Insights into the molecular mechanisms of ABC transporters and membrane insertases



Francesco Fiorentino

Kellogg College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

in Physical and Theoretical Chemistry

22nd July 2020

To the beloved memory of Ausilia and Fedele

Acknowledgements

First and foremost, I would like to thank my supervisor, Carol Robinson, for giving me the opportunity of being part of her group. I am truly indebted to her for her support throughout my DPhil; she always gave me confidence and freedom to pursue my ideas, even during moments where nothing worked. I credit her for both my scientific and personal growth.

I owe a huge thank you to Jani; I do not think I could have reached this level of expertise in biophysics and molecular biology without him. He taught me everything I know since my first day as a rotation student with a great deal of patience and contagious enthusiasm. He has been a fantastic teacher, advisor and friend during these years. I also thank Shahid for his help and teaching of native mass spectrometry and for bearing with a student who did not even know what native MS was. He has been a great desk mate and always ready to dispense science and life advice.

I would like to thank Rachel Davis, Zara Sands and Silvia Lovera from UCB for supporting my work and for providing useful suggestions throughout my whole PhD.

I thank Leo, not only for his "magic" detergents but also for all the fun in the lab and the advice close to the end of my PhD, he made the lab a more cheerful place. I also acknowledge Di, for his valuable help in lipidomics and proteomics and Shane for sharing his HDX knowledge; Joe for the continuous support with mass spec instrumentation and stimulating discussions and Mark for teaching me how to take care of chromatography systems. I thank Idi, for always offering a joke but also insightful perspectives on my work. Thanks to Bodhi, Tom, Tarick and Corinne for their company and help in the wet lab and good conversations (scientific and not).

Special thanks to Phil (Xingyu Qiu) who assisted me during the days (and nights) of HDX experiments. A crazy instrument helped us to foster true friendship. Thanks to my other fellow DPhils Daniel, Fabian (the two San Diego compañeros), Denis, Dom, and Siyun for the

laughter during the time spent together; their company has been one of the reasons I really enjoyed my time as a DPhil during both happy and tough times. I thank Josh, for keeping me company through constant Skype calls during these lockdown months.

I finally want to thank all the Robinson and Benesch group members; I think that each of them made my time as a DPhil generally enjoyable and it has been a pleasure to work with each one of you. Special mention to Vicky and Andy for taking care of 25 kids and keeping the lab up and running every day.

I owe a big thank you to my housemate and friend Sthitadhi, who not only helped me with coding but also was a great assistant in our kitchen shenanigans.

I am incredibly grateful to my friends who, although living far from me, have always been present in my life. Matteo, Luca B., Gabriele, Michele, Elia, Nicola, Luca C.: thanks for being a source of relief and laughter. I am proud to have such friends like you, knowing that I could rely on you any time, from any distance represented a great reassurance over these years.

I also thank Leonardo for being such a great friend (and food lover); some of the best months spent in Oxford are the ones spent with you.

I would like to thank my family: my parents, my brother Vincenzo and my sister Martina who unconditionally supported me during this journey. A great part of this achievement is shared with my parents, who always encouraged my love for science and gave me full freedom on every life choice. I could have never reached this accomplishment without your backing.

A special thank you to Maria Angela and Giuliano, my other half's parents, who welcomed me every time I was back in Rome and treated me like a son.

Last, but not least, I am incredibly grateful to Anna: her continuous encouragement throughout these years, full of highs and lows, made this achievement possible. I feel extremely fortunate to have you by my side; living 4 years apart was not easy, but you sustained me no matter what. Thank you because when I am with you the whole world has a different colour.

Abstract

Membrane proteins have a pivotal role in cellular homeostasis given their strategic position and involvement in pathways such as small molecules transport and interaction with ligands leading to downstream responses. In this thesis, I describe the application of mass spectrometry approaches to elucidate the structure, dynamics and regulation of membrane protein complexes, including ABC transporters and membrane insertases. The employment of native mass spectrometry (nMS) to membrane proteins has allowed the examination of non-covalent interactions, as described in Chapter 2 where the interaction between ABC importers and their cognate substrate binding proteins (SBPs) was analysed and the mechanism of transport of these complexes was clarified. This analysis indicated that substrate and nucleotide binding destabilise the interaction between the type II importer BtuCD and the SBP BtuF, while has no effect on the interaction between type I importer ModBC and the SBP ModA. Similarly, nMS allowed the elucidation of protein-lipid and protein-substrate interactions of the LPS transporter $LptB_2FG$, suggesting that cardiolipin competes for LPS binding and nucleotide binding increases affinity for the substrate. The combination of nMS and cryo-electron microscopy (cryo-EM) allowed further elucidation on the structure of this protein complex and its interactions with LptC and LptA. The combination of nMS with hydrogen-deuterium exchange mass spectrometry (HDX-MS) and molecular dynamics (MD) simulations yielded information on the conformational dynamics of the outer membrane LPS translocon LptDE, showing opposite effects of substrate and antimicrobial peptide thanatin. Indeed, while LPS induces opening of the periplasmic domain of LptD, thanatin stabilises this region, acting as a non-competitive inhibitor of LPS transport. In Chapter 5, nMS is again combined to cryo-EM, as well as lipidomics and *in vivo* studies to define the structure and subunit stoichiometry of the eukaryotic tail-anchored protein insertase complex, revealing the contribution of phosphatidylinositol binding on the stabilisation of the complex.

Declaration

This thesis describes research carried out between July 2017 and April 2020. The dissertation is in the form of an **integrated thesis** (as approved in the sub-department of Physical and Theoretical Chemistry from Hilary Term 2018). The thesis is a hybrid of conventional chapters (Chapters 1, 3, 6) and scientific papers that have been published or submitted for publication (Chapters 2, 4, 5). The data presented were acquired and analysed by myself unless explicitly stated; each chapter includes a section summarising the contributions to the described work. For stylistic reasons and since the work described in this thesis involved collaboration of varying degrees, I have used the third-person pronoun 'we' throughout the thesis.

This work has not been submitted for a degree, diploma or any other academic qualification at any other university or academic institution, in whole or in part.

Francesco Fiorentino

Oxford, 22nd July 2020

Contents

1 Background and methods for investigating membrane proteins 1

1.1 Introduction……1

1.1.1 Membrane protein complexes……..……2

1.1.2 Membrane lipids…….……..……5

1.1.3 Bacterial membrane and lipopolysaccharide…..……..……8

1.1.4 Protein-lipid interactions…..……...……10

1.2 Mass spectrometry.……11

1.2.1 Electrospray ionisation……...……..……..……13

1.2.2 ESI process mechanistic models....……...……..……15

1.2.3 Mass analysers …..……..……...…17

1.2.4 Quadrupole mass analyser…..…….…..……...…..17

1.2.5 Time of flight mass analyser…..……...……..19

1.2.6 Orbitrap mass analyser……..…….21

1.2.7 Ion detection…..……...…….23

1.2.8 Collision cell…..……...…….24

1.3 The application of mass spectrometry to structural biology……..25

1.3.1 Native mass spectrometry..…..……...……..……25

1.3.2 Evaluation of protein-ligand interactions using nESI-nMS……27

1.3.3 Hydrogen/deuterium exchange mass spectrometry (HDX-MS)……30

1.4 Mass spectrometry instrumentation...……37

1.4.1 Thermo Fisher Scientific Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer …….…….37

1.4.2 Thermo Fisher Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer……38

1.4.3 Waters Hybrid Q-ToF Synapt G2-Si Mass Spectrometer……..39

1.5 Aims of this dissertation……40

1.6 References……..……42

2 The different effects of substrates and nucleotides on the complex formation of ABC transporters 56

2.1 Introduction....……59

2.2 Results……...……61

2.2.1 Substrates have different effects on the formation of type I and type II transporter complexes....…….61

2.2.2 Vitamin B_{12} and cyanocobinamide destabilise the BtuCD-BtuF interaction……..……64

2.2.3 ATP binding acts synergistically with the substrate to decrease BtuCD-BtuF stability……..……66

2.2.4 ATP binding does not alter ModBC-ModA interactions…….68

2.3 Discussion…..……69

2.4 Supplementary Information...…..……73

2.5 Materials and methods……..…….……78

2.6 References....……..……..……83

3 Elucidating the mechanism of LPS extraction from the inner membrane by $LptB_2FGC$ 86

3.1 Introduction....……86

3.2 Results……...……90

3.2.1 Optimisation of $LptB_2FG$ purification and nMS analysis...……90

3.2.2 Substrate and lipid binding to the $LptB_2FG$ complex…..……92

3.2.3 ATP binding and hydrolysis have opposite effects on the $LptB_2FG$ substrate affinity..……..…….95

3.2.4 $LptB_2FGC$ purification and nMS analysis…….……97

3.2.5 Cryo-EM structure of $LptB_2FG$...……..…….98

3.2.6 Cryo-EM structure of $LptB_2FGC$…..……..……101

3.2.7 *pa*$LptB_2FGC$ is in the pre-loading state……103

3.2.8 The periplasmic domain of *pa*$LptB_2FG$/$LptB_2FGC$ is in a different conformation compared to its orthologues...……105

3.2.9 The periplasmic oligomeric protein LptA……107

3.2.10 $LptB_2FGC$ – LptA interaction and thanatin influence……108

3.3 Discussion...……110

3.4 Supplementary Information……114

3.4.1 Supplementary Figures..……114

3.4.2 Supplementary Tables...……126

3.5 Materials and methods…….……129

3.6 References…..…….…….……136

4 Concerted dynamics of an LPS translocon in the presence of substrate and an antimicrobial peptide 141

4.1 Introduction……144

4.2 Results…….……147

4.2.1 LPS and thanatin do not compete with each other for the same binding site on LptDE...……...……..…147

4.2.2 LPS binding to LptDE enhances β-taco opening and decreases β-barrel dynamics...149

4.2.3 Molecular dynamics simulations suggest an open LptDE conformation on substrate binding...152

4.2.4 Thanatin binding stabilises the LptD β-taco...154

4.2.5 Thanatin interferes with LPS-induced opening of LptD β-taco ...156

4.3 Discussion...159

4.4 Supplementary Information...164

4.4.1 Supplementary Figures...164

4.4.2 Supplementary Tables...195

4.5 Materials and methods...199

4.6 References...210

5 Structural basis of the ER membrane-associated steps in tail anchored protein insertion by the TRC40 (Get3)-insertase complex 215

5.1 Introduction...218

5.2 Results...222

5.2.1 The Get1/Get2 insertase forms an evolutionarily conserved heterotetramer...222

5.2.2 Get3 symmetric binding promotes heterotetramerisation of the Get1/Get2 insertase...222

5.2.3 The structure of CAMLΔN-WRB/TRC40 reveals a YidC-like fold in the membrane...225

5.2.4 The heterotetramer is stabilised by phosphatidylinositol binding...229

5.2.5 PI binds to functionally important interfacial residues...231

5.2.6 A conserved CAML helix α3' forms a functionally important gate with TRC40....234
5.3 Discussion.....236
5.3.1 A YidC-like mechanism for TA insertion involving a hydrophilic groove...236
5.3.2 The membrane heterotetramer increases the efficiency of TA insertion.....237
5.3.3 A dynamic model for TA insertion....238
5.3.4 Perspectives..240
5.4 Supplementary Information.....242
5.4.1 Supplementary Figures..242
5.4.2 Supplementary Tables...253
5.5 Materials and methods.....256
5.6 References.....267

6 Concluding remarks 273

List of abbreviations

ABC	ATP-binding cassette
ADP	Adenosine diphosphate
AMP-PNP	β-γ-imidoadenosine 5′-triphosphate
ATP	Adenosine triphosphate
C_8E_4	Tetraethylene glycol monooctyl ether
$C_{12}E_8$	Octaethylene glycol monododecyl ether
$C_{12}E_9$	Nonaethylene glycol monododecyl ether
CC	Correlation coefficient
CDL	Cardiolipin
CID	Collision-induced dissociation
CMC	Critical micelle concentration
CEM	Chain ejection model
CRM	Charged residue model
CD	Cytoplasmic domain
Cymal-5	5-Cyclohexylpentyl β-D-maltoside
Cymal-5 NG	5-Cyclohexylpentyl β-D-maltoside Neopentyl Glycol
Cymal-6	6-Cyclohexylhexyl β-D-Maltoside
(k)Da	(kilo)Dalton
DC	Direct current
DDM	n-Dodecyl β-D-maltoside
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylene diamine tetraacetic acid
EM	Electron microscopy
ER	Endoplasmic reticulum
(n)ESI	(Nano) Electrospray ionisation
ETD	Electron transfer dissociation
ECD	Extracellular domain
FRET	Förster resonance energy transfer
G1-OGD	[G1]-ether-C12 detergent mixture
GET	Guided entry of tail-anchored proteins
GFP	Green fluorescent protein
GPCR	G protein-coupled receptor
HCD	High-energy collision dissociation

HDX	Hydrogen/deuterium exchange
HPLC	High performance liquid chromatography
IEM	Ion evaporation model
IM	Ion mobility
IM	Inner membrane
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LDAO	n-Dodecyl-*N*,*N*-Dimethylamine *N*-Oxide
LMNG	Lauryl Maltose Neopentyl Glycol
LPS	Lipopolysaccharide
MALDI	Matrix assisted laser desorption ionisation
MCP	Micro-channel plate
MD	Molecular dynamics
MS	Mass spectrometry
MST	Microscale thermophoresis
m*/*z	Mass-to-charge ratio
NBD	Nucleotide-binding domain
NG	Neopentyl Glycol
NMR	Nuclear magnetic resonance
nMS	Native mass spectrometry
OGD	Oligoglycerol detergent
OGNG	Octyl glucose neopentyl glycol
OM	Outer membrane
O-PS	O-antigenic polysaccharide
PA	Phosphatidic acid
PAGE	Polyacrylamide gel electrophoresis
PC	Phosphatidylcholine
PDB	Protein Data Bank
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PL	Phospholipid
PS	Phosphatidylserine
PTM	Post-translational modification
RF	Radio frequency

RMSD	Root mean square deviation
SBP	Substrate binding protein
SBS	Substrate binding site
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SPR	Surface plasmon resonance
TA	Tail anchored
TEV	Tobacco etch virus
ToF	Time of flight
TMD	Transmembrane domain
UDM	n-Undecyl-β-D-Maltopyranoside
UHV	Ultra-high vacuum
WT	Wild type

CHAPTER 1

Background and methods for investigating membrane proteins

1.1 Introduction

Proteins represent a fundamental part of living organisms given their involvement in all the crucial cellular pathways. They regulate all aspects of life including, but not limited to, regulation of gene expression, catalysis of all biochemical reactions, signalling, transportation of molecules essential for cellular metabolism [1]. A common feature of these different tasks is the necessity for proteins of assembling in larger complexes, in order to finely regulate their functions. Moreover, another layer of protein function modulation is given by small molecule (substrates or cofactors) binding [2, 3], as well as post-translational modifications (PTMs) [4], which may affect both conformational states and oligomeric assembly of protein complexes.

Cell membranes are organised as lipid bilayers presenting many proteins that are either embedded (integral membrane proteins or simply membrane proteins) or associated (peripheral membrane proteins) to the inner or outer leaflet of the lipid bilayer (Fig. 1.1) [5]. Amongst different protein classes, membrane proteins, present a privileged location at the outermost cellular compartment, thus making them extremely important for cell integrity and internal homeostasis control [6].

In eukaryotic cells, plasma membranes act as dynamic barriers that not only determine cellular and organelle boundaries, but also play crucial roles in both physiological and pathological states given the multitude of their interactions [7]. In fact, membrane proteins are responsible for translocation of ions and small molecules (drugs, metabolites, neurotransmitters) and signal transduction, but also intercellular communication and attachment to the cytoskeleton and extracellular matrix. Their importance is proven by the fact that 20–40% of human genome encodes for membrane proteins [8, 9]. Furthermore, the majority of bacterial and plant cells are enclosed in rigid cell wall presenting different chemical composition and 3D-structure compared to lipid bilayers [10].

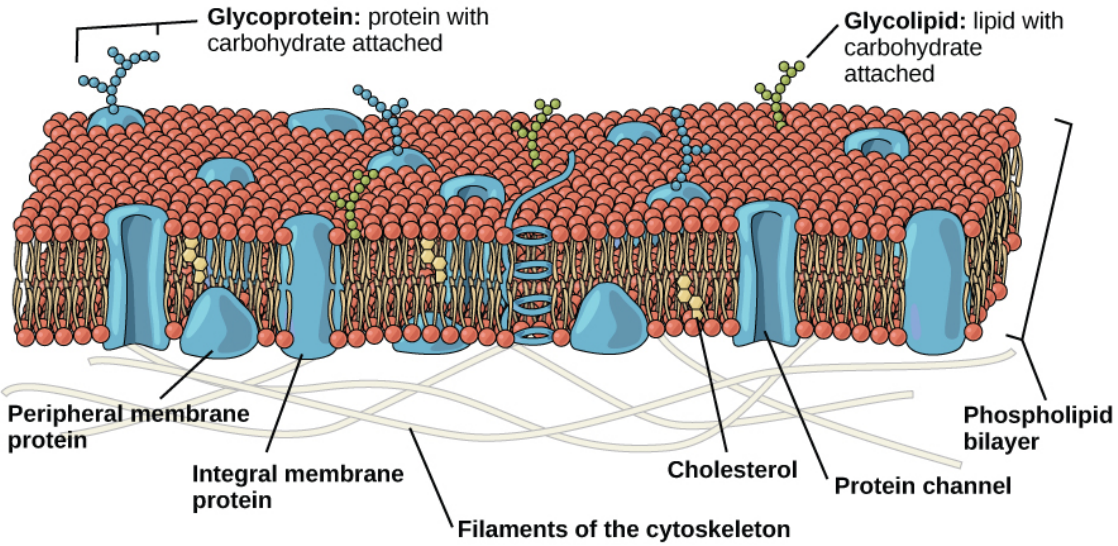


Figure 1.1 Cartoon representation of a eukaryotic cell membrane. The plasma membrane is organised as a lipid bilayer where proteins and lipids are dynamically interacting. Strong protein-protein and protein-lipid interactions introduce asymmetry within the lipid bilayer. Figure from OpenStax Biology [11].

1.1.1 Membrane protein complexes

Differently from soluble proteins, membrane proteins possess exposed hydrophobic residues that are in direct contact with the lipid environment. These residues are part of the transmembrane domains (TMDs) presenting both α-helix and β-barrel secondary structures [5]. TMDs are usually in contact with soluble domains that may be present on both sides of the membrane. These can be classified into cytoplasmic domains (CDs), extracellular domains (ECDs) or, in the case of bacteria, periplasmic domains [12]. Membrane proteins possess specialised structures related to their biological function and are classified as enzyme, receptors, channels and transporters [13] (Fig. 1.2).

An example of membrane-embedded enzymes is intramembrane proteases, that catalyse the hydrolysis of peptide bonds of peptide and protein species present in the cell membrane. These enzymes play important roles in peptide signalling and lipid metabolism as well as response to unfolded protein accumulation and cellular differentiation [14].

The largest family of receptors present in eukaryotic systems is represented by G

protein–coupled receptors (GPCRs). They share a seven trans-membrane helices structure and are responsive to a diverse range of stimuli comprising small molecules, peptides, but also other proteins and light. These stimuli induce conformational changes that determine the interaction with heterotrimeric G proteins (Gαβγ) or arrestins, thus initiating the downstream signalling cascade [15]. Interestingly, membrane lipids are also involved in determining G protein coupling [16], as well as mediating GPCRs oligomeric state [15].

Ion channels facilitate the passage of ions across cell membranes, thus enabling signal transmission and cellular homeostasis maintenance. According to their substrates ion channels are classified as ligand-gated, voltage-gated or mechanosensitive channels [17]. Similarly to what described for GPCRs, recent papers indicated lipid modulation for ion channels [18-20].

Transporters are responsible for the translocation of molecules through membranes and are powered by potential gradient of counter ions or ATP hydrolysis. Even in this case substrates range from ions, to small molecules (both hydrophilic, like carbohydrates or amino acids, and hydrophobic, like drugs or lipids) [21, 22] and large lipids such as lipopolysaccharides [23-25].

Given their multifaceted roles in human and microbial cell regulation, membrane proteins represent an attractive target for drug discovery, in fact they represent the target of more than 60% of currently marketed therapeutics [26, 27]. For instance, serotonin receptors 5HT1A and 5HT2A are GPCRs involved in neurological and psychiatric disorders like schizophrenia, depression, and Parkinson's disease and new drugs are being currently developed for these indications [15]. In the case of bacteria and cancer cells, multidrug efflux proteins are responsible for multidrug resistance development [28-30]. Moreover, membrane proteins play important roles in viral infections. In fact, some viruses present membrane proteins on their envelopes [31], while others have spike glycoproteins that bind eukaryotic membrane proteins to promote entry into cells [32].

A common feature shared by all membrane proteins is their relationship with the membrane environment, the complexity of which varies according to location, function, and diversity of the lipids within membranes. In the pursuit of clarifying the processes governing life,

understanding the structure and functions of membrane proteins is critically important. However, the analysis of membrane proteins has been historically challenging due to their inherent properties

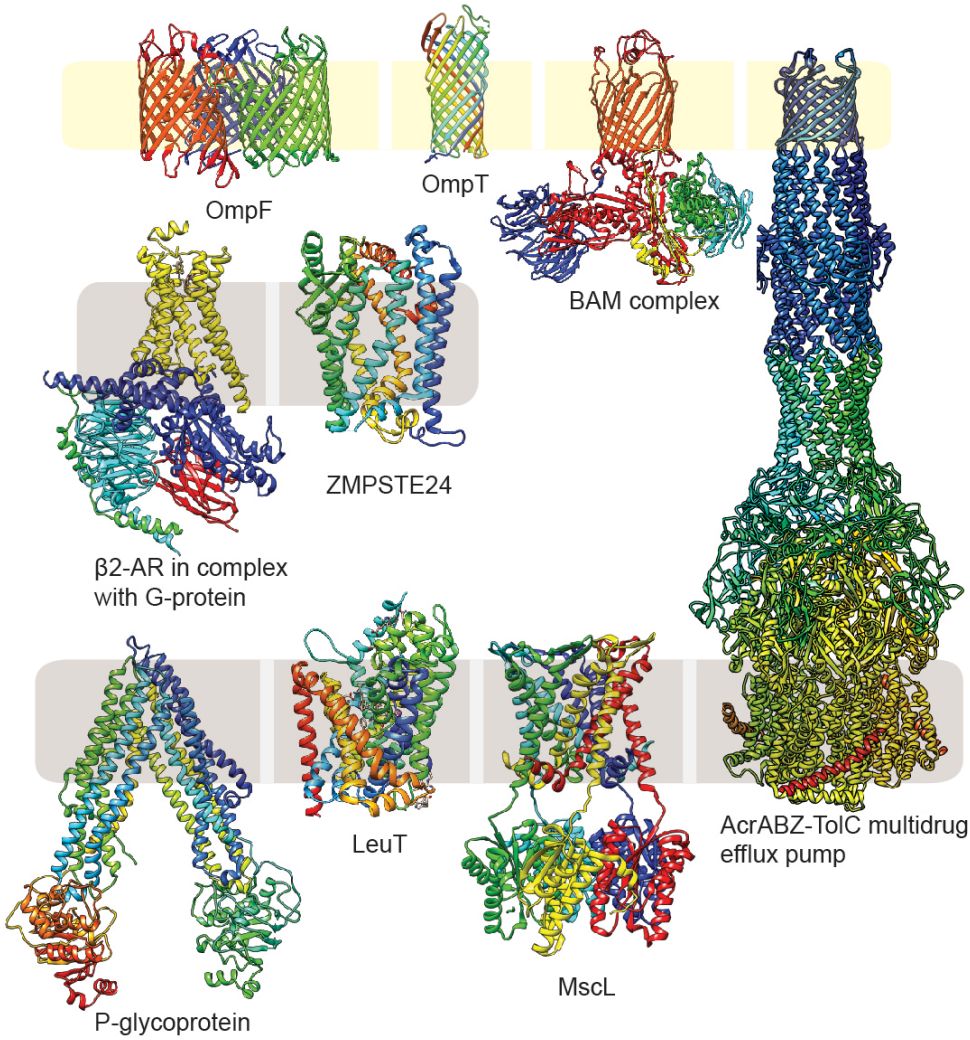


Figure 1.2 Representative examples of membrane proteins. AcrABZ–TolC (PDB ID: 5O66) [33] and P-glycoprotein (PDB ID: 4KSC) [34] are multidrug efflux transporters. LeuT (PDB ID: 3F3A) is a leucine transporter [35]. MscL (PDB ID: 4Y7K) is an example of a mechanosensitive channel [36]. ZMPSTE24 (PDB ID: 4AW6) is an example of a membrane-embedded metallopeptidase [37]. β2-adrnergic receptor (β2-AR) is an example of a GPCR and is shown complex with its cognate G protein (PDB ID: 3SN6) [16]. OmpF porin (PDB ID: 3K1B) [38], OmpT protease (PDB ID: 1I78) [39], and the BAM complex [40] (PDB ID: 5LJO) are examples of β-barrel bacterial outer membrane proteins. Figure adapted from Bolla *et al.* [41].

Firstly, while overexpression of soluble proteins usually results in high yields, in the case of membrane proteins, protein folding happens in the membrane, therefore the amount of protein that can be expressed is restricted by the membrane itself, as well as by protein folding and transport pathways that are limited in the case of overexpression [42]. Thus, the first bottleneck is obtaining enough protein material for characterisation [43, 44]. Secondly, the intrinsic flexibility of membrane proteins is another variable influencing the obtainment of atomistic details from X-ray crystallography and cryogenic-electron microscopy (cryo-EM) [45]. Finally, given the requirement for a hydrophobic environment, membrane proteins present low solubility and need a suitable amphipathic solubilisation system to preserve their structure, stability, and functionality [41, 46, 47]. Historically, detergents have been the most commonly used method to solubilise cell membranes and purify membrane proteins [48, 49]. Chemically, they are characterised by a polar head group and a hydrophobic chain, similarly to lipids. The presence of only one hydrophobic chain and their consequent conical geometry leads to the formation of micelles rather than bilayers, as in the case of membrane lipids. The micelle is formed when the detergent concentration reaches the critical micelle concentration (CMC) and is in equilibrium with individual detergent molecules. In the case of membrane proteins, the hydrophobic chains of detergent are oriented towards the transmembrane domains, thus creating an energetically favoured environment shielding the region from the aqueous environment [50]. Although detergent solubilisation lowers the sample heterogeneity, lipids may still co-purify with membrane proteins [47, 51], thus increasing the complexity of the system.

Although detergents are extensively used for purification and biophysical analysis of membrane proteins, cautions should be taken because micelle environments cannot fully replicate the native lipid bilayer environments, and may not be amenable for all purposes.

1.1.2 Membrane lipids

Lipids are amphipathic molecules that are organised in bilayers in biological membranes, accounting for up to 50% of their mass [12]. According to their structure, they are classified as

glycerophospholipids, sphingolipids, sterols, and glycolipids [52] (Fig. 1.3). Glycerophospholipids, or simply phospholipids (PLs), represent the most abundant class in biological membranes. All PLs share a phosphatidic acid (PA) moiety consisting of a glycerol backbone esterified with two fatty acid chains on positions 1 and 2 and bearing a phosphate group in position 3. PAs are biosynthetic precursors of many membrane lipids presenting different substituents on the phosphate group (Fig. 1.3a). These lipids include: phosphatidylglycerol (PG), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylcholine (PC), phosphatidylinositol (PI), and phosphorylated PI derivatives phosphatidylinositol bis- and tris-phosphate ($PI(4,5)P_2$ and $PI(3,4,5)P_3$). Cardiolipin (CDL) is a PG derivative where the two hydroxyl groups of phosphate-bound glycerol are esterified with fatty acids [41].

Sphingolipids are sphingosine derivatives, an amino alcohol that contains an unsaturated hydrocarbon chain (Fig. 1.3b). In sphingolipids, the amino group of the sphingosine backbone forms an amide bond with a fatty acid chain, while the primary hydroxyl group of sphingosines is esterified to phosphorylalcohol. For instance, esterification of the 1-hydroxy group with phosphorylcholine yields sphingomyelin [53].

Glycolipids consist of lipids with a carbohydrate attached by a glycosidic bond. In eukaryotic cells the sugar moiety is a short, branched oligosaccharide with fewer than 15 units (Fig. 1.3c). Their main function is signalling: they serve as markers for cellular recognition, and are crucial for transmitting growth and division signals [54]. For these reasons, they are key players in critical pathways like immune response [55].

Sterols are composed of four fused rings: three cyclohexane and one cyclopentane ring, with a hydroxyl group on the first cyclohexane available for esterification (Fig. 1.3d). Cholesterol is one of the predominant sterols present in higher eukaryotes but is absent from prokaryotes [56]. Given its and rigid structure of cholesterol, it is able to modify the fluidity of a membrane in a concentration-dependent manner [57].

Given the presence of the phosphate group most of PLs are zwitterionic or anionic the cell membrane is overall negatively charged at physiological pH. In addition, PLs head groups

determine the geometry of the lipids, thus influencing the packing and curvature of the lipid bilayer, while tails contribute to membrane topology, permeability ad fluidity [58, 59].

Lipid-lipid and protein-lipid interactions are determined by both head groups and tails [38, 59, 60]. In particular, hydrophobic interactions are the main drive for accumulation of certain lipids in microdomains, or lipid rafts [61, 62]. Moreover, the relative abundances of lipids are different amongst species but is also cell-dependant. For instance, PC represents ~40% of membrane lipids in mammals, while only ~10% in yeast. Furthermore, while abundance CDL is low on the plasma membrane, its concentration increases in mitochondrial membranes [53].

Interestingly, bacteria lack of many lipids present in eukaryotic cells (PC, PI, sphingolipids and cholesterol) and possess a completely different lipid distribution, as described in the next section.

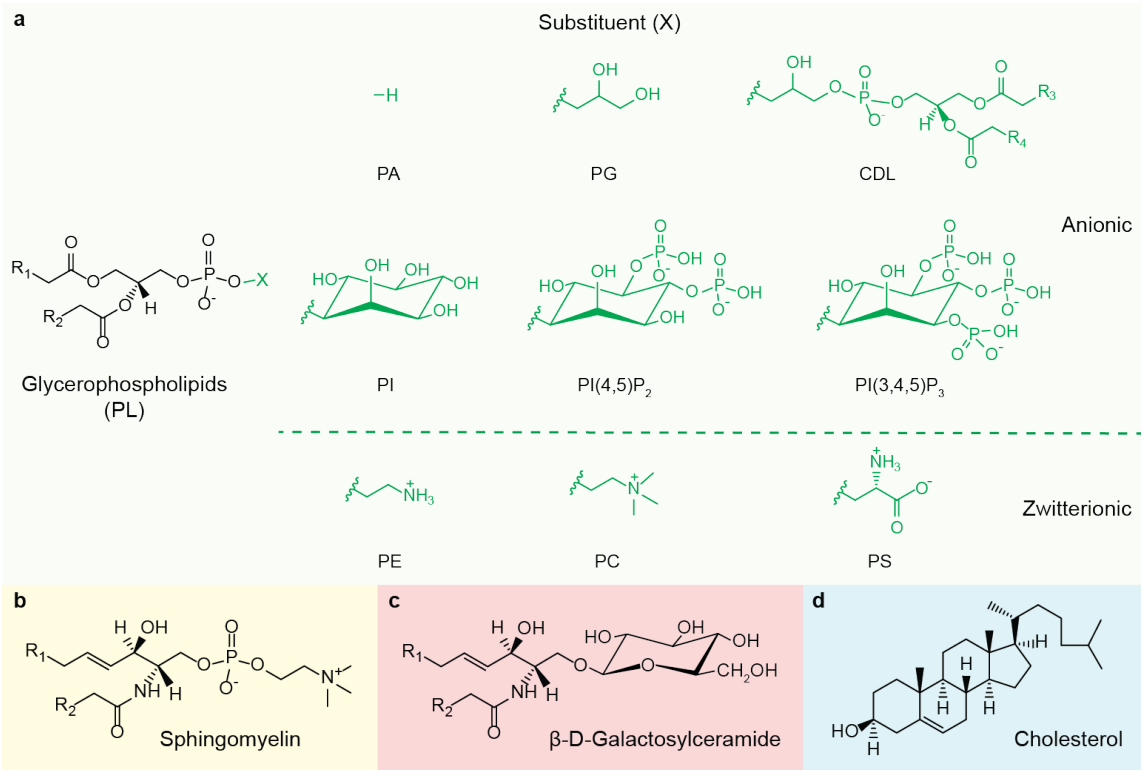


Figure 1.3 Structures of membrane lipids. (A) According to their overall charge glycerophospholipids (PLs) are classified as anionic (phosphatidic acid (PA), phosphatidylglycerol (PG), cardiolipin (CDL), phosphatidylinositol (PI), and its derivatives ($PI(4,5)P_2$ and $PI(3,4,5)P_3$)) and zwitterionic (phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylserine (PS)). (B) Sphingomyelin is an example of

sphingolipid. (C) β-D-galactosylceramide is an example of glycolipid. (D) Cholesterol, the most abundant sterol in mammals.

1.1.3 Bacterial membrane and lipopolysaccharide

Classically, bacteria are classified as either Gram-positive or Gram-negative based on the Gram staining technique, a differentiation based on cell wall properties. Gram-positive bacteria are characterised by a cytoplasmic membrane and a multi-layer peptidoglycan cell wall, whereas Gram-negative bacteria have two distinct membranes, inner and outer membrane (IM and OM, respectively), a thin peptidoglycan cell wall between them [63, 64] (Fig. 1.4a).

Escherichia coli, a Gram-negative bacteria species, has been the standard model for microbiology and its membrane properties have been deeply investigated. Three major membrane phospholipids have been reported: the zwitterionic PE (about 75% of membrane lipids), and the anionic lipids PG (about 20%) and CL [65]. Furthermore, Gram-negative bacteria also have a glycolipid embedded on the outer leaflet of their outer membrane (OM): lipopolysaccharide (LPS). Roughly 2×10^6 LPS molecules are present on Gram-negative bacteria outer membrane cell surface, covering ~75% of the cell surface [66]. Different LPS molecules present a conserved lipid A moiety which is glycosylated with a core oligosaccharide (core OS), which may be attached to a long, branched O-antigenic polysaccharide (O-PS, Figure 1.4b). Preservation of lipid A and inner core OS structure along different species is likely related to their key roles in preserving the integrity and permeability properties of the OM, which can be altered with the introduction of subtle changes [67]. Differently, O-PSs are variable in length and branching, which is thought to be connective to selective environmental pressure, such as host adaptive immunity and O-PS-specific bacteriophages [68, 69]. In fact, O-PS is recognised by the host immune system, thus triggering immune response, but is also associated with evasion from complement-mediated killing [70]. O-PS may contain up to 200 sugar moieties some of them bearing phosphate groups which interact with bivalent ions (for instance Mg^{2+}), thus allowing LPS molecules to create a tight network [71]. O-PS form a hydrophilic barrier

surrounding the OM which obstructs the entrance of antibiotics, which are mainly characterised by a hydrophobic structure [72].

Lipid A and the core OS are independently synthesised in the cytoplasm and transported to the periplasmic side of the IM trough two distinct transporters, MsbA [23] and Wzx [73], respectively. Following ligation and addition of O-PS on the IM, full LPS is transported to the outer membrane through a multi-protein complex consisting of 3 main sub-complexes: the IM ABC transporter $LptB_2FGC$, the periplasmic oligomeric complex $(LptA)_n$ and the OM LPS insertase LptDE [71]. The elucidation of LPS translocation mechanism from the IM to the OM will be the focus of Chapters 3 and 4.

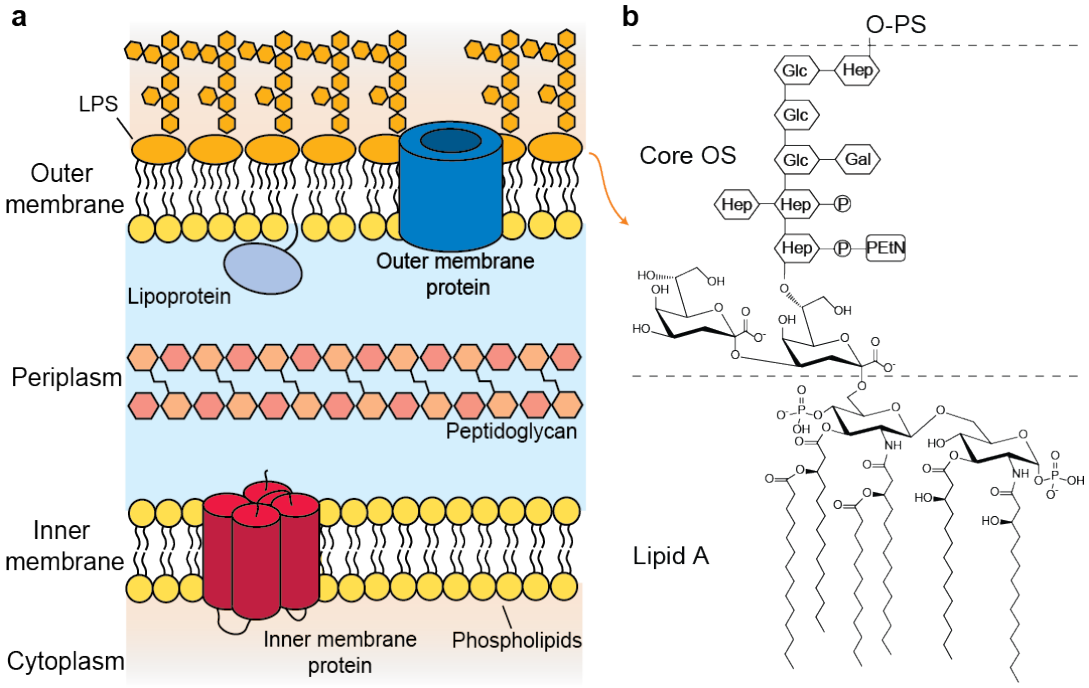


Figure 1.4 Gram-negative bacteria membrane. (A) Schematic of Gram-negative bacteria membrane organisation. The inner membrane is surrounded by a thin layer of peptidoglycan and an outer membrane having LPS on the external leaflet. (B) Structure of LPS: Lipid A is glycosylated with two keto-deoxyoctulosonate moieties (Kdo), which represent the minimal portion of the core OS. The O-PS is a highly variable polysaccharide with antigenic properties.

1.1.4 Protein-lipid interactions

As previously explained, small molecules (drugs, substrates, lipids) may modulate the structure and function of proteins. Ligand binding induces conformational changes that affect the interactions with partner proteins, thus leading to different downstream cellular processes. An example of the effect of substrate binding on protein-protein interactions is depicted in Chapter 2, where the interaction between bacterial ABC importers and their cognate substrate binding proteins is elucidated.

Since membrane proteins are surrounded by lipids their function is heavily affected by membrane composition. Lipids interacting with membrane proteins are classified as bulk, annular or non-annular based on their affinity and residence time [74]. Bulk lipids diffuse rapidly within the membrane with short residence time at the protein interface and do not interact with proteins. Annular lipids are in direct contact with the hydrophobic surface of membrane proteins. Although the annular lipid shell is in continuous with the bulk environment, their motions are restricted by tighter protein-lipid interactions. These lipids are important elements for the structural integrity, dynamics and activity of membrane proteins [41, 75-77]. Non-annular lipids are high affinity ligands that remain tightly bound to membrane proteins, even following purification in detergent micelles. Numerous studies indicated that lipids may act as agonists [78, 79], allosteric modulators [16, 38, 80-82], and can also mediate oligomeric state [35, 83], stability [84, 85], and translocation within the cell [81, 86].

Furthermore, some lipids are also substrates of membrane proteins such as lipid flipases, which translocate their lipid substrates from one leaflet of the bilayer to the other. For instance, the *E. coli* flippase MurJ transfers the peptidoglycan precursor lipid II from the inner to the outer leaflet of the OM, and is modulated by the non-annular lipid CDL [87]. Another example of membrane proteins with lipid substrate are the previously mentioned protein complexes involved in LPS transport from the IM to the OM: the ABC transporters MsbA [23] and $LptB_2FGC$ [88-90], and the OM translocon LptDE [91, 92].

1.2 Mass spectrometry

A wide range of techniques has become available for the investigation of protein complexes structure and modulation. These include classic structural biology methods such as X-ray crystallography, nuclear magnetic resonance (NMR), and the recently developed cryo-EM. Biophysical methods like isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), and microscale thermophoresis (MST) are employed to analyse and quantify biomolecular interactions, such as protein-protein, protein-DNA, protein-small molecule, and protein-lipid interactions.

X-ray crystallography, NMR and cryo-EM provide valuable information on the structure and interactions of protein complexes. However, each of them still presents limitations and challenges. For instance, both crystallography and NMR require large sample amounts, which represents a major drawback in the case of membrane proteins. In addition, while NMR is capable of characterising about protein dynamics, X-ray crystallography provides static structures of the protein in a single conformation [47, 93, 94]. On the other hand, X-ray crystallography offers unparalleled resolution, and does not have size limitations, while in the case of NMR the analysis of large macromolecules is particularly difficult, and a further challenge is introduced by sample heterogeneity of membrane proteins that leads to spectral crowding [94, 95]. Recent advances in instrumentation and software has allowed to obtain higher resolution for cryo-EM [96], however data acquisition is still not as automated as X-ray crystallography [95] and sample freezing may cause breaking of protein complexes interactions. Nonetheless, fast progress is being made to develop better data analysis softwares for cryo-EM, and the introduction of newer instruments may help extending this technique to smaller macromolecules and obtain atomic resolution comparable to crystallography [97].

ITC, SPR and MST are highly sensitive techniques to quantify binding affinities, and even kinetics in the case of SPR. While ITC is an in solution label-free technique, SPR necessitates immobilisation of one of the binding partners onto a surface and MST requires fluorescent labelling. On the other hand, ITC is highly sample consuming [98-100]. Furthermore,

since all these techniques provide information on the average thermodynamic properties of the system, none of them may inform on the influence of single ligand events or their influence on oligomeric states. Furthermore, these methods do not provide information regarding co-purified ligands, such as lipids in the case of membrane proteins [41, 101].

Along with improvements and breakthroughs of the previously described biophysical methods, each of them presenting their own advantages and drawbacks, other complementary techniques can provide valuable information [102-104]. Here I will focus on mass spectrometry (MS) and its ability to complement and inform these techniques by providing invaluable insights on protein complexes, particularly the more challenging membrane proteins, that are not accessible by other methods.

Ever since the observation of the first molecules in the gas phase, MS has been an indispensable technique for the observation of relatively small molecules (low thousands of Daltons) and has become an indispensable tool for proteomics, lipidomics, and metabolomics [105, 106]. The introduction of soft ionisation techniques and improvements in mass analysers and detectors enabled the transmission and detection of intact proteins. Further developments, including manipulation of pressure gradients within the instruments, allowed the preservation of non-covalent interactions in the gas phase [107, 108]. These discoveries led to the rise of a new technique, non-denaturing or native MS (nMS), that provides insights on subunit stoichiometry and composition of protein complexes, protein-ligand interaction, and analysis of heterogeneous samples. Moreover, instrument developments (such as high energy, ultra-high mass range mass spectrometers) [109-112] and the optimisation of solubilisation methods [113, 114] of membrane proteins have proven to be vital for the observation of intact membrane protein complexes in a mass spectrometer.

Another MS-based technique that can be coupled to nMS to obtain structural information on protein complexes and protein-ligand interactions is hydrogen/deuterium exchange mass spectrometry (HDX-MS). This technique will be further discussed later in this chapter.

