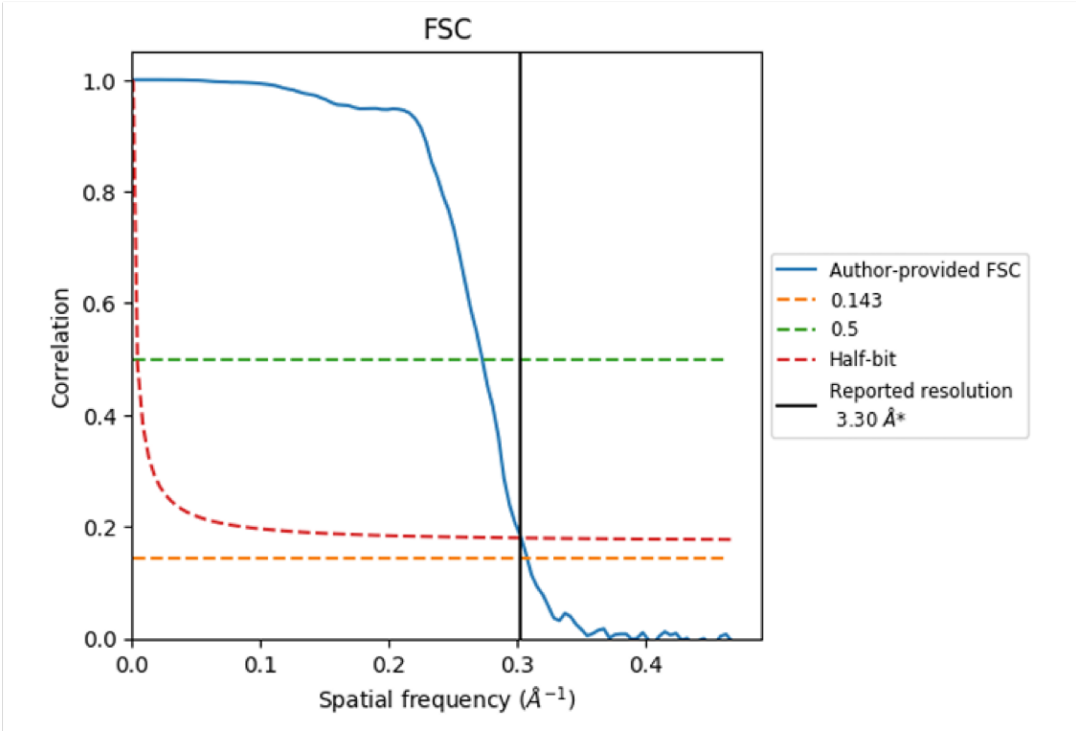
Appendix

Below are the Fourier shell correlation curves for maps used in model building and refinement in chapter 3.