1.2.1 Electrospray ionisation

The first step of MS analysis analyte ionisation and its transfer to the gas phase in the ions source. Molecules are ionised through protonation, deprotonation, formation of adducts with ions present in solution, or via injection/capture of electrons. Currently, different methods allow the formation of ions and are classified depending on their physicochemical mechanisms and the amount of energy transferred during the ionisation process. To preserve non-covalent interaction between the different subunits of protein complexes, "soft" ionisation techniques that can ionise high molecular weight molecules are required [115]. To date, the most widely employed techniques are matrix-assisted laser desorption ionisation (MALDI) [116, 117] and electrospray ionisation (ESI) [118]. Other soft ionisation approaches that are less routinely applied for nMS are desorption electrospray ionisation (DESI) [119, 120] and laser-induced liquid beam ion desorption (LILBID) [121]. Since ESI was the soft ionisation technique employed for all projects described in this dissertation, I will discuss the physico-chemical principles in the next section.

In ESI, a small volume of analyte solution is injected into a conductive capillary needle and inserted into a holder, where it possesses a weak flux. This set-up forms the source of the instrument, which is separated from a counter electrode by 0.3-2 cm. Typically, volatile buffers such as ammonium acetate or ammonium bicarbonate are used to increase the ionisation efficiency and reduce salt adduct formation [122]. A strong electric field is applied at atmospheric pressure by applying a potential difference (0.8-1.8 kV, either negative or positive depending on the analyte and type of experiment) across the source and the counter-electrode [123]. The liquid surface at the end of the capillary then accumulates charge and when the electrostatic forces are balanced by the surface tension (also known as the Rayleigh limit – eq. 1.1), the meniscus is distorted into a 'Taylor cone' [124].

$$z_R e = 8\pi(\varepsilon_0 \gamma R^3)^{\frac{1}{2}} \quad (1.1)$$

Where z_R is the charge on the droplet surface, ε_0 is the electrical permittivity, γ is the solution surface tension and R is the radius of the droplet.

Droplets are ejected from the Taylor cone apex as the potential further increases and the Rayleigh limit is exceeded, when the electrostatic forces exceed the surface tension and the solution-air interface is destabilised. Solvent evaporates from the droplets, charge-density increases at the surface, and coulombic repulsion exceeds surface tension so that each primary drop explodes to produce smaller droplets (also known as droplet fission, Fig. 1.5) [124]. It is estimated that offspring droplets have ~2% volume of parent droplets but ~15% charge, thus the charge per unit of volume increases by a factor of ~7.5 [125]. This process is cascaded and repeats to produce smaller droplets until gas-phase ions are produced. Further evaporation is also aided by collision with neutral gas molecules and helps to increase surface electric field and leads to ions desorption from the highly-charged droplets and entrance of the gas-phase ions in the mass spectrometer.

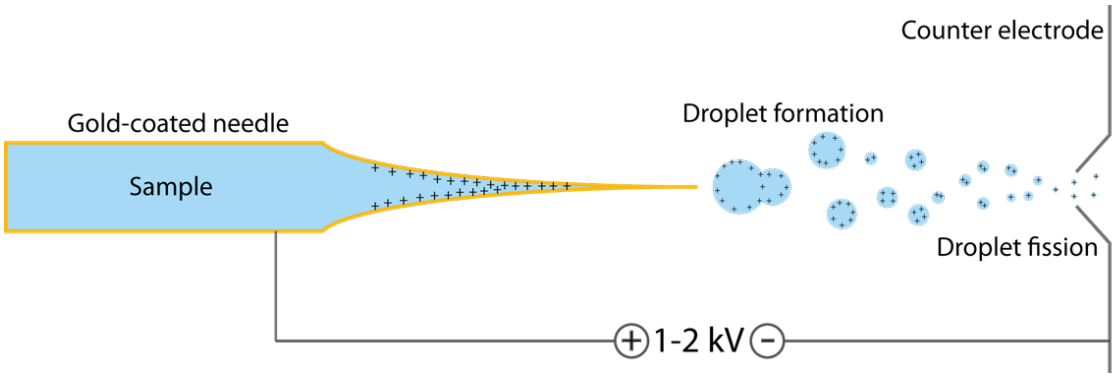


Figure 1.5 Schematic of the electrospray process. Sample solution is loaded into a conductive capillary (gold-coated needle) held at high voltage relative to the downstream electrode. Ion accumulation leads to the formation of Taylor cone. Following fission and evaporation events, charged ions of the macromolecule are produced and transmitted into the mass spectrometer for analysis.

Nano-electrospray ionisation (nESI) is a sub-type of ESI developed by Wilm and Mann in 1994 [126]. The principles underlying nESI are the same as ESI with the main differences being in the inner diameter of capillary needle orifice (1-5 μm in nESI, while in ESI it reaches up to 0.5 mm) and very slow flow rates, in the range of 10-50 nL/min [127]. These changes allow for low sample volumes, typically 1-3 μl, thus massively decreasing sample consumption, which

represents a great advantage when sample amount is limited, like in the case of biological samples. Furthermore, nESI is more amenable for studying protein complexes as it is a softer ionisation method compared to ESI. In fact, given the smaller orifice the initial droplet size is at least one order of magnitude smaller (150–200 nm), thus reducing the energy and the number of fission events necessary to yield gas-phase ions [115]. Furthermore, spectra obtained using nESI are characterised by better resolution and higher salt tolerance. The decreased number of fission events leads to less salt concentration, thus yielding spectra with less salt adducts, improving both resolution and signal intensity [115, 128]. One limitation of ESI partially overcome by nESI is incomplete macromolecule desolvation, which contributes to peak broadening and lower resolution. However, for high mass complexes, this might still be observed in nESI, where the mass increase can be up to 1% [129].

1.2.2 ESI process mechanistic models

The mechanism by which protein desorption occurs has been under debate over the years and three models have been proposed: ion evaporation model (IEM), charge residue model (CRM), and chain ejection model (CEM).

According to the IEM, emitted droplets contain pre-charged ions and the variables involved in the process are the electric field on droplet surface, and the chemical properties of the ions, comprising the molecular structure [130] (Fig. 1.6a). Following solvent evaporation, the droplet shrinks and the charge density increases until the electric field at the surface overcomes solvation forces, thus ejecting the ions from the droplet [131]. This mechanism can explain the charge state distributions of small molecules, however for larger ones (> 1000 Da) it seems to deviate from experimental evidence since it underestimates their charge state [132].

The CRM is generally accepted to explain the ionisation process of large macromolecules, such as folded proteins [131, 133] (Fig. 1.6b). According to this model, the final charge state of the analyte is independent on the molecule's charge state in solution but rather on the droplet size and the mass of the molecule. This model proposes that the droplets

undergo multiple evaporation and fission cycles until they approximately reach the size of the analyte. Further evaporation removed neutral solvent molecules to finally generate the gas-phase charged ion. The final charge on the molecule is closely related to the Rayleigh limit (thus to the diameter of the droplet) and the size of the molecule (which, for proteins, is roughly proportional to their mass) [134]. This model was confirmed by experimental observations for a diverse range of proteins, regardless their surface chemistry [135].

The third model, CEM, better applies to disordered polymers with solvent-accessible hydrophobic residues, such as unfolded proteins [133, 136] (Fig. 1.6c). Since the exposure of hydrophobic regions to charged solvent is energetically unfavourable, unfolded protein termini migrate to the surface and the protein is eventually expelled because of Coulombic repulsion. Given the higher surface exposure of unfolded proteins compared to folded ones, they carry a higher and broader charge state distribution.

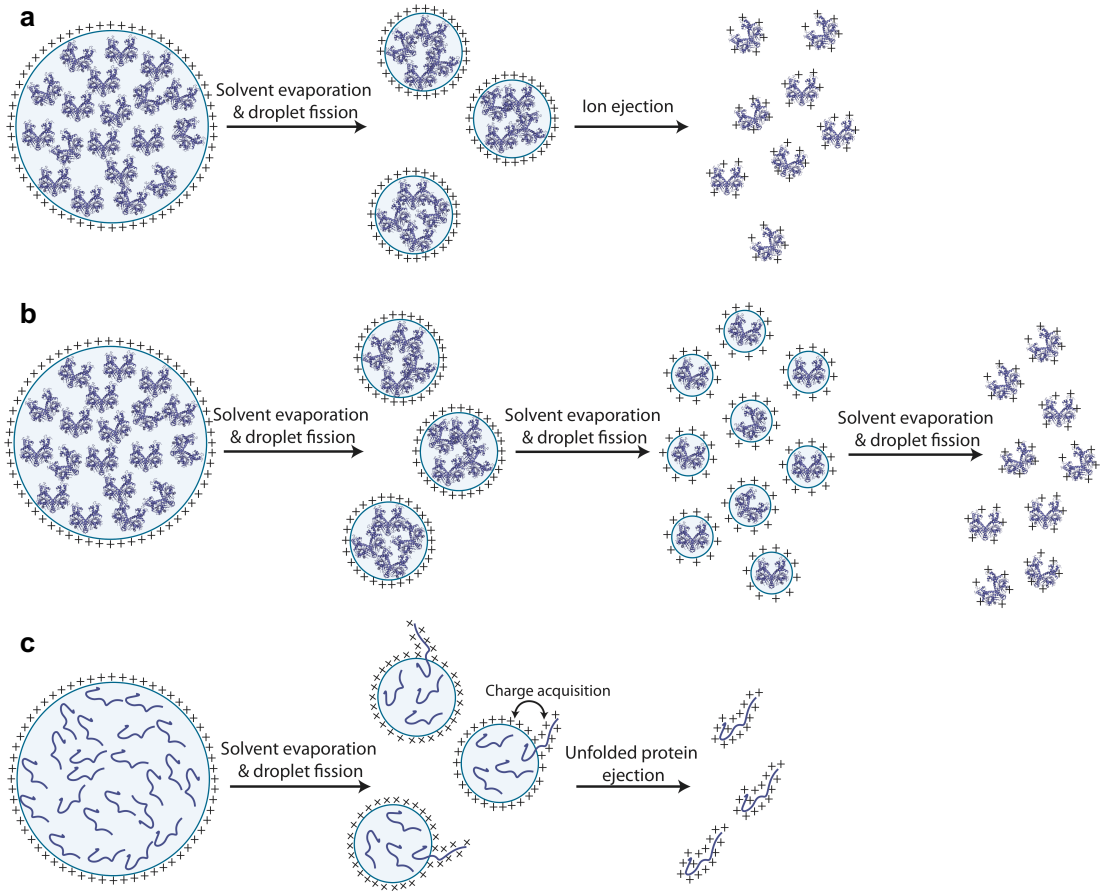


Figure 1.6 Models of electrospray ionisation. A) Ion evaporation model (IEM). B) Charge residue model (CRM). C) Chain ejection model (CEM). The protein illustrated in A) and B) is *S. pombe* Get3 (PDB ID: 2WOO).

1.2.3 Mass analysers

The main ability of mass spectrometry is to separate and analyse ions according to their mass to charge ratio (*m/z*). A wide range of mass analysers have been developed which discriminate ions on the basis of their flight time, trajectory stability, kinetic energy, resonance frequency and momentum. Each mass analyser has different feature and the most important factors to consider are resolution, mass accuracy, *m/z* range, as well cost. For the aim of this dissertation, I will describe three types of mass analysers: quadrupole, time-of-flight, and orbitrap mass analyser.

1.2.4 Quadrupole mass analyser

Quadrupole mass analysers are composed of four parallel rods (cylindrical or hyperbolic), with two sets of opposing pairs that are connected diagonally (Fig. 1.7). They separate ions according to their trajectory stability within an oscillating electric field (Φ_0). This is generated by applying opposite voltages to the rod pairs, in particular these voltages are composed of a direct current (DC) component of magnitude *U* and an alternating current (AC) component of magnitude *V* at a radiofrequency (RF) *f*, as indication by equation 1.2 [137]:

$$\Phi_0 = \pm (U - V cos\omega t) \tag{1.2}$$

The quadrupolar electric field deflects the ions which travel following helical trajectories following the Mathieu equation. The trajectory is determined by the inner diameter distance between rods (r_0), the DC potential U and the RF amplitude V. Thus, for a quadrupole, assuming constant r_0 and RF angular frequency ($\omega = 2\pi f$), the x and y positions of an ion during a certain time frame (*t*) can be determined as follows:

$$\frac{d^2u}{d\xi^2} + (a_u - 2q_u cos2\xi)u = 0 \quad (1.3)$$

$$a_u = a_x = -a_y = \frac{8zeU}{m\omega^2 r_0^2} \quad (1.4)$$

$$q_u = q_x = -q_y = \frac{4zeV}{m\omega^2 r_0^2} \quad (1.5)$$

Where u may be *x* or *y* as it is the displacement, a_u and q_u are additional stability parameters, and $\xi = \frac{\omega t}{2}$. This equation establishes a relationship between the position of the ion and time. If the *x* and *y* values of an ion remain lower than r_0, then it will be successfully transferred through the quadrupole [138].

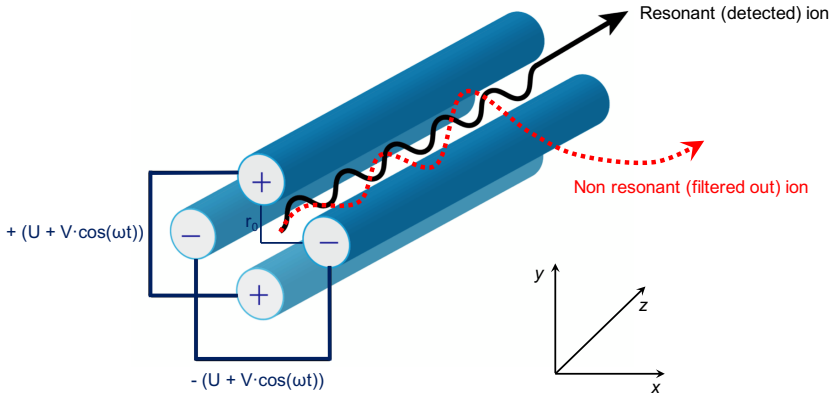


Figure 1.7 Schematic of the quadrupole mass analyser. Opposing rods are electrically connected (applied voltages as shown) and 180° out of phase with respect to the adjacent ones. Resonant ions (black) are stable within the field are transmitted, while non-resonant ones (red) have unstable trajectories and collide to the rods.

Quadrupoles may be employed in both RF-scanning mode and RF-only mode. In RF-scanning mode RF frequency is kept constant, while DC voltage and RF amplitude are modulated to allow transmission of a certain *m/z* range. By plotting a_u against q_u one can outline areas of stability (Fig. 1.8a) on each dimension. The overlay of these two stability graphs (Fig. 1.8b) allows to find areas of stability in the (x, y) plane. In these areas, U and V are such that x and y do not exceed r_0, i.e. that the ions do not clash on the rods. The rearrangement of equations 1.4 and 1.5 for *U* and *V* and plotting *U* against *V* gives proportional triangles for ions with different

m/z ratios (Fig. 1.8c) Thus, a constant *U/V* ratio (red line in Fig. 1.8c), that crosses through different *m/z* stability areas allows detection of ions with different masses [139]. Introducing a high *U/V* ratio improves the resolving power (the peaks of the triangles do not overlap). In RF-only mode the DC voltage is set to 0 and the quadrupole acts as ion guide. This allows to focus and transmit ions in a wide m/z range and the maximum *m/z* that can be transmitted is described by equation [140]:

$$\left(\frac{m}{z}\right)_{max} = \frac{4 \times 10^6 V}{f^2 r_0^2} \tag{1.6}$$

It is worth noticing that ions with large *m/z* and mass experience poor focusing and smaller ions present unstable trajectories, causing both types of analyses to crash into the rods. Quadrupoles can be turned into linear ion traps by adding a potential to end electrodes, to give a confining potential within the axial direction [141].

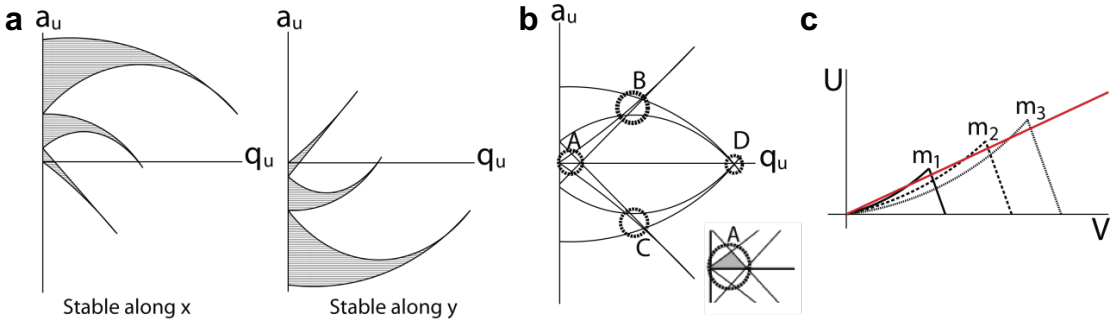


Figure 1.8 Ions stability based on the Mathieu equation. A) Areas of stability defined by a_u and q_u along the x- and y-axes. B) The overlap of plots obtained in A) yields the areas of stability on the (x, y) plane. C) Change in m/z causes a change in the stability area on the U-V space, producing proportional triangles.

1.2.5 Time of flight mass analyser

Time of flight (ToF) analysers record the time ions take to drift in a field-free flight tube. They are composed of two parts: the pusher, that accelerates the incoming ions, and the field-free region that terminates with the detector [142]. ToF analysers present different geometries, but for clarity the principles of ToF will be explained using the linear ToF as an example (Fig.

1.9a). A packet of ions is accelerated in the direction of the flight tube by applying a potential (V), thus as the velocity (v) of the ions is given by equation 1.7 [143]:

$$v = \left(\frac{2Vz}{m}\right)^{1/2} \tag{1.7}$$

Since the length of the drift tube (L) is known, the time that each ion takes to cross the field-free region is converted into *m*/*z* value by rearranging equation 1.7 into equation 1.8:

$$\frac{m}{z} = \frac{2Vt^2}{L^2} \tag{1.8}$$

Since there is a direct proportion between *m*/*z* and t^2, theoretically there is no upper mass limit [144]. However, from a practical point of view it is worth noticing that the transmission of larger ions is limited by the dispersion of the ion beam. Moreover, there are further deviations from theory since ions arriving at the pusher have a distribution of kinetic energies, thus contributing to resolution decrease since isobaric (same *m*/*z*) ions arrive at different times [145]. To overcome this issue, the ToF mass analyser can be equipped with a reflecton, consisting of a series of lenses with fixed electrostatic field positioned orthogonally with respect to the beam axis (Fig. 1.9b). The reflecton serves as an ion mirror where ions with higher kinetic energy penetrate more into the lenses, thus compensating the initial difference in kinetic energy with a higher flight time and improving the resolution [145].

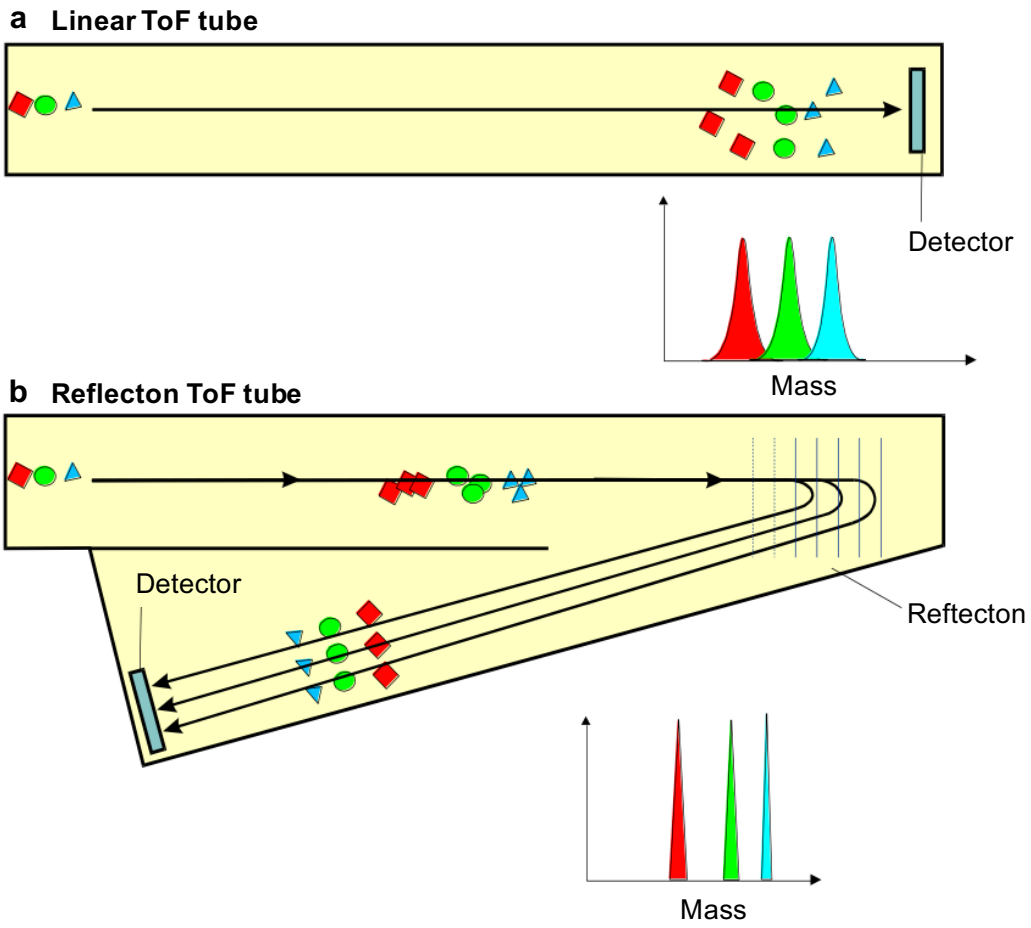


Figure 1.9 Schematic of linear vs reflecton ToF. A) In the linear ToF configuration ions with the same *m/z* may experience different extraction pulses due to their different initial potential energies leading to slightly different arrival times and a reduction in resolution (broadening of the peaks). B) With the addition of reflecton ions with higher kinetic energy penetrate further in the lenses, thus having a longer path that compensate for the initial difference in kinetic energy. This helps refocussing of the ions thus increasing the resolution. Figure adapted from [146].

1.2.6 Orbitrap mass analyser

The Orbitrap is an ion trap invented by Makarov [147], as a development of the Kingdon trap [148], using the principle of orbital trapping. The Orbitrap consists of an external barrel-like electrode and an internal coaxial spindle-shaped electrode, across which a potential difference is applied (Fig. 1.10). The ions are orbitally trapped purely by electrostatic fields [149], and the potential is described by the following equation:

$$U(r,z) = \frac{kz^2}{2} - \frac{kr^2}{4} + \frac{k}{2}(R_m)^2 ln\left[\frac{r}{R_m}\right] + C \quad (1.9)$$

where k is the field curvature, r is the radial coordinate, R_m is the characteristic radius, and C is a constant. Prior to the entrance in the Orbitrap, ions are focussed in a C-trap and then injected perpendicular to the z-axis. Ion packets follow trajectories around the spindle that can be described as two independent motions: a radial orbital motion around the electrode and a harmonic oscillation in the z-direction. The frequency of the harmonic motion (ω) is independent of the energy and spatial separation of the ions [150] and is related to the *m/z* ratio through the following equation:

$$\omega = \left(k\frac{z}{m}\right)^{\frac{1}{2}} \quad (1.10)$$

As explained before, ions are injected in the orbitrap as packets, thus effects of difference kinetic energies for isobaric ions and spatial spread on the measurement are minimal [149]. Moreover, since the oscillating frequency can be measured multiple times the mass accuracy and sensitivity are further increased.

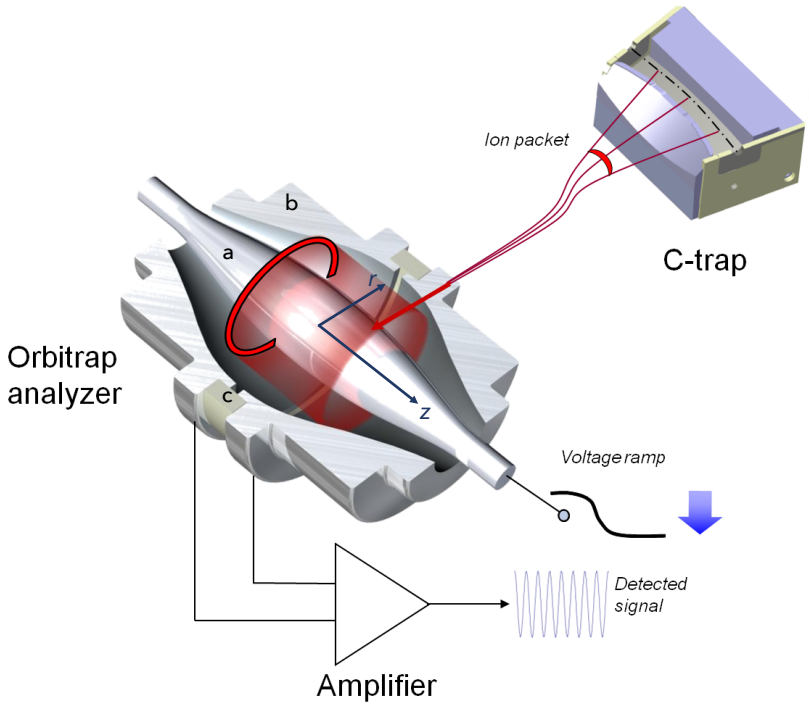


Figure 1.10 Schematic of the Orbitrap mass analyser. Ions are focussed in the C-trap and the injected orthogonally in the Orbitrap analyser. Ions orbit around the central electrode (a) and the oscillations around the z axis are independent of initial kinetic energy. The outer electrode (b) is

split into two by an insulating ring (c). The image current received by the outer electrodes is Fourier transformed to produce a spectrum in the *m/z* domain. Figure adapted from Thermo Scientific.

1.2.7 Ion detection

The last step of a MS acquisition is ion detection. The Orbitrap can also function as mass detector using the image current generated on the outer electrode by the ions oscillating around the internal spindle. The coherent oscillations of ions with a series of *m/z* values will cause a superposition of oscillating image currents with several frequencies. A fast Fourier Transformation (FFT) algorithm is then employed to convert the time-domain signal to a frequency-domain signal from which the *m/z* spectrum is achieved [147, 151].

Differently from Orbitrap, other mass analysers, such as the previously described quadrupole and ToF, need a separate detection system. One of the most commonly used detectors is the microchannel plate (MCP) detector consisting of parallel cylindrical channels functioning as a continuous dynode electron multiplier [152] (Fig. 1.11). The collision of ions on the detector leads to the emission of secondary electrons, which in turn impact on the channels causing signal amplification with an amplification efficiency that goes up to a factor of 10^5.

It is important to notice that ESI ionisation and transmission efficiency affect spectral intensity, thus spectral intensity does not necessarily represent the abundance of species in solution. Given the semi-quantitative nature of MS, isotopic labelling and/or internal standards are used to yield quantitative measurements [153].

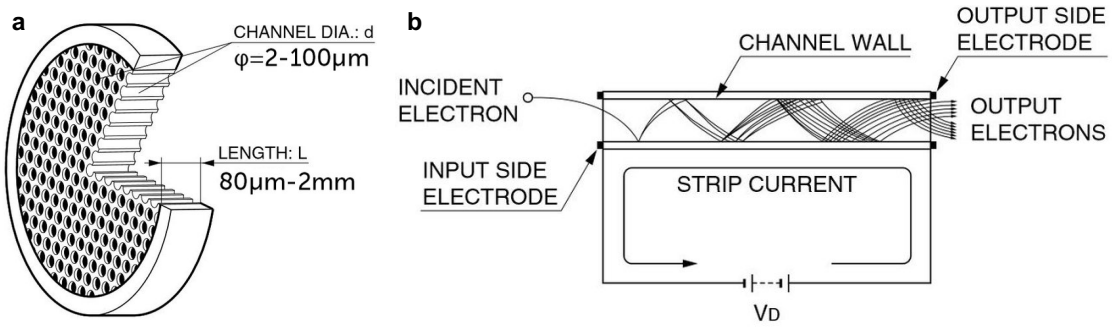


Figure 1.11 Schematic of a MCP. A) Section showing the architecture of a microchannel plate. B) Schematic of an individual channel: collision of an electron causes a cascaded release of secondary electrons, leading to signal amplification. Figure adapted from [128].

1.2.8 Collision cell

The collision cell is frequently used in order to apply collisional activation to analyte ions. This process causes the thermal activation of the molecules in the gas-phase enabling desolvation, removal of bound adducts, fragmentation, structural rearrangements, and in the case of protein oligomers, complex dissociation (collisional induced dissociation or CID). The cell is filled with inert gas (such as nitrogen or argon) with which the sample may collide, resulting in conversion of kinetic energy into internal energy [154]. Over a period of tens of microseconds, analyte may undergo thousands of collisions and the internal energy can be distributed to the molecules' many degrees of freedom and to the environment [115]. As the internal energy increases, the complex passes through distinct stages: cleaning, restructuring, unfolding, dissociation, and finally fragmentation [110]. For protein complexes, information such as composition and organisation of subunits within the complex can be obtained. For small molecules or peptides, particularly in the fields of proteomics, lipidomics or metabolomics, the fragmentation pattern allows the identification of the species present in solution.

For MS of intact membrane protein complexes collisional activation is crucial to remove detergent from the membrane protein-detergent complex and yield resolved *m/z* charge state distribution. To achieve this result, it is necessary to fine tune the applied energy in order to strip the detergent, but still maintain an intact, folded protein complex [36, 113, 155].

Recently, the development of a novel class of detergents allowing the preservation of non-covalent interactions, but also easy micelle removal in the mass spectrometer, has represented a significant advance for the study of membrane protein complexes using nMS [156]. For instance, the use of oligoglycerol detergent (OGD) G1-OGD allowed the investigation on the substrate and lipid-binding properties of the LPS transporter $LptB_2FG$, as described in Chapter 3.

1.3 The application of mass spectrometry to structural biology

1.3.1 Native mass spectrometry

As mentioned previously, the development of soft ionisation techniques and following instrumental improvements allowed the observation of intact proteins and the preservation of non-covalent interactions in the gas phase [99, 100]. In previous years MS has allowed the observation of large complexes up to the mega-Dalton size [157], including ribosomes [158] and intact viruses [159]. Further instrumental improvements and optimisation of solution conditions allowed the observation of the first membrane protein complex, the *E. coli* tetrameric ABC transporter BtuCD. The complex was kept in n-dodecyl β-D-maltoside (DDM) detergent micelle, which was gradually stripped off the protein in the mass spectrometer through collision with inert gas, as described previously. The latest advancement in the nMS field has been the application of the Orbitrap technology to intact protein complexes, thus entering the era of high resolution nMS, allowing the identification of small molecules bound to protein complexes [160, 161].

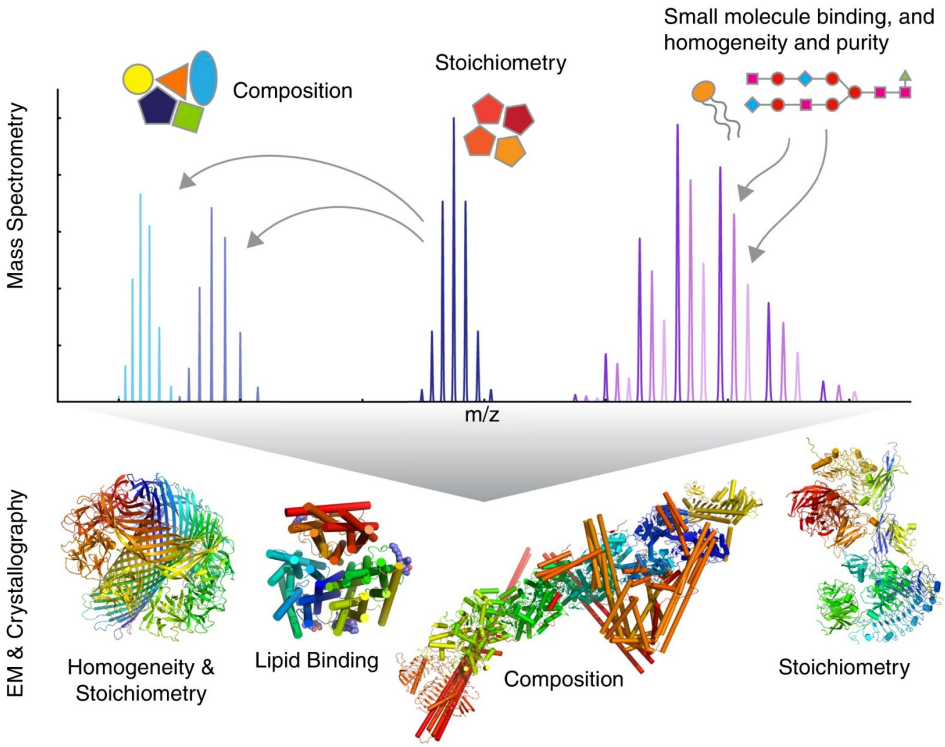


Figure 1.12 Applications of nMS in structural biology. NMS can be used to identify the stoichiometry and composition of protein complexes, to assess small molecule binding and to optimise conditions for structural techniques, such as cryo-EM and X-ray crystallography. Figure obtained from Liko et al. [162].

The immediate application of nMS consists on the determination of subunit stoichiometry, which can be further assessed through CID experiments. This ability has helped the resolution of controversies regarding complexes [163] subunit stoichiometry and has shown its potential in assisting solving ambiguous densities in in X-ray crystallography or cryo-EM [164-166] (Fig. 1.12).

The higher resolution of MS, compared to other techniques like size exclusion chromatography (SEC) or multi angle light scattering (MALS), allows accurate insights on purified proteins, such as co-purified small molecules, including lipids, thus helping in the assignment of ligand densities in structural methods [167, 168]. Lipid identification is a key step for understanding membrane protein structure and regulation. It has been shown that lipid binding may influence membrane proteins oligomeric state [35] as well as function [87]. A

powerful technique that may be coupled to nMS to identify lipids is liquid chromatography–mass spectrometry (LC-MS)–based lipidomics, that allows the detailed analysis of low mass (700–1,500 Da) lipid entities [169]. Furthermore, through exogenous addition of small molecules, nMS allows the evaluation of ligand binding on protein-protein interactions [16] or on enzymatic activity of membrane enzymes [37].

nMS may also be coupled with HDX-MS to gain information on the dynamics of protein complexes and protein-ligand interactions. This is a solution-based technique that exploits the ability of backbone amide hydrogen to exchange for deuterium upon dilution into D_2O-based buffer. In the following section I will discuss some of the HDX-MS basic principles that may be relevant for the understanding and interpretation of data described in Chapter 4.

1.3.2 Evaluation of protein-ligand interactions using nESI-nMS

Native MS coupled to nESI has been widely employed to detect and quantify protein-ligand interactions, including small molecules, peptides and, particularly in the case of membrane proteins, lipids [16, 25, 34, 37, 38, 87]. Despite many successful examples in literature, care should be taken when analysing native mass spectra displaying non-covalent protein-ligand interactions as solution phase equilibria may be altered upon ionisation and transfer in the gas phase, the two main steps of the electrospray process. These limitations may lead to false positives, resulting in the formation of non-specific protein-ligand complexes during the nESI process that are sufficiently long-lived to be detected in the gas phase. On the other hand, false negatives are also possible as a consequence of the voltages applied to focus the protein ions or CID, hence leading to a decrease or suppression of the signal. This is quite common in the case of small molecules given their low molecular weight and, consequently fewer interactions with the protein partner.

The phenomenon of non-specific binding involving macromolecules during nESI-nMS may be rationalised in the context of the charged residue model (CRM) of ESI, the mechanism that is widely accepted to be responsible for the formation of gaseous ions of macromolecular

complexes [132, 134]. As previously explained, following multiple evaporation and fission cycles the droplets reach the size of the analyte, then further solvent evaporation ultimately generates the gas-phase charged ion. During this process, one or more ligand molecules dissolved in solution may be present in the same droplet as the protein and droplet evaporation to dryness leads to the formation of non-specific complexes. The probability of the droplets containing more than one ligand molecule increases with concentration, hence one way to reduce non-specific adducts formation is conduction experiments at low ligand concentration. Conversely, at high concentrations (> 50 μM) non-specific binding always plays a major role in protein-ligand interaction [170].

The tendency of forming non-specific adducts is related to the chemistry and structure of the ligands. The presence of hydrophilic groups makes molecules more prone to participate in intermolecular H-bonds with the protein. In addition, the kinetic stability of non-specific association positively correlates with size as the number of non-covalent interactions is increased [171].

In the case of membrane proteins, although hydrophobic interactions play a role, polar amino acids are usually positioned in more accessible areas of the protein surface that could interact with lipid headgroups, rather than in hydrophobic transmembrane regions [5]. Hence, protein-lipid interactions are mainly governed by the chemistry of lipid headgroups and the acid/basic nature of surface protein residues [38], although tails still may play a role. For instance, CDL and PG are both anionic lipid, however a comparison of their binding affinity to a certain membrane protein only based on nMS may lead to erroneous conclusions since CDL presents a polar head with 2 phosphate groups, compared to the single charged polar head of PG. In addition, nMS polarity may play a role in the observation of lipid binding to membrane proteins, as indicated in a recent report [38], whereby PG binding to the OM protein OmpF was observed only in positive nESI polarity, while PC binding was detected only in negative nESI polarity.

These examples indicate that lipid binding to membrane proteins and, more generally protein-ligand interactions, should be validated using complementary techniques, ranging from

ITC and SPR for thermodynamic and kinetic validation to cryo-EM and X-ray crystallography for structural observations [25, 172]. Ion mobility (IM) coupled to nMS enables the distinction between specific and non-specific binding events based on the ability of a certain ligand (small molecule or lipid) to increase protein resistance to unfolding in the gas phase, hence stabilising the protein structure [37, 172]. HDX-MS and MD simulations allow to gain insights into the ligand-induced dynamics of the protein target, thereby aiding the differentiation between specific and non-specific interactions.

In addition, when calculating dissociation constants using nESI-nMS, appropriate statistical methods can be applied to account for non-specific interactions arising during the ESI process [173].

Apart from analysis using different techniques, appropriate controls should be run, for instance include amino acid mutations. Mutation of residues that are thought to interact with a certain ligand in solution should determine changes in binding affinity for specific ligands, while non-specific interactions are expected to be unaffected [16, 25, 38, 87, 173]. This strategy can be coupled with the above described complementary approaches, in order to yield a complete picture of ligand binding (non)specificity. For instance, in Chapter 5 I will describe the application of this strategy, along with *in vivo* experiments, to assess the specific binding of PI to the GET pathway proteins Get1 and CAML.

In the case of lipid binding to membrane proteins, a recently developed MS-based method allows the distinction between specific and non-specific lipids based on their different ability to exchange with detergent in solution. This strategy consists of adding an excess of detergent to the protein-lipid solution just before nMS analysis. Whilst detergent outcompetes weaker non-specific lipid binding, specific interactions are retained [174]. This method, along with validation using HDX-MS and MD simulations has been employed to investigate the interaction between the LPS transport protein LptDE and Re-LPS, an LPS substructure (Chapter 4).

1.3.3 Hydrogen/deuterium exchange mass spectrometry (HDX-MS)

As explained above, HDX-MS is a solution technique where the protein is incubated with deuterated buffer to initiate the labelling reaction, i.e. the exchange between amide hydrogens and deuterium. Following labelling the protein is digested in peptide fragments and the increase in mass of each peptide, corresponding to different regions of the protein, is measured with MS [175]. In HDX-MS we monitor the H→D exchange of backbone hydrogens, however these hydrogens are not the only ones exchanging with solvent. Indeed, hydroxyl, sulfhydryl, amine, and amide groups present in the side chains (i.e. Ser, Cys, Lys, Gln etc.) may also exchange with solvent. Nevertheless, side chains are highly dynamic and the rate of exchange of these moieties is so fast that they completely back-exchange during the quenching and HPLC steps [176]. Furthermore, -CH bonds do not undergo HDX on an experimentally relevant time scale, with the possible exception of the histidine C(ε)1 – H in the imidazole side chain [177] (Fig. 1.13).

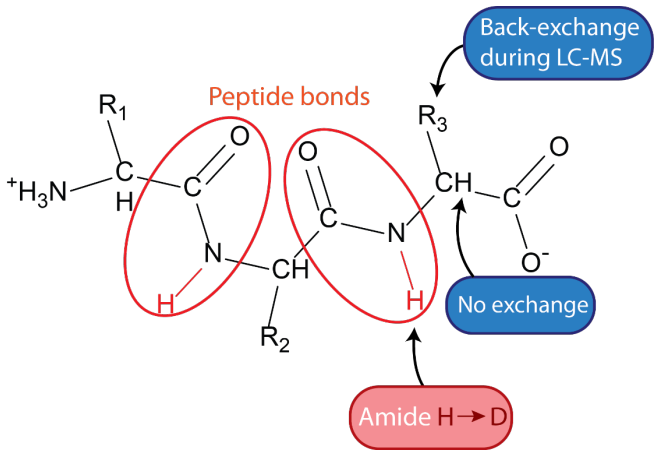


Figure 1.13 Different types of hydrogens in proteins. Proteins possess three kinds of hydrogens theoretically capable of exchanging to deuterium. Side chain hydrogens quickly back-exchange during the LC-MS step; hydrogens bound to C exchange so slowly that the exchange rate is not relevant on experimental time scales. On the other hand, peptide bond amides exchange rates can be monitored and yield valuable information about proteins conformational dynamics.

Backbone hydrogens participate in hydrogen bonds determining the secondary structures, thus the exchange H→D rate is an indicator of the stability of such hydrogen bonds and water accessibility of the backbone amide, thus it is representative of the dynamics of the protein. Since every residue (except for proline and the first amino acid of the protein) possess an amide N-H group, virtually HDX allows the elucidation of the dynamics of every protein domain.

The HDX events are enabled by local fluctuations of the protein. Specifically, for labelling to occur hydrogen bonds should be weakened and this happens when a certain protein segment reaches a transient partially unfolded (open) conformation, following the reaction described below [178]:

$$P\text{-}H_{cl} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} P\text{-}H_{op} \xrightarrow{k_{ch}} P\text{-}D_{op} \underset{k_{op}}{\overset{k_{cl}}{\rightleftharpoons}} P\text{-}D_{cl} \tag{1.11}$$

Where $P\text{-}H_{cl}$ and $P\text{-}H_{op}$ indicate the unlabelled protein in the folded (closed) and unfolded (open) states, respectively; $P\text{-}D_{cl}$ and $P\text{-}D_{op}$ indicate the deuterated protein in the folded (closed) and unfolded (open) states, respectively. The rate constants are indicated by k_{op} (unfolding, or opening, rate), k_{cl} (refolding, or closing, rate), and k_{ch} (chemical labelling rate). Given this equation we can outline a steady state analysis, in order to provide some theoretical background [179].

If we group the unlabelled proteins states $P\text{-}H_{cl}$ and $P\text{-}H_{op}$ and collectively indicate them as U, we can state that:

$$\frac{d[U]}{dt} = -k_{ch}\,[P\text{-}H_{op}] \quad where\ [U] = [P\text{-}H_{op}] + [P\text{-}H_{op}] \tag{1.12}$$

but also:

$$\frac{d[P\text{-}H_{op}]}{dt} = -(k_{cl}+k_{ch})\ [P\text{-}H_{op}] + k_{op}\ [P\text{-}H_{cl}] \tag{1.13}$$

The rate of formation and consumption of $P\text{-}H_{op}$ can be assumed to be approximately equal, such that:

$$\frac{d[P\text{-}H_{op}]}{dt} \approx 0 \;\Rightarrow\; (k_{cl}+k_{ch})\,[P\text{-}H_{op}] = k_{op}\,[P\text{-}H_{cl}] = k_{op}\,[U] - [P\text{-}H_{op}]$$

$$\Rightarrow \quad [P\text{-}H_{op}] = \frac{k_{op}}{k_{cl}+k_{ch}+k_{op}}[U] \tag{1.14}$$

We can now insert this in equation (1.1) to obtain the pseudo first order equation:

$$\frac{d[U]}{dt} = -\frac{k_{ch}k_{op}}{k_{cl}+k_{ch}+k_{op}}[U] \tag{1.15}$$

which can be solved:

$$[U] = e^{-k_{HDX}\,t} \quad where \quad k_{HDX} = \frac{k_{ch}k_{op}}{k_{cl}+k_{ch}+k_{op}} \tag{1.16}$$

Experimentally, k_{HDX} indicates the observed HDX rate, which of course depends on all the kinetic constants involved in the labelling reaction. For proteins in their native state we can assume that the different exchange sites are predominantly in their closed state, so that $k_{op} << k_{cl}$, thus:

$$k_{HDX} = \frac{k_{ch}k_{op}}{k_{cl}+k_{ch}} \tag{1.17}$$

In most physiological conditions unfolding/refolding rates are much faster than the chemical exchange rate (k_{op}, $k_{cl} >> k_{ch}$) [179-181], i. e. the probability of exchange happening during a single opening event is small and numerous rapid opening/closing are necessary to obtain H to D exchange. This leads to a rearrangement of equation (1.7):

$$k_{HDX} = \frac{k_{op}}{k_{cl}}\,k_{ch} \tag{1.18}$$

The k_{op}/k_{cl} ratio is the opening reaction equilibrium constant (K_{op}), such that:

$$k_{HDX} = K_{op}k_{ch} \tag{1.19}$$

Macroscopically, this translates in a single isotopic distribution gradually shifting to higher *m/z* over time.

In some cases, the opening/closing motions may occur at much lower rates than the chemical exchange reaction ($k_{cl} << k_{ch}$). In these situations, portions of the protein undergo

unfolding events such that residues are concertedly exposed to solvent and uptake deuterium before the protein returns to the closed state. In such kinetic regime equation (1.8) becomes:

$$k_{HDX} = k_{op} \tag{1.20}$$

indicating that the opening rate is equal to the observed exchange rate. This cooperative unfolding event leads to a bimodal isotopic distribution (Fig. 1.14). The low-mass envelope represents those residues that have not yet visited the state capable of exchange, while the high-mass population represents multiple backbone hydrogens that have already undergone the H-D exchange following the correlated conformational fluctuation. EX1 kinetics may be induced by denaturation [182-184] or high pH [185-187], however multiple studies also demonstrated that some protein domains may physiologically undergo fluctuations following the EX1 regime [188-192]. It is important to notice that he two kinetic regimes described above are two ideal, limiting situations. In fact, proteins may experience mixed EX2/EX1 behaviour, referred as EXX kinetics, leading to overlap of high- and low-mass populations [193].

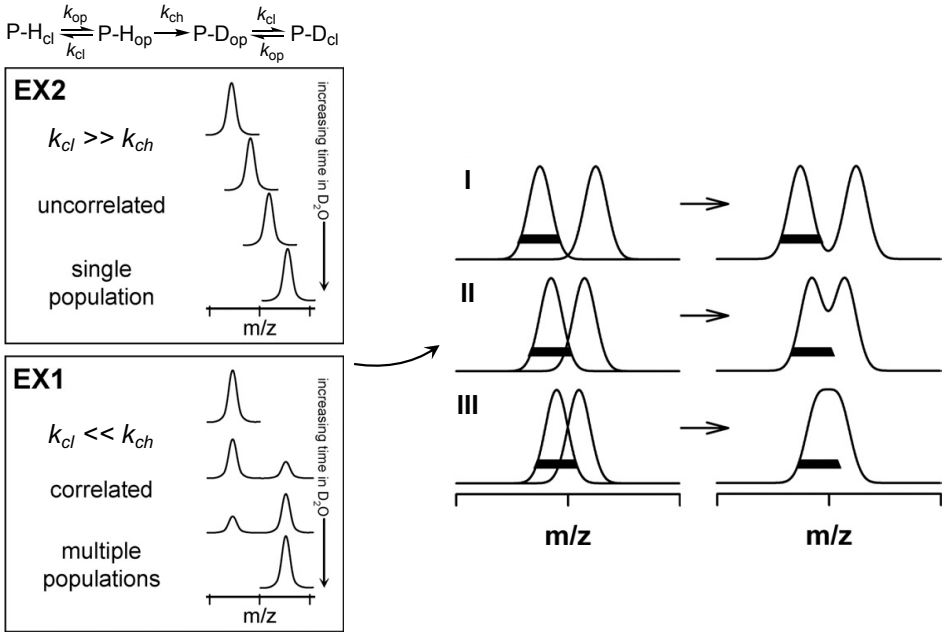


Figure 1.14 Different HDX kinetics. The left panel summarises the difference between EX2 and EX1 kinetic regimes. In the EX2 kinetic regime the peptide isotopic distribution gradually shifts towards higher *m/z* over time. In the EX1 kinetics a bimodal distribution is present, with the intensity of the higher *m/z* population rising over time. The right panel shows a situation

where a peptide is subject to both EX2 and EX1 kinetics (called EXX), with increasing overlap of the mass distributions. Figure adapted from Weis et al. [181].

From a more practical point of view, an HDX experiment proceeds as follows: dilution into deuterated buffer, reaction quenching, proteolytic digestion, high-pressure liquid chromatography (HPLC) separation, MS analysis (Fig. 1.15). Typically, a protein solution in H_2O-based buffer is diluted (10 to 20-fold) using a buffer with more than 99% deuterium content, so that the final deuterium content is >90%, and with pD adjusted to physiological value or varied, if required [194]. The high deuterium content of the labelling buffer serves to force the labelling direction in one direction. In continuous labelling experiments, like the ones described in this dissertation, the protein sample diluted in deuterated buffer and, after different time intervals, the reaction is quenched by mixture with a quenching solution in a 1:1 volume ratio. Usually, the reaction is carried out at room temperature (20-25 °C), while quenching is performed by lowering pH to 2 and temperature to 0 °C. The quenching solution may also contain reducing agents, usually tris(2-carboxyethyl)phosphine (TCEP), to facilitate unfolding through disulfide bonds reduction [195, 196].

Determination of deuterium levels of the whole protein does not provide any information on local conformational dynamics, therefore, to increase resolution proteolytic digestion is employed [176]. Because digestion of the deuterium labelled protein complex must occur under quench conditions, acid proteases are used. To date, the most widely acid-protease for these purposes is pepsin, a non-specific protease cleaving at hydrophobic residues. To assess of the identity of each peptic peptide yielded after pepsin digestion, it is necessary to sequence those that are generated. Most commonly, this is accomplished using tandem MS techniques [175].

Following pepsin digestion, peptides are desalted and separated using reverse phase HPLC. Even under acid conditions, the occurrence of some amide back-exchange (D → H) during digestion and LC is inevitable. To minimising this effect the solution temperature is kept to 0 °C during all LC steps. Following LC separation peptides are injected in the mass

spectrometer with online ESI-MS or MALDI-MS [180, 197]. The ionisation techniques used in mass spectrometry will be discussed in the following section.

Once the masses of different peptides are obtained, the deuteration level of each segment as a function of labelling time t can be calculated using the following equation:

$$d(t) = \frac{m(t) - m_0}{m_{100} - m_0} \tag{1.21}$$

Where *m(t)* Is the centroid mass of the peptide of interest, m_{100} is the mass of the fully deuterated peptide and m_0 is the mass of the undeutreated peptide.

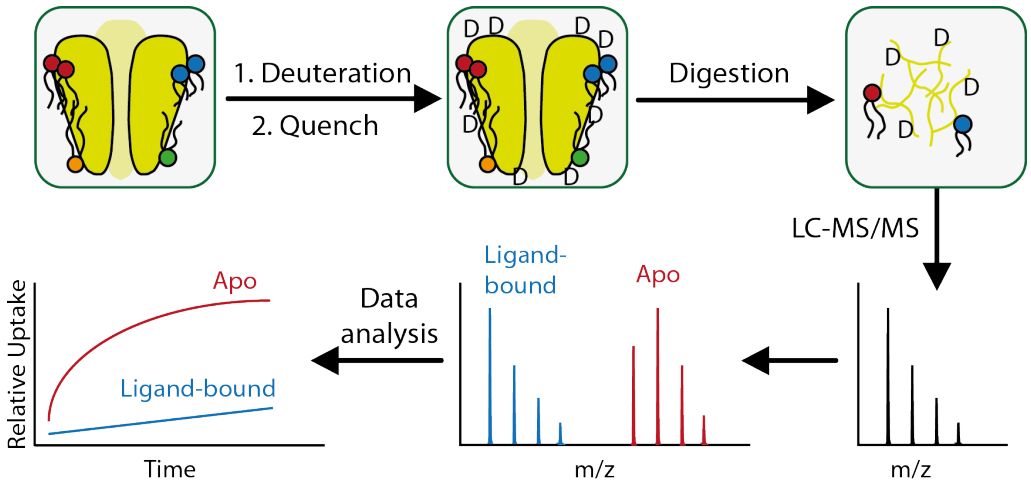


Figure 1.15 Representation of HDX workflow to interrogate ligand-induced conformational changes. The protein complex (either apo or ligand-bound) is diluted into D_2O-based buffer and the labelling reaction is quenched at low pH after a given time interval. The deuterated protein is digested and the resulting fragments are then separated and characterised by LC-MS. The protection level of each peptide is obtained by plotting D uptake at each time point for the two different states.

The time points typically covered range from minutes to hours (sometimes days), during which the mass of the protein gradually increases according to the kinetic parameters of equation (1.1). the domain that spends a considerable amount of time in open conformations undergoes faster exchange than those that are more tightly folded or interacting with binding partners.

Studies of the HDX of membrane proteins remain relatively limited in the number given technical challenges associated with poor sequence coverage and the low level of deuterium labelling in transmembrane regions. Low sequence coverage is related to low proteolytic efficiency, as well as the fact that many peptides may not elute from the LC column given their high hydrophobicity. Nonetheless, substantial progresses have been made lately and recent studies employing various membrane mimetics have shown the power of HDX-MS in elucidating the structural dynamics of membrane proteins. Different solubilisation strategies have been used to investigate membrane proteins. Some studies employing detergent micelles [190, 198-201] elucidated the effect of drugs, substrates or protein-protein interaction on the conformational dynamics of the membrane proteins. Further reports showed the possibility to employ lipid bilayers, such as styrene-maleic acid lipid particles (SMALP) solubilised protein [202] or lipid nanodiscs [77, 203, 204], to investigate the influence of lipid composition on solution accessibility and conformational dynamics of membrane proteins.

A key step on the interpretation of HDX-MS is the comparison of the deuterium uptake of a given peptide in different states (i.e. apo vs ligand-bound protein). Strictly speaking, the data obtained from this analysis provides evidence about the stability of the hydrogen bond networks. Thus, to extract the right information about protein dynamics it is necessary to couple this method with complementary techniques, such as molecular dynamics (MD) simulations that can deliver atomistic details on conformational changes and key interactions of the protein complex [77, 190, 204].

1.4 Mass spectrometry instrumentation

1.4.1 Thermo Fisher Scientific Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer

All nMS experiments performed in this dissertation were obtained on a Thermo Fisher Scientific Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer (simply referred as Q-Exactive, Fig. 1.16). This mass spectrometer has a nESI ion source; following ionisation the analyte is directed to the cone and guided through a stacked ion guide (S-lenses) towards an injection flatapole that, compared to the previous versions of the instrument, has been upgraded to enable ion trapping. This modification converts the flatapole in a linear ion trap that allows fragmentation of trapped ions through CID [205]. In this way, oligomeric protein complexes may be dissociated and the monomer species can be selected in the quadrupole according to their *m/z* [206]. Furthermore, this allows detergent removal before the quadrupole, thus allowing transmission of protein liberated from micelle in the quadrupole consenting selection of ligand-bound species for further HCD experiments. The injection flatapole is followed by a bent flatapole operating in RF-only mode that precludes the transmission of neutral species. The next step of the ions path is the first mass analyser represented by a quadrupole that allows the selection of ions for tandem MS analysis. Ions are then transmitted to the C-trap that allows storage before the injection in the higher-energy collisional dissociation (HCD) cell. Here, they undergo collisional activation that, in the case of membrane protein complexes, allows the removal of detergent micelles or subunit and ligand dissociation. The current instrumentation enables ultra-high mass transmission [207], moreover the higher pressure achievable in the HCD cell compared to previous instruments cell allows for improved ion trapping and activation [162]. Pressure of the ultra-high vacuum (UHV) changes from the standard 10^{-11}–10^{-10} mbar to ~ 10^{-9} mbar.

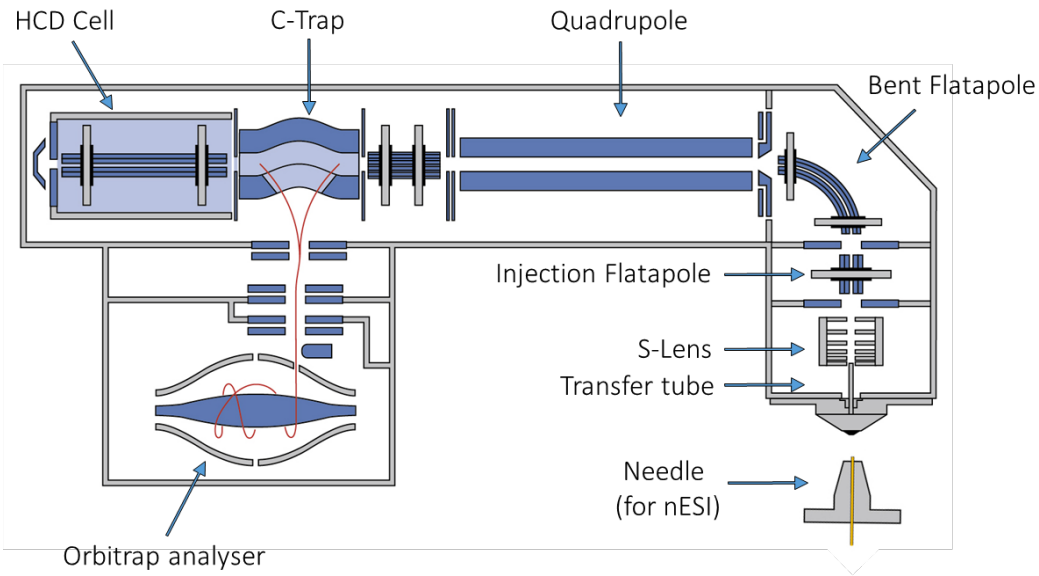


Figure 1.16 Schematic of the Thermo Scientific Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer. Ions produced from nESI are focussed in the S-lens and transmitted to through the flatapole and the quadrupole which may operate in scanning or in mass-selection mode. Then ions enter the C-trap where they are focussed and injected in the HCD for collisional activation, or directly in the Orbitrap. Figure adapted from Gault et al. [162].

1.4.2 Thermo Fisher Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer

The Hybrid Ion Trap-Orbitrap Mass Spectrometer (LTQ Orbitrap XL, Fig. 1.17) was used for proteomics and lipidomics analysis. Like the Q-Exactive mass spectrometer, ions are generated through nESI, however selection happens in linear ion trap, where ions may also be fragmented by CID. The MS^2 fragmentation pattern allow the identification of peptides and lipids present in solution for proteomics and lipidomics, respectively. The instrument also has s C-trap for ion storage, a HCD cell for further activation, and the Orbitrap for mass analysis and detection.

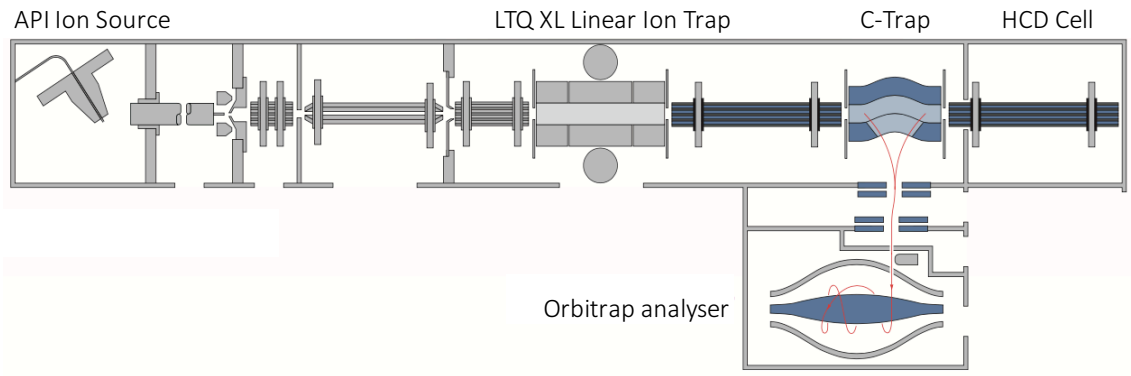


Figure 1.17 Schematic of the Thermo Fisher Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer. Similar to the Q-Exactive, ions are generated through nESI but then transmitted in a linear ion trap before the C-trap and Orbitrap for mass analysis and detection. Figure from Thermo Scientific.

1.4.3 Waters Hybrid Q-ToF Synapt G2-Si Mass Spectrometer

The peptide identification performed in the HDX-MS experiments was executed on a Waters Hybrid Q-ToF Synapt G2-Si Mass Spectrometer (Fig. 1.18). Peptides eluted from HPLC are ionised in the source region through nESI. In this case, the nESI capillary is positioned orthogonal to the cone, giving rise to the so-called Z-spray configuration, to reduce the transmission of neutral species. A multiple set of lenses allows ion focussing to the quadrupole mass analyser where they can be either mass selected or scanned. Subsequently, ions are transmitted to a second quadrupole, operating at RF-only mode and filled with neutral gas (usually Ar or N_2), that acts as a collision (trap) cell for CID. Finally, the ion beam is sent to the ToF coupled to a MCP-type detector for mass analysis and detection. The instrument is equipped with an IM cell (not used in this thesis), that allows determination of collisional cross sections of peptides or proteins (if instrument used for nMS).

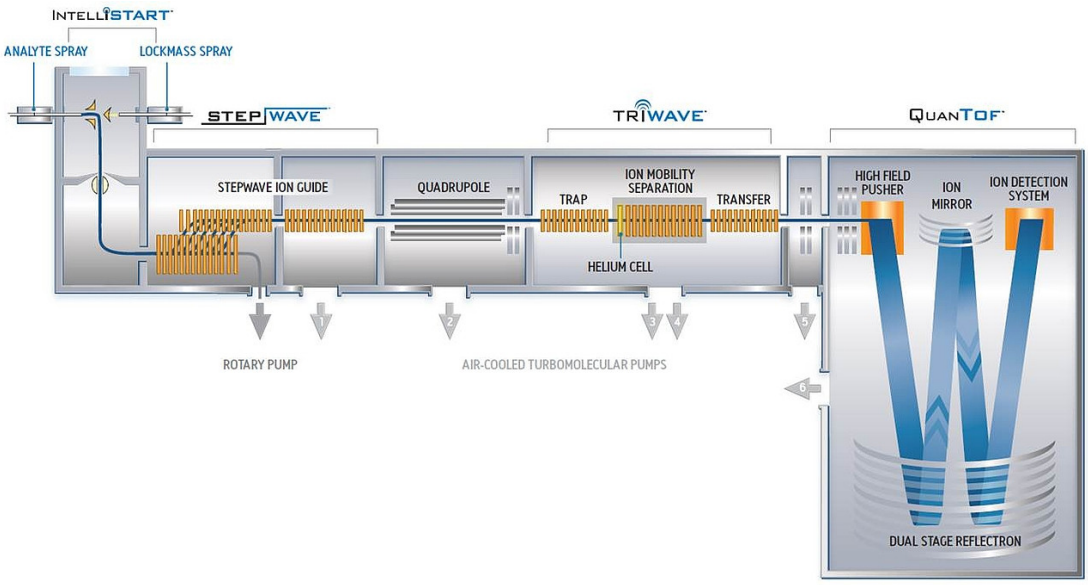


Figure 1.18 Schematic of the Waters Hybrid Q-ToF Synapt G2-Si Mass Spectrometer. Proteins are ionised using a nESI source and are transmitted through the different lenses, focussing the beam to the quadrupole and the trap (where CID is applied) before reaching the ToF for mass analysis, coupled to a MCP for mass detection. In this case, the ToF is equipped with a W-shaped, dual stage reflecton for improved mass analysis. The ion detection system is represented by a MCP. Figure from Waters Corp.

1.5 Aims of this dissertation

The recent developments of nMS have opened a new era in MS guided structural biology [163]. In particular, the utilisation of nMS to membrane proteins has allowed the interrogation of the multiple non-covalent interactions governing biological membranes. At the same time, the field of HDX-MS has been greatly expanding in recent years and has found application to membrane proteins. The combination of MS-based techniques with other biophysical and structural biology methods (MD simulations, cryo-EM but also *in vivo* studies) allows a detailed investigation on the roles of protein complexes and protein-protein interaction networks in the context of cellular homeostasis.

Great interest has been dedicated to the elucidation of the mechanism of action of membrane transporters. This family of proteins is the emphasis of this dissertation and a diverse

range of transporter is explored. The previously mentioned capabilities of nMS are employed to elucidate the structure and modulation of different membrane protein complexes.

In Chapter 2 nMS is employed to elucidate the influence of substrate and ATP binding on the complex formation of bacterial ABC importers, thus facilitating the clarification of their mechanism of transport [208].

Having established a method to investigate ABC transporters, in Chapter 3 we investigated the mechanism of LPS extraction by the atypical ABC transporter $LptB_2FG$, which intimately interacts with the single pass membrane protein LptC. We analysed the influence of substrate, lipids and nucleotide binding on $LptB_2FG$ using nMS and then integrated nMS with cryo-EM to study the interactions between $LptB_2FG$ and LptC. This membrane protein complex also interacts with a soluble periplasmic protein: LptA. Using nMS we also probed $LptB_2FGC$-LptA interaction, as well as the effect of antimicrobial peptides on LptA and $LptB_2FGC$-A complex.

Following the interest on the LPS transport system, in Chapter 4 we combined nMS, HDX-MS, and MD simulations to examine the influence of substrate, lipids and inhibitor binding on the conformational dynamics of the LPS insertase LptDE, in order to obtain molecular details on the LPS insertion process in the OM.

Finally, in Chapter 5, nMS and lipidomics are combined to cryo-EM and *in vivo* experiments to clarify the structure and subunit stoichiometry of the eukaryotic tail-anchored protein insertase complex, revealing the contribution of lipid binding on the stabilisation of the protein complex.

In summary, a great variety of small molecule – protein interactions are explored: from ions to nucleotides; from to phospholipids to the high molecular weight heterogeneous LPS, including its truncated substructures. The aim of the research presented here is to investigate the effects of these ligands on protein structure and dynamics, and to identify crucial lipid binding hotspots to membrane proteins.

1.6 References

1. Eisenberg, D., et al., *Protein function in the post-genomic era*. Nature, 2000. **405**(6788): p. 823-836.
2. Biswas, A., et al., *Interaction of alpha-crystallin with some small molecules and its effect on its structure and function*. Biochim. Biophys. Acta, 2016. **1860**(1 Pt B): p. 211-221.
3. Schlessinger, J., *Ligand-induced, receptor-mediated dimerization and activation of EGF receptor*. Cell, 2002. **110**(6): p. 669-672.
4. Duan, G. and D. Walther, *The roles of post-translational modifications in the context of protein interaction networks*. PLoS Comput. Biol., 2015. **11**(2): p. e1004049.
5. von Heijne, G., *Membrane-protein topology*. Nat. Rev. Mol. Cell. Biol., 2006. **7**(12): p. 909-918.
6. Lingwood, D. and K. Simons, *Lipid rafts as a membrane-organizing principle*. Science, 2010. **327**(5961): p. 46-50.
7. Bissig, C. and J. Gruenberg, *Lipid sorting and multivesicular endosome biogenesis*. Cold Spring Harb. Perspect Biol., 2013. **5**(10): p. a016816.
8. Stevens, T.J. and I.T. Arkin, *Do more complex organisms have a greater proportion of membrane proteins in their genomes?* Proteins, 2000. **39**(4): p. 417-420.
9. Vinothkumar, K.R. and R. Henderson, *Structures of membrane proteins*. Q. Rev.Biophys., 2010. **43**(1): p. 65-158.
10. Reece, J.B., Campbell, N. A., *Membrane Structure and Function*, in *Campbell Biology (11th Ed.)*. 2016.
11. Clark, M.A., Choi, J., Douglas, M., *Biology 2e*. 2018, https://openstax.org/details/books/biology-2e.
12. Lodish, H., et al., *Membrane Proteins*, in *Molecular Cell Biology*. 2000, W. H. Freeman.
13. Wallin, E. and G. von Heijne, *Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms*. Protein Sci., 1998. **7**(4): p. 1029-1038.
14. Freeman, M., *Rhomboid proteases and their biological functions*. Annu. Rev. Genet., 2008. **42**: p. 191-210.
15. Hauser, A.S., et al., *Trends in GPCR drug discovery: new agents, targets and indications*. Nat. Rev. Drug Discov., 2017. **16**(12): p. 829-842.
16. Yen, H.Y., et al., *PtdIns(4,5)P2 stabilizes active states of GPCRs and enhances selectivity of G-protein coupling*. Nature, 2018. **559**(7714): p. 423-427.
17. Wickenden, A., B. Priest, and G. Erdemli, *Ion channel drug discovery: challenges and future directions*. Future Med. Chem., 2012. **4**(5): p. 661-679.
18. Hong, K.H. and C. Miller, *The lipid-protein interface of a Shaker K(+) channel*. J. Gen. Physiol., 2000. **115**(1): p. 51-58.

19. Hughes, T.E.T., et al., *Structural basis of TRPV5 channel inhibition by econazole revealed by cryo-EM*. Nat. Struct. Mol. Biol., 2018. **25**(1): p. 53-60.
20. Gao, Y., et al., *TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action*. Nature, 2016. **534**(7607): p. 347-351.
21. Wolfersberger, M.G., *Uniporters, symporters and antiporters*. J. Exp. Biol., 1994. **196**: p. 5-6.
22. Locher, K.P., *Mechanistic diversity in ATP-binding cassette (ABC) transporters*. Nature Structural & Molecular Biology, 2016. **23**: p. 487-493.
23. Mi, W., et al., *Structural basis of MsbA-mediated lipopolysaccharide transport*. Nature, 2017. **549**(7671): p. 233-237.
24. Luo, Q., et al., *Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG*. Nat. Struct. Mol. Biol., 2017. **24**(5): p. 469-474.
25. Dong, H., et al., *Structural and functional insights into the lipopolysaccharide ABC transporter LptB2FG*. Nat. Commun., 2017. **8**(1): p. 222.
26. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *How many drug targets are there?* Nat. Rev. Drug Discov., 2006. **5**(12): p. 993-996.
27. Wymann, M.P. and R. Schneiter, *Lipid signalling in disease*. Nat. Rev. Mol. Cell Biol., 2008. **9**(2): p. 162-176.
28. Miller, S.I., *Antibiotic Resistance and Regulation of the Gram-Negative Bacterial Outer Membrane Barrier by Host Innate Immune Molecules*. mBio, 2016. **7**(5).
29. Chen, Z., et al., *Mammalian drug efflux transporters of the ATP binding cassette (ABC) family in multidrug resistance: A review of the past decade*. Cancer Lett., 2016. **370**(1): p. 153-164.
30. Amaral, L., et al., *Inhibitors of bacterial efflux pumps that also inhibit efflux pumps of cancer cells*. Anticancer Res., 2012. **32**(7): p. 2947-2957.
31. Pielak, R.M. and J.J. Chou, *Influenza M2 proton channels*. Biochim. Biophys. Acta, 2011. **1808**(2): p. 522-529.
32. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein*. Cell, 2020. **S0092-8674**(20): p. 30262-2.
33. Janganan, T.K., et al., *Evidence for the assembly of a bacterial tripartite multidrug pump with a stoichiometry of 3:6:3*. Journal of Biological Chemistry, 2011. **286**(30): p. 26900-26912.
34. Marcoux, J., et al., *Mass spectrometry reveals synergistic effects of nucleotides, lipids, and drugs binding to a multidrug resistance efflux pump*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(24): p. 9704-9709.

35. Gupta, K., et al., *The role of interfacial lipids in stabilizing membrane protein oligomers*. Nature, 2017. **541**(7637): p. 421-424.

36. Reading, E., et al., *The role of the detergent micelle in preserving the structure of membrane proteins in the gas phase*. Angew. Chem. Int. Ed. Engl., 2015. **54**(15): p. 4577-4581.

37. Mehmood, S., et al., *Mass spectrometry captures off-target drug binding and provides mechanistic insights into the human metalloprotease ZMPSTE24*. Nat. Chem., 2016. **8**(12): p. 1152-1158.

38. Liko, I., et al., *Lipid binding attenuates channel closure of the outer membrane protein OmpF*. Proceedings of the National Academy of Sciences of the United States of America, 2018. **115**(26): p. 6691-6696.

39. Leney, A.C., et al., *Amphipathic polymers enable the study of functional membrane proteins in the gas phase*. Analytical Chemistry, 2012. **84**(22): p. 9841-9847.

40. Iadanza, M.G., et al., *Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM*. Nature Communications, 2016. **7**.

41. Bolla, J.R., et al., *Membrane Protein-Lipid Interactions Probed Using Mass Spectrometry*. Annu. Rev. Biochem., 2019. **88**: p. 85-111.

42. Wagner, S., et al., *Rationalizing membrane protein overexpression*. Trends Biotechnol., 2006. **24**(8): p. 364-371.

43. Lacapere, J.J., et al., *Determining membrane protein structures: still a challenge!* Trends Biochem. Sci., 2007. **32**(6): p. 259-270.

44. Zorman, S., et al., *Advances and challenges of membrane-protein complex production*. Curr. Opin. Struct. Biol., 2015. **32**: p. 123-130.

45. Rawson, S., et al., *The changing landscape of membrane protein structural biology through developments in electron microscopy*. Mol. Membr. Biol., 2016. **33**(1-2): p. 12-22.

46. Marsh, D., *Protein modulation of lipids, and vice-versa, in membranes*. Biochim. Biophys. Acta, 2008. **1778**(7-8): p. 1545-1575.

47. Barrera, N.P., M. Zhou, and C.V. Robinson, *The role of lipids in defining membrane protein interactions: insights from mass spectrometry*. Trends Cell Biol., 2013. **23**(1): p. 1-8.

48. Drew, D., et al., *Optimization of membrane protein overexpression and purification using GFP fusions*. Nat. Methods, 2006. **3**(4): p. 303-313.

49. Newby, Z.E., et al., *A general protocol for the crystallization of membrane proteins for X-ray structural investigation*. Nat. Protoc., 2009. **4**(5): p. 619-637.

50. Prive, G.G., *Detergents for the stabilization and crystallization of membrane proteins*. Methods, 2007. **41**(4): p. 388-397.
51. Bechara, C. and C.V. Robinson, *Different modes of lipid binding to membrane proteins probed by mass spectrometry*. J. Am. Chem. Soc., 2015. **137**(16): p. 5240-5247.
52. Fahy, E., et al., *A comprehensive classification system for lipids*. Journal of Lipid Research, 2005. **46**(5): p. 839-861.
53. van Meer, G., D.R. Voelker, and G.W. Feigenson, *Membrane lipids: where they are and how they behave*. Nat. Rev. Mol. Cell Biol., 2008. **9**(2): p. 112-124.
54. Voet D., V.J., Pratt C., *Fundamentals of Biochemistry Life at the Molecular Level (4th Ed.)*, ed. I. John Wiley & Sons. 2013.
55. Amon, R., et al., *Glycans in immune recognition and response*. Carbohydr. Res., 2014. **389**: p. 115-122.
56. Andreyev, A.Y., et al., *Subcellular organelle lipidomics in TLR-4-activated macrophages*. Journal of Lipid Research, 2010. **51**(9): p. 2785-2797.
57. De Meyer, F. and B. Smit, *Effect of cholesterol on the structure of a phospholipid bilayer*. Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(10): p. 3654-3658.
58. Drin, G., *Topological regulation of lipid balance in cells*, in *Annu. Rev. Biochem.* 2014. p. 51-77.
59. Pomorski, T.G., T. Nylander, and M. Cardenas, *Model cell membranes: discerning lipid and protein contributions in shaping the cell*. Adv. Colloid Interface Sci., 2014. **205**: p. 207-220.
60. Mbamala, E.C., A. Ben-Shaul, and S. May, *Domain formation induced by the adsorption of charged proteins on mixed lipid membranes*. Biophys J., 2005. **88**(3): p. 1702-1714.
61. Sezgin, E., et al., *The mystery of membrane organization: composition, regulation and roles of lipid rafts*. Nat. Rev. Mol. Cell Biol., 2017. **18**(6): p. 361-374.
62. Pike, L.J., *Rafts defined: a report on the Keystone Symposium on Lipid Rafts and Cell Function*, in *J. Lipid Res.* 2006: United States. p. 1597-8.
63. Raetz, C.R. and C. Whitfield, *Lipopolysaccharide endotoxins*. Annu. Rev. Biochem., 2002. **71**: p. 635-700.
64. Sohlenkamp, C. and O. Geiger, *Bacterial membrane lipids: diversity in structures and pathways*. FEMS Microbiol. Rev., 2016. **40**(1): p. 133-159.
65. Raetz, C.R., Dowhan, W., *Biosynthesis and Function of Phospholipids in Escherichia Coli*. J. Biol. Chem., 1990. **265**(3): p. 1235-1238.
66. Nikaido, H., *Outer membrane*. Escherichia coli and Salmonella: Cellular and Molecular Biology, 1996: p. 29-47.

67. Whitfield, C. and M.S. Trent, *Biosynthesis and export of bacterial lipopolysaccharides*. Annu. Rev. Biochem., 2014. **83**: p. 99-128.

68. Stenutz, R., A. Weintraub, and G. Widmalm, *The structures of Escherichia coli O-polysaccharide antigens*. FEMS Microbiology Reviews, 2006. **30**(3): p. 382-403.

69. Liu, B., et al., *Structural diversity in Salmonella O antigens and its genetic basis*. FEMS Microbiology Reviews, 2014. **38**(1): p. 56-89.

70. Joiner, K.A., *Complement evasion by bacteria and parasites*, in *Annual Review of Microbiology*. 1988. p. 201-230.

71. Okuda, S., et al., *Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model*. Nat. Rev. Microbiol., 2016. **14**(6): p. 337-45.

72. Ruiz, N., D. Kahne, and T.J. Silhavy, *Transport of lipopolysaccharide across the cell envelope: the long road of discovery*. Nat. Rev. Microbiol., 2009. **7**(9): p. 677-83.

73. Islam, S.T. and J.S. Lam, *Wzx flippase-mediated membrane translocation of sugar polymer precursors in bacteria*. Environ. Microbiol., 2013. **15**(4): p. 1001-15.

74. Lee, A.G., *How lipids affect the activities of integral membrane proteins*. Biochimica et Biophysica Acta - Biomembranes, 2004. **1666**(1-2): p. 62-87.

75. Powl, A.M., J.M. East, and A.G. Lee, *Importance of direct interactions with lipids for the function of the mechanosensitive channel MscL*. Biochemistry, 2008. **47**(46): p. 12175-12184.

76. Bechara, C., et al., *A subset of annular lipids is linked to the flippase activity of an ABC transporter*. Nat. Chem., 2015. **7**(3): p. 255-262.

77. Martens, C., et al., *Direct protein-lipid interactions shape the conformational landscape of secondary transporters*. Nat Commun, 2018. **9**(1): p. 4151.

78. Hansen, S.B., X. Tao, and R. MacKinnon, *Structural basis of PIP2 activation of the classical inward rectifier K+ channel Kir2.2*. Nature, 2011. **477**(7365): p. 495-498.

79. Maccarrone, M., *Missing Pieces to the Endocannabinoid Puzzle*. Trends Mol. Med., 2020. **26**(3): p. 263-272.

80. Dawaliby, R., et al., *Allosteric regulation of G protein-coupled receptor activity by phospholipids*. Nature Chemical Biology, 2016. **12**(1): p. 35-39.

81. Dainese, E., et al., *Membrane lipids are key modulators of the endocannabinoid-hydrolase FAAH*. Biochem J., 2014. **457**(3): p. 463-472.

82. Patrick, J.W., et al., *Allostery revealed within lipid binding events to membrane proteins*. Proceedings of the National Academy of Sciences of the United States of America, 2018. **115**(12): p. 2976-2981.

83. Nury, H., et al., *Structural basis for lipid-mediated interactions between mitochondrial ADP/ATP carrier monomers*. FEBS Lett., 2005. **579**(27): p. 6031-6036.

84. Schwall, C.T., V.L. Greenwood, and N.N. Alder, *The stability and activity of respiratory Complex II is cardiolipin-dependent*. Biochim. Biophys. Acta, 2012. **1817**(9): p. 1588-1596.

85. Laganowsky, A., et al., *Membrane proteins bind lipids selectively to modulate their structure and function*. Nature, 2014. **510**(7503): p. 172-175.

86. Sharpe, H.J., T.J. Stevens, and S. Munro, *A comprehensive comparison of transmembrane domains reveals organelle-specific properties*. Cell, 2010. **142**(1): p. 158-169.

87. Bolla, J.R., et al., *Direct observation of the influence of cardiolipin and antibiotics on lipid II binding to MurJ*. Nat. Chem., 2018. **10**(3): p. 363-371.

88. Li, Y., B.J. Orlando, and M. Liao, *Structural basis of lipopolysaccharide extraction by the LptB2FGC complex*. Nature, 2019. **567**(7749): p. 486-490.

89. Tang, X., et al., *Cryo-EM structures of lipopolysaccharide transporter LptB2FGC in lipopolysaccharide or AMP-PNP-bound states reveal its transport mechanism*. Nat Commun, 2019. **10**(1): p. 4175.

90. Owens, T.W., et al., *Structural basis of unidirectional export of lipopolysaccharide to the cell surface*. Nature, 2019. **567**(7749): p. 550-553.

91. Qiao, S., et al., *Structural basis for lipopolysaccharide insertion in the bacterial outer membrane*. Nature, 2014. **511**(7507): p. 108-111.

92. Dong, H., et al., *Structural basis for outer membrane lipopolysaccharide insertion*. Nature, 2014. **511**(7507): p. 52-56.

93. Acharya, K.R. and M.D. Lloyd, *The advantages and limitations of protein crystal structures*. Trends Pharmacol. Sci., 2005. **26**(1): p. 10-14.

94. Montelione, G.T., et al., *Protein NMR spectroscopy in structural genomics*. Nat. Struct. Biol., 2000. **7 Suppl**: p. 982-985.

95. Monaco, S., et al., *Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF*. J. Appl. Crystallogr., 2013. **46**(Pt 3): p. 804-810.

96. Carroni, M. and H.R. Saibil, *Cryo electron microscopy to determine the structure of macromolecular complexes*. Methods, 2016. **95**: p. 78-85.

97. Higgins, M.K. and S.M. Lea, *On the state of crystallography at the dawn of the electron microscopy revolution*. Curr. Opin. Struct. Biol., 2017. **46**: p. 95-101.

98. Freyer, M.W. and E.A. Lewis, *Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions*. Methods Cell Biol., 2008. **84**: p. 79-113.

99. Jerabek-Willemsen, M., et al., *Molecular interaction studies using microscale thermophoresis*. Assay Drug Dev. Technol., 2011. **9**(4): p. 342-353.

100. Olaru, A., et al., *Surface plasmon resonance (SPR) biosensors in pharmaceutical analysis*. Crit. Rev. Anal. Chem., 2015. **45**(2): p. 97-105.

101. Saliba, A.E., I. Vonkova, and A.C. Gavin, *The systematic analysis of protein-lipid interactions comes of age*, in *Nat. Rev. Mol. Cell Biol.* 2015: England. p. 753-61.

102. Kalinin, S., et al., *A toolkit and benchmark study for FRET-restrained high-precision structural modeling*. Nat. Methods, 2012. **9**(12): p. 1218-1225.

103. Micsonai, A., et al., *Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy*. Proc. Natl. Acad. Sci. USA, 2015. **112**(24): p. E3095-103.

104. Puchner, E.M. and H.E. Gaub, *Force and function: probing proteins with AFM-based force spectroscopy*. Curr. Opin. Struct. Biol., 2009. **19**(5): p. 605-614.

105. Lossl, P., M. van de Waterbeemd, and A.J. Heck, *The diverse and expanding role of mass spectrometry in structural and molecular biology*. Embo J., 2016. **35**(24): p. 2634-2657.

106. Robinson, C.V., *From molecular chaperones to membrane motors: through the lens of a mass spectrometrist*. Biochem. Soc. Trans., 2017. **45**(1): p. 251-260.

107. Katta, V., Chait, B. T., *Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry*. J. Am. Chem. Soc., 1991. **113**: p. 8534-8535.

108. Ganem, B., Y.T. Li, and J.D. Henion, *Detection of noncovalent receptor-ligand complexes by mass spectrometry*. J. Am. Chem. Soc., 2002. **113**: p. 6294-6296.

109. Breuker, K. and F.W. McLafferty, *Stepwise evolution of protein native structure with electrospray into the gas phase, 10(-12) to 10(2) s*. Proc. Natl. Acad. Sci. USA, 2008. **105**(47): p. 18145-18152.

110. Benesch, J.L., *Collisional activation of protein complexes: picking up the pieces*. J Am Soc Mass Spectrom, 2009. **20**(3): p. 341-8.

111. Sobott, F., et al., *Subunit exchange of multimeric protein complexes. Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry*. J. Biol. Chem., 2002. **277**(41): p. 38921-38929.

112. Benesch, J.L., F. Sobott, and C.V. Robinson, *Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry*. Anal. Chem., 2003. **75**(10): p. 2208-2214.

113. Barrera, N.P. and C.V. Robinson, *Advances in the mass spectrometry of membrane proteins: from individual proteins to intact complexes*. Annu. Rev. Biochem., 2011. **80**: p. 247-271.

114. Morgner, N., et al., *Mass spectrometry--from peripheral proteins to membrane motors*. J. Mol. Biol., 2012. **423**(1): p. 1-13.