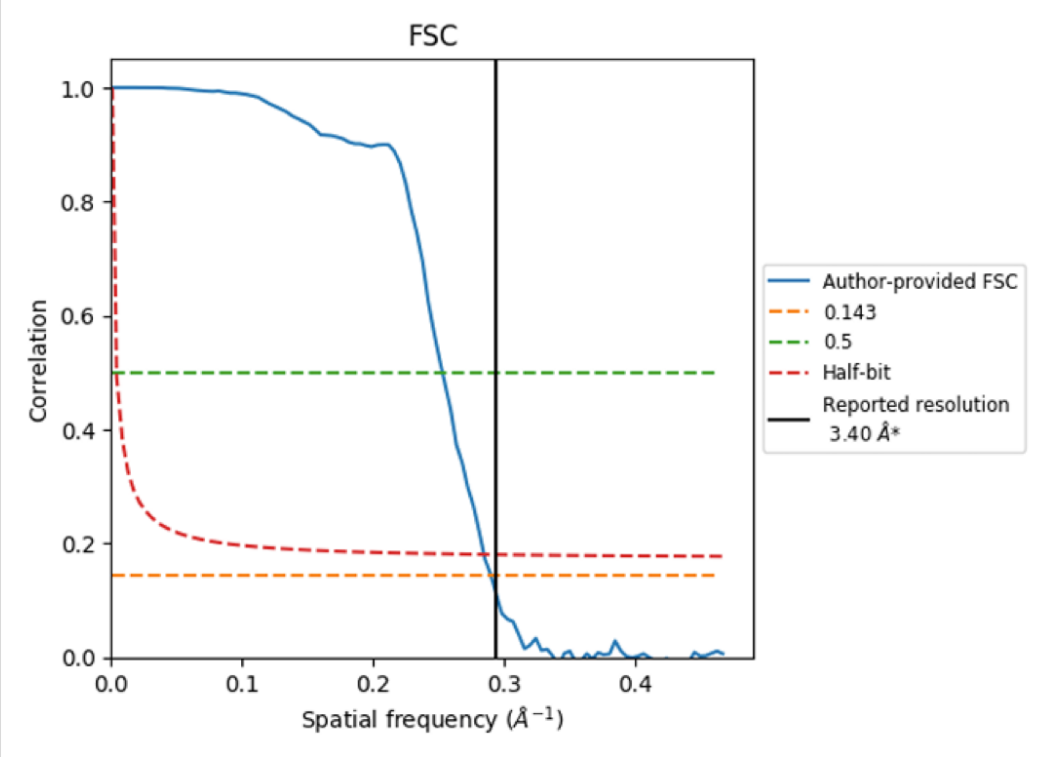
RagAB: close-closed conformation



RagAB: open-closed conformation

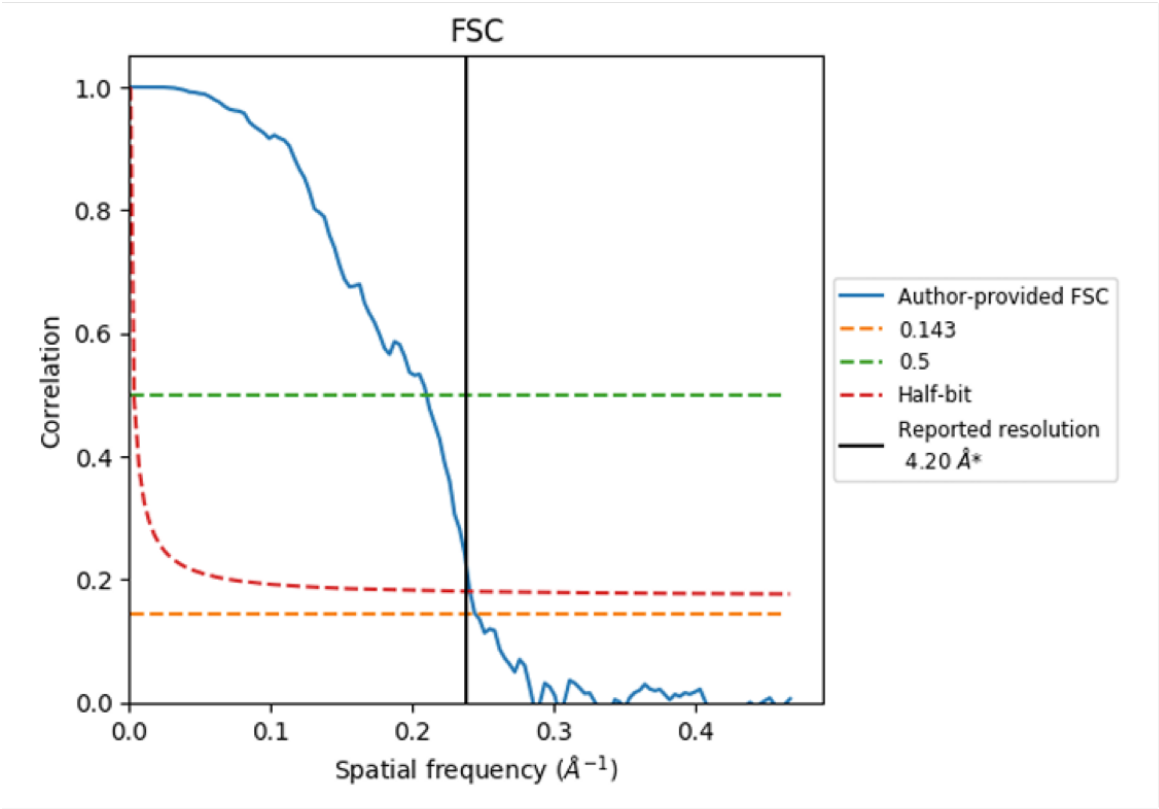