115. Benesch, J.L., et al., *Protein complexes in the gas phase: technology for structural genomics and proteomics*. Chem Rev, 2007. **107**(8): p. 3544-67.
116. Madler, S., E. Boeri Erba, and R. Zenobi, *MALDI-ToF mass spectrometry for studying noncovalent complexes of biomolecules*. Top Curr. Chem., 2013. **331**: p. 1-36.
117. Spengler, B., Kirsch, D., Kaufmann, R., Karas, M., Hillenkamp, F., Giessmann, U., *The detection of large molecules in matrix-assisted UV-laser desorption*. Rap. Comm. Mass Spec., 1990. **4**(9): p. 301-305.
118. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules*. Science, 1989. **246**(4926): p. 64-71.
119. Ambrose, S., et al., *Native Desorption Electrospray Ionization Liberates Soluble and Membrane Protein Complexes from Surfaces*. Angew. Chem. Int. Ed. Engl., 2017. **56**(46): p. 14463-14468.
120. Takats, Z., et al., *Mass spectrometry sampling under ambient conditions with desorption electrospray ionization*. Science, 2004. **306**(5695): p. 471-473.
121. Mager, F., et al., *LILBID-mass spectrometry of the mitochondrial preprotein translocase TOM*. J. Phys. Condens. Matter, 2010. **22**(45): p. 454132.
122. Laganowsky, A., et al., *Mass spectrometry of intact membrane protein complexes*. Nat. Protoc., 2013. **8**(4): p. 639-651.
123. Cech, N.B. and C.G. Enke, *Practical implications of some recent studies in electrospray ionization fundamentals*. Mass Spectrom Rev, 2001. **20**(6): p. 362-87.
124. Rohner, T.C.L., N.; & Girault, H. H., *Electrochemical and theoretical aspects of electrospray ionisation*. Physical Chemistry Chemical Physics, 2004. **6**(12): p. 3056.
125. Tang, K. and R.D. Smith, *Physical/chemical separations in the break-up of highly charged droplets from electrosprays*. J. Am. Soc. Mass Spectrom., 2001. **12**(3): p. 343-347.
126. Wilm, M.S.M.M., *Electrospray and Taylor-cone theory, Dole's beam of macromolecules at last?* . International Journal of Mass Spectrometry and Ion Processes, 1994. **136**(2-3): p. 167-180.
127. Wilm, M. and M. Mann, *Analytical properties of the nanoelectrospray ion source*. Anal. Chem., 1996. **68**(1): p. 1-8.
128. de Hoffmann, E. and V. Stroobant, *Mass Spectrometry: Principles and Applications*. 3rd ed. 2007.
129. McKay, A.R., et al., *Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes*. J. Am. Chem. Soc., 2006. **128**(35): p. 11433-11442.
130. Iribarne, J.V.T., B. A., *On the evaporation of small ions from charged droplets*. J. Chem. Phys., 1976. **64**(6): p. 2287-294.

131. Konermann, L., et al., *Unraveling the mechanism of electrospray ionization*. Anal. Chem., 2013. **85**(1): p. 2-9.

132. Iavarone, A.T. and E.R. Williams, *Mechanism of charging and supercharging molecules in electrospray ionization*. J. Am. Chem. Soc., 2003. **125**(8): p. 2319-2327.

133. Mehmood, S., T.M. Allison, and C.V. Robinson, *Mass spectrometry of protein complexes: from origins to applications*. Annu. Rev. Phys. Chem., 2015. **66**: p. 453-474.

134. de la Mora, J.F., *Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism*. Anal. Chim. Acta, 2000. **406**(1): p. 93-104.

135. Felitsyn, N., Peschke, M. & Kebarle, P., *Origin and number of charges observed on multiply-protonated native proteins produced by ESI*. . Int. J. Mass spectrom., 2002. **219**: p. 39-62.

136. Konermann, L., A.D. Rodriguez, and J. Liu, *On the formation of highly charged gaseous ions from unfolded proteins by electrospray ionization*. Anal. Chem., 2012. **84**(15): p. 6798-6804.

137. March, R.E., *Quadrupole ion traps*. Mass Spectrom. Rev., 2009. **28**(6): p. 961-989.

138. Hofstadler, S.A., Bakhtiar, R. & Smith, R. D. , *Electrospray Ionization Mass Spectroscopy: Part I. Instrumentation and Spectral Interpretation*. . J. Chem. Educ. , 1996. **73**: p. A82.

139. Henchman, M.S., C., *Understanding the quadrupole mass filter through computer simulation*. . J. Chem. Educ., 1998. **75**: p. 1049.

140. Sobott, F., et al., *A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies*. Anal. Chem., 2002. **74**(6): p. 1402-1407.

141. Douglas, D.J.F., A. J.; Mao, D., *Linear ion traps in mass spectrometry*. Mass Spectrom. Rev., 2005. **24**(1): p. 1 - 29.

142. Chernushevich, I.V., A.V. Loboda, and B.A. Thomson, *An introduction to quadrupole-time-of-flight mass spectrometry*. J. Mass Spectrom., 2001. **36**(8): p. 849-865.

143. Cotter, R.J., *Time-of-flight mass spectrometry for the structural analysis of biological molecules*. Anal. Chem., 1992. **64**(21): p. 1027a-1039a.

144. Benesch, J.L., et al., *Quadrupole-time-of-flight mass spectrometer modified for higher-energy dissociation reduces protein assemblies to peptide fragments*. Anal. Chem., 2009. **81**(3): p. 1270-1274.

145. Doroshenko, V.M. and R.J. Cotter, *Ideal velocity focusing in a reflectron time-of-flight mass spectrometer*. J. Am. Soc. Mass Spectrom., 1999. **10**(10): p. 992-999.

146. Patel, A. 2017: https://www.slideshare.net/AashishPatel14/mass-spectrometry-finalpptx.

147. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis*. Anal. Chem., 2000. **72**(6): p. 1156-1162.

148. Kingdon, K.H., *A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures*. Phys. Rev., 1923. **21**(4): p. 408.

149. Hu, Q., et al., *The Orbitrap: a new mass spectrometer*. J Mass Spectrom, 2005. **40**(4): p. 430-43.

150. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis*. Anal Chem, 2000. **72**(6): p. 1156-62.

151. Hu, Q., et al., *The Orbitrap: a new mass spectrometer*. J. Mass Spectrom., 2005. **40**(4): p. 430-443.

152. Fraser, G.W., *The ion detection efficiency of microchannel plates (MCPs)*. Int. J. Mass Spectrom., 2002. **215**: p. 13-30.

153. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review*. Anal. Bioanal. Chem., 2007. **389**(4): p. 1017-1031.

154. Jennings, K.R., *The changing impact of the collision-induced decomposition of ions on mass spectrometry*. International Journal of Mass Spectrometry, 2000. **200**(1-3).

155. Barrera, N.P., et al., *Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions*. Nat. Methods, 2009. **6**(8): p. 585-587.

156. Urner, L.H., et al., *Modular detergents tailor the purification and structural analysis of membrane proteins including G-protein coupled receptors*. Nat. Commun., 2020. **11**(1): p. 564.

157. van Berkel, W.J., et al., *Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry*. .Protein Sci, 2000. **9**(3): p. 435-439.

158. Rostom, A.A., et al., *Detection and selective dissociation of intact ribosomes in a mass spectrometer*. Proc. Natl. Acad. Sci. USA, 2000. **97**(10): p. 5185-5190.

159. Tito, M.A., et al., *Electrospray Time-of-Flight Mass Spectrometry of the Intact MS2 Virus Capsid*. J. Am. Chem. Soc., 2000. **122**(14): p. 3550-3551.

160. Rose, R.J., et al., *High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies*. Nature Methods, 2012. **9**(11): p. 1084-1086.

161. Gault, J., et al., *High-resolution mass spectrometry of small molecules bound to membrane proteins*. Nat. Methods, 2016. **13**(4): p. 333-336.

162. Liko, I., et al., *Mass spectrometry guided structural biology*. Curr. Opin. Struct. Biol., 2016. **40**: p. 136-144.

163. Aguilar-Gurrieri, C., et al., *Structural evidence for Nap1-dependent H2A–H2B deposition and nucleosome assembly*. EMBO Journal, 2016. **35**(13): p. 1465-1482.

164. Rozen, S., et al., *CSNAP Is a Stoichiometric Subunit of the COP9 Signalosome*. Cell Reports, 2015. **13**(3): p. 585-598.

165. Jackson, V.A., et al., *Super-complexes of adhesion GPCRs and neural guidance receptors*. Nature Communications, 2016. **7**.

166. Abbas, Y.M., et al., *Structure of V-ATPase from the mammalian brain*. Science, 2020. **367**(6483): p. 1240-1246.

167. Liko, I., et al., *Dimer interface of Bovine cytochrome c oxidase is influenced by local posttranslational modifications and lipid binding*. Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(29): p. 8230-8235.

168. Su, C.C., et al., *MmpL3 is a lipid transporter that binds trehalose monomycolate and phosphatidylethanolamine*. Proc. Natl. Acad. Sci. USA, 2019. **116**(23): p. 11241-11246.

169. Gupta, K., et al., *Identifying key membrane protein lipid interactions using mass spectrometry*. Nature Protocols, 2018. **13**(5): p. 1106-1120.

170. Kitova, E. N., Soya, N. and Klassen, S., *Identifying Specific Small-Molecule Interactions Using Electrospray Ionization Mass Spectrometry*. Anal Chem, 2011. **83**(13): p. 5160-5167.

171. Wang, W., Kitova, E. N., and Klassen, S., *Nonspecific Protein-Carbohydrate Complexes Produced by Nanoelectrospray Ionization. Factors Influencing Their Formation and Stability*. Anal Chem, 2005. **77**(10): p. 3060-3071.

172. Laganowsky, A., et al., *Membrane proteins bind lipids selectively to modulate their structure and function*. Nature, 2014. **510**(7503): p.172–175.

173. Shimon, L., Sharon, M., Horovitz, A., *A Method for Removing Effects of Nonspecific Binding on the Distribution of Binding Stoichiometries: Application to Mass Spectroscopy Data*. Biophys J, 2010. **99**(5): p. 1645-1649.

174. Bolla, J. R., et al., *A Mass-Spectrometry-Based Approach to Distinguish Annular and Specific Lipid Binding to Membrane Proteins*. Angew. Chem. Int. Ed. 2020. **59**(9): p. 3523-3528

175. Wales, T.E. and J.R. Engen, *Hydrogen exchange mass spectrometry for the analysis of protein dynamics*. Mass Spectrometry Reviews, 2006. **25**(1): p. 158-170.

176. Zhang, Z. and D.L. Smith, *Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation*. Protein Science, 1993. **2**(4): p. 522-531.

177. Niimura, N., et al., *Hydrogen and hydration in proteins*. Cell Biochemistry and Biophysics, 2004. **40**(3): p. 351-369.

178. Hvidt, A. and S.O. Nielsen, *Hydrogen Exchange in Proteins*, in *Advances in Protein Chemistry Volume 21*. 1966. p. 287-386.

179. Konermann, L., X. Tong, and Y. Pan, *Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches*. J Mass Spectrom, 2008. **43**(8): p. 1021-36.

180. Konermann, L., J. Pan, and Y.H. Liu, *Hydrogen exchange mass spectrometry for studying protein structure and dynamics*. Chem Soc Rev, 2011. **40**(3): p. 1224-34.

181. Weis, D.D., et al., *Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis*. J Am Soc Mass Spectrom, 2006. **17**(11): p. 1498-1509.

182. Deng, Y. and D.L. Smith, *Identification of unfolding domains in large proteins by their unfolding rates*. Biochemistry, 1998. **37**(18): p. 6256-6262.

183. Ferraro, D.M. and A.D. Robertson, *EX1 Hydrogen Exchange and Protein Folding*. Biochemistry, 2004. **43**(3): p. 587-594.

184. Maier, C.S., M.I. Schimerlik, and M.L. Deinzer, *Thermal denaturation of Escherichia coli thioredoxin studied by hydrogen/deuterium exchange and electrospray ionization mass spectrometry: Monitoring a two-state protein unfolding transition*. Biochemistry, 1999. **38**(3): p. 1136-1143.

185. Swint-Kruse, L. and A.D. Robertson, *Temperature and pH dependences of hydrogen exchange and global stability for ovomucoid third domain*. Biochemistry, 1996. **35**(1): p. 171-180.

186. Houliston, R.S., et al., *pH and urea dependence of amide hydrogen-deuterium exchange rates in the β-trefoil protein hisactophilin*. Biochemistry, 2002. **41**(4): p. 1182-1194.

187. Arrington, C.B. and A.D. Robertson, *Microsecond to minute dynamics revealed by EX1-type hydrogen exchange at nearly every backbone hydrogen bond in a native protein*. Journal of Molecular Biology, 2000. **296**(5): p. 1307-1317.

188. Merkle, P.S., et al., *Substrate-modulated unwinding of transmembrane helices in the NSS transporter LeuT*. Sci. Adv., 2018. **4**(5): p. eaar6179.

189. Zhou, J., et al., *Conformational dynamics of 1-deoxy-d-xylulose 5-phosphate synthase on ligand binding revealed by H/D exchange MS*. Proc Natl Acad Sci U S A, 2017. **114**(35): p. 9355-9360.

190. Nielsen, A.K., et al., *Substrate-induced conformational dynamics of the dopamine transporter*. Nat Commun, 2019. **10**(1): p. 2714.

191. Wales, T.E. and J.R. Engen, *Partial unfolding of diverse SH3 domains on a wide timescale*. Journal of Molecular Biology, 2006. **357**(5): p. 1592-1604.

192. Li, M.J., M. Guttman, and W.M. Atkins, *Conformational dynamics of P-glycoprotein in lipid nanodiscs and detergent micelles reveal complex motions on a wide time scale*. J Biol Chem, 2018. **293**(17): p. 6297-6307.

193. Xiao, H., et al., *Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment*. Protein Sci., 2005. **14**(2): p. 543-557.

194. Man, P., et al., *Defining the interacting regions between apomyoglobin and lipid membrane by hydrogen/deuterium exchange coupled to mass spectrometry*. J. Mol. Biol., 2007. **368**(2): p. 464-472.

195. Yan, X., et al., *Hydrogen/deuterium exchange and mass spectrometric analysis of a protein containing multiple disulfide bonds: Solution structure of recombinant macrophage colony stimulating factor-beta (rhM-CSFbeta)*. Protein Sci., 2002. **11**(9): p. 2113-2124.

196. Zhang, H.M., et al., *Simultaneous reduction and digestion of proteins with disulfide bonds for hydrogen/deuterium exchange monitored by mass spectrometry*. Analytical Chemistry, 2010. **82**(4): p. 1450-1454.

197. Engen, J.R. and T.E. Wales, *Analytical Aspects of Hydrogen Exchange Mass Spectrometry*, in *Annual Review of Analytical Chemistry*. 2015. p. 127-148.

198. Merkle, P.S., et al., *Substrate-modulated unwinding of transmembrane helices in the NSS transporter LeuT*. Science Advances, 2018. **4**(5).

199. Ahdash, Z., et al., *HDX-MS reveals nucleotide-dependent, anti-correlated opening and closure of SecA and SecY channels of the bacterial translocon*. Elife, 2019. **8**.

200. Du, Y., et al., *Assembly of a GPCR-G Protein Complex*. Cell, 2019. **177**(5): p. 1232-1242 e11.

201. Zhang, X., et al., *Dynamics of the β2-adrenergic G-protein coupled receptor revealed by hydrogen-deuterium exchange*. Analytical Chemistry, 2010. **82**(3): p. 1100-1108.

202. Reading, E., et al., *Interrogating Membrane Protein Conformational Dynamics within Native Lipid Compositions*. Angew. Chem. Int. Ed. Engl., 2017. **56**(49): p. 15654-15657.

203. Moller, I.R., et al., *Conformational dynamics of the human serotonin transporter during substrate and drug binding*. Nat Commun, 2019. **10**(1): p. 1687.

204. Martens, C., et al., *Integrating hydrogen-deuterium exchange mass spectrometry with molecular dynamics simulations to probe lipid-modulated conformational changes in membrane proteins*. Nat. Protoc., 2019. **14**(11): p. 3183-3204.

205. Belov, M., *Method and Apparatus for Mass Spectrometry of Macromolecular Complexes*. 2015: US Patent Application 20150340213.

206. Ben-Nissan, G., et al., *Triple-Stage Mass Spectrometry Unravels the Heterogeneity of an Endogenous Protein Complex*. Anal. Chem., 2017. **89**(8): p. 4708-4715.

207. van de Waterbeemd, M., et al., *High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles*. Nat. Methods, 2017. **14**(3): p. 283-286.

208. Fiorentino, F., et al., *The Different Effects of Substrates and Nucleotides on the Complex Formation of ABC Transporters*. Structure, 2019. **27**(4): p. 651-659.e3.

CHAPTER 2

The different effects of substrates and nucleotides on the complex formation of ABC transporters

Adapted from a paper published in *Structure*, **2019**, 27(4):651-659

Francesco Fiorentino,[1] Jani Reddy Bolla,[1] Shahid Mehmood,[1,2] Carol V. Robinson[1,3,*]

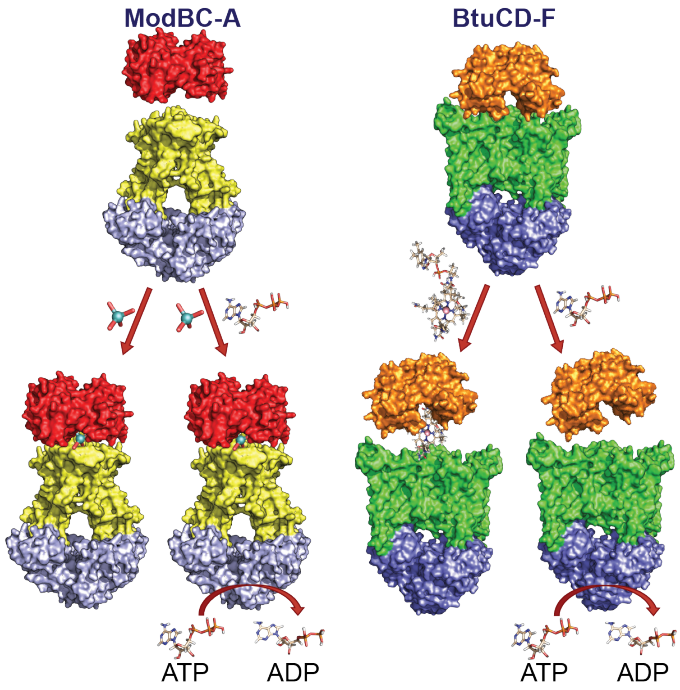
[1]Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK

[2]Present address: The Francis Crick Institute, 1 Midland Road, London NW1 1ST, UK

[3]Lead Contact

*Correspondence: carol.robinson@chem.ox.ac.uk

Graphical Abstract



Acknowledgments

The research in the Robinson laboratory is supported by an MRC programme grant (MR/N020413/1), an ERC Advanced Grant ENABLE (695511) and a Wellcome Trust Investigator Award (104633/Z/14/Z). F.F. holds a SABS CDT Studentship supported by the EPSRC and the MRC (EP/L016044/1). BtuCD and ModBC plasmids were a kind gift from Kaspar Locher (ETH Zurich). A. Howes and D. Wu (University of Oxford) are acknowledged for proofreading the manuscript and help in lipidomics analysis, respectively.

Author contributions

F.F., J.R.B., S.M. and C.V.R. designed the study. F.F., J.R.B. prepared BtuF and ModA constructs. F.F. expressed and purified all proteins and performed all MS measurements. F.F. analysed the data with the help of J.R.B., and S.M. F.F. and C.V.R. wrote the paper with support from J.R.B and S.M.

Declaration of interest

C.V.R. is a co-founder and consultant of OMass Therapeutics.

Abstract

The molybdate importer (ModBC-A of *A. fulgidus*) and the vitamin B_{12} importer (BtuCD-F of *E. coli*) are members of the type I and type II ABC importer families. Here we study the influence of substrate and nucleotide binding on complex formation and stability. Using native mass spectrometry we show that the interaction between the periplasmic substrate-binding protein (SBP) ModA and the transporter ModBC is dependent upon binding of molybdate. By contrast, vitamin B_{12} disrupts interactions between the transporter BtuCD and the SBP BtuF. Moreover, while ATP binds cooperatively to BtuCD-F, and acts synergistically with vitamin B_{12} to destabilise the BtuCD-F complex, no effect is observed for ATP binding on the stability of ModBC-A. These observations not only highlight the ability of MS to capture these importer-SBP complexes but allow us to add molecular detail to proposed transport mechanisms.

2.1 Introduction

ATP-binding cassette (ABC) transporters are a super-family of membrane proteins that couple the hydrolysis of ATP to the translocation of a diverse range of molecules across lipid bilayers of biological membranes. The basic ABC transporter architecture consists of four core domains: two transmembrane domains (TMDs), and two nucleotide-binding domains (NBDs). The two TMDs interact to form a central path, allowing the substrate to be either imported into, or exported out of, the cytoplasm. The NBDs contain ATP-binding pockets with ATPase activity: ATP hydrolysis provides the energy required for conformational changes that enable passage of substrates through the TMDs [1]. ABC exporters are typically homodimers, in which TMDs and NBDs are encoded by single protein chain [2]. In contrast, importers have a tetrameric structure with their TMDs and NBDs expressed as separate proteins, present exclusively in internal membranes of prokaryotes and archaea. Whereas NBDs are structurally similar and contain highly conserved motifs, the TMDs of different transporters vary widely in sequence and structure, a feature linked to different mechanisms of transport.

Depending on the fold of the TMDs, ABC importers can be classified as type I, type II or energy-coupling factor transporters identified recently [3-5]. Type I and type II importers require a periplasmic substrate-binding protein (SBP) for their function. SBPs capture substrates and deliver them to the transporters allowing the passage of the ligand through the lipid bilayer [6]. The molybdate transporter ModBC-A [7] along with the maltose transporter $MalFGK_2$ [8] and vitamin B_{12} transporter BtuCD-F [9] are the most studied type I and type II importers, respectively. In addition to different numbers of TMDs, another feature that distinguishes the two classes of importers is a lack of substrate binding sites (SBS) in the TMDs of type II importers [10].

Type I importers mediate the translocation of small molecules such as ions, amino acids, small peptides, mono- and oligosaccharides, and are characterised by a broad range of binding affinities [11]. Structural and biophysical studies have shown that type I importers employ an alternating-access transport model in which conformational changes expose the substrate binding site to alternate faces either side of the membrane [12]. In the resting state, the transporter has an

inward-facing conformation, which changes to outward-facing once the substrate-loaded binding protein interacts with the periplasmic face. This change of conformation allows substrate release from the SBP, and passage through the TMDs into the SBS. ATP hydrolysis then powers the opening of the NBDs, and the substrate is released into the cytoplasm. Previous studies have indicated that there are subtle differences in the transport mechanism, even within the same class of importers. The type I maltose ABC importer $MalFGK_2$ was shown to bind to its cognate SBP both in the presence and in the absence of maltose, and in both apo and ADP-bound states [13]. By contrast, surface plasmon resonance (SPR) studies indicate that while molybdate is required to initiate the binding of the SBP in *Archaeoglobus fulgidus* ModBC-A, this is not the case for *Haemophilus influenzae* ModBC-A, resulting in substrate-dependent mechanistic differences between the two transporters [14].

Type II importers are responsible for the high-affinity uptake of larger molecules, such as haem and vitamin B_{12}, as well as smaller ones, including like inorganic ions [15, 16] and possess a more complex mechanism of transport than type I importers. The X-ray structure of the entire BtuCD-F complex revealed critical interactions between the transporter and the SBP and shed light on the post-translocation conformation of the complex [17]. The crystal structures of AMP-PNP-bound BtuCD-F and BtuCD, provide mechanistic details of the ATP-powered release of vitamin B_{12} into the cytoplasm [10, 18]. These structures suggest that upon ATP binding the TM helices undergo a rearrangement to create a cleft, which temporarily is able to accommodate substrate. The absence of substrate-bound co-crystal structures, however, limits detailed knowledge of the transport cycle. Moreover, recent studies provided insights into the effects of ATP binding and hydrolysis on complex formation which led to differing conclusions. Specifically, microscale thermophoresis assays showed that ATP binding and hydrolysis strengthen the interaction between BtuCD and BtuF [18]. By contrast, single-molecule fluorescence imaging revealed that membrane-embedded BtuCD formed a stable complex with BtuF and found that addition of ATP alters the interaction between the two proteins [19].

Given the fact that the molecular details of the coupled effect of substrate and ATP binding remain unclear [20-22] we set out to resolve these effects for these two importers using

native mass spectrometry (MS), a technique that enables the analysis of the subunit stoichiometry and ligand binding properties of membrane proteins encapsulated in detergents micelles [23, 24]. Native MS can also be used to report on the population of different complexes in solution by comparing the intensities of ligand-bound charge states within a particular protein system [25-28]

The first native MS experiments were developed and applied to BtuCD and revealed that the non-covalent tetramer could be preserved after release from detergent micelles with post-translational modifications and nucleotide binding intact [23]. At that time, it was not possible however to capture interactions with the substrate binding protein BtuF. To do this, MS conditions needed to be further developed to maintain interactions between transmembrane subunits with both cytoplasmic and periplasmic subunits. To maintain interactions either side of the membrane we optimised further MS conditions and selected ModBC-A from *Archaeoglobus fulgidus* and BtuCD-F from *Escherichia coli* as representatives of type I and type II importer families respectively (Fig. 2.1.1a). Following purification of wild-type ModBC, BtuCD, and their respective SBPs, we investigated their lipid binding properties, as well as the effect of substrate, and ATP binding and hydrolysis on complex formation and dissociation of these ABC importers.

2.2 Results

2.2.1 Substrates have different effects on the formation of type I and type II transporter complexes

We expressed and purified BtuCD following established protocols (see Methods section) [9]. We then performed a detergent screen to establish conditions that preserve the complex for native MS [29]. Selecting detergents from different classes, we found that tetraethylene glycol monooctyl ether (C_8E_4) both solubilised the membrane protein complex and yielded well-resolved charge states corresponding to the mass of intact BtuCD (Fig. S2.1a and Table 2.1). Adducts were also observed consistent with the presence of co-purified lipids, particularly lipopolysaccharides (LPS), an observation typical of the native mass spectra of membrane proteins [27, 30] and in particular of ABC transporters [31, 32].

Having optimised the MS conditions to detect BtuCD, we then purified BtuF [17] (Fig. S2.3a) and studied its binding properties to BtuCD. To form the BtuCD-F complex, we incubated BtuF with BtuCD and introduced this solution to the mass spectrometer. We observed a new charge state series corresponding to the intact BtuCD-F (>90%) complex (Fig. 2.1.1b). Associated with the charge state series for the intact complex is a series of satellite peaks indicating that the complex has retained binding to membrane lipids.

For ModBC-A, we expressed and purified ModBC as described [7] and performed a detergent screen to obtain well-resolved mass spectra. We selected C_8E_4 to replicate the MS conditions used for BtuCD and similarly the experimental mass is consistent with the intact tetramer of ModBC with various phospholipids and LPS associated (Fig. S2.1b). Despite the fact that BtuCD and ModBC were purified using different detergents the same type of bulk lipids from *E. coli* remained bound to the transporters. (Fig. S2.2).

To record the mass spectrum of the entire ModBC-A complex we purified ModA [7] (Fig. S2.3b) and incubated it with a solution containing ModBC. The most abundant species was ModBC; only a small population had formed a complex with ModA. Moreover, mass spectra also indicated the presence of free ModA in solution, confirming that the complex had not formed (Fig. S2.1c) and implying that additional components are required to achieve full complexation in solution. We reasoned that substrate may be required and added an excess of sodium molybdate to a solution of ModA before buffer exchange and then incubated this solution with ModBC. Under these conditions, we observed a different pattern of charge states whose deconvolution yields a mass consistent with the presence of 100% full ModBC-A complex (Fig. 2.1.1c). Considering the deconvoluted masses of the individual proteins and the mass of the full complex we calculated a difference of 168 Da, consistent with the presence of molybdate ions bound to the intact complex (calculated as 160 Da). Additional satellite peaks, similar to those observed for ModBC, were assigned to lipid-bound species, associated with the full ModBC-A·MoO_4^{2-} complex.

We suggest that type I importers, exemplified here by ModBC-A, have low affinity for their cognate SBP and only the addition of the substrate triggers the formation of the full

complex. By contrast, type II importers, exemplified here by BtuCD-F, bind readily to their cognate SBP, even in the absence of substrate. Taken together, these results imply that the two transporters examined here have different affinities towards their respective SBPs.

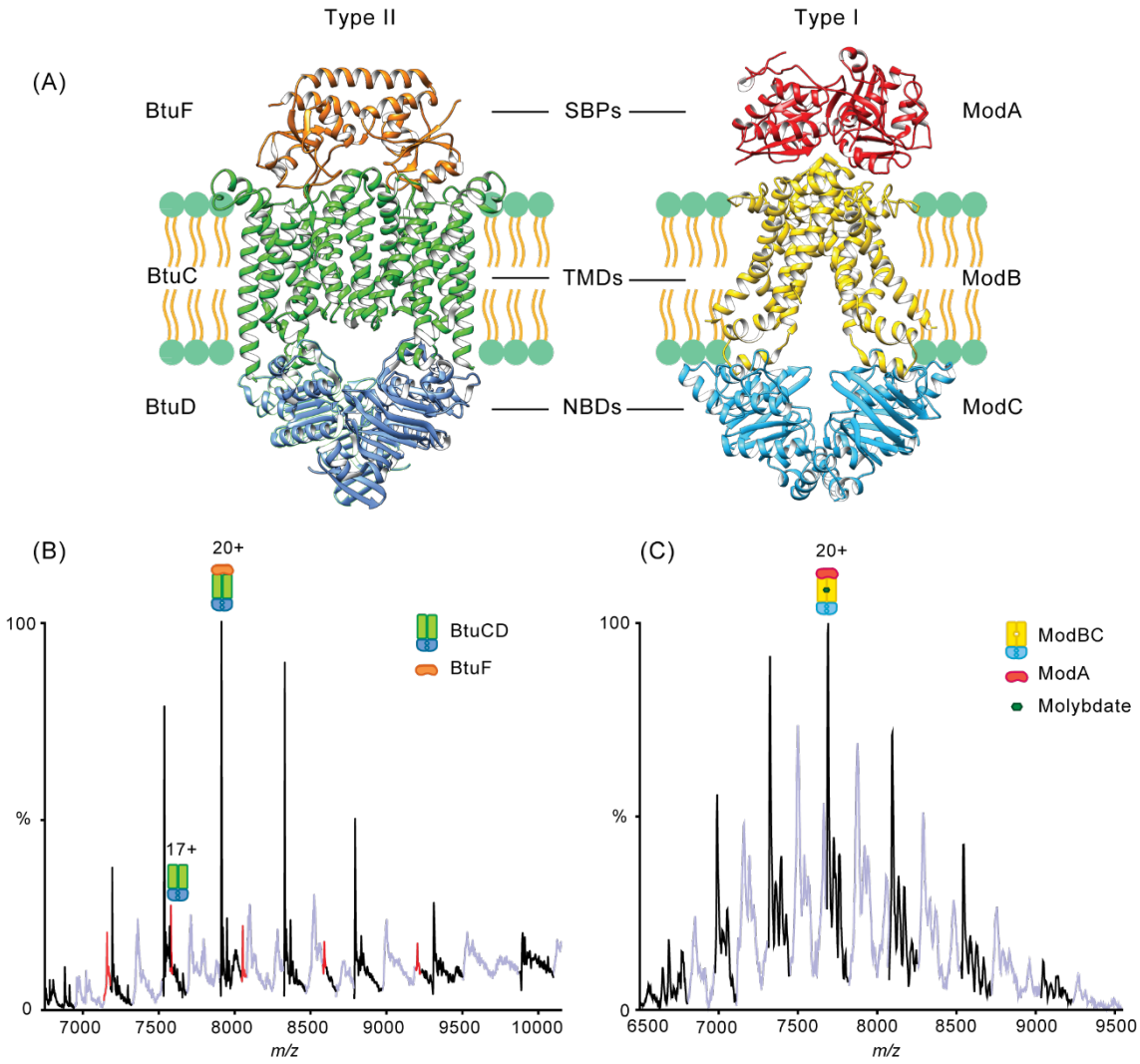


Figure 2.1 Structures of BtuCD-F and ModBC-A and their respective mass spectra. (a) X-ray crystal structures of BtuCD-F (left, PDB ID 2QI9) and ModBC-A (right, PDB ID 2ONK) visualised using PyMol (Schrödinger). (b) Mass spectrum of BtuCD-F reveals a charge state series consistent with the mass of the tetrameric transporter ($BtuC_2D_2$) bound to the SBP (BtuF) ($BtuC_2D_2$–F) a smaller population of uncomplexed transporter ($BtuC_2D_2$ red peaks) is also apparent. Different lipid species, primarily LPS, are bound to the transporter both in the presence and the absence of the SBP (light blue peaks). (c) Mass spectrum of ModBC-A shows a charge state series consistent with the tetrameric transporter bound to SBP and molybdate with different

lipid-bound species primarily LPS (light blue peaks). Theoretical and observed molecular masses of all species are presented in Table 2.1.

2.2.2 Vitamin B_{12} and cyanocobinamide destabilise the BtuCD-BtuF interaction

Having demonstrated that BtuCD and BtuF form a stable complex, the next question was to examine whether vitamin B_{12} had an effect on the stability of the BtuCD-F complex. We supplemented BtuF with an excess of vitamin B_{12} and recorded a mass spectrum of the BtuF:vitamin B_{12} complex (Fig. 2.2a, inset). We then incubated the vitamin B_{12}-loaded BtuF with a solution of BtuCD. Interestingly, vitamin B_{12} binding was not detected in either complex (BtuCD or BtuCD-F) (Fig. 2.2a). Absence of vitamin B_{12} is in line with the previous observation that the TMDs of BtuCD lack a high affinity substrate binding site [10]. The binding of vitamin B_{12} is likely transient, making it difficult to detect with biophysical methods. We noticed, however, an increase in the intensity of charge states corresponding to uncomplexed BtuCD (45%), compared with BtuCD-F in the absence of substrate wherein only 10% of BtuCD was observed (Fig. 2.2a and Fig.1b, red peaks). The presence of vitamin B_{12}, though not detected directly within the complex, destabilises the interactions between the SBP and the transporter.

Recently, it has been reported that cyanocobinamide, a closely related analogue of vitamin B_{12}, is imported by BtuCD-F and the co-crystal structure of BtuF-cyanocobinamide has been determined [33]. To further investigate this interaction using native MS we added an excess amount of dicyanocobinamide to BtuF alone and observed two series of charge states, corresponding to apo-BtuF and BtuF-cyanocobinamide (Fig. 2.2b, inset), confirming binding to BtuF in line with previous results [33]. We then incubated dicyanocobinamide-loaded BtuF with a solution of BtuCD and observed that the BtuCD and BtuCD-F peak intensities had increased to similar value as those observed when vitamin B_{12} was added (Fig. 2.2b), indicating that the complex was destabilised.

The destabilisation effect of vitamin B_{12} on the complex was further confirmed by using a molecule with a different structure, vitamin B_9 (folic acid, Fig. 2.2b). Under analogous conditions to those used for vitamin B_{12}, vitamin B_9 did not affect the stability of the complex

BtuCD-F with little dissociation of BtuCD (15%) (Fig. 2.2c). The absence of any effects of vitamin B_9 highlights the specificity of the effect for vitamin B_{12} and cyanocobinamide, which, upon binding, induce dissociation of BtuF. We conclude therefore that type II importer BtuCD-F exists in solution as a stable complex with its SBP, even in the absence of ligands. Addition of substrates to this type II importer has a disruptive effect on protein interactions leading to reduced stability of the substrate binding protein.

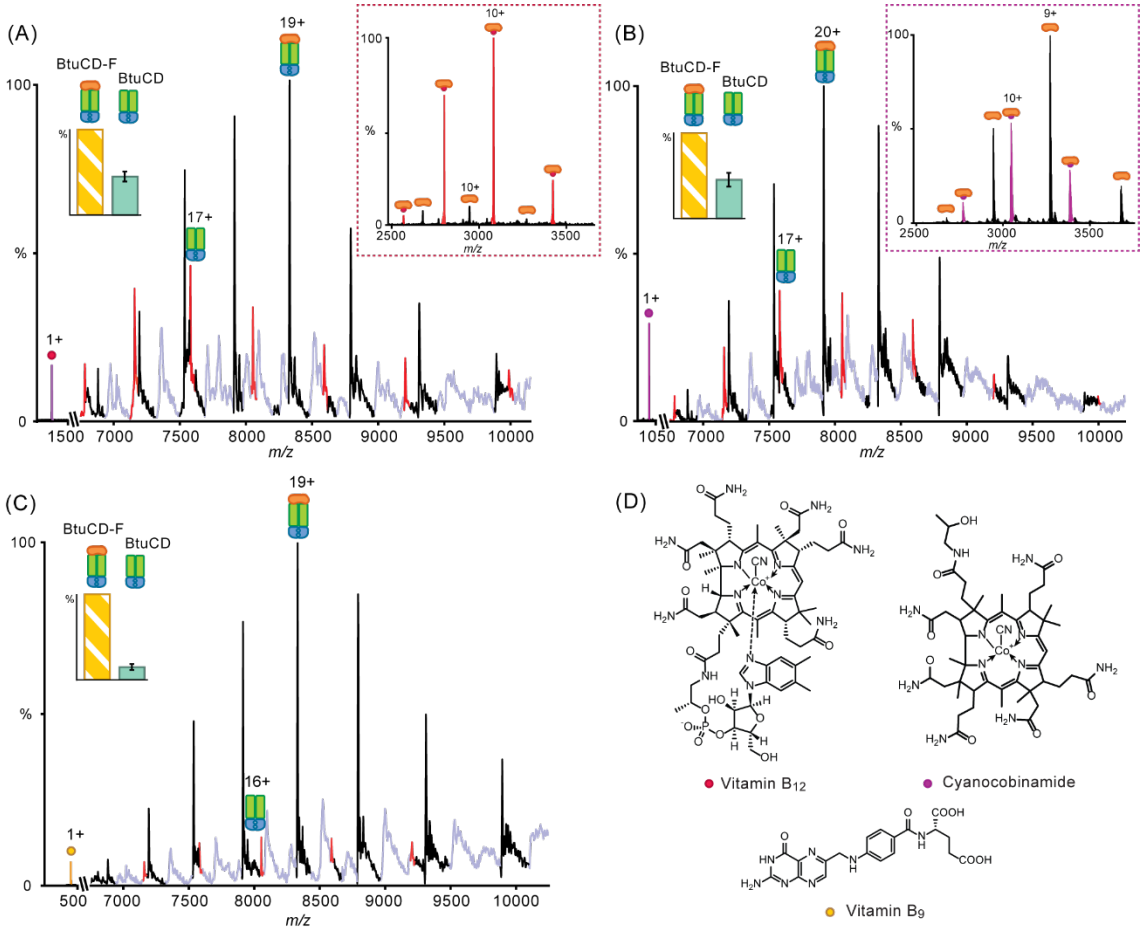


Figure 2.2 BtuCD-F is destabilised by vitamin B_{12} and cyanocobinamide. (a) Mass spectrum of BtuCD-F recorded after addition of 2 mM vitamin B_{12} to BtuF before buffer exchange. An increase in the intensity of the charge states corresponding to uncomplexed BtuCD (red peaks 45% with respect to BtuCD-F 100%) is observed in the presence of vitamin B_{12} (purple peak) compared with in the absence of ligand (cf Fig. 2.2.1b). The inset shows the mass spectrum of BtuF-Vitamin B_{12} complex after addition of 2mM vitamin B_{12}. (b) Mass spectrum of BtuCD-F after addition of 2 mM dicyanocobinamide to BtuF before buffer exchange. The ratio of BtuCD

(red peaks 44% with respect to BtuCD-F 100%) is closely similar to the extent of complex formation in the presence of vitamin B_{12}. The inset shows the mass spectrum of BtuF-Cyanocobinamide complex after addition of 2mM dicyanocobinamide. (c) Mass spectrum of BtuCD-F after addition of 2 mM vitamin B_9 to BtuF before buffer exchange. The ratio of BtuCD (red peaks 15% with respect to BtuCD-F 100%) in the presence of vitamin B_9 is closely similar to the extent of complex formation in the absence of ligand. Data are represented as mean ± standard deviation (*n*=3). (d) Chemical structures of the molecules used for the binding experiments.

2.2.3 ATP binding acts synergistically with the substrate to decrease BtuCD-BtuF stability

To gain further insights into the vitamin B_{12} translocation mechanism we studied the effects of ATP binding and turnover on complex stability. We added an excess of ATP/Mg^{2+} to a solution of BtuCD, prior to buffer exchange into MS buffer, and incubated it with a solution of BtuF following the same protocol as above. After a 10-minute incubation time the mass spectrum indicates the presence of three BtuCD-F species in solution: the apo form and two nucleotide-bound states with one and two ADP molecules bound, referred as apo-BtuCD-F, BtuCD-F·1ADP and BtuCD-F·2ADP, respectively (Fig. 2.3a and Fig. S2.4a). In order to distinguish the effect of binding from hydrolysis, we carried out analogous experiments in the presence of AMP-PNP, a non-hydrolysable ATP analogue. We observed apo-BtuCD-F as well as a population assigned to BtuCD-F·2AMP-PNP (Fig. 2.3b and S2.4b). Interestingly, while BtuCD-F·2ADP is predominant, apo-BtuCD-F is more abundant than BtuCD-F·1ADP, and in the presence of AMP-PNP no population of a BtuCD-F bound to a single AMP-PNP species was detected. This binding pattern of nucleotides was found to be insensitive to longer incubation periods suggesting that equilibrium had been reached.

Overall, our findings suggest that binding of the first nucleotide increases the affinity of the protein for the second one, a characteristic feature of cooperative binding [34]. Moreover, in both cases, BtuCD was bound to the two different nucleotides in a similar fashion to that observed for the full complex. The overall population of free BtuCD and corresponding adducts was found

to increase after addition of nucleotides, with the BtuCD/BtuCD-F ratio being in the range of 40% to 45% in the presence of ATP or AMP-PNP (Fig. 2.3a and b, Fig. S2.4a and b, Fig. S2.5). Such an increase in BtuCD-related charge states is analogous to the situation observed for vitamin B_{12}, thus indicating that a destabilising effect is also induced by nucleotide binding.

To study the combined effect of substrate and nucleotide, we added solutions of ATP and vitamin B_{12} simultaneously to the BtuCD-F complex. Interestingly, we noticed that the increase in the intensity of free BtuCD charge states was even more significant when both ligands are present (Figs. S2.4a and c, Fig. S2.5), implying a synergistic effect of binding both ATP and vitamin B_{12}. Our data also show that the cooperativity in ATP binding observed above is retained in the presence of vitamin B_{12} (Fig. S2.4c). The results obtained from the addition of ATP, AMP-PNP and ATP-vitamin B_{12} suggest a mechanism whereby ATP binding causes a conformational change that leads to a decrease in affinity for the substrate binding protein. While the AMP-PNP experiments cannot exclude a contribution from ATP hydrolysis, in destabilising the complex *in vivo,* they do demonstrate that nucleotide binding alone is sufficient to destabilise the interaction between BtuCD and BtuF.

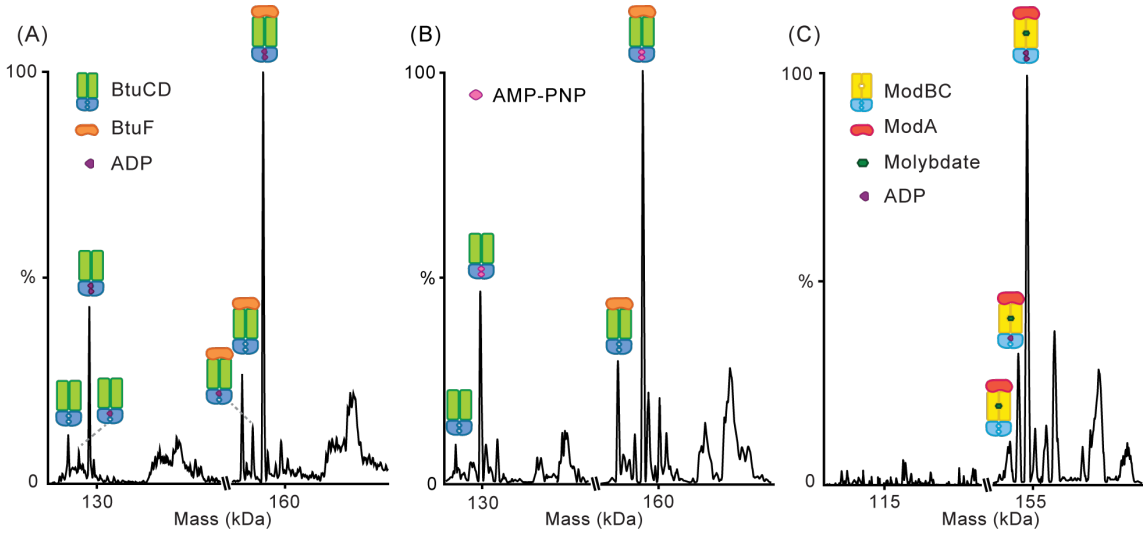


Figure 2.3 Effect of nucleotides on BtuCD-F and ModBC-A complex formation. (a) Deconvoluted masses of the spectrum of BtuCD-F recorded after addition of 5 mM ATP to BtuCD before buffer exchange. The presence of ADP-bound species of both BtuCD-F and BtuCD are observed with similar ratios. The relative abundance of BtuCD in this spectrum is

40% of the BtuCD-F signal. (b) Deconvoluted masses of the spectrum of BtuCD-F recorded after addition of 5 mM AMP-PNP to BtuCD before buffer exchange. Also in this case, AMP-PNP-bound species are detected. The relative abundance of BtuCD in this spectrum is 43% of the BtuCD-F signal. (c) Deconvoluted masses of the spectrum of ModBC-A recorded following the addition of 10 μM ATP to ModBC after buffer exchange. ModBC-A species bound to zero, one and two ADP molecules are observed and no ModBC alone is detected.

2.2.4 ATP binding does not alter ModBC-ModA interactions

To evaluate the impact of ATP binding and hydrolysis on the ModBC-A complex we first added an excess of ATP to a solution of ModBC before exchanging into an MS compatible buffer. We then incubated this solution with ModA. The resulting mass spectrum did not display any nucleotide binding, implying that all ATP had been removed during buffer exchange. To observe nucleotide binding, we added excess ATP to ModA pre-loaded with sodium molybdate and recorded the mass spectrum under optimised conditions. Three populations of ModBC-A·MoO_4^{2-} are observed: apo form and the species bound to one and two ADP molecules (Fig. 2.3c and Fig. S2.6). The distribution of the three species is different from BtuCD-F; indeed, the apo form is less abundant, thus no cooperativity is observed in ATP/ADP binding to ModBC-A·MoO_4^{2-}, indicating little or no cooperativity of the two ATP binding events. Moreover, we did not observe additional charge states corresponding to ModBC, implying that ATP did not influence the ModBC-A/ModBC interactions significantly.

Based on our results we propose that nucleotide binding has a different impact on type I and type II transporters examined here. In the type I transporter, ATP did not perturb interactions between the transporter and its cognate SBP. Nucleotide binding was found to be independent of the effect of substrate, the latter being pivotal for complex formation. By contrast, ATP binding and hydrolysis modulates interactions between the transporter and the SBP in the transport cycle of the type II importer. In addition, ATP binding follows different patterns in the two complexes: while we observed cooperative ATP binding in the case of BtuCD-F, we did not observe cooperativity for ModBC-A. While we cannot rule out cooperativity of nucleotide binding for

type I importers we speculate that differences in the TMD structures of the two transporters may influence interactions and cross-talk within the NBDs.

2.3 Discussion

In the current study, we employed native MS to observe membrane protein complexes developing the approach further to capture ABC importers bound to their cognate substrate binding proteins. Capturing these complexes allowed us to explore the influence of ligands on their stability. Specifically, we showed that the two ABC importers were associated with similar lipids even when extracted and purified in different detergents. Moreover, the cohort of bound lipids did not change significantly when binding of ATP or substrates or during complex formation. These observations suggest that the transporters do not have a specific preference for the bulk lipids that surround them, in contrast to the flippase TmrAB investigated previously, wherein negatively charged lipids bound tightly to the protein [31]. We also demonstrated that binding of BtuCD-F substrates, vitamin B_{12} and cyanocobinamide, led to disruption of the BtuCD-BtuF complex, whereas molybdate binding is required to promote docking of ModA onto ModBC. The dissociation of BtuCD-F complex is explained by the fact that once vitamin B_{12}-loaded BtuF docks on to BtuCD insufficient space is available to accommodate the bulky B_{12} substrate (Fig. 2.2b). In addition, the ModBC-A complex is stable, regardless of the presence of ATP, while BtuCD-F is destabilised by binding of ATP or its non hydrolysable analogue AMP-PNP. Moreover, the disruptive effect of vitamin B_{12} is enhanced by the presence of nucleotides, consistent with the ligands acting synergistically to promote substrate release. We also highlighted further differences in ATP binding mechanisms between type I and type II transporters. While no significant cooperativity was detected for ModBC-A, when binding to nucleotides, for BtuCD-F the first ATP binding event facilitates the second in accordance with typical cooperative binding models.

On the basis of our results, and with contributions from previous structural and biophysical data [10, 14, 17-22], we present two different transport models for type I and type II ABC importers.

For the type I importer ModBC-A our data show that in the absence of substrate the extent of SBP binding onto the transporter is very low; essentially, there is no complex formation (Fig. 2.4a, state I). Binding of substrate results in increased complex formation, triggering the docking of the SBP onto the transporter, trapping the substrate and beginning of the transport cycle (Fig. 2.4a, state II).

This interaction is then thought to initiate a conformational rearrangement, converting the TMDs from inward to outward-facing conformation [35] and likely forming a tunnel which allows the translocation of the substrate from the SBP to the SBS, located in the TMDs (Fig. 2.4a, state III and Fig. 2.2.1b). Subsequent ATP binding induces a further rearrangement to an inward-facing conformation causing dissociation of the substrate from the SBS and its release into the cytoplasm (Fig. 2.4a, state IV). The translocation of the substrate and dissociation of ADP from the NBDs reset the transporter to the initial state (Fig. 2.4a, state I). Together these steps are consistent with the alternating access transport mechanism in which conformational changes expose the substrate binding site to alternate faces either side of the membrane [12, 36].

The type II importer (BtuCD-F) studied here follows a different cycle wherein the transporter complex with the SBP is stable in solution, even in the absence of substrate (Fig. 2.4b state I and Fig. 2.2.1b). Following the cooperative binding of ATP, the TMDs undergo conformational changes leading to a disruption of the interactions with the SBP. The reduction in the level of complexation does not necessarily imply that appreciable dissociation takes place *in vivo,* rather it suggests that the complex is undergoing conformational changes whereby binding of SBP to the TMDs is perturbed. We propose a model whereby the SBP-bound transporter partially opens upon nucleotide binding consistent with the increased levels of uncomplexed transporter observed by MS (Fig. 2.4b state II) [19]. Moreover, earlier studies showed that type II importers are characterised by a high level of futile ATP hydrolysis, with a high ATP:vitamin B_{12} transport ratio [37, 38].

Once the complex opens, substrate is bound (Fig. 4b, state III) and partially accommodated in a cleft formed by the helices present in the TMDs [18] as a consequence of nucleotide binding. However, the presence of substrate and the full occupancy of nucleotide

binding sites has a synergistic effect rendering the full complex more susceptible to dissociation, leading to the dissociation of ADP and Pi, which triggers the subsequent translocation of the substrate into the cytoplasm (Fig. 4b, state IV). The presence of substrate decreases the stability of the complex; therefore, a critical factor in the release of the substrate to the cytoplasm is the achievement of the state of minimal energy represented by the SBP-bound transporter (Fig. 4b state I).

Over the past years, biochemical techniques such as surface plasmon resonance [14, 39] and single-molecule FRET [40, 41] have been used to study the binding of ligands on ABC transporters, providing valuable information about their influence on the mechanism of transport.

The MS approach we have applied here, therefore, complements earlier studies [18-20] by providing evidence of the role of substrate and ATP in complex formation and substrate transport across the lipid bilayer. These results further indicate that there is an essential difference between type I and type II ABC importers examined here. Moreover, ABC transporters have a pivotal role in bacterial life and therefore represent optimal targets for the development of antimicrobial drugs [42]. Given the increasing number of multidrug-resistant bacteria, there is an urgent need for new strategies targeting less explored pathways. Hence, better understanding of ABC importers mechanisms may prove useful for the development of novel therapeutics.

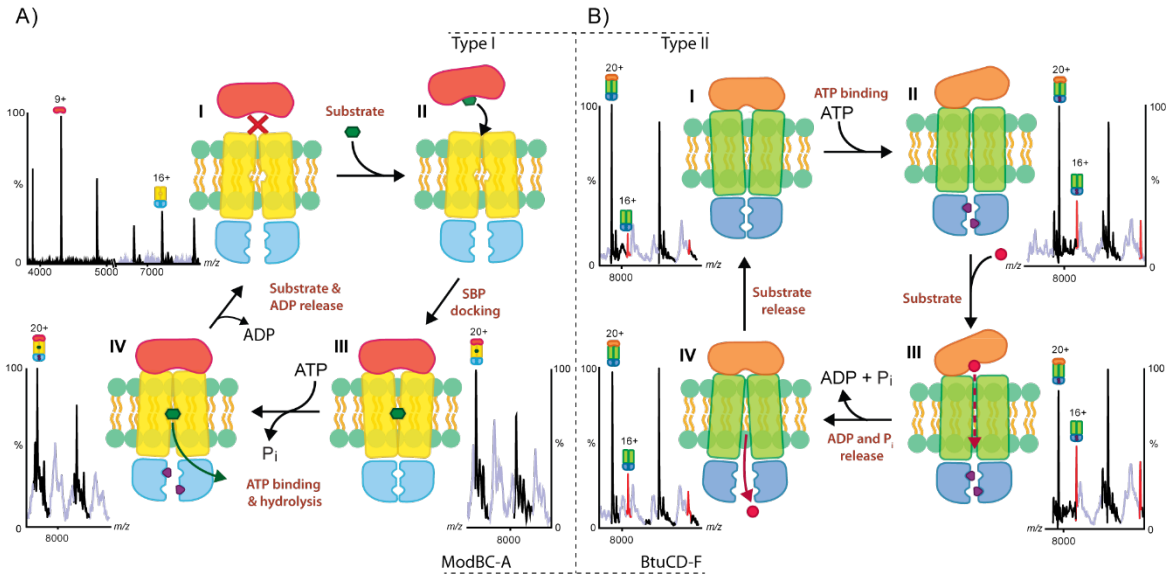


Figure 2.4 Transport models for type I importer ModBC-A and type II importer BtuCD-F. (a) Transport model for ModBC-A. The formation of the complex in the absence of molybdate

is unfavourable (state I). Binding of molybdate triggers the docking of ModA onto ModBC (state II). Following a rearrangement of the TMDs, molybdate moves towards the SBS (state III). ATP binding and hydrolysis triggers a further conformational change that translocate the substrate into the cytoplasm (state IV). The release of ADP resets the transporter to the resting state (state I). (b) Transport model for BtuCD-F. BtuCD and BtuF form a stable complex even in the absence of vitamin B_{12} (state I). ATP binds cooperatively, leading to a partial displacement of BtuF from BtuCD (state II). When vitamin B_{12} is available, a transient BtuCD-F·2ADP·B_{12} complex is formed (state III). This unstable conformation is relaxed through ADP and Pi dissociation, followed by vitamin B_{12} release (state IV) leading to the resetting of the transporter.

Species	Theoretical mass (Da)	Observed mass (Da)	Difference (Da)
BtuCD	128839	128837	-2
ModBC	115011	115015	+4
BtuF	29414	29412	-2
BtuF·Vitamin B12	30770	30768	-2
BtuF·Cyanocobinamide	30430	30427	-3
ModA	38584	38583	-1
BtuCD-F	158253	158251	-2
BtuCD-F·2ADP	159107	159105	-2
BtuCD-F·2AMP-PNP	159313	159320	+7
ModBC-A·MoO42-	153763	153766	+3
ModBC-A·MoO42-·2ADP	154615	154623	+8

Table 2.1 Theoretical and observed masses of all species presented in the current study. Theoretical masses were calculated assuming the loss of the first methionine on BtuCD, ModBC and ModA.

2.4 Supplementary Information

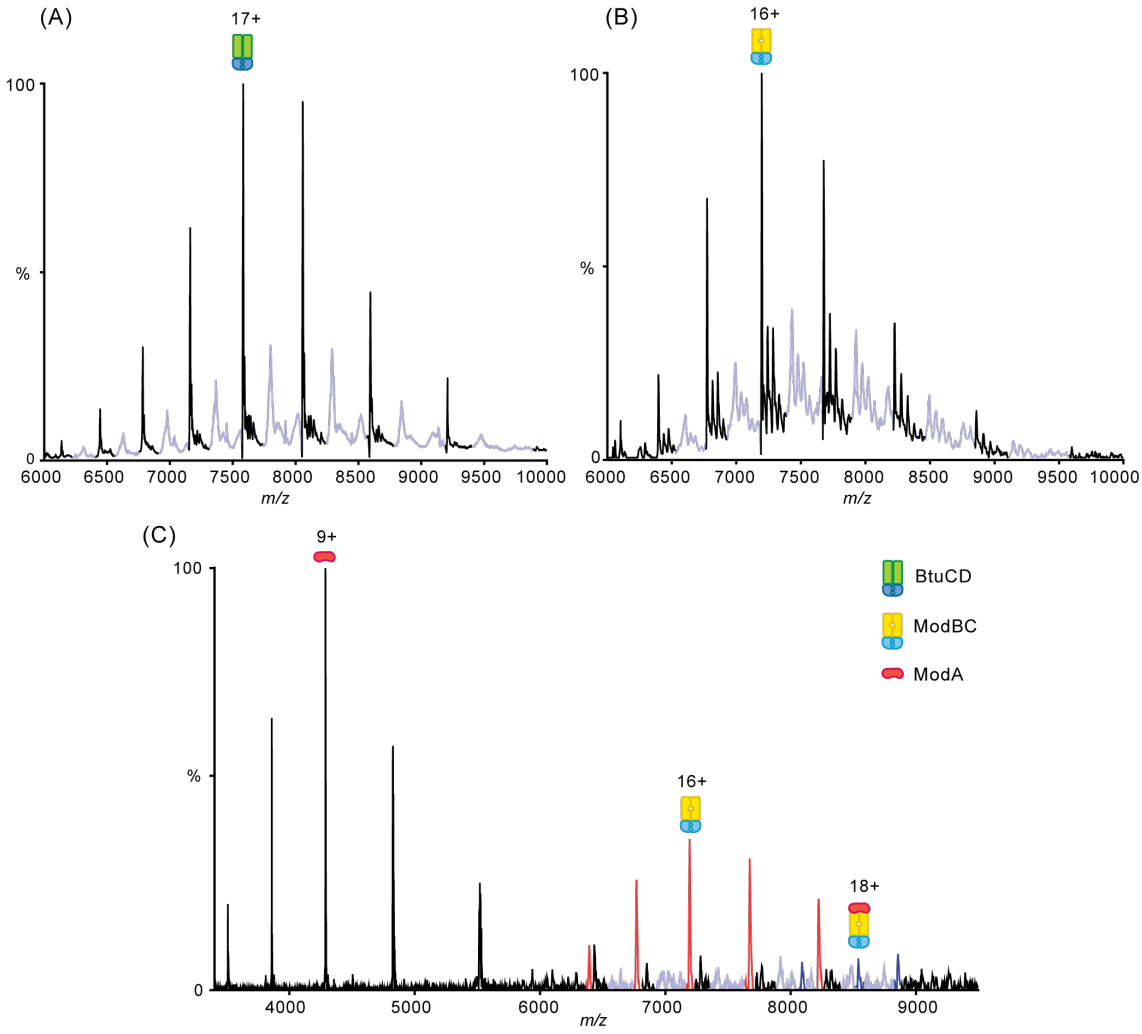


Figure S2.1 Mass spectra of uncomplexed BtuCD and ModBC, Related to Figure 1. (a) Mass spectrum of BtuCD in 0.50% (w/v) C_8E_4 reveals a charge state series consistent with the tetrameric transporter with different lipid-bound species, mostly LPS (light blue peaks). (b) Mass spectrum of ModBC in 0.50% (w/v) C_8E_4 reveals a charge state series consistent with the tetrameric transporter bound to different lipid species, mainly LPS (light blue peaks). (c) Mass spectrum of ModBC and ModA acquired in the absence of molybdate shows only a low population of the full ModBC-A complex (blue peaks).

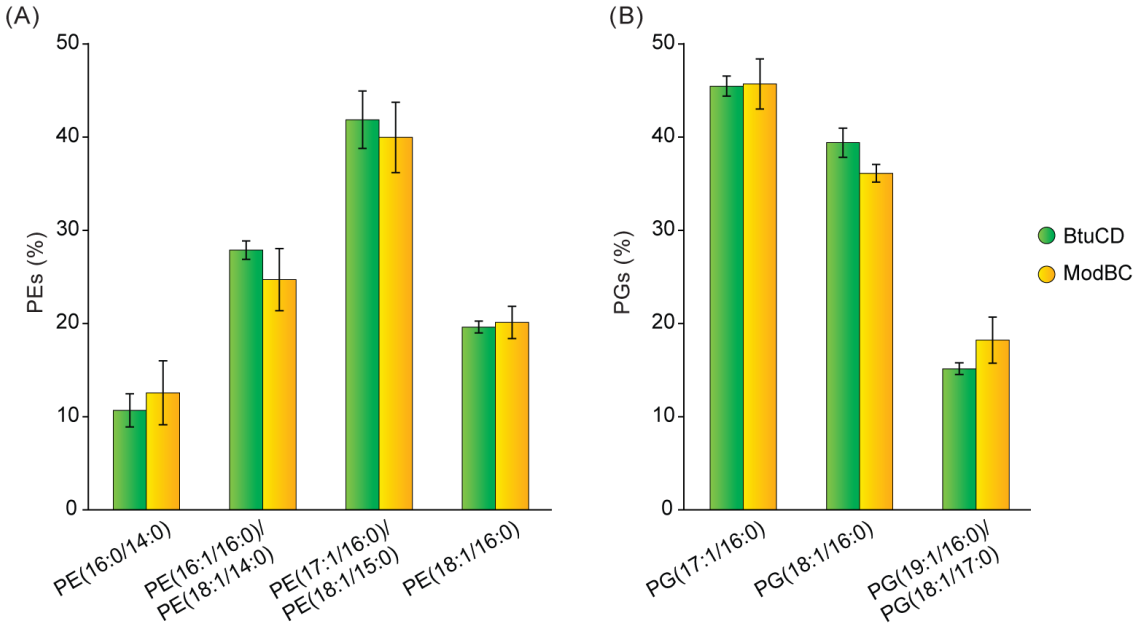


Figure S2.2 Lipidomics analysis for BtuCD (green) and ModBC (yellow), Related to Figure 1. (a) Relative abundance of phosphatidylethanolamines (PEs). (b) Relative abundance of phosphatidylglycerols (PGs). Error bars represent standard deviations (*n*=3).

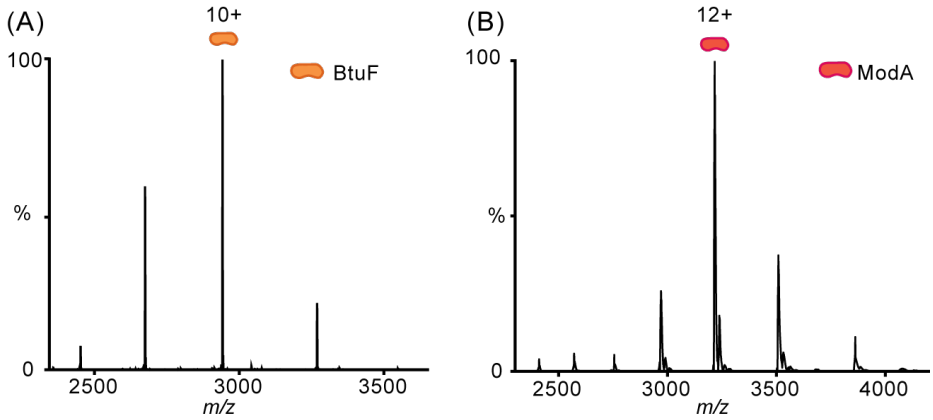


Figure S2.3 Mass spectra of BtuF (a) and ModA (b), Related to Figure 1. Both proteins were separately purified and analysed in the absence of any detergent. Theoretical and observed masses and recorded in Table 2.1.

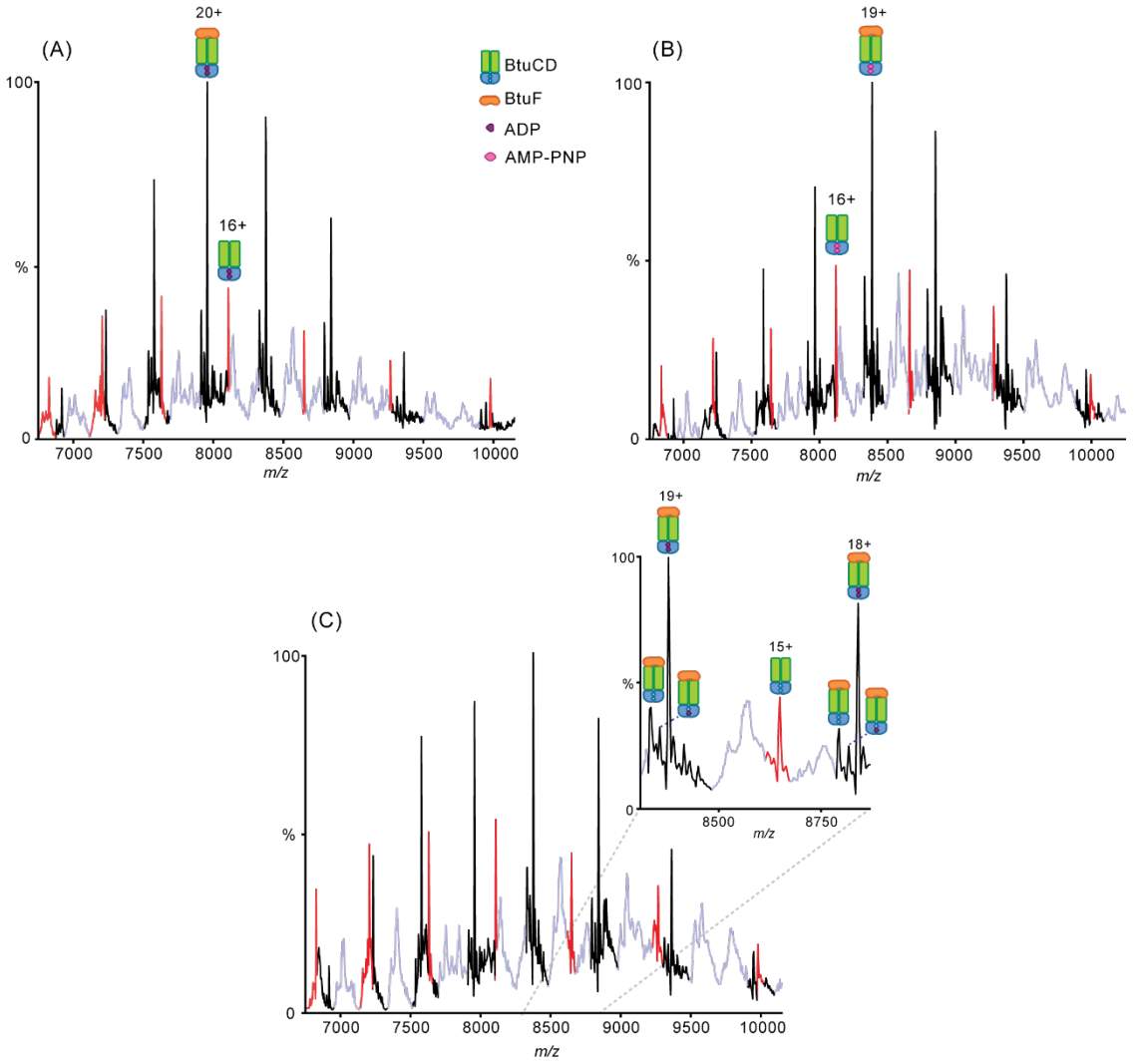


Figure S2.4 Nucleotides destabilise BtuCD-BtuF interaction, Related to Figure 3. (a) Mass spectrum of BtuCD-F recorded after addition of 5 mM ATP to BtuCD before buffer exchange. ADP-bound species are detected with binding indicative of cooperativity along with an increase in the intensity of the BtuCD charge states (red peaks). (b) Mass spectrum of BtuCD-F recorded after addition of 5 mM AMP-PNP to BtuCD before buffer exchange. Also in this case, AMP-PNP-bound species are detected with binding indicative of cooperativity along with an increase in the intensity of the BtuCD charge states (red peaks). (c) Mass spectrum of BtuCD-F recorded after addition of 5 mM ATP to BtuCD and 2mM vitamin B_{12} to BtuF before buffer exchange. The spectrum shows a further increase in the BtuCD charge states evident of a synergistic effect upon the transport of B_{12} and the binding of nucleotides (red peaks).

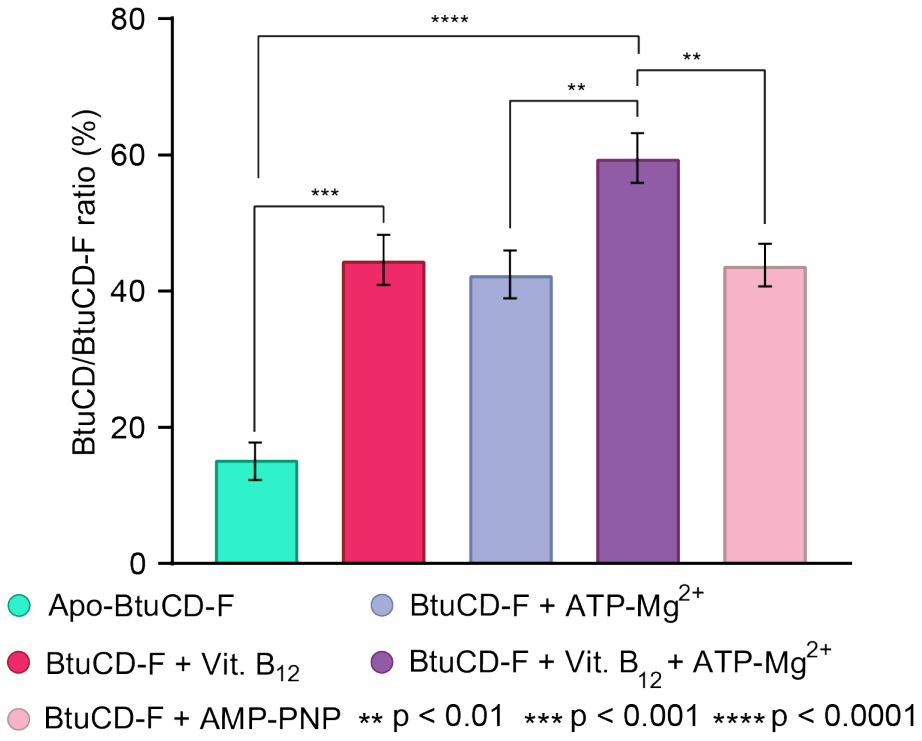


Figure S2.5 BtuCD/BtuCD-F molar ratio in the presence of vitamin B_{12} and/or nucleotides compared to apo state. Bar charts summarising BtuCD/BtuCD-F molar ratio for apo-BtuCD-F (green), BtuCD-F + vitamin B12 (red), BtuCD-F + ATP (light blue), BtuCD-F + vitamin B12 + ATP (violet), BtuCD-F + AMP-PNP (pink). The bar charts indicate significant increase of free BtuCD following addition of substrates or nucleotides compared to apo protein. The concentration of free BtuCD is further significantly augmented in the presence of both ligands.

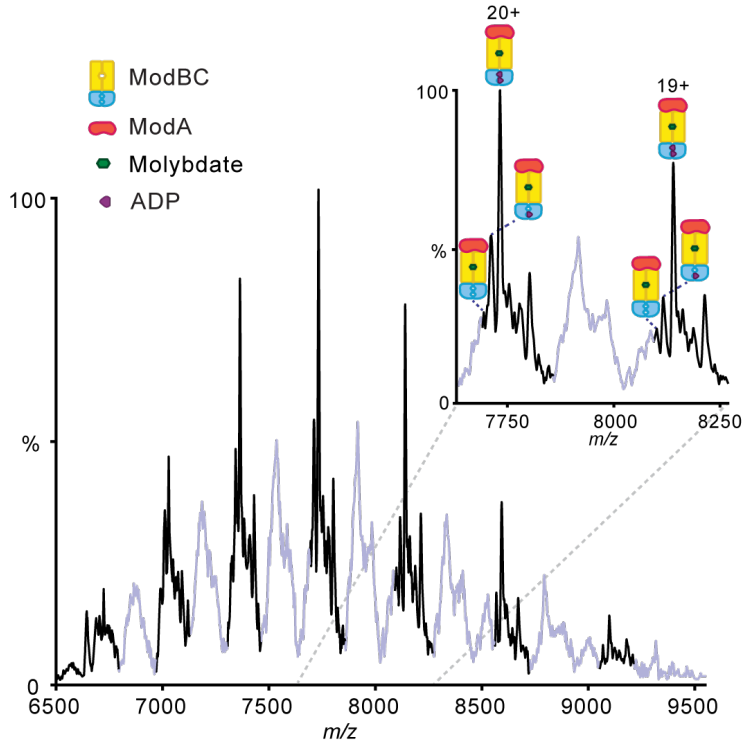


Figure S2.6 Mass spectrum of ADP-bound form of ModBC-A, Related to Figure 3. Mass spectrum of ModBC-A following the addition of 10 µM ATP to ModBC after buffer exchange.

2.5 Materials and methods

2.5.1 Plasmid preparation and protein expression

ModA plasmid was obtained by inserting the gene fragment encoding the predicted mature part of the periplasmic binding-protein (amino acids 32-342) into a modified pET-28a expression vector (Novagen) between BamHI and XhoI cloning sites, using an In-Fusion cloning kit (Clonetech). The modified pET-28a contained an N-terminal decahistidine affinity tag followed by a TEV protease cleavage site. The BtuF plasmid was constructed by inserting the gene fragment containing amino acids 23-266 into pET-22b expression vector (Novagen) between NcoI and XhoI cloning sites. The plasmids used for over-expression of *E. coli* BtuCD and *A. fulgidus* ModBC were a kind gift from Kaspar Locher (ETH Zurich). All plasmids were amplified by transforming them into *E. coli* Stellar Competent Cells (Takara) and the DNA sequences were verified by Sanger sequencing.

For each protein, the plasmid was transformed in *E. coli* BL21(DE3) (New England Biolabs). Several colonies were inoculated into 100 ml LB media and grown overnight at 37 °C. One litre of LB in 2 litre shaker flasks was inoculated with 7 ml of overnight culture and grown at 37 °C until the culture reached OD600 nm (OD600) between 0.6 and 0.8. Isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to the culture at a final concentration of 0.5 mM and grown for 3 h at 37 °C. Cells were collected by centrifugation at 5,000xg for 10 min at 4 °C. Cell pellets were resuspended in buffer containing 150mM NaCl, 20 mM Tris (pH 7.5) and stored at -80 °C.

2.5.2 Purification of soluble proteins

ModA. Resuspended cells were thawed and supplemented with an EDTA-free protease inhibitor cocktail (Roche). The cell suspension was passed several times through an M-110 PS microfluidizer (Microfluidics) at 15,000 psi. Insoluble material was pelleted by centrifugation at 20,000xg for 20 min at 4 °C. The supernatant was filtered before loading onto a 5 ml HisTrap-HP column (GE Healthcare, Piscataway, NJ) equilibrated in 200 mM NaCl, 20 mM Hepes (pH

7.5), 10% glycerol and 20 mM imidazole. After the clarified supernatant was loaded, the column was initially washed with 50 ml of 200 mM NaCl, 20 mM Hepes (pH 7.5), 10% glycerol and 20 mM imidazole, and washed again with 50 ml of 200 mM NaCl, 20 mM Hepes (pH 7.5), 10% glycerol and 80 mM imidazole. The bound protein was eluted with 200 mM NaCl, 20 mM Hepes (pH 7.5), 10% glycerol and 500 mM imidazole. Peak fractions were pooled, incubated with TEV protease and dialysed against 200 mM NaCl, 20 mM Hepes (pH 7.5), and 10% glycerol. The protein was then concentrated and loaded onto the Superdex 200 size exclusion chromatography (SEC) column in 200 mM NaCl, 20 mM Hepes (pH 7.5), and 10% glycerol.

BtuF. The protein was purified following the same protocol as ModA, except that the final buffer contains also 10 mM DTT.

2.5.3 Purification of membrane proteins

Resuspended cells were thawed and supplemented with EDTA-free protease inhibitor cocktail (Roche). The cell suspension was passed several times through an M-110 PS microfluidizer (Microfluidics) at 15,000 psi. Insoluble material was pelleted by centrifugation at 20,000xg for 20 min at 4 °C. The supernatant was ultracentrifuged at (200,000xg for 1h), and the membrane fractions were collected. Membranes were resuspended in ice-cold buffer containing 150 mM NaCl, 20 mM Tris (pH 7.5) and 20% glycerol. Resuspended membranes were used either directly or flash frozen in liquid nitrogen and stored at -80 °C.

ModBC. The protein was solubilised from the membrane fraction with 150 mM NaCl, 20 mM Tris (pH 7.5), 20% glycerol, 1% n-Dodecyl-β-D-Maltopyranoside (DDM, Anatrace) and 1% octaethylene glycol monododecyl ether ($C_{12}E_8$, Anatrace) for 1h at 4 °C. Extracted membrane proteins were clarified by centrifugation at 20,000xg for 20 min at 4 °C. Supernatant was filtered before loading onto a 5 ml HisTrap-HP column (GE Healthcare, Piscataway, NJ) equilibrated in 150 mM NaCl, 20 mM Tris (pH 7.5), 10% glycerol, 20 mM imidazole and 0.01% $C_{12}E_8$. After the clarified supernatant was loaded, the column was initially washed with 50 ml of 150 mM NaCl, 20 mM Tris (pH 7.5), 10% glycerol, 20 mM imidazole and 0.01% $C_{12}E_8$, and washed again with 50 ml of 150 mM NaCl, 20 mM Tris (pH 7.5), 10% glycerol, 80 mM imidazole and 0.01%

$C_{12}E_8$. The bound protein was eluted with 150 mM NaCl, 20 mM Tris (pH 7.5), 10% glycerol, 500 mM imidazole and 0.01% $C_{12}E_8$. The protein was concentrated to 2.5 ml an Amicon Ultra-15 concentrator unit (Millipore) with a molecular cut-off of 100 kDa and buffer exchanged to 150 mM NaCl, 20 mM Tris (pH 7.5), 0.5 mM EDTA, 10% glycerol and 0.01% $C_{12}E_8$ using PD-10 desalting column. The protein was further concentrated and loaded onto the Superdex 200 size exclusion chromatography (SEC) column in 150 mM NaCl, 20 mM Tris (pH 7.5), 0.5 mM EDTA, 10% glycerol and 0.01% $C_{12}E_8$.

BtuCD. The protein was purified following the same protocol as ModBC, except that protein was solubilised in 1% n-Dodecyl-*N,N*-Dimethylamine *N*-Oxide (LDAO, Anatrace) and all the buffers were supplemented with 0.05% LDAO.

2.5.4 Native MS experiments

Purified membrane proteins were buffer exchanged into MS Buffer (two times the CMC of detergent of interest and 200 mM ammonium acetate) using a centrifugal buffer exchange device (Micro Bio-Spin 6, Bio-Rad) as previously described (Laganowsky et al., 2013). The best quality mass spectra of all membrane proteins were obtained using 0.5% C_8E_4 (Anatrace) as detergent.

Soluble proteins were buffer exchanged into 200 mM ammonium acetate (Sigma Aldrich) using the same method. The freshly buffer-exchanged proteins were kept on ice, with protein concentration measured as before. The protein samples were diluted as desired in 200 mM ammonium acetate buffer with detergent as necessary and loaded into a gold-coated capillary Clark borosilicate capillary (Harvard Apparatus) prepared in the laboratory (Hernández and Robinson, 2007). The experiment was performed using a modified Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany), optimised for analysing protein complexes of high mass and membrane proteins (Gault et al., 2016). The instrument was calibrated using caesium iodide solution. Typically, 2 µl of buffer exchanged protein solution was electrosprayed from gold-plated borosilicate capillaries prepared in house. The instrument parameters for MS are: 1.2kV capillary voltage, S-lens RF 100%,

quadrupole selection from 1,000 to 15,000 *m/z* range, collisional activation in the HCD cell 0-300 V, argon UHV pressure $1.2*10^{-9}$ mbar and 30 to 200 °C capillary temperature, resolution of the instrument at 17,500 at *m/z* = 200 (a transient time of 64 ms) and ion transfer optics (injection flatapole, inter-flatapole lens, bent flatapole, transfer multipole: 8, 7, 6, 4 V respectively). The noise level was set at 3 rather than the default value of 4.64. Where required, baseline subtraction was performed to achieve a better-quality mass spectrum. Data were analysed using the Xcalibur 3.0 (Thermo Scientific) and UniDec (www.unidec.chem.ox.ac.uk) software packages. Theoretical and observed molecular masses of all species are described in Table 2.1.

Complex formation experiments. To allow the observation of BtuCD-F spectrum, it was necessary to incubate solutions of the two proteins in a 1:2 molar ratio (BtuF:BtuCD). In order to observe the effect of vitamin B_{12} on complex formation, 2 mM vitamin B_{12} (Sigma Aldrich) was added to a solution of BtuF before buffer exchange, then the solutions containing vitamin B_{12}-loaded BtuF and BtuCD were incubated and analysed as described above. Experiments with cyanocobinamide and vitamin B_9 (Sigma Aldrich) were performed in the same way as described above. To analyse the impact of ATP binding and hydrolysis, 5 mM ATP (Sigma Aldrich) and 5 mM $MgCl_2$ (Sigma Aldrich) were added to a solution of BtuCD before buffer exchange; then solutions containing BtuF and ADP-bound BtuCD were mixed and analysed as described above. The AMP-PNP experiments were performed by adding 5 mM AMP-PNP (Sigma Aldrich) to a solution of BtuCD before buffer exchange; then solutions containing BtuF and AMP-PNP-bound BtuCD were mixed and analysed as previously described. To assess the effect of both vitamin B_{12} binding and ATP binding and hydrolysis, 2 mM vitamin B_{12} was added to BtuF and 5 mM ATP was added to BtuCD before buffer exchange; then the solutions were mixed and analysed as mentioned above.

To detect the spectrum of ModBC-A in a 1:1 molar ratio was used (ModA:ModBC). Sodium molybdate (Na_2MoO_4, Sigma Aldrich) was added to a solution of ModA at a concentration of 5 mM, before buffer exchange. To analyse the influence of ATP-binding and hydrolysis, 5 mM ATP and 5mM $MgCl_2$ were added to a solution of ModBC before buffer exchange, then the solutions were incubated and analysed following the protocol described

above. The spectrum obtained did not show ADP binding. The experiment was therefore repeated using the following method: to a solution of 7 µM ModBC in MS buffer, ATP and $MgCl_2$ were added to a final concentration of 10 µM. Then, MoO_4^{2-}-bound ModA and ADP-bound ModBC were incubated and analysed as described above.

2.5.5 Lipidomics analysis

For LC-MS/MS analysis, the phospholipids were separated on a C18 column (Acclaim PepMap 100, C18, 75 µm × 15 cm; Thermo Scientific) by Dionex UltiMate 3000 RSLC nano System connected to a hybrid LTQ Orbitrap mass spectrometer (Thermo Scientific) via a dynamic nanospray source using PicoTip Emitter SilicaTip (New Objective). A binary buffer system was used with buffer A of ACN: H_2O (60:40), 10 mM ammonium formate, 0.1% formic acid and buffer B of IPA: ACN (90:10), 10 mM ammonium formate, 0,1% formic acid. The phospholipids were separated at 40 °C with a gradient of 32% to 99% buffer B at a flow rate of 300 nl/min over 30 min. Typical MS conditions were: spray voltage of 1.8 kV and capillary temperature of 175 °C. The LTQ-Orbitrap XL was operated in negative ion mode and in data-dependent acquisition with one MS scan followed by five MS/MS scans. Survey full-scan MS spectra were acquired in the Orbitrap (*m/z* 350−2,000) with a resolution of 60,000. Collision-induced dissociation (CID) fragmentation in the linear ion trap was performed for the five most intense ions at an automatic gain control target of 30,000 and a normalised collision energy of 38% at an activation of q = 0.25 and an activation time of 30 ms. Data were analysed by extracting ion chromatogram (XIC) and its area under the curve (AUC) of each cardiolipin was processed and integrated in Xcalibur3.0 (Thermo Scientific)) with 50 ppm 4 mass tolerance and 7 point Gaussian smoothing. The relative abundances of PEs and PGs in the samples were calculated based on the ratio of their AUCs to the sum of the AUCs of the phospholipids belonging to the same class

2.5.6 Quantification and statistical analysis

All the experiments described in the paper were performed three times (n=3). Average values and standard deviations were calculated from at least 4 charge states in each of the three independent repeats and plotted using Microsoft Excel. P values were obtained from unpaired Welch's t-test and plotted using GraphPad Prism 8.2.

2.6 References

1. Rees, D.C., E. Johnson, and O. Lewinson, *ABC transporters: the power to change.* Nat. Rev. Mol. Cell. Biol., 2009. **10**(3): p. 218-227.
2. Dawson, R.J. and K.P. Locher, *Structure of a bacterial multidrug ABC transporter.* Nature, 2006. **443**(7108): p. 180-185.
3. Locher, K.P., *Mechanistic diversity in ATP-binding cassette (ABC) transporters.* Nature Structural & Molecular Biology, 2016. **23**: p. 487-493.
4. Swier, L.J.Y.M., Slotboom, D., Poolman, B., *ABC Importers.* , in *ABC Transporters - 40 Years on.* A.M. George, Editor. 2015, Springer. p. 3-36.
5. Rempel, S., Stanek W. K., Slotboom D. J., *ECF-Type ATP-Binding Cassette Transporters.* Annu. Rev. Biochem., 2019. **88**: p. 551-576.
6. Maqbool, A., et al., *The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity.* Biochem Soc Trans, 2015. **43**(5): p. 1011-1017.
7. Hollenstein, K., D.C. Frei, and K.P. Locher, *Structure of an ABC transporter in complex with its binding protein.* Nature, 2007. **446**: p. 213-216.
8. Chen, J., et al., *Trapping the transition state of an ATP-binding cassette transporter: evidence for a concerted mechanism of maltose transport.* Proc. Natl. Acad. Sci. USA, 2001. **98**: p. 1525-1530.
9. Locher, K.P., A.T. Lee, and D.C. Rees, *The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism.* Science, 2002. **296**: p. 1091-1098.
10. Korkhov, V.M., S.A. Mireku, and K.P. Locher, *Structure of AMP-PNP-bound vitamin B12 transporter BtuCD-F.* Nature, 2012. **490**: p. 367-372.
11. Berntsson, R.P.A., et al., *A structural classification of substrate-binding proteins.* FEBS Lett., 2010. **584**: p. 2606-2617.
12. Chen, J., *Molecular mechanism of the Escherichia coli maltose transporter.* Curr Opin Struct Biol, 2013. **23**(4): p. 492-498.
13. Bohm, S., et al., *Conformational plasticity of the type I maltose ABC importer.* Proc. Natl. Acad. Sci. USA, 2013. **110**(14): p. 5492-5497.