RagAB: open-open conformation

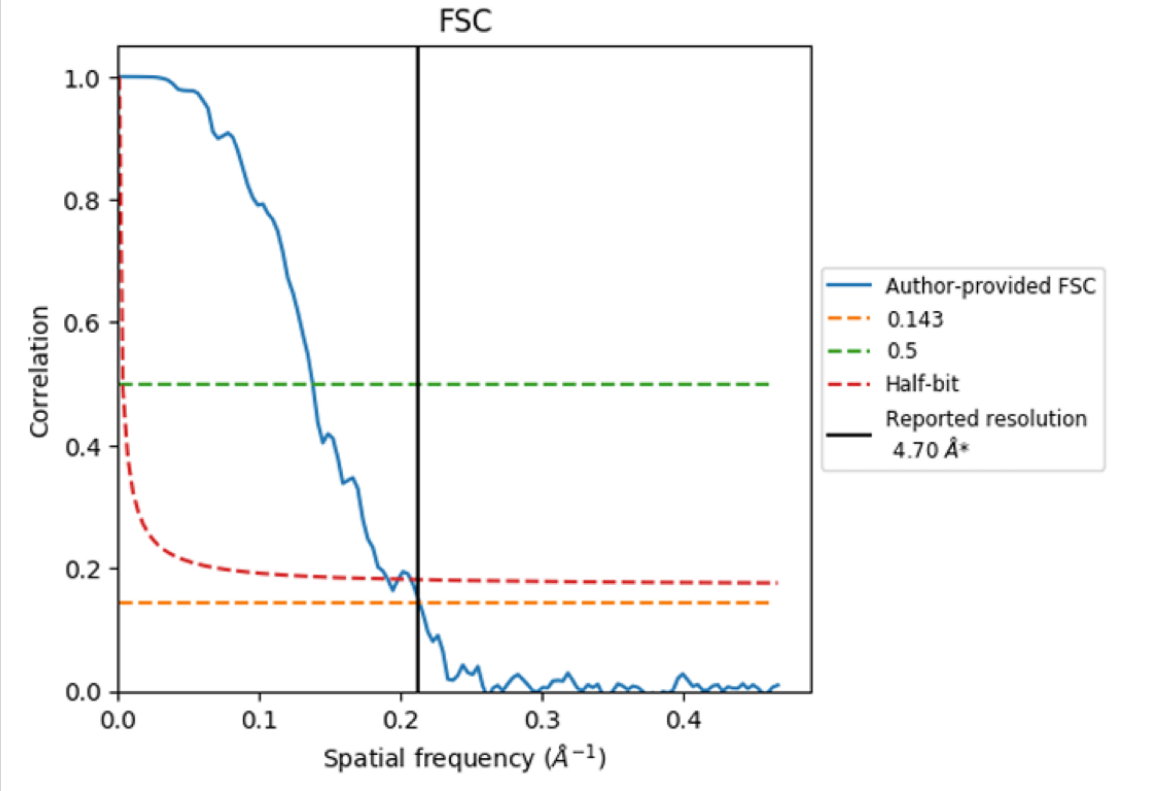


Below are the Fourier shell correlation curves for maps used in model building and refinement in chapter 4.