14. Vigonsky, E., E. Ovcharenko, and O. Lewinson, *Two molybdate/tungstate ABC transporters that interact very differently with their substrate binding proteins.* Proc. Natl. Acad. Sci. USA, 2013. **110**(14): p. 5440-5445.
15. Pinkett, H.W., et al., *An inward-facing conformation of a putative metal-chelate-type ABC transporter.* Science, 2007. **315**(5810): p. 373-377.
16. Rice, A.J., et al., *EPR spectroscopy of MolB2C2-a reveals mechanism of transport for a bacterial type II molybdate importer.* J. Biol. Chem., 2013. **288**(29): p. 21228-21235.
17. Hvorup, R.N., et al., *Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF.* Science, 2007. **317**(5843): p. 1387-1390.
18. Korkhov, V.M., et al., *Structure of AMP-PNP-bound BtuCD and mechanism of ATP-powered vitamin B12 transport by BtuCD-F.* Nat. Struct. Mol. Biol., 2014. **21**: p. 1097-1099.
19. Goudsmits, J.M.H., D.J. Slotboom, and A.M. van Oijen, *Single-molecule visualization of conformational changes and substrate transport in the vitamin B12 ABC importer BtuCD-F.* Nat. Commun., 2017. **8**(1): p. 1652.
20. Lewinson, O., et al., *A distinct mechanism for the ABC transporter BtuCD-BtuF revealed by the dynamics of complex formation.* Nat. Struct. Mol. Biol., 2010. **17**(3): p. 332-338.
21. Joseph, B., et al., *Transmembrane gate movements in the type II ATP-binding cassette (ABC) importer BtuCD-F during nucleotide cycle.* J. Biol. Chem., 2011. **286**: p. 41008-41017.
22. Joseph, B., et al., *Conformational cycle of the vitamin B12 ABC importer in liposomes detected by double electron-electron resonance (DEER).* J. Biol. Chem., 2014. **289**(6): p. 3176-3185.
23. Barrera, N.P., et al., *Micelles protect membrane complexes from solution to vacuum.* Science, 2008. **321**(5886): p. 243-246.
24. Gault, J., et al., *High-resolution mass spectrometry of small molecules bound to membrane proteins.* Nat. Methods, 2016. **13**(4): p. 333-336.
25. Patrick, J.W., et al., *Allostery revealed within lipid binding events to membrane proteins.* Proc. Natl. Acad. Sci. USA, 2018. **115**(12): p. 2976-2981.
26. Mehmood, S., et al., *Mass spectrometry captures off-target drug binding and provides mechanistic insights into the human metalloprotease ZMPSTE24.* Nat. Chem., 2016. **8**(12): p. 1152-1158.
27. Bolla, J.R., et al., *Direct observation of the influence of cardiolipin and antibiotics on lipid II binding to MurJ.* Nat. Chem., 2018. **10**(3): p. 363-371.
28. Yen, H.Y., et al., *PtdIns(4,5)P2 stabilizes active states of GPCRs and enhances selectivity of G-protein coupling.* Nature, 2018. **559**(7714): p. 423-427.

29. Laganowsky, A., et al., *Mass spectrometry of intact membrane protein complexes.* Nat. Protoc., 2013. **8**(4): p. 639-651.
30. Reading, E., et al., *The role of the detergent micelle in preserving the structure of membrane proteins in the gas phase.* Angew. Chem. Int. Ed. Engl., 2015. **54**(15): p. 4577-4581.
31. Bechara, C., et al., *A subset of annular lipids is linked to the flippase activity of an ABC transporter.* Nat. Chem., 2015. **7**(3): p. 255-262.
32. Mehmood, S., et al., *Structural and Functional Basis for Lipid Synergy on the Activity of the Antibacterial Peptide ABC Transporter McjD.* J. Biol. Chem., 2016. **291**(41): p. 21656-21668.
33. Mireku, S.A., et al., *Conformational Change of a Tryptophan Residue in BtuF Facilitates Binding and Transport of Cobinamide by the Vitamin B12 Transporter BtuCD-F.* Sci. Rep., 2017. **7**: p. 41575.
34. Cubrilovic, D., et al., *Determination of protein-ligand binding constants of a cooperatively regulated tetrameric enzyme using electrospray mass spectrometry.* ACS Chem. Biol., 2014. **9**(1): p. 218-226.
35. Oldham, M.L. and J. Chen, *Crystal structure of the maltose transporter in a pretranslocation intermediate state.* Science, 2011. **332**(6034): p. 1202-1205.
36. Khare, D., et al., *Alternating access in maltose transporter mediated by rigid-body rotations.* Mol. Cell, 2009. **33**(4): p. 528-536.
37. Borths, E.L., et al., *In vitro functional characterization of BtuCD-F, the Escherichia coli ABC transporter for vitamin B12 uptake.* Biochemistry, 2005. **44**(49): p. 16301-16309.
38. Tal, N., E. Ovcharenko, and O. Lewinson, *A single intact ATPase site of the ABC transporter BtuCD drives 5% transport activity yet supports full in vivo vitamin B12 utilization.* Proc. Natl. Acad. Sci. USA, 2013. **110**(14): p. 5434-5439.
39. Qasem-Abdullah, H., et al., *ATP binding and hydrolysis disrupt the high-affinity interaction between the heme ABC transporter HmuUV and its cognate substrate-binding protein.* J. Biol. Chem., 2017. **292**(35): p. 14617-14624.
40. Husada, F., et al., *Conformational dynamics of the ABC transporter McjD seen by single-molecule FRET.* Embo J., 2018. **37**(21).
41. Yang, M., et al., *Single-molecule probing of the conformational homogeneity of the ABC transporter BtuCD.* Nat. Chem. Biol., 2018. **14**(7): p. 715-722.
42. Mireku, S.A., et al., *Structural basis of nanobody-mediated blocking of BtuF, the cognate substrate-binding protein of the Escherichia coli vitamin B12 transporter BtuCD.* Sci. Rep., 2017. **7**(1): p. 14296.

CHAPTER 3

Elucidating the mechanism of LPS extraction from the inner membrane by $LptB_2FGC$

Author contributions

F. F. carried out protein expression, purification and nMS experiments. Chih-Chia Su (Case Western Reserve University School of Medicine) collected and processed EM data and built the structural models. F. F. analysed and interpreted the structural models and performed structural comparisons described in this chapter with inputs from Jani R. Bolla.

3.1 Introduction

Gram-negative bacteria are characterised by a double-membrane envelope: the inner membrane (IM) which surrounds the cytoplasm is mainly composed of phospholipids, whereas the outer membrane (OM) has an asymmetric bilayer with lipopolysaccharide (LPS) molecules on the outer leaflet and phospholipids in the inner leaflet [1]. While the lipidic portion of the OM is necessary to avoid the passage of polar molecules, LPS plays an essential role in preventing toxic hydrophobic compounds, such as antibiotics, from entering the cell [2, 3], which is possible due to its amphipathic nature.

The LPS molecule is composed of three moieties: lipid A, core oligosaccharide and an O-antigen glycan which itself varies in its composition. The O-antigen contains up to 200 sugars, and some of them bear phosphate groups which interact with bivalent ions, thus allowing LPS molecules to create a tight network [4]. The lipid A core oligosaccharide complex and the O-antigen units are synthesised independently in the cytoplasm and then transported to the periplasmic side of the IM by MsbA [5] and Wzx [6], respectively.

Mature LPS is then extracted from the IM and transported through the periplasm to be inserted into the outer leaflet of the OM by seven essential proteins LptA-G (Figure 3.1a) [4, 7]. The IM sub-complex comprises the atypical ATP-binding cassette (ABC) transporter $LptB_2FG$

extracts LPS from the IM to the periplasmic components of the Lpt system. The hydrolysis of ATP provides the energy to allow the necessary conformational changes for LPS translocation. $LptB_2FG$ forms a stable sub-complex with the single pass membrane protein LptC [8, 9], which in turn interacts with the periplasmic protein LptA, which presents a β-taco fold and acts as a bridge to the OM across the periplasm [10-12]. The OM sub-complex is constituted by the heterodimer LptDE: the soluble β-taco domain of LptD represents the last portion of the bridge that connects to the OM β-barrel domain where the exit gate of LPS is present. LptE is a lipoprotein inserted in the lobe of the β-barrel that is thought to contribute to the extrusion of LPS from the complex [13-15].

Structurally, $LptB_2FG$ is a heterotetramer resembling the spatial organisation of previously described ABC importers [16-20], although its function is quite different and presents specific structural features. Similar to other ABC transporters, the complex has two nucleotide-binding domains (NBDs), the LptB subunits, that capture and hydrolyse ATP, and two transmembrane domains (TMDs), represented by LptF and LptG, that create a cavity and a translocation path for LPS and present a periplasmic taco domain. Differently from canonical ABC transporters, $LptB_2FG$ does not allow the passage of the substrate across the membrane but extracts LPS directly from the lipid bilayer (Figure 3.1b). Furthermore, recent reports have indicated that the transporter does not act as a tetramer, but it intimately interacts with LptC. This protein presents a transmembrane helix and a periplasmic β-taco soluble domain that interacts with LptF and LptA. The transmembrane helix of LptC inserts between the two TMDs of LpB_2FG (specifically TM1 of LptG and TM5 of LptF) and regulates the ATPase activity and LPS extraction efficiency of the protein complex [21-23]. Thus, the helix is involved in the unidirectional transport from the IM to the central cavity of the ABC transporter, while the β-taco domain of LptC directly interacts with a similar domain present in LptF and LptA, forming a continuous path for LPS from the ABC transporter system to the periplasmic bridge formed by the LptA oligomer [4, 21-24].

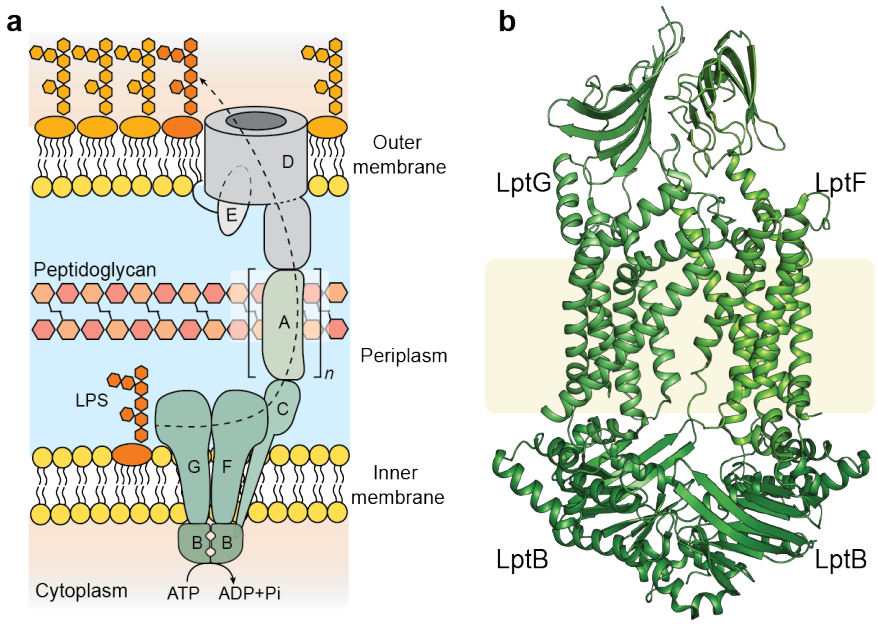


Figure 3.1 Schematic of the LPS transport system and structures of $LptB_2FG$ and $LptB_2FGC$. (a) Schematic of the Lpt system: LPS is transported from the IM to OM through a multiprotein complex comprising 7 proteins. It is extracted from the IM from the ABC transporter $LptB_2FG$ which interacts which the single transmembrane protein LptC. LptC regulates LPS translocation and forms a periplasmic bridge with the soluble oligomeric LptA oligomer. The C-terminus of LptA interacts with the soluble N-terminal domain of LptD. LPS is transported through the periplasmic within the protein bridge and finally inserted in the outer leaflet of the OM by the heterodimer LptDE. (b) X-ray crystal structure of *P. aeruginosa* $LptB_2FG$ (PDB ID: 5X5Y) [25].

Among different Gram-negative bacteria, *Pseudomonas aeruginosa* has raised particular interest given its rising multidrug resistance to established antibiotics causing high levels of morbidity [26]. *P. aeruginosa* is a human opportunistic pathogen infecting immunocompromised patients, including those suffering from burn wounds, cancer, AIDS and cystic fibrosis. Remarkably, LPS from *P. aeruginosa* regulates the interaction of the bacterium with the host receptor cystic fibrosis transmembrane regulator (CFTR) [27] as well as biofilm formation [28]. It has also been shown that LPS has a cooperative association with the type III secretion system (TTSS), a surface-attached needle-like complex that mediates the injection of cytotoxins directly

into eukaryotic cells, causing cellular damage [29]. Consequently, LPS transport is an ideal target pathway for the development of novel drugs against *P. aeruginosa* infections and, generally, Gram-negative pathogens.

Many efforts in elucidating the mechanism of LPS biogenesis in *P. aeruginosa* have yielded to the structures of the $LptB_2FG$ complex [25], LptA [12] and LptDE [15], though only the structure of LptD transmembrane region has been solved. Interestingly, the proteins belonging to *P. aeruginosa* share low sequence similarity with other species orthologues. For instance, *pa*LptDE shows less than 28% identity to its orthologues from other Gram-negative bacteria, [15], including *E. coli* (Fig. S3.1). In line with this, *pa*LptA possesses only 28% sequence identity with *ec*LptA [12] and *pa*LptC shares only 18% identical residues with *ec*LptC (Fig. S3.2). Similarly, *pa*LptF and *pa*LptG share 29% and 38% sequence identity with their *E. coli* orthologues, respectively, while *pa*LptB sequence similarity is higher at ~66% (Fig. S3.3). Despite the low sequence homology with other species, the solved structures of Lpt proteins have displayed similar three-dimensional architectures to their orthologues. On the other hand, *pa*LptD possesses a periplasmic domain ~90 residues longer compared to its orthologues as well as two extra loops, thus suggesting that the architecture of the multi-protein complex in *P. aeruginosa* may be different from other species. In addition, the structure of the fully functional LPS transporter *pa*$LptB_2FG$ in complex with *pa*LptC is currently unknown and the influence of nucleotide binding to protein-substrate interaction has not been explored. Moreover, given the lipid nature of LPS, we hypothesise that membrane lipids may play an active role in substrate binding and extraction, hence modulating LPS transport.

To this end, we integrated native mass spectrometry (nMS) with cryo-electron microscopy (cryo-EM) to investigate the structure and modulation of *P. aeruginosa* $LptB_2FGC$ and we expanded our investigation also to *pa*LptA, in order to obtain further insights on $LptB_2FGC$-LptA interaction.

3.2 Results

3.2.1 Optimisation of $LptB_2FG$ purification and nMS analysis

We initially expressed $LptB_2FG$ from *P. aeruginosa* in *E. coli* strain C43(DE3) and purified it following a classic two-step purification protocol consisting of immobilized metal affinity chromatography (IMAC) and size-exclusion chromatography (SEC), as described in the methods section. The protein complex was purified in a buffer containing 0.03% (w/v) n-dodecyl-β-D maltopyranoside (DDM) and screened to establish the ideal conditions for nMS analysis [34]. We initially selected tetraethylene glycol monooctyl ether (C_8E_4) as nMS detergent and observed resolved charge states corresponding to the mass of intact $LptB_2FG$ (Figure 3.2a). We also detected adducts correspondent to co-purified phospholipids (PLs), as well as a broad adduct around 3.5 kDa corresponding to endogenous LPS [20, 35], which is not surprising given the fact that LPS is the substrate of the protein complex. In an effort to optimise nMS analysis of $LptB_2FG$ and evaluate the influence of lipids and nucleotides on substrate binding, we sought to remove the co-purified lipids through an on-column delipidation protocol [36] as described in the methods section. The method involved the use of 2% DDM + 2% octyl glucose neopentyl glycol (OGNG) in the membrane solubilisation step and 50 column volumes (CVs) of a buffer containing 1% DDM + 1% OGNG during the IMAC wash step. This procedure allowed partial removal of co-purified PLs, however, we could still observe LPS adducts, indicating a high degree of affinity for the protein (Figure 3.2b).

To increase the chances of removing the LPS adduct, we then expressed $LptB_2FG$ from ClearColi BL21(DE3), an *E. coli* strain holding different gene knockouts that stop the biosynthetic pathway of LPS to a precursor, lipid IVA. Initial purification following the classic procedure resulted in $LptB_2FG$ bound to up to five lipid IVA molecules (Figure 3.2c). In fact, we were not able to detect the apo form of the protein complex, but could see multiple 1.4 kDa adducts, consistent with the mass of lipid IVA (1,405 Da), although we cannot exclude the presence of cardiolipin, whose mass in the same range. The application of the delipidation

protocol to $LptB_2FG$ expressed from ClearColi strain yielded fully delipidated protein, indeed nMS analysis revealed no detectable lipid IVA bound to the complex (Figure 3.2d).

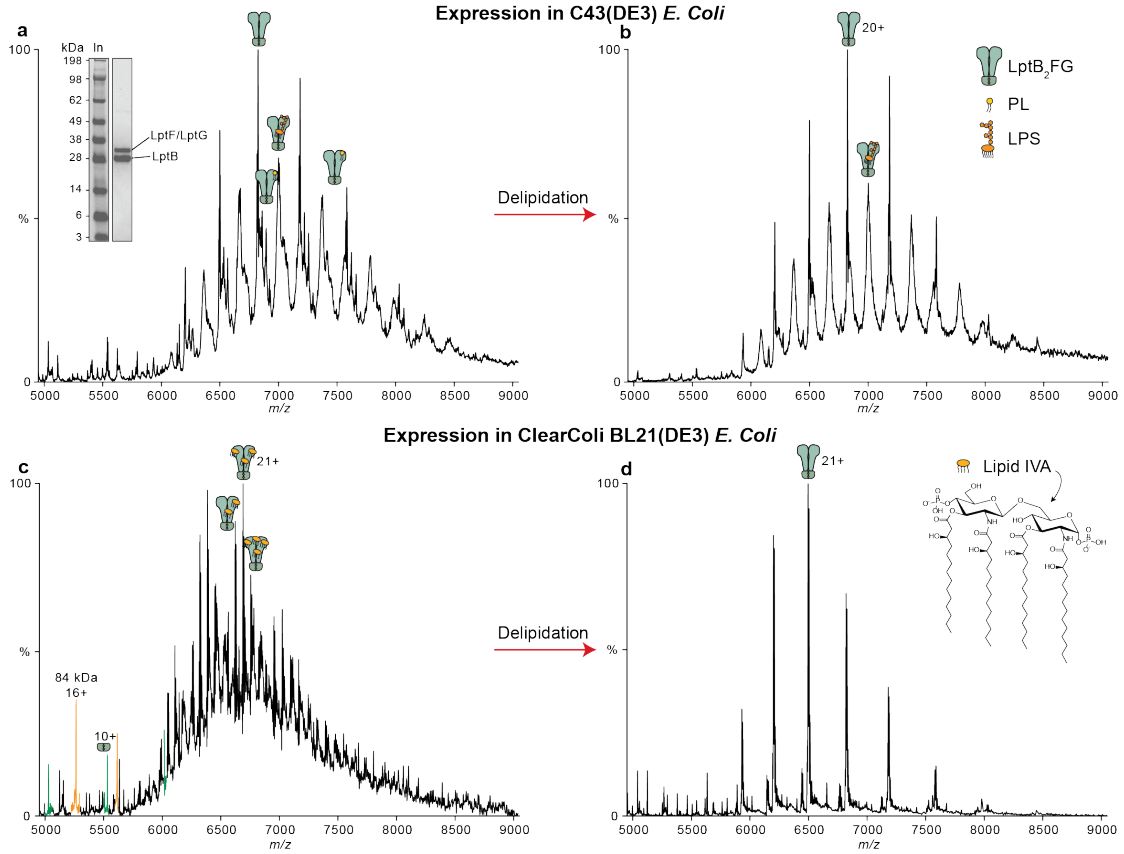


Figure 3.2 nMS analysis of $LptB_2FG$ under different expression and purification conditions. (a) nMS analysis of $LptB_2FG$ expressed in *E. coli* strain C43(DE3) and purified using classic protocols, along with SDS-PAGE gel showing the purified protein complex subunits. nMS indicates the presence of apo-$LptB_2FG$, as well as phospholipid (PL) and LPS-bound $LptB_2FG$. (b) nMS analysis of $LptB_2FG$ expressed in *E. coli* strain C43(DE3) and purified following delipidation protocol shows the presence of apo-$LptB_2FG$ and LPS-bound $LptB_2FG$. (c) nMS analysis of $LptB_2FG$ expressed in in ClearColi BL21(DE3) and purified using classic protocols denotes the presence of multiple lipid IVA molecules bond to the protein complex, while apo-$LptB_2FG$ is not detectable. (d) nMS analysis of $LptB_2FG$ expressed in ClearColi BL21(DE3) and purified following delipidation protocol shows the successful purification of 'clean' protein, without any lipid species bound.

Having found the best condition for LPS-free $LptB_2FG$ expression and purification, we re-evaluated different classes of detergents for nMS analysis. Apart from the previously shown spectra in C_8E_4, we obtained well-resolved spectra of $LptB_2FG$ in n-dodecyl-*N,N*-dimethylamine-*N*-oxide (LDAO) and OGNG using negative mode. Recently, a new library of oligoglycerol detergents (OGDs) has been described and demonstrated wide application for purification and nMS of different membrane proteins, allowing the observation of protein complexes at relatively low activation energies [37]. We employed the [G1]-ether-C12 detergent mixture (simply referred to as G1-OGD) which has shown the ability of retaining non-covalent interactions and easy detergent removal from the proteomicelles [37]. We obtained well resolved spectra of delipidated $LptB_2FG$ in negative mode (Figure S3.4). Lower energy conditions were required to observe the release the protein from the micelles compared to C_8E_4, LDAO or OGNG, thus we decided to carry out ligand binding experiments using G1-OGD as nMS detergent.

3.2.2 Substrate and lipid binding to the $LptB_2FG$ complex

Having established the protocols for $LptB_2FG$ delipidation and nMS analysis, we decided to probe the influence of different ligands on the substrate-binding properties of the complex. To study the LPS-binding properties, yet retaining well resolved spectra, we employed the analogue Re-LPS, an LPS substructure consisting of the lipid A moiety and two ketodeoxyoctonic acids (Kdo) residues belonging to the core oligosaccharide (Figure 3.3a), that has already been used to investigate Lpt system functionality in previous studies [38, 39]. We incubated a solution of 3 μM $LptB_2FG$ in 200 μM ammonium acetate (pH 8.0) and 0.02% (w/v) G1-OGD with 2.5 μM Re-LPS and recorded mass spectra under our optimised conditions. We could detect a small percentage (less than 5%) of apo-$LptB_2FG$, but observed multiple 2.2 kDa adducts, consistent with the mass of Re-LPS (2239 Da), indicating up to 3 Re-LPS molecules bound to $LptB_2FG$. When Re-LPS concentration was decreased to 0.4 μM, we could observe ~70% apo protein and up to 2 Re-LPS adducts (Figure 3.3b).

We then proceeded to investigate the structure-activity relationship (SAR) for protein-substrate interaction in order to elucidate the essential moieties for LPS binding. We incubated

$LptB_2FG$ with 0.4 μM and 2.5 μM lipid A-diphosphate (LA-diP), a substructure of Re-LPS lacking the two Kdo groups (Figure 3.3a). Using nMS we could observe LA-diP bound to $LptB_2FG$ at both concentrations measured, and up to three LA-diP were bound to $LptB_2FG$. Quantification of the relative abundance of lipid-bound protein indicated that LA-diP binding to $LptB_2FG$ was significantly lower compared to Re-LPS in both cases (Figure 3.3c). These observations suggest that the Kdo residues in the inner core of LPS important moieties for the interaction between $LptB_2FG$ and its substrates.

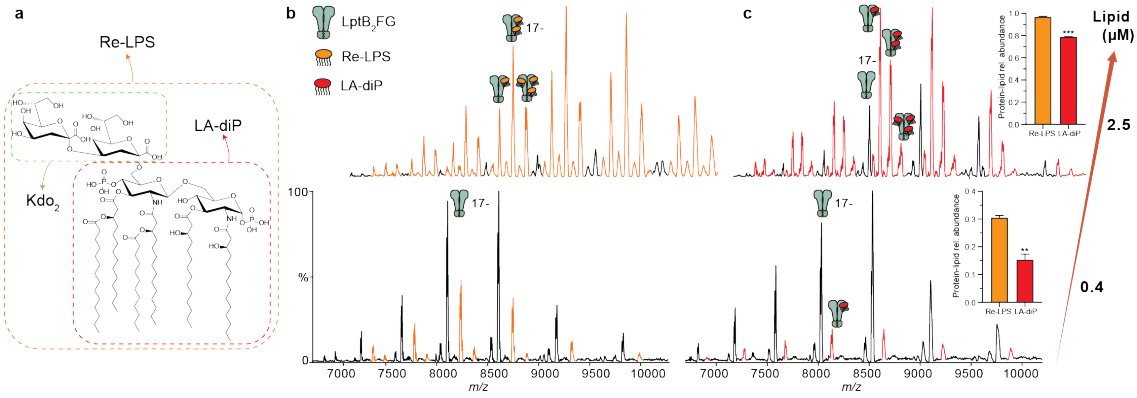


Figure 3.3 Quantification of the interaction between $LptB_2FG$ and lipid A derivatives. (a) Molecular structure of Re-LPS (orange) and LA-diP (red). (b) nMS analysis of $LptB_2FG$:Re-LPS interaction. $LptB_2FG$ (3 μM) was mixed with increasing concentrations of Re-LPS (0.4 μM and 2.5 μM). Charge state series indicating Re-LPS binding were observed and the total amount of Re-LPS bound protein complex was calculated. (c) nMS analysis of $LptB_2FG$:LA-diP interaction. The same procedure used for Re-LPS was used in this case. Quantification of total lipid-bound protein indicates the preference of $LptB_2FG$ for Re-LPS binding over LA-diP (inset). Error bars indicate standard deviations (n= 3). A Welch's T-test was used to evaluate statistical significant differences (** $p < 0.01$; **** $p < 0.0001$).

To broaden the investigation on the lipid binding properties of $LptB_2FG$, we performed similar lipid binding experiments using three lipid species commonly present in Gram-negative bacteria inner membrane: phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and cardiolipin (CDL). We chose 14:0 PE (DMPE), 14:0 PG (DMPG), and 14:0 CDL to be consistent

with the 14 carbon lipid tails of Re-LPS. In the case of DMPE (Figure S3.5a) and DMPG (Figure S3.5b), we could not observe any binding event at 5 µM concentration. Differently, CDL could bind $LptB_2FG$ more tightly: at 0.4 µM the fraction of CDL-bond protein, was similar to the previously described LA-diP (~ 15% lipid-bound protein), while at 2.5 µM concentration ~50% protein was bound to CDL (Figure 3.4a,b).

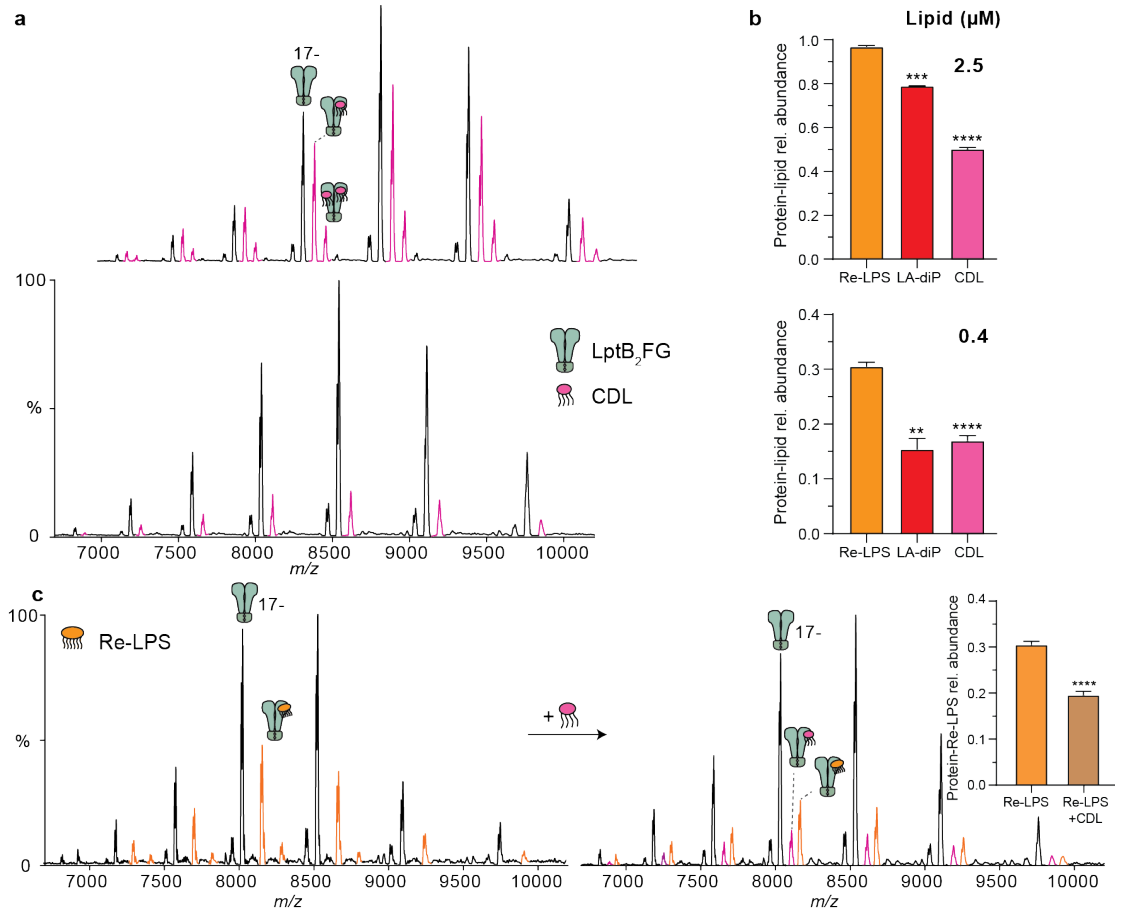


Figure 3.4 Quantification of the interaction between $LptB_2FG$ and CDL and its influence on Re-LPS binding. (a) nMS analysis of $LptB_2FG$:CDL interaction. $LptB_2FG$ (3 µM) was mixed with increasing concentrations of CDL (0.4 µM and 2.5 µM). At 0.4 µM one binding event was observed, at 2.5 µM binding of up to 3 CDL molecules was observed indicating higher affinity of the complex for CDL compared to other bacterial IM lipids. (b) Quantification of total lipid-bound protein indicates that CDL binds $LptB_2FG$ similar to LA-diP at lower concentrations, but less binding is observed at higher concentrations. (c) Re-LPS (0.4 µM) as added to $LptB_2FG$ (4 µM) and the resulting solution was mixed with CDL (0.4 µM); resultant mass spectra indicate

the presence of both $LptB_2FG$:CDL and $LptB_2FG$:Re-LPS complexes in solution. Inset: Quantification of total Re-LPS bound protein indicates that CDL addition to the solution mixture caused a significant decrease of Re-LPS-$LptB_2FG$ binding. Error bars indicate standard deviations (n= 3). A Welch's T-test was used to test statistical significant differences (** $p < 0.01$; *** $p < 0.01$; **** $p < 0.0001$).

We found these lipid binding properties intriguing, thus to test the possibility that the CDL influences the substrate binding properties of $LptB_2FG$ we added 14:0 CDL (0.4 μM) to a solution containing $LptB_2FG$:Re-LPS complex obtained by the addition of 0.4 μM Re-LPS to 3 μM protein (Figure 3.4c). Following CDL addition, the amount of Re-LPS bound to the protein decreased significantly; also, we did not observe the $LptB_2FG$:Re-LPS:CDL tertiary complex. CDL possesses the same essential moieties of PG (a phosphoglycerol head and lipid tails), in fact, 14:0 CDL resembles two DMPG molecules yet the binding properties of the two molecules are quite different since DMPG displays low binding even at higher concentrations. The contrast between DMPG and 14:0 CDL binding properties led us to think that CDL interacts with $LptB_2FG$ in a specific way. Moreover, the effect of CDL on Re-LPS binding, comprising the decrease of bound Re-LPS and the absence of a tertiary complex suggests that CDL and Re-LPS compete with each other and the two lipids may share the same binding site.

3.2.3 ATP binding and hydrolysis have opposite effects on the $LptB_2FG$ substrate affinity

To gain insights into the LPS translocation mechanism, we investigated the effects of ATP binding and hydrolysis on $LptB_2FG$:Re-LPS interaction. To a solution containing 3 μM $LptB_2FG$ in 200 μM ammonium acetate (pH 8.0) and 0.02% G1-OGD, we added ATP/Mg^{2+} to a final concentration of 20 μM. The mass spectrum denoted the presence of apo-$LptB_2FG$ as well as the two nucleotide-bound states with one and two ADP molecules bound, indicated as $LptB_2FG$•1ADP and $LptB_2FG$•2ADP, respectively (Figure S3.6a).

Having established a method to analyse protein-nucleotide binding and having observed that the protein is able to hydrolyse ATP in this *in vitro* experiment, we added 20 µM ATP-Mg^{2+} to a solution containing the $LptB_2FG$:Re-LPS complex obtained following incubation of 3 µM protein with 0.4 µM Re-LPS. We found that following ATP binding and hydrolysis the total amount of Re-LPS bound $LptB_2FG$ was slightly decreased (Figure 3.5).

To trap the protein complex in its ATP-bound state we carried out analogous experiments in the presence of β-γ-imidoadenosine 5′-triphosphate (AMP-PNP), a non-hydrolysable analogue of ATP. Following the addition of AMP-PNP to $LptB_2FG$, we observed the presence of nucleotide-bound species ($LptB_2FG$•1AMP-PNP and $LptB_2FG$•2AMP-PNP (Figure S3.6b). Remarkably, when AMP-PNP was added to $LptB_2FG$:Re-LPS complex it determined a significant increase of lipid-bound species (Figure 3.5).

Combined, these experiments indicate that $LptB_2FG$ in its ATP-bound state has a higher affinity for Re-LPS compared to the apo form, while the ADP-bound state of $LptB_2FG$ is in a conformation presenting lower affinity for the substrate.

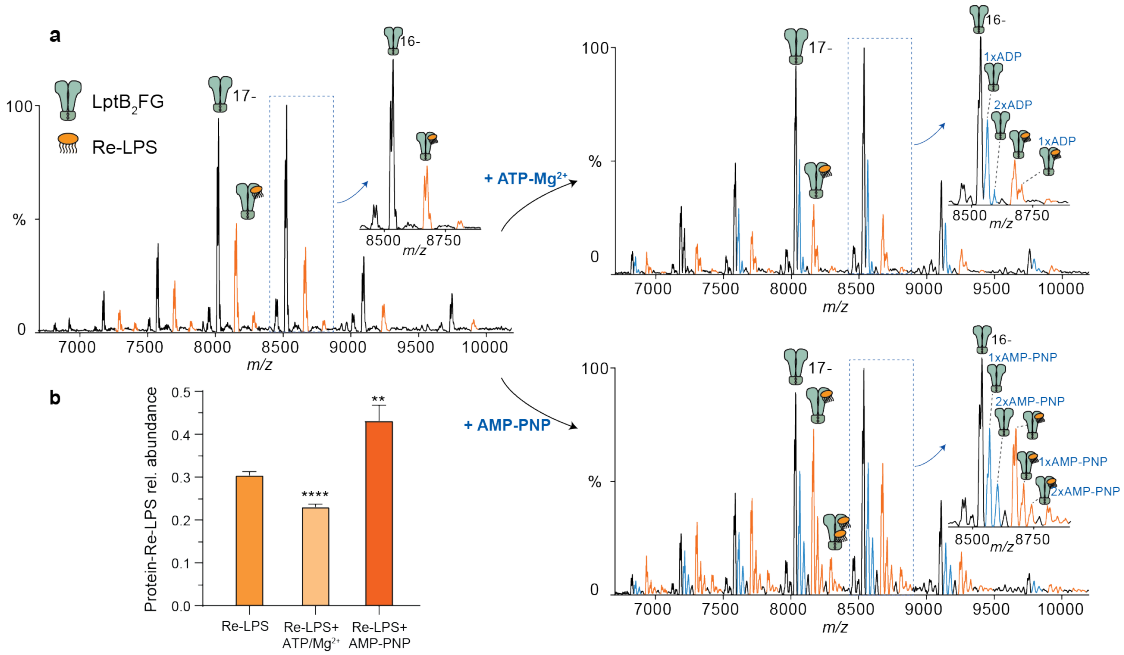


Figure 3.5 Quantification of the influence of nucleotide binding on $LptB_2FG$:Re-LPS interaction. (a) Re-LPS (0.4 µM) as added to $LptB_2FG$ (4 µM) and the resulting solution was mixed with ATP-Mg^{2+} (top right) or AMP-PNP (bottom right) at a final nucleotide concentration

of 20 μM. In the case of ATP-Mg^{2+}, nMS shows ADP binding to both $LptB_2FG$ and $LptB_2FG$:Re-LPS complexes. Similarly, when AMP-PNP was added, it is possible to observe up to 2 AMP-PNP molecules bound to both $LptB_2FG$ and $LptB_2FG$:Re-LPS complexes. (b) Quantification of Re-LPS binding to LpB_2FG in different states indicates that ATP binding and subsequent hydrolysis to ADP decreases Re-LPS:LpB_2FG interaction. Conversely, AMP-PNP increases the affinity for Re-LPS. Error bars indicate standard deviations (n= 3). A Welch's T-test was used to test statistical significant differences (** $p < 0.01$; **** $p < 0.0001$ vs Re-LPS – $LptB_2FG$ relative abundance).

3.2.4 $LptB_2FGC$ purification and nMS analysis

Recently, high-resolution structures for $LptB_2FG$ in complex with LptC have been released, revealing the structure of LptC and its modulation of the ATPase activity for an efficient LPS extraction from the IM [21-23]. To gain further insights on the protein machinery involved in LPS extraction from the IM, we decided to co-express $LptB_2FG$ with LptC and analyse the complex using nMS. To this end, we followed the same protocol described previously, consisting of overexpressing the different subunits of the complex in ClearColi strain and proceeding with delipidation procedure as before. We confirmed with SDS-PAGE that the LptC was successfully co-expressed with the other complex subunits, then proceeded with nMS analysis (Figure 3.6). Interestingly, we detected the presence in solution of both $LptB_2FG$ and $LptB_2FGC$ complexes, the reason may be ascribed to an in-solution equilibrium between the two complexes or lower expression levels of LptC compared to the other subunits. Furthermore, differently from the previous construct, although the complex underwent similar purification protocol we observed different adducts, probably co-purified lipids consisting of a mixture of lipid IVA and phospholipids. This observation is quite intriguing, indeed the only difference between the two purification procedures is the additional presence of LptC, hence the presence of an extra helix between the two TM domains of LptF and LptG may create additional surfaces for lipid binding.

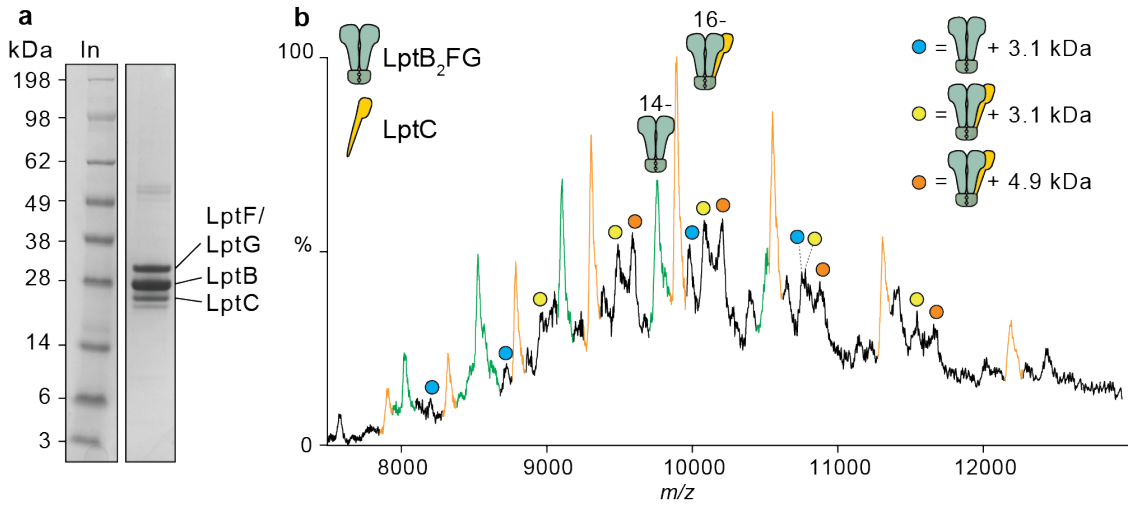


Figure 3.6 SDS-PAGE and nMS analysis of $LptB_2FGC$. (a) SDS-PAGE showing successful co-purification of $LptB_2FG$ and LptC. (b) mass spectrum of $LptB_2FGC$ expressed in ClearColi BL21(DE3) and purified following delipidation protocol. Both $LptB_2FG$ and $LptB_2FGC$ are present in solution, as well as adducts ranging from 3.1 to 4.9 kDa, corresponding to a mixture of lipids (most likely PLs and lipid IVA) co-purified with the protein complex.

3.2.5 Cryo-EM structure of $LptB_2FG$

To gain further insights on the above observations at the atomistic level, we set out to solve structures of the complexes. To this end, we reconstituted purified $LptB_2FGC$ in lipid nanodiscs using the protocol described in the methods section. As indicated by nMS experiments (Fig. 3.6), we could observe both $LptB_2FG$ and $LptB_2FGC$ in solution (Fig. S3.7a-c) and we solved the structure of both complexes. Two-dimensional class averages of $LptB_2FG$ cryo-EM particle images presented clear structural features (Fig. S3.7b), showing a quaternary complex shaped like a milk jug. Image processing yielded a final map at overall 3.11 Å resolution (Fig. S3.7d-S3.8a).

The cryo-EM maps showed clear side chain densities in both NBDs and TMDs, allowing us to unambiguously fit models from the previously published *pa*$LptB_2FG$ (PDB ID: 5X5Y) [25] (Fig. S3.8b-c). On the cytoplasmic side, we could observe the LptB dimer forming the NBD, while on the transmembrane portion the six TM helices of LptF and LptG constitute the TMDs

(TM1$_F$-TM6$_F$ and TM1$_G$-TM6$_G$). The TMDs are organised to form a central cavity with two lateral gates between TM1$_F$ - TM5$_G$ and TM1$_G$-TM5$_F$, while a coupling helix connects TM2 and TM3 (Fig. 3.7). We could also resolve the periplasmic β-taco domains of LptF and LptG, comprising nine and seven β-strands, respectively, that extend from TM3 and TM4 (Fig. 3.7).

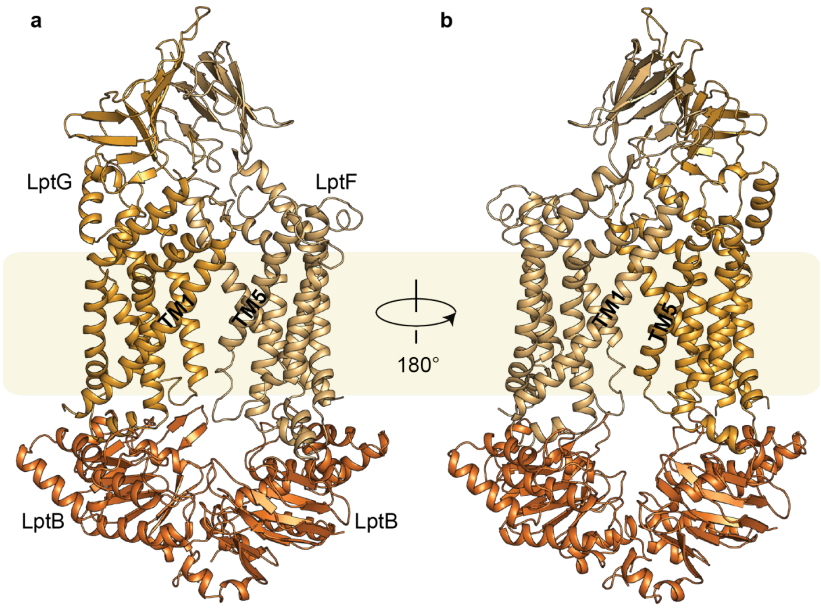


Figure 3.7 Cartoon representation of the LptB$_2$FG complex. (a) Cartoon representation of LptB$_2$FG. TM1$_G$ and TM5$_F$, forming the lateral gate are labelled. (b) Rotation of 180° along the y-axis relative to (a).