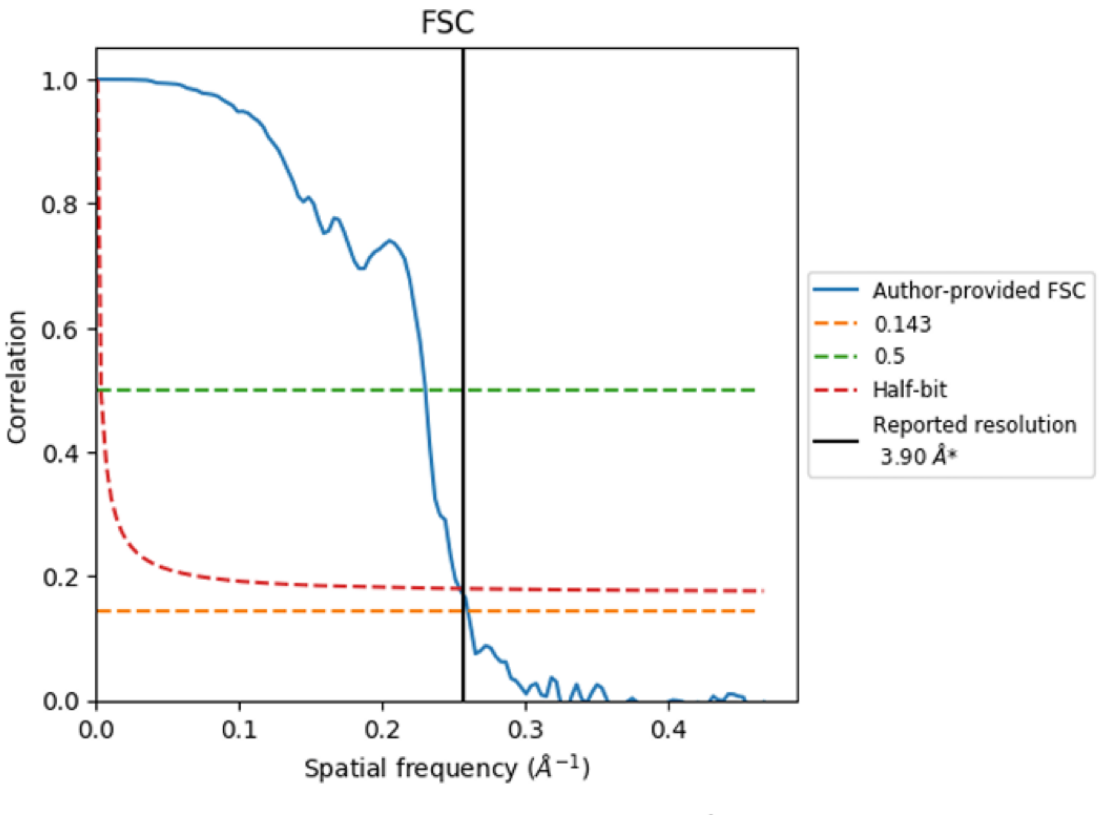
Levan SusCD: closed-closed conformation



Levan SusCD: open-closed conformation

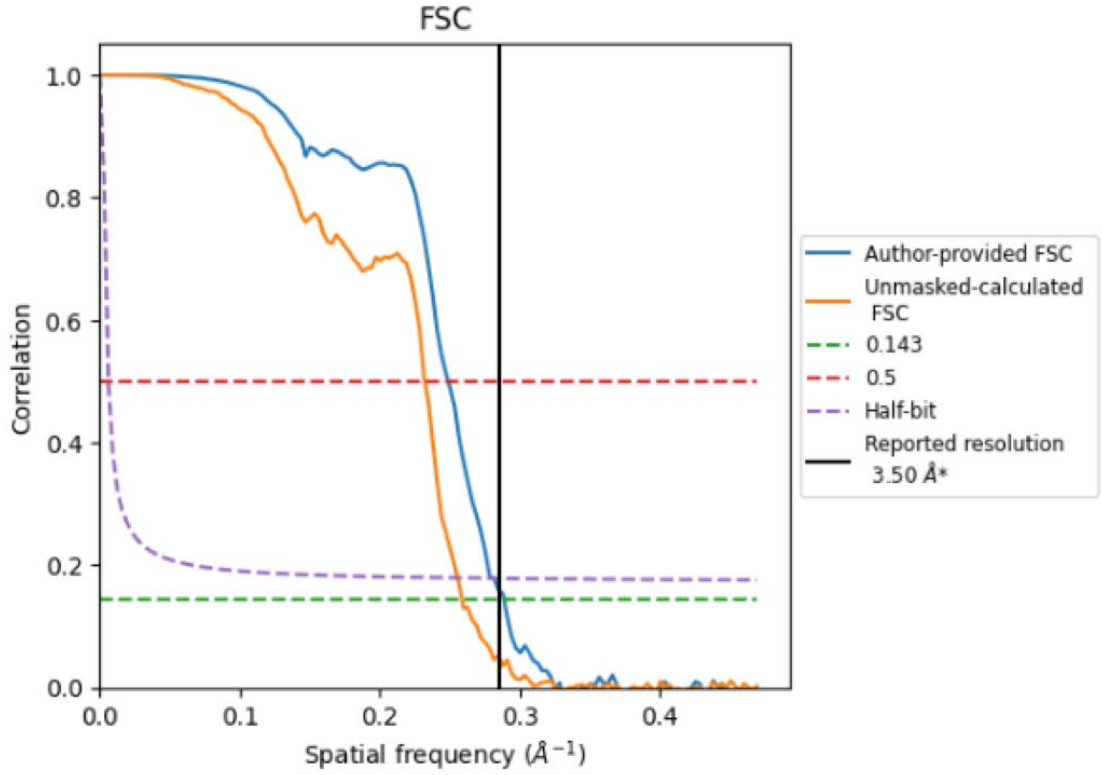


Levan SusCD: closed-closed conformation

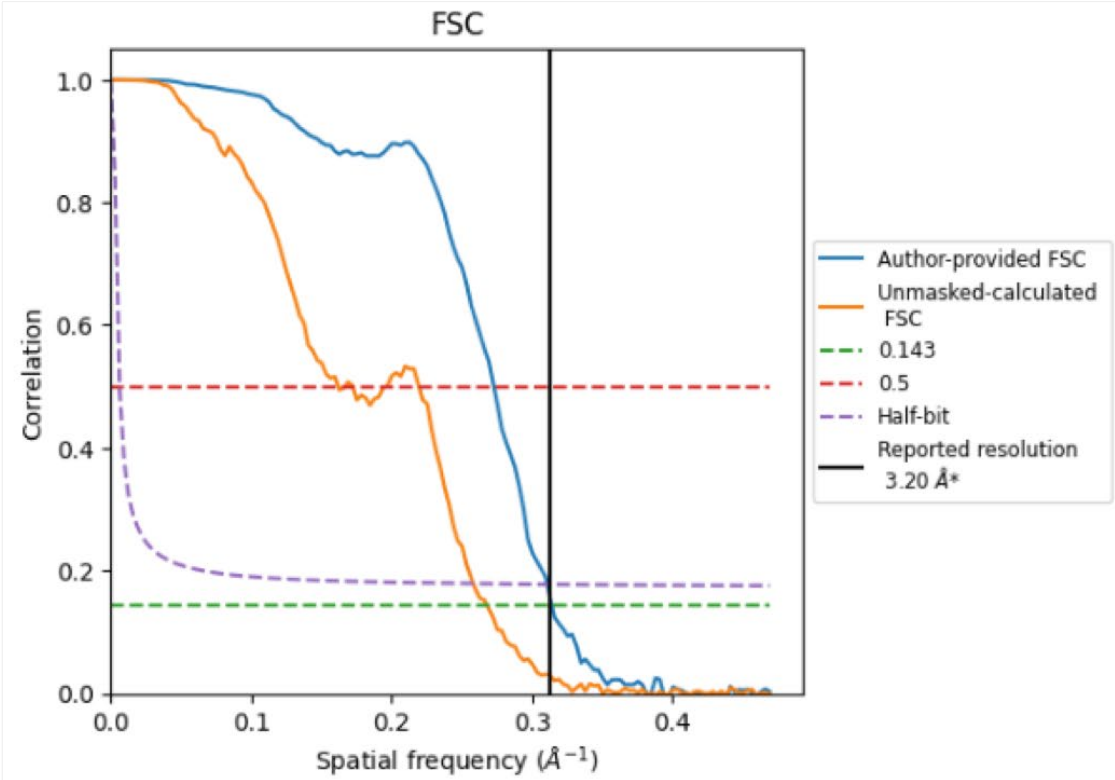


Below are the Fourier shell correlation curves for maps used in model building and refinement in chapter 5.