Compared to the previously published *pa*LptB$_2$FG (PDB ID: 5X5Y) [25], our structure presents significant changes in the conformation of the TMDs with a root mean square deviation (RMSD) of 4.13 Å over 1120 residues (Fig. 3.8a,b). While the NBDs seem to have similar conformations, the two lateral gates display different features given by the distorted arrangement of the TMD forming the lateral gates (TM5$_F$, TM1$_G$, TM1$_F$ and TM5$_G$). In fact, the TM5$_F$ – TM1$_G$ lateral gate is more opened in our structure as a consequence of the outward rotation of both TM5$_F$ and TM1$_G$, hence increasing the distance between the two α-helices up to 4 Å compared to previously published LptB$_2$FG (Fig. 3.8c). For instance, the C$_\alpha$ distance between Leu26$_G$ and Leu310$_F$ increases from 9.7 Å in Luo et al. *pa*LptB$_2$FG to 12.2 Å in our *pa*LptB$_2$FGC, similarly and Ile33$_G$ - Leu317$_F$ C$_\alpha$ distance increases from 5.9 Å to 9.9 Å our *pa*LptB$_2$FG (Fig. 3.8c).

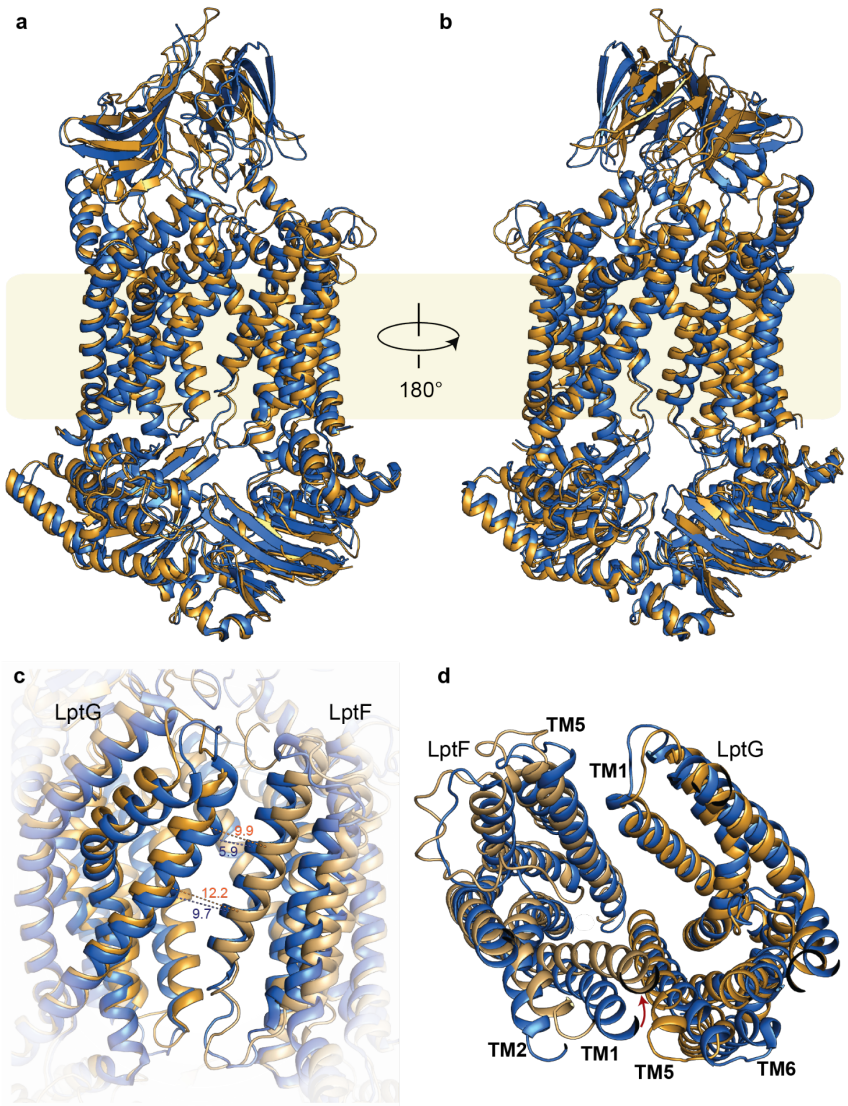


Figure 3.8 Structure alignment of two structures of *pa*LptB$_2$FG. (a) Superimposition of our *pa*LptB$_2$FG (orange) with Luo et al. *pa*LptB$_2$FG (PDB ID: 5X5Y, blue) (b) Rotation of 180° along the y-axis relative to (a). (c) Focus on TM5$_G$ – TM1$_F$ gate, showing the Cα distances between Leu26$_G$ and Leu310$_F$ and Ile33$_G$ and Leu317$_F$ (d) Periplasmic view of LptF and LptG TMDs showing the conformational changes from Luo et al. *pa*LptB$_2$FG to our *pa*LptB$_2$FG, such as the anti-clockwise rotation of TM1$_F$.

On the opposite side, the distance between TM1$_F$ and TM5$_G$ is markedly decreased in our *pa*LptB$_2$FG compared to the previous structure, indeed both helices rotate towards the centre of the cavity accompanied by the neighbouring TM2$_F$ and TM6$_G$. These conformational changes are the consequence of opposite rotations in the two TMDs of LptF and LptG; as shown in Fig.

3.8d, from a periplasmic view, LptF undergoes anti-clockwise rotation, as opposed to the clockwise rotation underwent by LptG. The reason for the different conformation of the $LptB_2FG$ structures relies on the different purification protocols used to purify the two complexes. Indeed, we expressed the complex in an LPS-deficient *E. coli* strain and purified it following an extensive delipidation protocol. On the other hand, when we followed a similar protocol to the one described by Luo et al., we obtained highly lipidated $LptB_2FG$ bound to LPS (Fig. 3.2a). Hence, our *pa*$LptB_2FG$ may be in a state representing the initial phase of the transportation cycle, while the previously published one represents LPS-bound state.

Moreover, the fact that the gate between $TM1_F$ - $TM5_G$ is open in Luo et al. $LptB_2FG$ while it is closed in our $LptB_2FG$ suggests that the preferential LPS entry gate may be the cavity between $TM5_F$ and $TM1_G$, which is aligned with the β-taco tunnel formed by LptF and LptC (Fig. S3.9a,b). These conformational fluctuations are also reflected in the positioning of the periplasmic domains, particularly LptF β-taco. In our $LptB_2FG$ structure, LptF β-taco is twisted from the original position in $LptB_2FG$, thus influencing on the orientation of LptC β-taco (see next section), which is in direct contact with LptF (Fig. S3.9c-d).

3.2.6 Cryo-EM structure of $LptB_2FGC$

We also obtained the cryo-EM structure of *pa*$LptB_2FGC$ following nanodisc reconstitution as described previously. Two-dimensional class averages of $LptB_2FGC$ cryo-EM particle indicated the presence of an extended portion above the nanodiscs consistent with the attachment of LptC β-taco on the top of LptF (Fig. S3.7c). Image processing yielded a final $LptB_2FGC$ map at overall 3.17 Å resolution (Figs. S3.7d, S3.10.

Based on our the previously published *ecl*$LptB_2FGC$ (PDB ID: 6MIT) [21], we could model the NBDs, TMDs and LptC for this structure (Fig. S3.10b,c). In the case of LptC, most of the particles did not show density for the transmembrane helix (TM_C) probably given the high flexibility of this domain. On the other hand, we could resolve the periplasmic β-taco domain of LptC in direct contact with LptF β-taco region (Figs. 3.9a,b; S3.10b,c). LptF and LptC form a

tunnel via a head-to-tail interaction between their C- and N-terminal β-strands, while there are limited contacts between side chains on the convex surfaces of the LptF and LptC β-taco strands (Fig. 3.9).

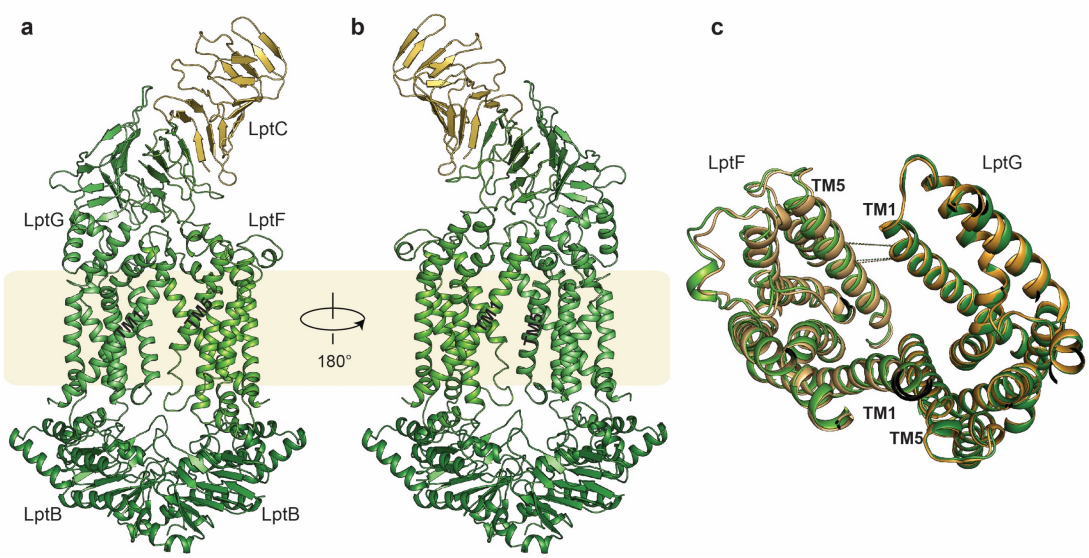


Figure 3.9 Cartoon representation of the LptB$_2$FGC complex. (a) Cartoon representation of LptB$_2$FGC. TM1$_G$ and TM5$_F$, forming the lateral gate are labelled. (b) Rotation of 180° along the y-axis relative to (a). (c) Superimposition of our *pa*LptB2FGC (green) and our *pa*LptB2FG (orange) structures.

Compared to our *pa*LptB$_2$FG structure, *pa*LptB$_2$FGC presents minimal changes in the conformation with an RMSD of 1.03 Å over 704 residues. We did not observe differences in the NBDs, while minimal changes in conformation are present in the TMDs. Indeed, in *pa*LptB$_2$FG the distance between TM5$_F$ and TM1$_G$ is slightly lower compared to *pa*LptB$_2$FGC (Fig. 3.9c). For instance, the Ile33$_G$ - Leu317$_F$ C$_\alpha$ distance increases from 9.9 Å (*pa*LptB$_2$FG) to 11.2 Å (*pa*LptB$_2$FGC); in addition, the Ala34$_G$ - Asp316$_F$ C$_\alpha$ distance increases from 8.9 Å (*pa*LptB$_2$FG) to 10.7 Å (*pa*LptB$_2$FGC). On the other side, the TM1$_F$ - TM5$_G$ gate is shut in both structures (Fig. 3.9c).

3.2.7 *pa*LptB$_2$FGC is in the pre-loading state

Notably, whilst the differences in TMDs conformation between our *pa*LptB$_2$FG and *pa*LptB$_2$FGC structures are relatively small, previous reports indicate a marked conformational opening of the TM5$_F$ - TM1$_G$ in LptB$_2$FGC from other Gram-negative bacteria species, compared to their correspondent LptB$_2$FG [21, 23]. In these structures, the increased aperture of the lateral gate is a consequence of the presence TM$_C$ between TM5$_F$ and TM1$_G$. Moreover, these complexes were LPS-bound, while we did not observe any density. This is not surprising since we expressed LptB$_2$FGC in an LPS-deficient *E. coli* strain and purified it following delipidation protocol. As shown in Fig. 3.6, a mixture of lipids was co-purified with the complex; however, these lipids could not be observed in cryo-EM.

Structure alignment between *pa*LptB$_2$FGC and LPS-bound *sf*LptB$_2$FGC (PDB ID: 6S8N, RMSD of 0.97 Å over 472 residues) indicates that the TM5$_F$ – TM1$_G$ distance is decreased in our structure (Fig. 3.10). While the LptF structures from the two complexes are perfectly aligned, the TMDs of LptG undergo an anti-clockwise rotation in *pa*LptB$_2$FGC (Fig. 3.10c), which also causes a complete closure of the other gate defined by TM1$_F$ and TM5$_G$.

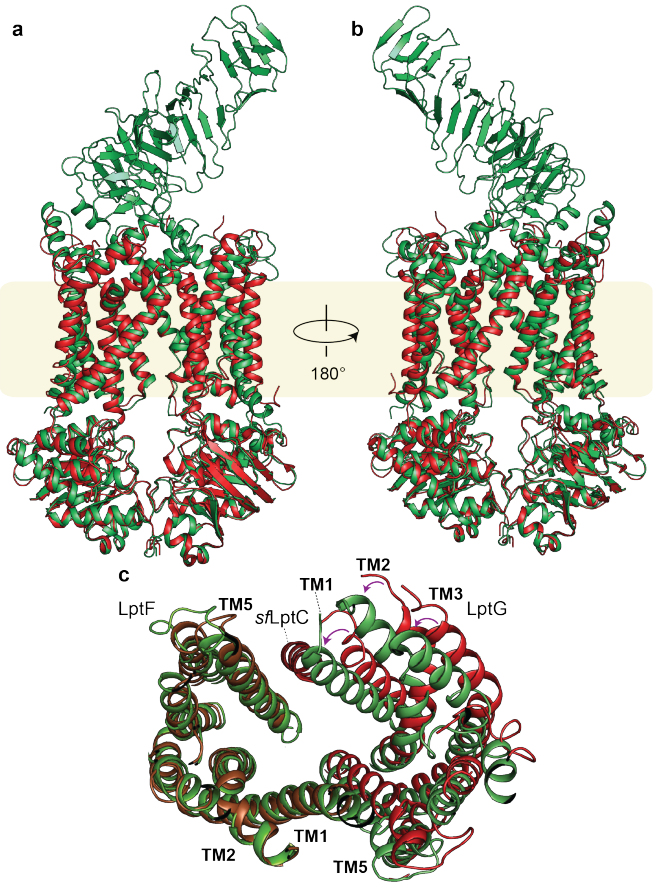


Figure 3.10 Structure alignment of *pa*LptB$_2$FGC and *sf*LptB$_2$FGC. (a) Superimposition of *pa*LptB$_2$FGC (green) with *sf*LptB$_2$FG (PDB ID: 6S8N, red) (b) Rotation of 180° along the y-axis relative to (a). (c) Periplasmic view of the cavity formed by the TMDs showing the conformational differences between *pa*LptB$_2$FGC and *sf*LptB$_2$FGC, such as the anti-clockwise rotation of *pa*TM1$_G$ and other *pa*LptG TMDs, in particular *pa*TM2$_G$ and *pa*TM3$_G$, hence decreasing the width of TM5$_F$– TM1$_G$ gate.

We observed similar features when comparing *pa*LptB$_2$FGC to the crystal structures of LptB$_2$FGC from *E. coli* (*ec*LptB$_2$FGC, PDB ID: 6MI7, RMSD of 2.34 Å over 448 residues) [22], but also *Enterobacter cloacae* (*ecl*LptB$_2$FGC, PDB ID: 6MIT, RMSD 5.05 Å over 480 residues) and *Vibrio cholerae* (*vc*LptB$_2$FGC, PDB ID: 6MJP RMSD 2.47 Å over 464 residues) [21], where LptF and LptG are rotated in opposite directions to given space to TM$_C$ (Fig. S3.11).

These observations suggest that the presence of LPS is somehow related to a wider aperture of both gates, in particular TM5$_F$ – TM1$_G$. Interestingly, when *sf*LptB$_2$FGC and

*ec*LptB$_2$FGC were trapped in their ATP-bound state using AMP-PNP or ADP-vanadate, respectively, TM5$_F$ – TM1$_G$ were in tight contact with each other and TM$_C$ was not visible despite the fact that most particle images showed the β-taco region, suggesting that TM$_C$ becomes disordered and moves away from TM5$_F$ – TM1$_G$ interface [22, 23]. Moreover, no LPS density was observed in the cavity formed by the TMDs, indicating that it had been extruded away from the transmembrane region. Hence the presence of LPS affects the position of TM$_C$ within the structure; LPS - TM$_C$ interactions stabilise the α-helix, while the absence of the substrate causes TM$_C$ to move away from TM5$_F$ - TM1$_G$ and become disordered.

Considering these observations, we may conclude that our *pa*LptB$_2$FGC structure represents a state of the transporter whereby, given the absence of bound LPS, TM$_C$ has slightly moved away from the TM5$_F$ – TM1$_G$ gate and becomes disordered. Consequently, TM5$_F$ and TM1$_G$ get closer to each other, assuming a conformation similar to what we observed in LptB$_2$FG. Since we observed lipids bound to the protein complex using nMS (Fig. 3.6), we cannot exclude that these influenced the conformation of the transporter, somehow mimicking LPS binding.

3.2.8 The periplasmic domain of *pa*LptB$_2$FG/LptB$_2$FGC is in a different conformation compared to its orthologues

Apart from the conformational differences concerning the TMDs, we also noticed structural changes on the β-taco portions of LptG and, in particular, LptF. In this case, the differences are likely not due to the absence of LPS bound, but rather are species-specific peculiarities of *P. aeruginosa*. In fact, a comparison between the previously reported *pa*LptB$_2$FG (PDB ID: 5X5Y) and any of LptB$_2$FG/LptB$_2$FGC structure shows that the β-taco domains are tilted to the side of LptG, different from the upright positioning of these domains in the crystal structure of *pa*LptB$_2$FG. Fig. S3.12 shows the superstition of *pa*LptB$_2$FG with LptB$_2$FG from *Klebsiella pneumoniae* (*kp*LptB$_2$FG, RMSD 2.68 Å over 608 residues), *ecl*LptB$_2$FGC (RMSD 2.72 Å over 472 residues) and *vc*LptB$_2$FGC (RMSD 2.45 Å over 472 residues).

In line with this, a comparison between our *pa*LptB$_2$FGC cryo-EM structure and *ecl*LptB$_2$FGC shows a unique structural variation between these orthologues (Fig. 3.11). The head-to-tail β-taco tunnel formed by LptC and LptF is indeed tilted of ~33° (Fig. 3.11b), thus suggesting a different arrangement of the periplasmic bridge in *P. aeruginosa*. A similar analysis indicated that the *pa*LptF-C β-taco is rotated by ~46° compared to the *V. cholerae* equivalent (Fig. 3.11d).

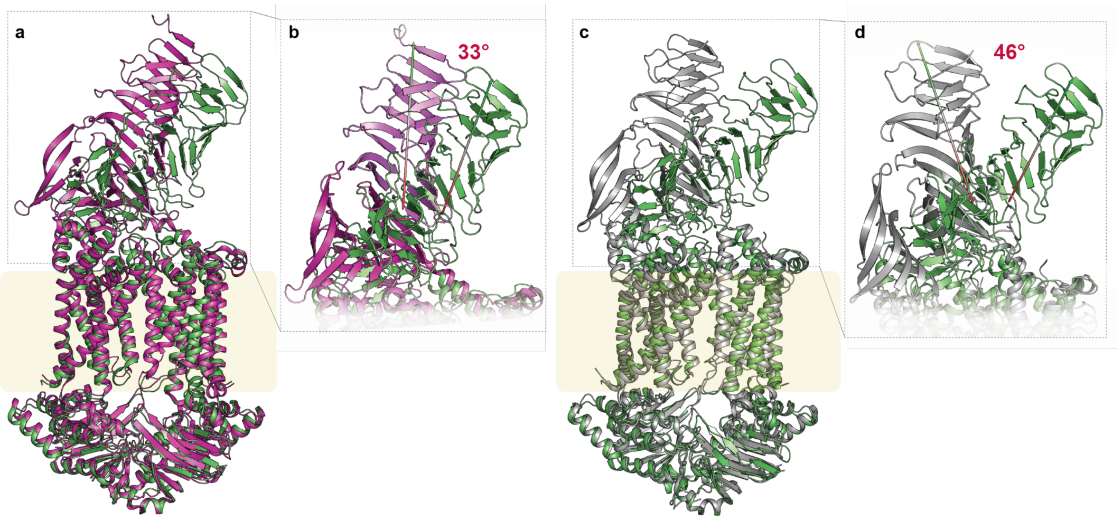


Figure 3.11 Structure alignment of *pa*LptB$_2$FGC and LptB$_2$FGC from *E. cloacae* and *V. cholerae*. (a) Superimposition of *pa*LptB$_2$FGC (green) with *ecl*LptB$_2$FG (PDB ID: 6MIT, magenta). (b) Superimposition of *pa*LptB$_2$FGC (green) with *vc*LptB$_2$FG (PDB ID: 6MJP, grey). (c) and (d) Focus on LptFC β-taco tunnel indicating the magnitude of rotation of *pa*LptFC β-taco compared to *ec*LptFC (c) and *vc*LptFC (d).

Another important feature of LptB$_2$FGC is the opening state of the β-taco domain; Owens et al. showed that *ecl*LptF and *vc*LptF are in two different conformations: indeed *ecl*LptF β-taco tunnel is open, while *vc*LptF is closed [21]. Superimposition of *pa*Lpt β-taco domain from our cryo-EM structure with both *ecl*LptF and *vc*LptF reveals that it is open, as indicated by the similar orientation of the key loop between strands β3 and β4 (residues 179-189, Fig. S3.13) in *pa*LptF and *ecl*LptF. In addition a smaller loop between β1 and β2 (residues 153-156) is also pointing outside in *pa*LptF.

Sequence alignment of LptF and LptG orthologues from different Gram-negative bacteria shows that many residues in the inner cavity of the transporter are conserved (Fig. S3.14). Interestingly, residues in LptF and LptG that have been previously indicated to point towards the LPS polar moieties (e.g. the phosphorylated glucosamine) [22, 23] are mutated in *P. aeruginosa*. Conversely, amino acids interacting with the hydrophobic acyl chains are conserved or mutated to chemically equivalent amino acids (Figs. S3.14-S3.15); these include I25 and I26 in $TM1_F$, but also L29 in $TM1_G$, M63 and M66 in $TM2_G$, and F317 and F319 in $TM5_G$. These findings suggest that LPS binding mode may be altered in *Pa*$LptB_2FG$, in particular, the interactions with the polar head of LPS may be different, while the contacts with the fatty acid moieties are retained. In addition, most of the residues in $TM5_F$ engaging hydrophobic interactions with TM_C (L305, L309, L310, L312) are highly conserved, confirming the importance of LptF for the interaction with LptC as already demonstrated by the higher number of $TM_C - TM5_F$ contacts compared to $TM_C - TM1_G$.

3.2.9 The periplasmic oligomeric protein LptA

Following successful expression and purification of $LptB_2FGC$, we set out to add another building block of the Lpt system, the periplasmic protein LptA. We expressed *Pa*LptA (also called LptH) [12] in *E. coli* strain BL21(DE3) and purified it following established protocols, as described in the methods section. Well-resolved mass spectra indicated the prevalence in solution dimeric LptA, although we could observe charge states series corresponding to monomer, trimer and traces of tetramer. Therefore, under our experimental conditions, the dimer was the most abundant form (Figure S3.16).

We decided to assess the influence of β-hairpin peptides on the oligomeric state of LptA. Murepavadin is a β-hairpin peptidomimetic that has been shown to specifically interact with *Pa*LptDE [40]. Given the similar β-taco fold of LptA and LptD, we wondered whether murepavadin could also influence LptA oligomeric state and, consequently, on the interaction between LptA and LptC. Analysis carried out using nMS indicated that murepavadin does not interact with LptA (Figure S3.17), thus the only target of the Lpt system is LptDE.

We then performed a similar experiment on thanatin, a naturally occurring antimicrobial peptide that has been shown to interact with *E. coli* LptA and LptDE [41], but also has a broad spectrum antimicrobial activity [42]. In this case, we could detect additional charge states series corresponding to thanatin binding both monomeric and dimeric LptA (Figure 3.12). In the presence of thanatin, the amount of dimeric LptA decreases in favour of monomeric LptA, thus indicating that thanatin is acting as a protein-protein interaction disruptor.

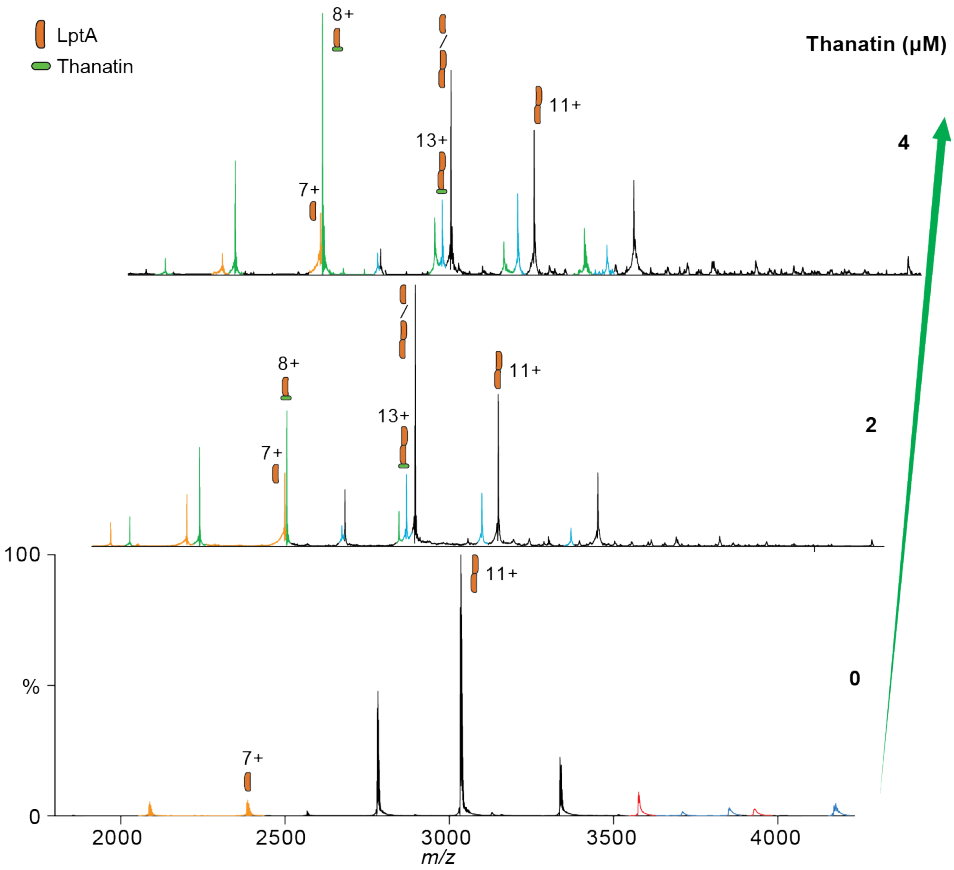


Figure 3.12 Effect of thanatin on LptA oligomeric state. LptA (2 μM) was mixed with increasing concentrations of Re-LPS (2 μM and 4 μM). Mass spectra show thanatin binding to both monomeric and dimeric LptA and indicate an increase of the abundance of monomeric LptA in the presence of thanatin.

3.2.10 LptB$_2$FGC – LptA interaction and thanatin influence

To broaden the scope of our investigation, we decided to analyse the interaction between t LptB$_2$FGC and LptA. Thus, we co-purified LptB$_2$FGC and LptA by adding an excess of LptA

before SEC, in order to saturate the available binding sites on LptC and remove the excess LptA given the difference in elution time of the two complexes. Indeed, SDS-PAGE analysis indicated that a fraction of LptA was co-eluted with $LptB_2FGC$ (Figure 3.13a). We then concentrated the eluted fractions and analysed the resulting solution using nMS. We could still observe the charge states series corresponding to $LptB_2FG$, but we also detected a new charge series with the mass corresponding to the $LptB_2FGC$-A complex, thus suggesting that all the available $LptB_2FGC$ complexes were saturated with LptA. Intriguingly, although LptA is prevalently a dimer in the tested concentrations when mixed with $LptB_2FGC$, we could only see a single LptA molecule interacting with the complex (Figure 3.13b). To exclude the possibility that the observed complex is a SEC artefact, we mixed previously purified LptA with $LptB_2FGC$ and obtained a similar result. further confirm this observation, we added LptA to $LptB_2FGC$-A, but could not observe any additional LptA subunit interacting with the complex (Figure 3.13c). These results suggest that once interacting with LptC, LptA assumes a conformation that precludes the interaction with other LptA molecules, thus preventing uncontrolled oligomerisation. Consequently, it is likely that the presence of the OM protein complex LptDE makes the formation of the periplasmic bridge more energetically favourable *in vivo*.

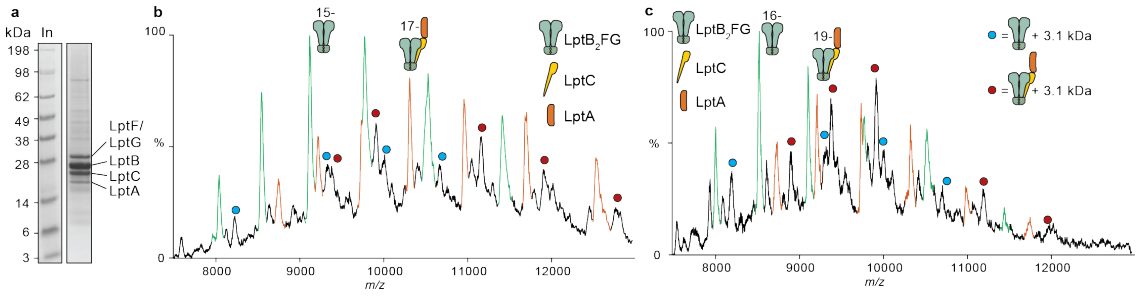


Figure 3.13 SDS-PAGE and nMS analysis of $LptB_2FGC$-A. (a) SDS-PAGE showing successful co-purification of $LptB_2FG$, LptC and LptA. (b) $LptB_2FGC$ and LptA were mixed before SEC, the eluted complex was concentrated and buffer exchanged and analysed through nMS. The spectrum indicates the presence of $LptB_2FG$ and LptB2FGC-A in solution, as well as lipid adducts. (c) $LptB_2FGC$ and LptA were mixed after SEC and just before nMS analysis. The spectrum indicates the presence of the same species; lipid-bound species are more evident, but a quantification is not possible given the heterogeneity of the sample.

We also investigated the influence of thanatin on the interaction between $LptB_2FGC$ and LptA. Interestingly, when we added thanatin to $LptB_2FGC$-A we could not see any interaction and the complex was not disrupted by the presence of the peptide. These results suggest that, once the complex is formed, it assumes a stable conformation that is more resistant to disruption compared to the LptA oligomer (Figure S3.18).

3.3 Discussion

In the current study, we integrated nMS and cryo-EM to perform a structural and functional analysis of the LPS transporter $LptB_2FG$ and its binding partners LptC and LptA. We carried out an extensive optimisation for the purification of the $LptB_2FG$ complex to completely delipidate the protein and allow the investigation of its ligand binding properties. Analysis carried out using nMS on $LptB_2FG$ indicate a very tight interaction with its substrate (Fig. 3.3) and suggest a critical role for the sugar moieties in the inner core for substrate binding, as indicated by decreased binding affinity of LA-diP, which lacks the Kdo moieties (Fig. 3.3). Amongst membrane lipids, CDL presents a much higher affinity for $LptB_2FG$ compared to PG and PE (Figs. S3.5 and 3.4a,b). Intriguingly, not only CDL tightly binds $LptB_2FG$, but also seems to have an active role in mediating substrate binding; indeed, we observed that addition of CDL to Re-LPS-bound $LptB_2FG$ causes a significant decrease of Re-LPS binding (Fig. 3.4c) and did not detect the formation of the ternary complex. These data suggest a competitive mode of action for CDL, whereby it competes with the substrate for binding, indicating a regulatory role played by this membrane lipid. This feature has already been observed for other inner membrane proteins transporting lipid substrates, like the lipid II flippase MurJ [43]. Moreover, CDL has been shown to modulate the oligomeric state and function of both bacterial [36] and mammalian (particularly mitochondrial) [44, 45] membrane proteins.

We also elucidated the role of nucleotide binding showing that the ADP-bound state of the protein complex presents lower affinity for the substrate; conversely, using AMP-PNP we

demonstrated that $LptB_2FG$ has a higher affinity for Re-LPS when trapped in its ATP-bound state (Fig. 3.5). These observations suggest a transport cycle whereby the complex captures the substrate in its ATP-bound state, while ATP hydrolysis causes a conformational change that helps pushing LPS from the substrate binding pocket in the TMDs to the β-taco tunnel formed by LptF, LptC and LptA. This explanation is in agreement with previously described LPS extraction model for *pa*$LptB_2FG$, where LPS loading is ATP binding dependant while LPS extraction is driven by ATP hydrolysis [25]. On the other hand, newly proposed models, based on the full $LptB_2FGC$ complex structure, advocate for an ATP binding independent LPS loading and ATP-dependent extraction, although there are discrepancies on the role of ATP hydrolysis [21, 23]. One explanation for these differences lie on the choice of compounds used to mimic the ATP-bound state (AMP-PNP vs ADP-vanadate) but also inter-species alterations need to be considered. In particular, *P. aeruginosa* Lpt proteins have very low sequence identity compared to other bacterial orthologues (Fig. 3.3, 3.4, 3.11) and both $LptB_2FG$ [25] and $LptB_2FGC$ structures present conformational differences (Fig. 3.9-S3.11). Hence, the LPS transport mechanism may present discrepancies from the ones described in other Gram-negative bacteria, though further investigation is necessary.

To dissect the role of LptC in the structure and conformation of the transporter, we expressed and purified $LptB_2FGC$. Interestingly, we observed a great amount of co-purified lipids (Fig. 3.6), although we followed the same protocol used for $LptB_2FG$ where we obtained successful delipidation. Since the only difference introduced in the system in LptC, it is likely that the presence of TM_C introduces additional surfaces for lipid binding that may be less accessible to detergent. In line with these, in the cryo-EM structure of *ec*$LptB_2FGC$ a lipid density modelled as POPG was observed between TM_C, $TM5_F$ and $TM6_F$ [22], hence indicating that the presence of this extra transmembrane helix creates patches for lipid binding. It is possible that lipids are necessary to stabilise the LptC-LptFG interaction, hence they are resistant to delipidation protocols. We also observed the presence of $LptB_2FG$ in solution and correspondent lipid-bound adducts. This can be justified by the fact that lipids co-purifying with $LptB_2FGC$ are dissolved in mixed lipid-detergent micelles and may interact with $LptB_2FG$.

In our $LptB_2FG$/$LptB_2FGC$ cryo-EM structures we observed rotation of $TM1_F$ and $TM2_F$ and consequent closing of the $TM1_F$-$TM5_G$ gate (Figs. 3.8, 3.9) suggesting that preferred LPS entrance may be from the $TM5_F$ - $TM1_G$ gate; this cavity is also perfectly aligned with the convex surface of the β-taco tunnel formed by LptF and LptC (Fig. S3.9a,b). These analyses are in line with the crystal structures of *ecl*$LptB_2FGC$ and *vc*$LptB_2FGC$ [21]. On the other hand, Tang et. al observed one acyl tail of LPS and a DDM molecule interacting with $TM1_F$-$TM5_G$ helices in *sf*$LptB_2FGC$ [23], hence they concluded that the $TM1_F$-$TM5_G$ gate may still represent an LPS entry route. Given these observations we can conclude that the $LptB_2FGC$ captures LPS from both sides, although there is a preference for the $TM5_F$ - $TM1_G$ gate given the wider opening and the alignment with the β-taco tunnel.

Comparison of *pa*$LptB_2FGC$ with other $LptB_2FGC$ structures indicated that the above-mentioned $TM5_F$ - $TM1_G$ gate is slightly tighter in our structure, (Fig. 3.10-S3.11). All the published $LptB_2FGC$ structures are bound to LPS which interacts with TM_C, thus stabilising the helix. Conversely, in our *pa*$LptB_2FGC$ TM_C assumes is disordered and moves away from $TM5_F$ - $TM1_G$ gate, thus justifying the decreased distance between the two helices. A similar phenomenon has been described for *sf*$LptB_2FGC$ and *ec*$LptB_2FGC$ in their ATP-bound states, whereby LPS was not observed inside TMDs cavity [22, 23]. This suggests that our *pa*$LptB_2FGC$ is in a conformation where LPS is not bound (pre-loading state) but still influenced by surrounding lipids, as indicated by our nMS data (Fig. 3.6).

Whilst this conformational difference may be ascribed to the absence of LPS, the orientation of the β-taco domain of LptF is likely a specific feature of *P. aeruginosa*. Indeed, we noticed that LptF β-taco is tilted in $LptB_2FG$ crystal structure (Fig. 3.11); in addition, in our $LptB_2FG$ cryo-EM structure, given presence of LptC, we clearly show that LptF tilting determines a drastic change in the orientation of the periplasmic bridge (Fig. S3.13). This is probably the greatest difference between *pa*$LptB_2FGC$ and its orthologues and may be representative of specific spatial orientation of the whole Lpt system in *P. aeruginosa*. In line with this, *pa*LptD presents a longer periplasmic domain (~90 extra residues, the equivalent of 5-6 β-taco strands) compared to its

orthologues (Fig. 3.2), which contributes to a different organisation of the periplasmic bridge. All these observations, along with the low sequence similarity between Lpt proteins and their bacterial orthologues, point towards a different structure and regulation of *P. aeruginosa* Lpt system. Further studies are necessary to shed light on the structural details of these discrepancies, in particular substrate and nucleotide binding studies on $LptB_2FGC$ (for instance, using nMS and cryo-EM) are necessary to further clarify the transport mechanism.

Our nMS analysis also demonstrated that dimeric LptA is able to interact with $LptB_2FGC$ even in the absence of LPS or nucleotides, and this is the first time the complex is observed using structural methods (Fig. 3.13). Our spectra showed only one copy of LptA binding to the complex, although LptA is a dimer in solution, suggesting that LptC-LptA interaction is tighter than LptA-LptA interaction. On the other hand, the periplasmic bridge is made of multiple copies of LptA and the presence of OM protein LptDE may be necessary to build the bridge spanning between the two membrane protein complexes, thus explaining the absence of multimeric LptA bound to $LptB_2FGC$. Notably, while thanatin disrupted the LptA dimer (Fig. 3.12), addition of thanatin was not capable of disrupting $LptB_2FGC$-A (Fig. S3.18), in line with the observation that LptC-LptA interaction is stronger than LptA-LptA. However, thanatin could still impair interaction with LptC during the early stages of Lpt complex formation, i.e. before the IM Lpt sub-complex physically interacts with the periplasmic bridge.

In summary, this study sets the ground for further investigation on the IM extraction of LPS in *P. aeruginosa*, we revealed an unprecedented role for CDL in the modulation of the sub-complex $LptB_2FG$ and obtained the high-resolution structure of the $LptB_2FG$ and $LptB_2FGC$ complexes. The comparison of the conformation of our *pa*$LptB_2FG$ with the previously published *pa*$LptB_2FG$ revealed changes related to LPS binding. Similarly, comparison with other Gram-negative bacteria orthologues provided unprecedented information on TMDs conformational differences between pre-loading and LPS-bound state, but also yielded insights on the β-taco spatial orientation in different bacterial species.

Future work will be devoted to the functional evaluation of CDL and nucleotide binding to $LptB_2FGC$ using both nMS and cryo-EM, as well as conformational analysis of fully lipidated

$LptB_2FGC$ vs delipidated $LptB_2FGC$. Moreover, we will optimise $LptB_2FGC$-A purification in order to gain molecular details on this interaction through structural and biophysical methods. In addition, the influence of substrate or CDL binding to the protein complex, as well as the effect of LptA docking on LptC may be investigated using HDX-MS.

3.4 Supplementary Information

3.4.1 Supplementary Figures

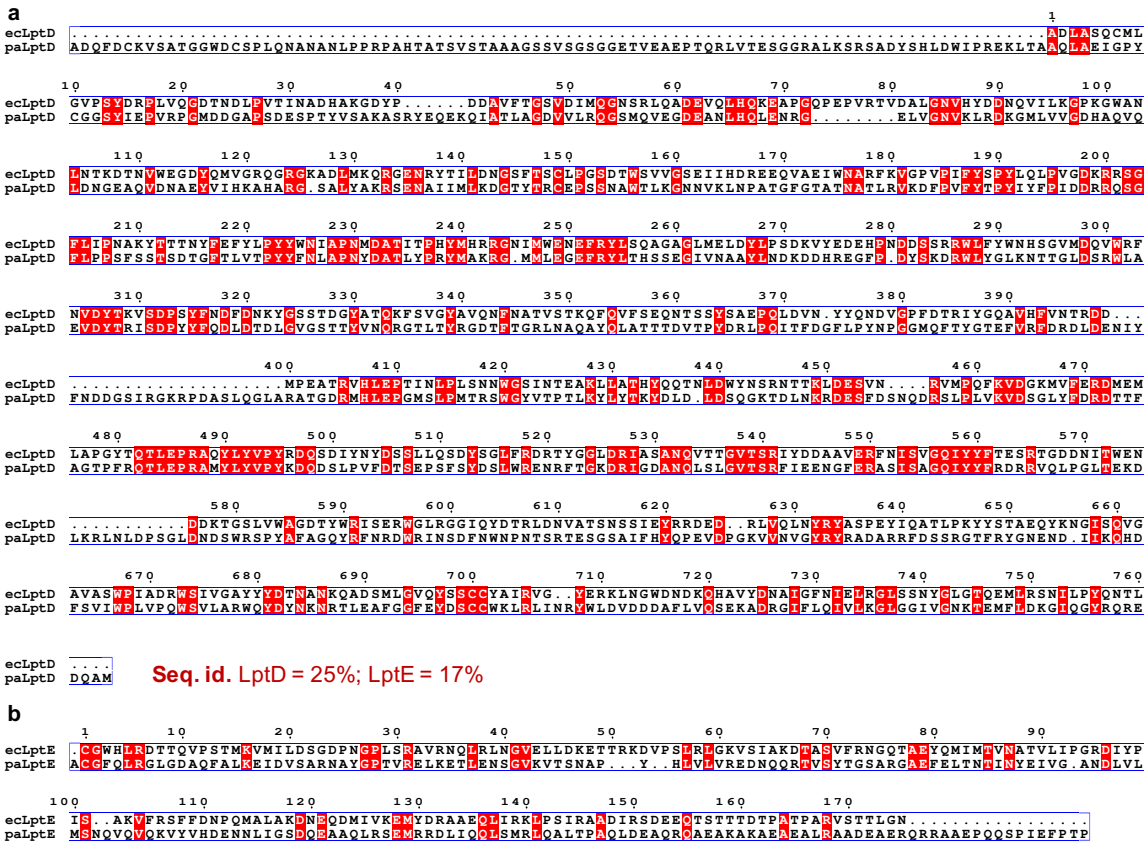


Figure S3.1 Sequence alignment of *ec*LptDE-*pa*LptDE. (a) Sequence alignment for LptD. (b) Sequence alignment for LptE. Red colour indicates conserved residues, yellow indicates highly similar residues. Alignments generated with NPS@ClustalW [31, 32] and ESPript 3.0 [33].