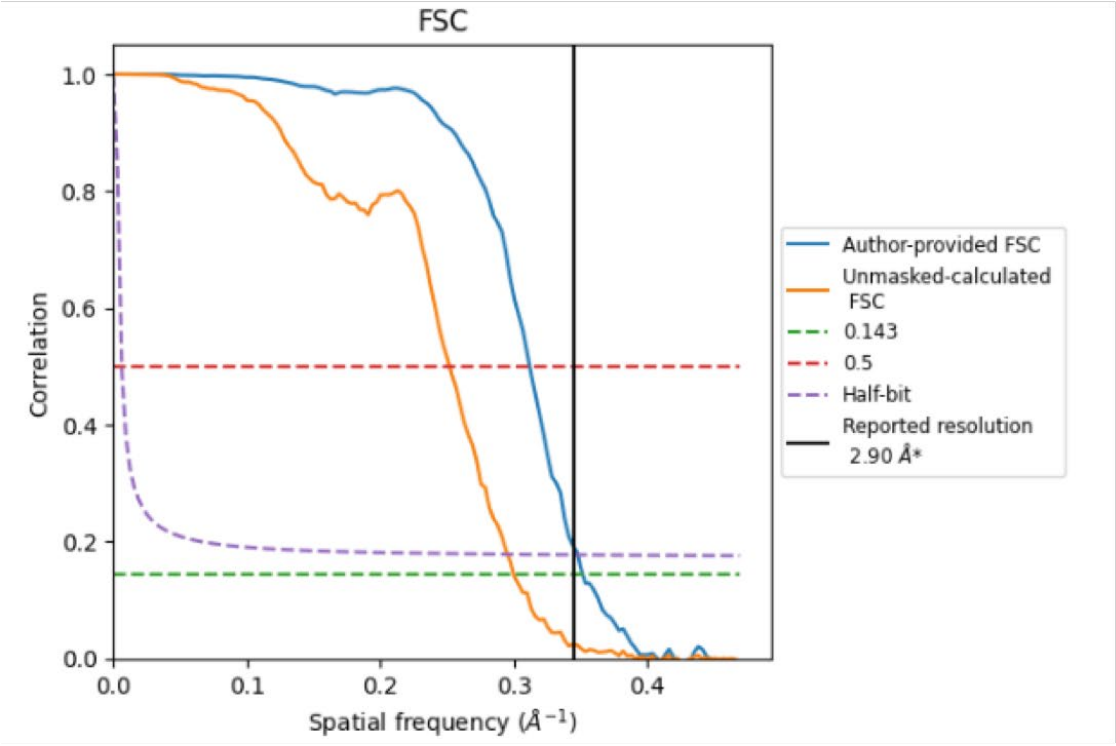
Substrate free levan utilisome



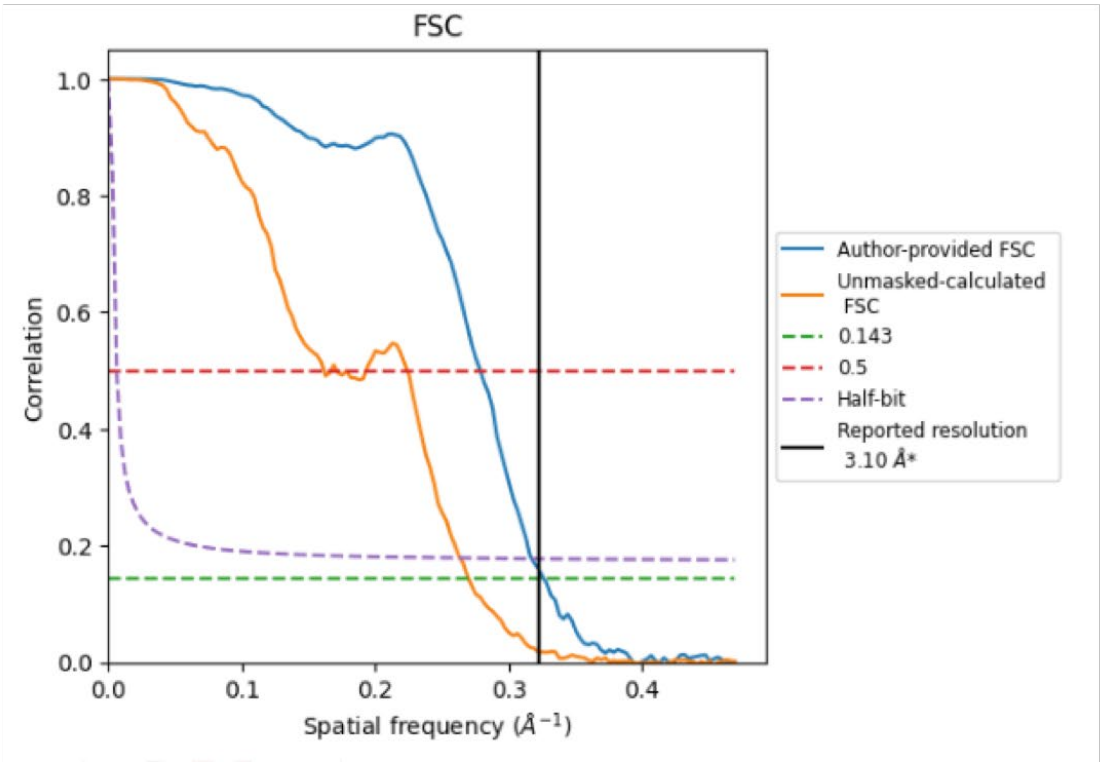
Levan utilisome with FOS DP 8-12



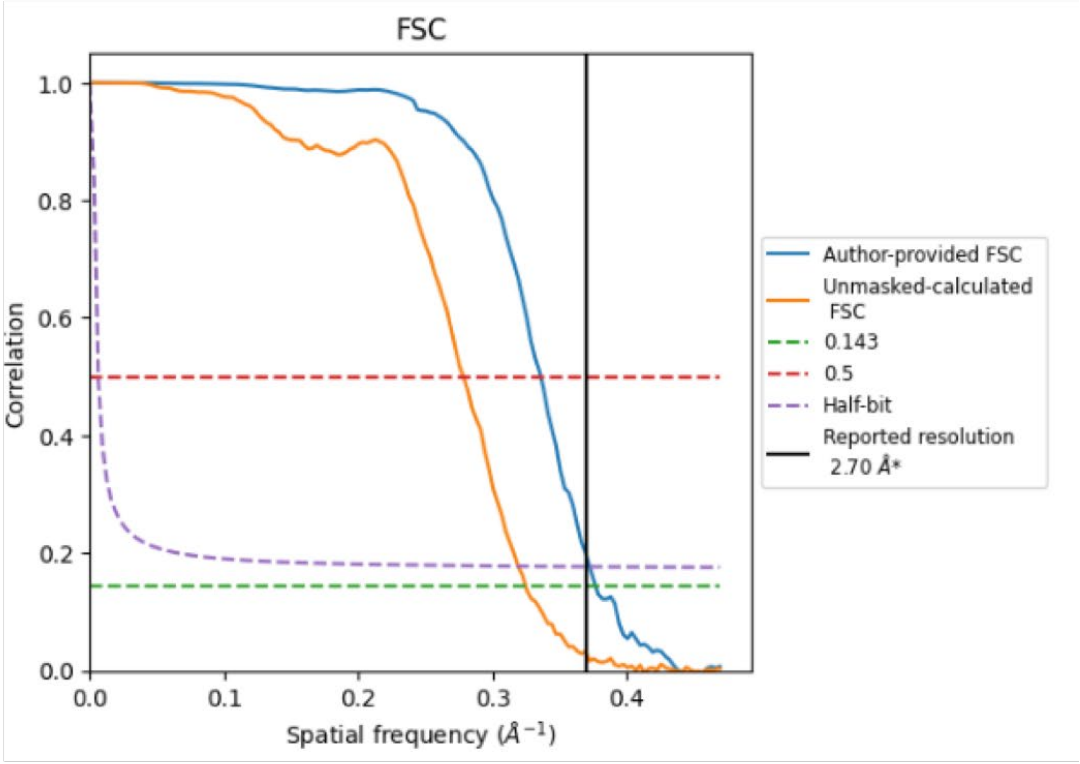
Core SusCD components of levan utilisome with FOS DP8-12



Inactive levan utilisome with FOS DP15-25



Core SusCD components of utilisome with FOS DP15-25



SusC components of the dextran utilisome