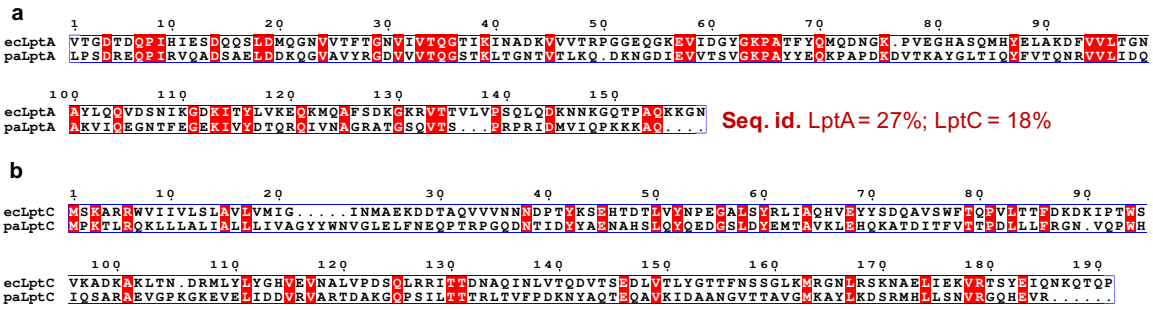


Figure S3.2 Sequence alignment of *ec*LptA-*pa*LptA and *ec*LptC-*pa*LptC. (a) Sequence alignment for LptA. (b) Sequence alignment for LptC. Legend as in Fig. 3.2.

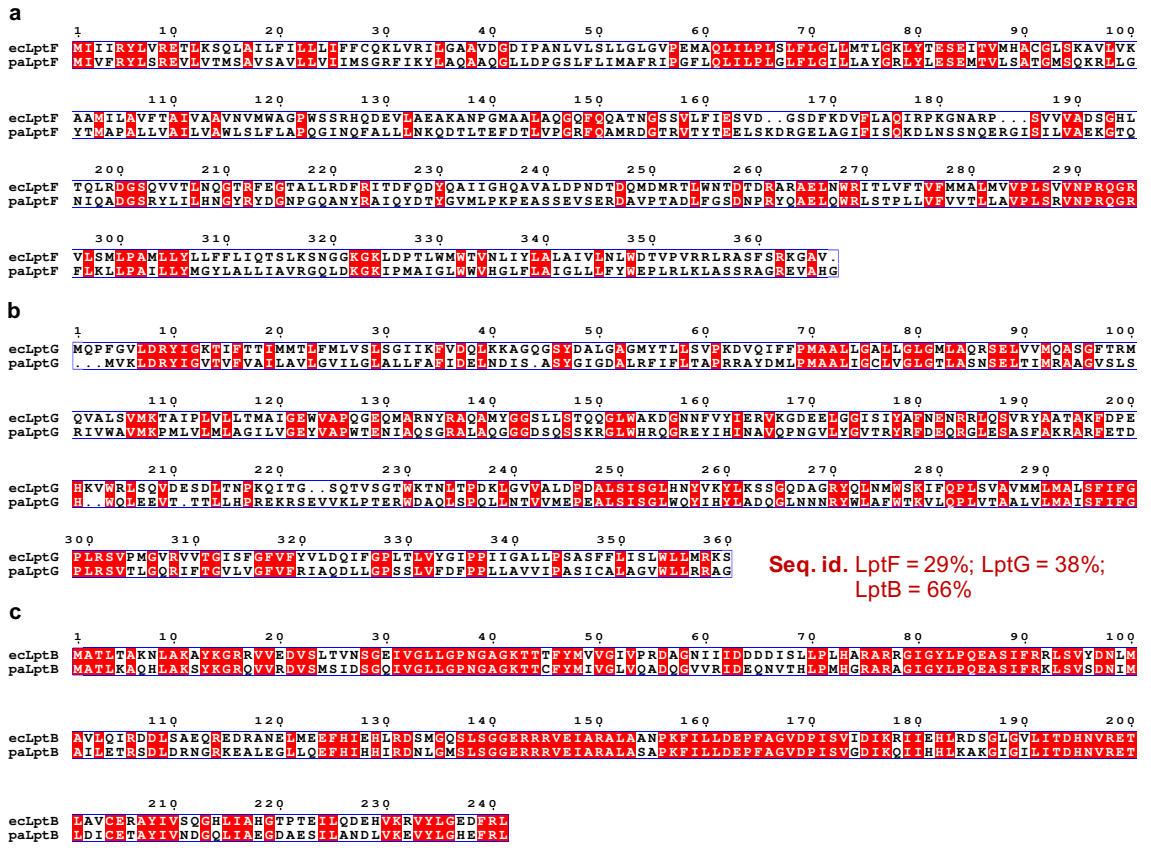


Figure S3.3 Sequence alignment of *ec*LptB$_2$FG-*pa*LptB$_2$FG. (a) Sequence alignment for LptF. (b) Sequence alignment for LptG. (c) Sequence alignment for LptB. Legend as in Fig. 3.2.

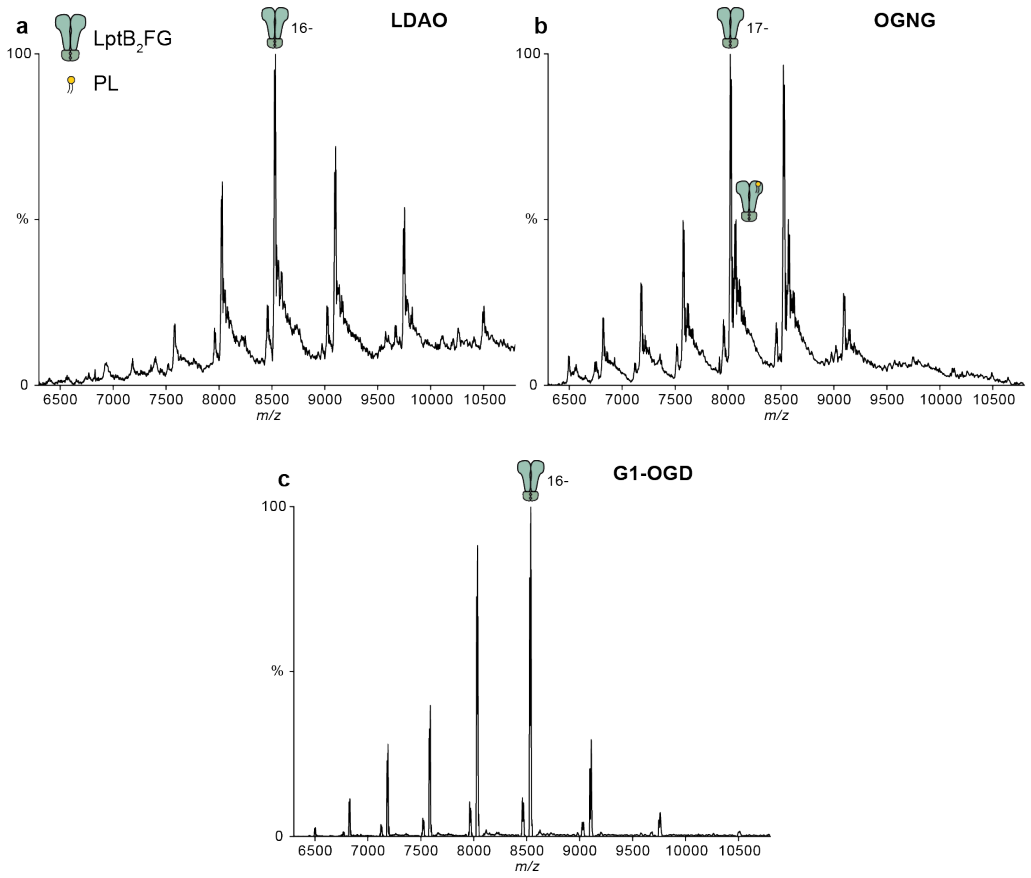


Figure S3.4 Optimisation of MS conditions and detergent for further $LptB_2FG$ analysis. Mass spectra of $LptB_2FG$ in 200 mM ammonium acetate supplemented with 0.05% LDAO (a), 0.12% OGNG (b), 0.02% G1-OGD (c). The minor peaks in (c) correspond to the 1.4 kDa adduct.

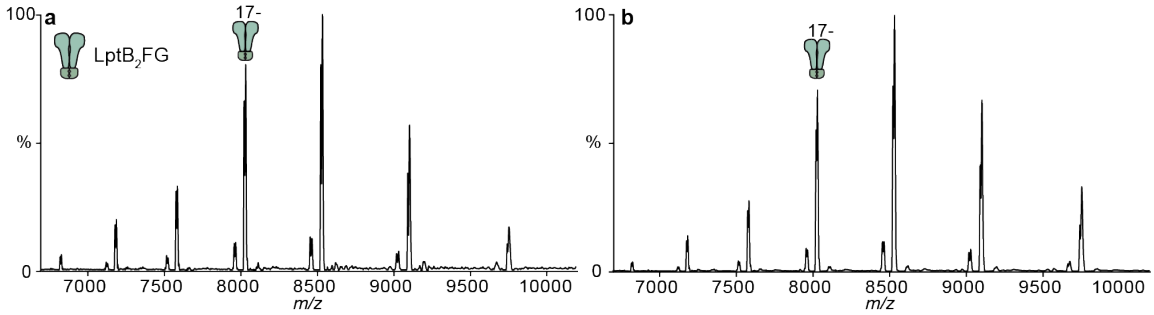


Figure S3.5 Mass spectra of $LptB_2FG$ in the presence of DMPE and DMPG. $LptB_2FG$ (3 μM) was mixed with 5 μM of of DMPE (a) or DMPG (b), but no significant binding was observed.

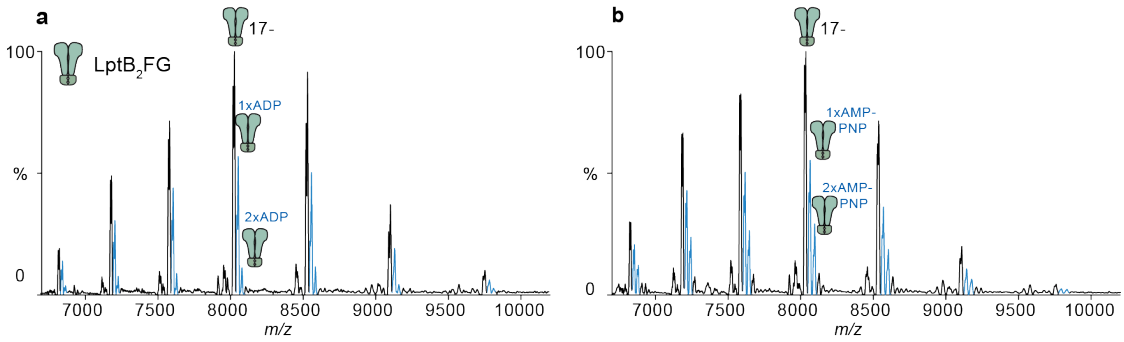


Figure S3.6 nMS analysis of nucleotide-$LptB_2FG$ interaction. (a) A solution of $LptB_2FG$ (4 μM) was mixed with ATP-Mg^{2+} (20 μM), the resultant spectrum shows binding of 1 and 2 ADP molecules to $LptB_2FG$, indicating that the protein complex is enzymatically active. (b) A solution of $LptB_2FG$ (4 μM) was mixed with AMP-PNP (20 μM), nMS analysis indicates binding of 1 and 2 AMP-PNP molecules to $LptB_2FG$.

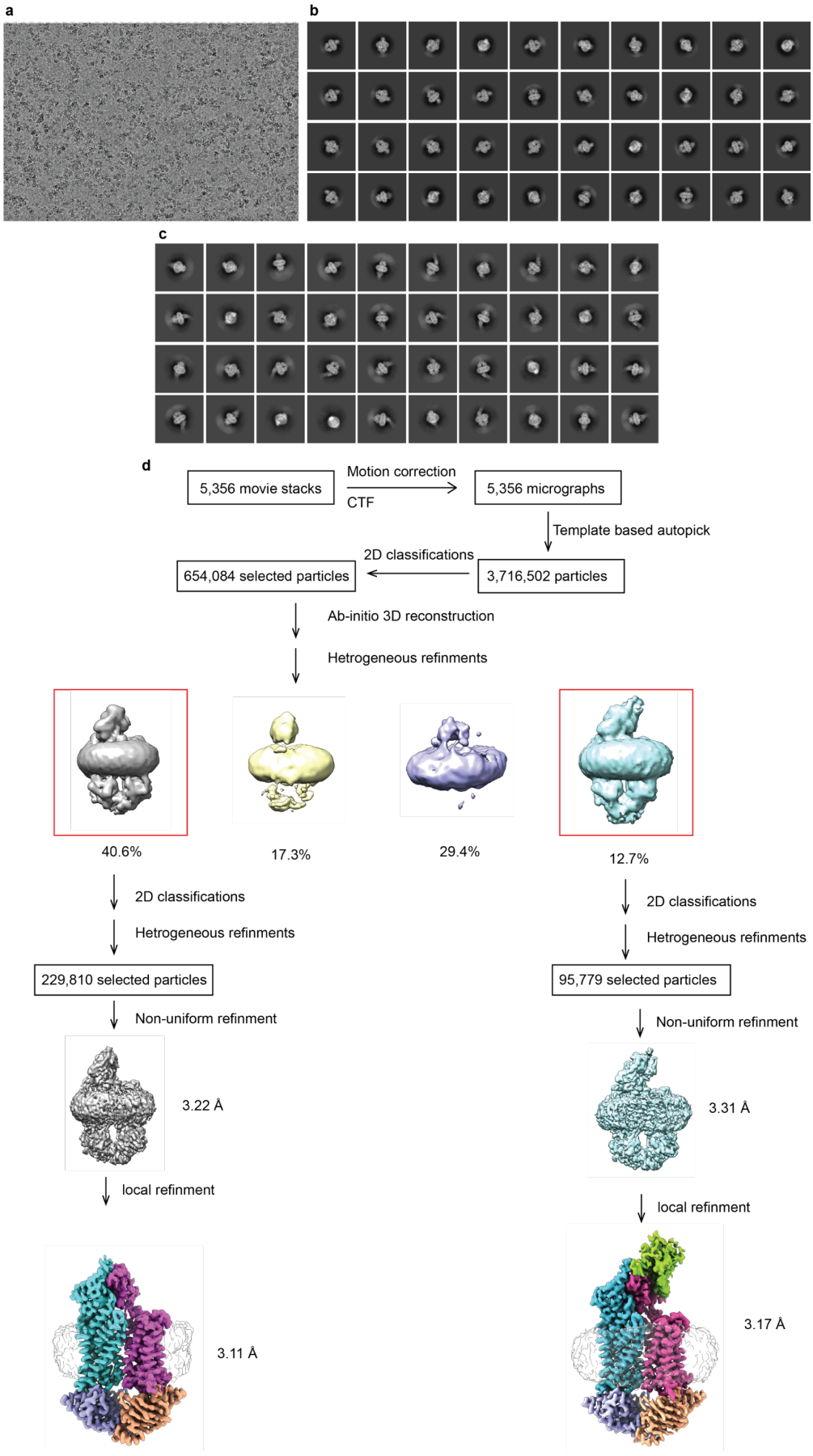

a
b
c
d
5,356 movie stacks
Motion correction
CTF
5,356 micrographs
Template based autopick
3,716,502 particles
2D classifications
654,084 selected particles
Ab-initio 3D reconstruction
Hetrogeneous refinments
40.6%
17.3%
29.4%
12.7%
2D classifications
Hetrogeneous refinments
229,810 selected particles
Non-uniform refinment
3.22 Å
local refinment
3.11 Å
2D classifications
Hetrogeneous refinments
95,779 selected particles
Non-uniform refinment
3.31 Å
local refinment
3.17 Å

Figure S3.7 Image processing for the cryo-EM data of nanodisc-embedded $LptB_2FG$ and LptB2FGC. (a) Representative cryo-EM image of $LptB_2FG$/$LptB_2FGC$ in nanodiscs. (b) 2D class averages for $LptB_2FG$ particles. (c) 2D class averages for $LptB_2FGC$ particles. (d) Workflow for the 3D classification and refinement of cryo-EM particle images. After the first round of 3D classification, all the particles classified into the two best classes, corresponding to $LptB_2FG$ and $LptB_2FGC$, were kept for further processing, yielding to the final models for $LptB_2FG$ and $LptB_2FGC$.

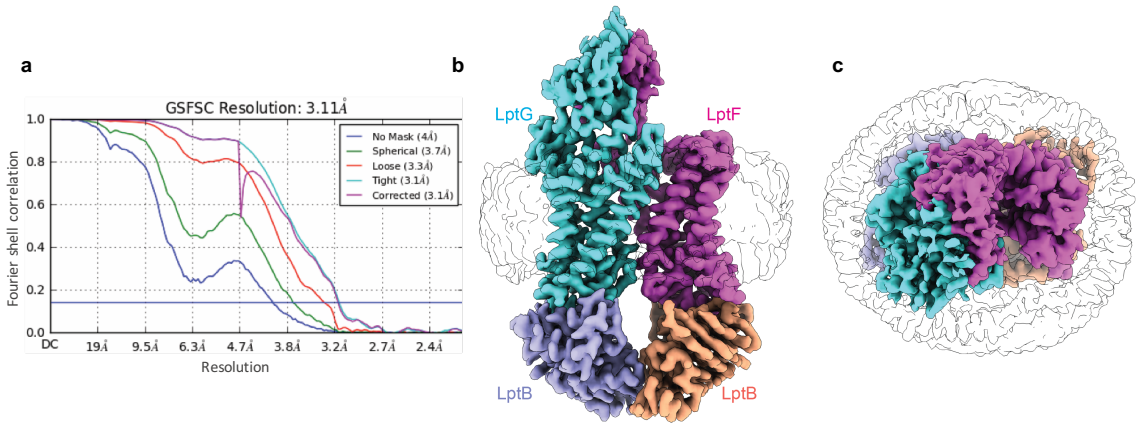


Figure S3.8 Cryo-EM single-particle data processing for $LptB_2FG$. (a) Gold-standard FSC curves of the final EM maps of $LptB_2FG$. (b) Surface view of the cryo-EM map of $LptB_2FG$, filtered to 3.11 Å resolution (c) Periplasmic view of $LptB_2FG$.

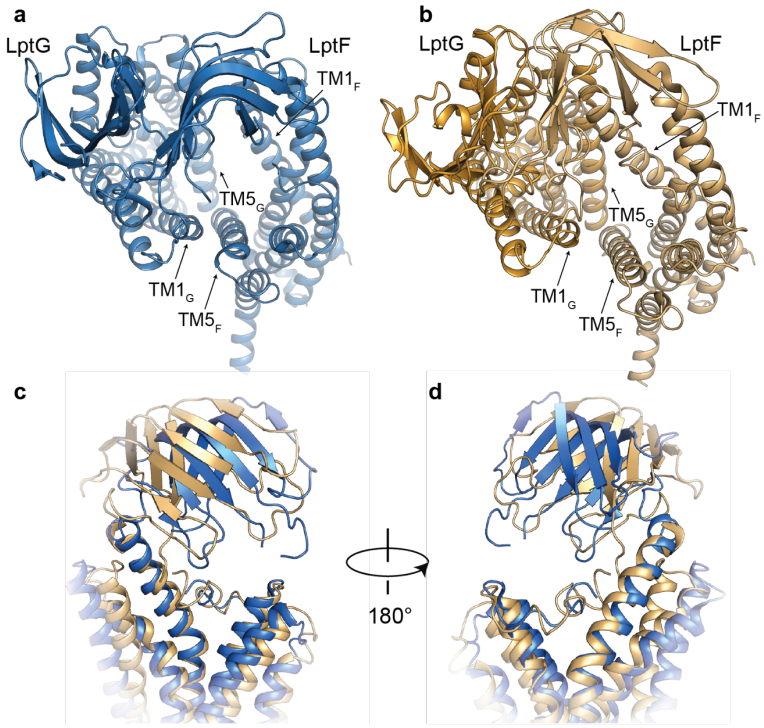


Figure S3.9 Further comparison between *pa*LptB$_2$FG structures. (a) Periplasmic view of LptF and LptG of *pa*LptB$_2$FG (PDB ID: 5X5Y) showing both β-taco domains and TMDs. (b) Same as (a) applied on our *pa*LptB$_2$FG. (c) Focus on LptF β-taco domain presenting a different spatial orientation in our *pa*LptB$_2$FG compared Luo et al. *pa*LptB$_2$FG. LptF β-taco is twisted towards the central axis of the transporter and has a more disordered structure. (d) Rotation of 180° along the y-axis relative to (a).

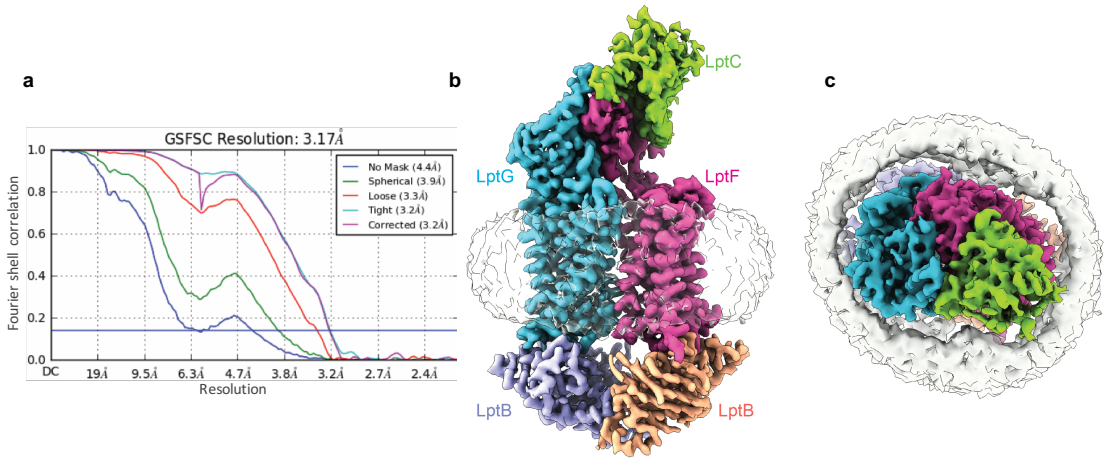


Figure S3.10 LptB$_2$FGC complex cryo-EM structure. (a) Gold-standard FSC curves of the final EM maps of LptB$_2$FGC. (b) Surface view of the cryo-EM map of LptB$_2$FGC, filtered to 3.17 Å resolution (c) Periplasmic view of LptB$_2$FGC.

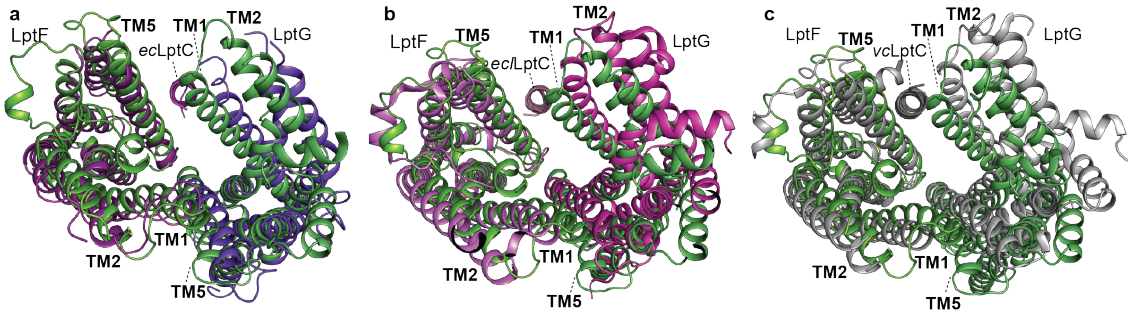


Figure S3.11 Structure alignment of *pa*LptB$_2$FGC and LptB$_2$FGC from other Gram-negative bacteria, periplasmic view. (a) Superimposition of *pa*LptB$_2$FGC (green) with *ec*LptB$_2$FGC (PDB ID: 6MI7, purple). (b) Superimposition of *pa*LptB$_2$FGC (green) with *ecl*LptB$_2$FG (PDB ID: 6MIT, magenta). (c) Superimposition of *pa*LptB$_2$FGC (green) with *ec*LptB$_2$FG (PDB ID: 6MJP, grey). The periplasmic views of all three structure alignments show a similar anti-clockwise rotation of *pa*LptG TMDs, thus a smaller aperture on the TM5$_F$– TM1$_G$ LPS entry gate.

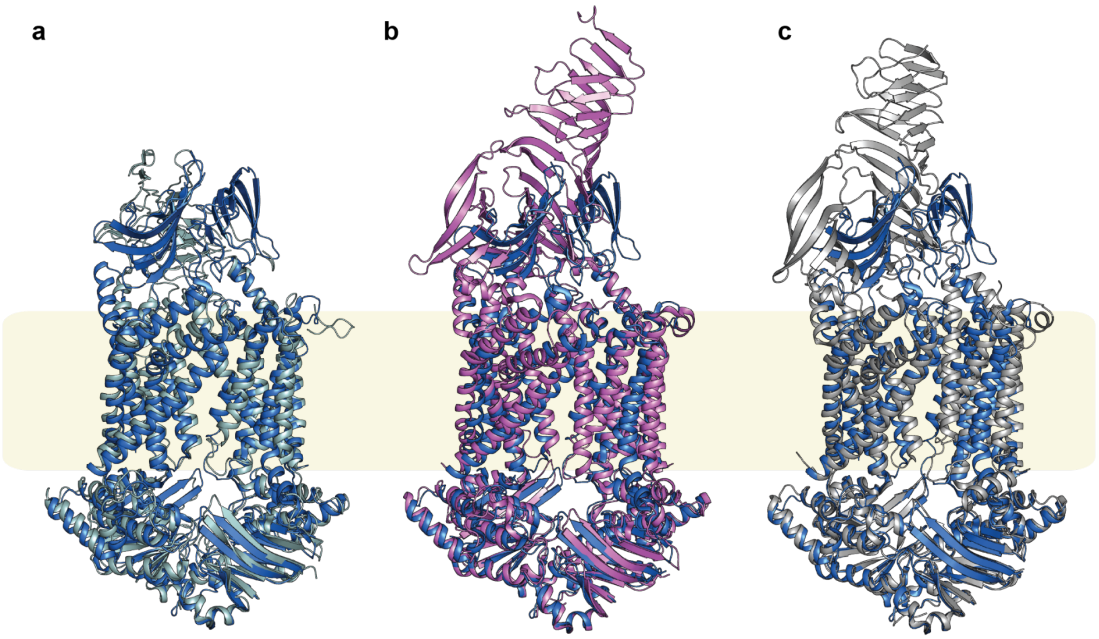


Figure S3.12 Structure alignment of *pa*LptB$_2$FG and LptB$_2$FG/LptB$_2$FGC from other Gram-negative bacteria. (a) Superimposition of *pa*LptB$_2$FG (pdb ID: 5X5Y, blue) with *kp*LptB$_2$FG (PDB ID: 5L75, cyan) (b) Superimposition of *pa*LptB$_2$FG (pdb ID: 5X5Y, blue) with *ecl*LptB$_2$FG (PDB ID: 6MIT, magenta). (c) Superimposition of *pa*LptB$_2$FG (pdb ID: 5X5Y, blue) with *vc*LptB$_2$FG (PDB ID: 6MJP, grey). In all three cases *pa*LptB$_2$FG β-taco domain is tilted towards LptF, forming a smaller angle with the membrane.

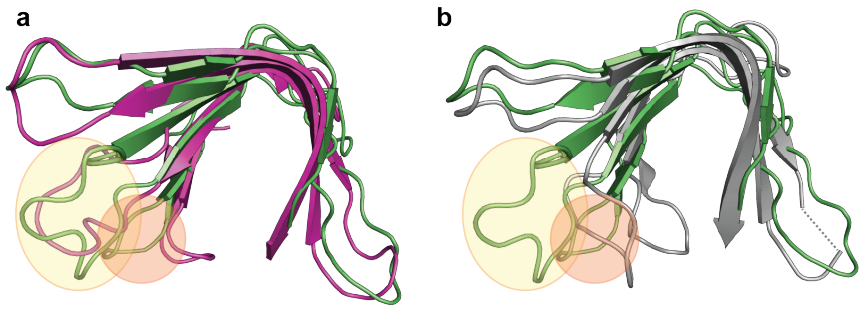


Figure S3.13 Structure alignment of *pa*LptF β-taco with *E. cloacae* and *V. cholera* homologues. (a) Superimposition of *pa*LptF β-taco (green) with *ecl*LptF β-taco (PDB ID: 6MIT, magenta). The two structures present a similarly open conformation: loop 179-189 (yellow circle) points towards outside in both structures, while loop 153-156 (orange circle) is more open in

*pa*LptF. (b) Superimposition of *pa*LptF β-taco (green) with *vc*LptF β-taco (PDB ID: 6MJP, grey). In this case, all loops in cause a more open conformation in *pa*LptF compared to *vc*LptF.

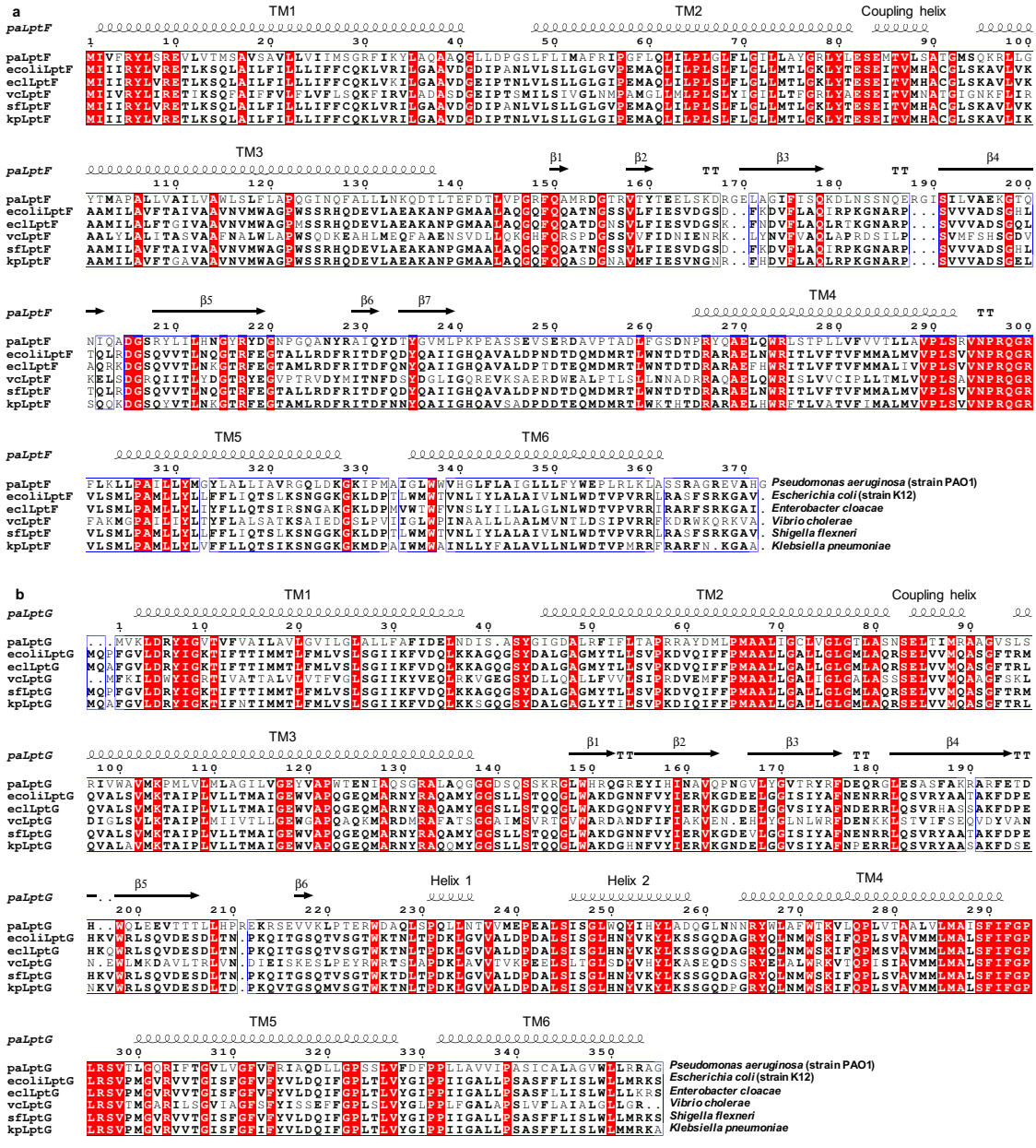


Figure S3.14 Sequence alignment of *pa*LptF and *pa*LptG with their orthologues from other Gram-negative bacteria. (a) Sequence alignment for LptF. (b) Sequence alignment for LptG. We selected for alignment the five Gram-negative bacteria species that we mentioned in this paper. Red colour indicates conserved residues, yellow indicates highly similar residues.

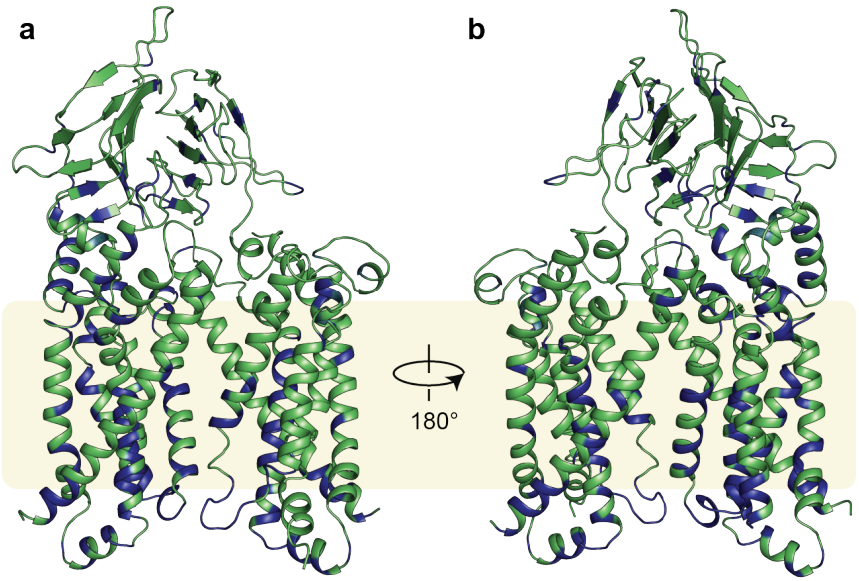


Figure S3.15 Mapping of highly conserved residues on *pa*LptFG structures. Conserved residues are coloured in dark blue on the structures. Most of the conserved regions are in the internal part of the cavity formed by the TMDs.

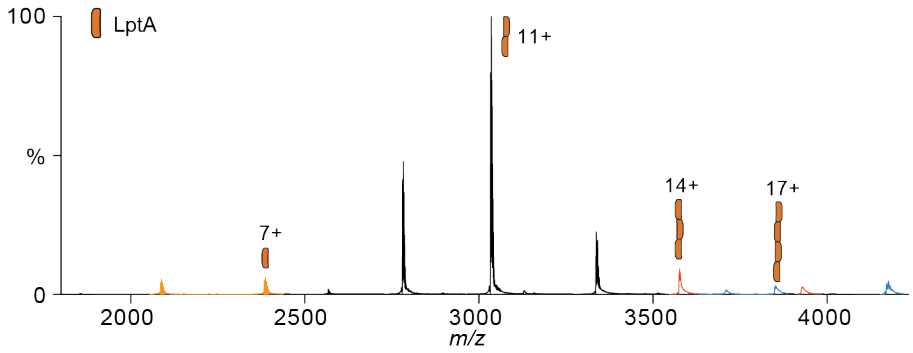


Figure S3.16 nMS analysis of LptA. Mass spectrum of LptA indicating the presence in solution of monomer, dimer, trimer and tetramer, with a clear prevalence of dimeric species.

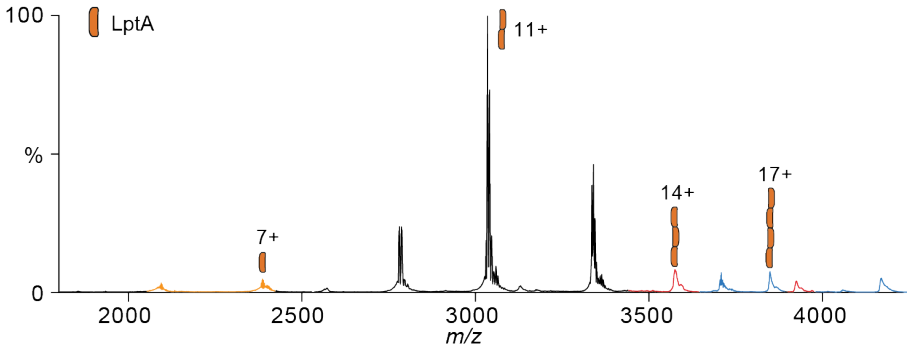


Figure S3.17 nMS analysis of murepavadin effect on LptA oligomeric state. Murepavadin (8 μM) was added to LptA (2 μM), but no binding or effect on LptA oligomeric state was observed.

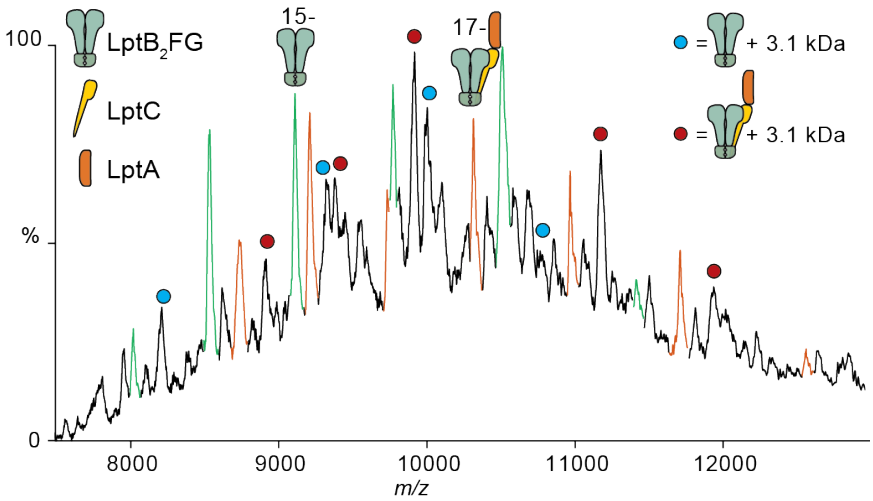


Figure S3.18 nMS analysis of thanatin effect on $LptB_2FGC$:LptA interaction. Thanatin (10 μM) was added to $LptB_2FGC$-A complex (4 μM) and analysed through nMS. The spectrum indicates the presence of $LptB_2FG$ and LptB2FGC-A in solution, as well as lipid adducts, but no thanatin-bound adducts were detected.

3.4.2 Supplementary Tables

Figure	Complex	Experimental molecular mass ± s.d. (Da)	Theoretical molecular mass (Da)
Figure 3.2a	$LptB_2FG$	136,485 ± 3.9	136,661
	$LptB_2FG$ + 1x PL	137,174 ± 2.8	N/A
	$LptB_2FG$ + 2x PL	137,884 ± 16	N/A
	$LptB_2FG$ + LPS	139,984 ± 11	N/A
Figure 3.2b	$LptB_2FG$ + 2x Lipid IVA	139,092 ± 1.8	139,471
	$LptB_2FG$ + 3x Lipid IVA	140,494 ± 2.5	140,876
	$LptB_2FG$ + 4x Lipid IVA	141,902 ± 11	142,282
Figure S3.4a	$LptB_2FG$	136,552 ± 7.9	136,661
Figure S3.4b	$LptB_2FG$	136,547 ± 4.3	136,661
	$LptB_2FG$ + 1x PL	137,293 ± 3.2	N/A
Figure S3.4c	$LptB_2FG$	136,632 ± 18	136,661
Figure 3.3b	$LptB_2FG$	136,617 ± 12	136,661
	$LptB_2FG$ + 1x Re-LPS	138,852 ± 10	138,900
	$LptB_2FG$ + 2x Re-LPS	141,129 ± 8.4	141,139
	$LptB_2FG$ + 3x Re-LPS	143,396 ± 16	143,378
Figure 3.3c	$LptB_2FG$	136,639 ± 13	136,661
	$LptB_2FG$ + 1x LA-diP	138,403 ± 7.9	138,459
	$LptB_2FG$ + 2x LA-diP	140,207 ± 11	140,257
	$LptB_2FG$ + 3x LA-diP	142,013 ± 12	142,055
Figure S3.5	$LptB_2FG$	136,636 ± 20	136,661
Figure 3.4	$LptB_2FG$	136,646 ± 12	136,661
	$LptB_2FG$ + 1x CDL	137,892 ± 19	137,902
	$LptB_2FG$ + 2x CDL	139,136 ± 15	139,143
Figure S3.6	$LptB_2FG$	136,599 ± 12	136,661
	$LptB_2FG$ + 1x ADP	137,036 ± 21	137,088
	$LptB_2FG$ + 2x ADP	137,467 ± 18	137,515
	$LptB_2FG$ + 1x AMP-PNP	137,099 ± 8.7	137,167
	$LptB_2FG$ + 2x AMP-PNP	137,634 ± 8.8	137,673
Figure 3.5	$LptB_2FG$	136,602 ± 8.5	136,661
	$LptB_2FG$ + 1x ADP	137,039 ± 10	137,088

	$LptB_2FG$ + 2x ADP	137,461 ± 10	137,515
	$LptB_2FG$ + 1x AMP-PNP	137,135 ± 13	137,167
	$LptB_2FG$ + 2x AMP-PNP	137,651 ± 12	137,673
	$LptB_2FG$ + 1x Re-LPS	138,843 ± 5.2	138,900
	$LptB_2FG$ + 1x Re-LPS + 1x ADP	139,278 ± 8.4	139,327
	$LptB_2FG$ + 1x Re-LPS + 1x AMP-PNP	139,476 ± 11	139,406
Figure 3.6	$LptB_2FG$	136,543 ± 72	136,661
	$LptB_2FG$ + Lipid adduct 1	139,669 ± 109	N/A
	$LptB_2FGC$	158,236 ± 58	158,289
	$LptB_2FGC$ + Lipid adduct 1	161,281 ± 69	N/A
	$LptB_2FGC$ + Lipid adduct 2	163,106 ± 82	N/A
Figure S3.16-S3.17	LptA monomer	16,693 ± 1.4	16,680
	LptA dimer	33,382 ± 1.8	33,360
Figure 3.12	LptA monomer + thanatin	19,131 ± 2.6	19,116
	LptA dimer + thanatin	35,818 ± 1.6	35,796
Figure 3.13-S3.18	$LptB_2FG$	136,606 ± 70	136,661
	$LptB_2FGC$-A	175,002 ± 63	174,969
	$LptB_2FG$ + Lipid adduct	139,719 ± 35	N/A
	$LptB_2FGC$-A + Lipid adduct	178,114 ± 88	N/A

Table S3.1 Molecular masses determined by nMS.

Data set	**$LptB_2FG$ and $LptB_2FGC$ reconstituted in nanodiscs (1D1)**	
Ligand	**$LptB_2FG$**	**$LptB_2FGC$**
Data collection and processing		
Magnification	81,000	
Voltage (kV)	300	
Electron Microscope	Krios-GIF-K3	
Defocus range (μm)	-1.0 to -2.25	
Total exposure time (s)	3.4	
Energy filter width (eV)	20	

Pixel size (Å)	1.08	
Total dose (e^-/ Å^2)	50	
Number of frames	40	
Does rate (e^-/ phys. Pixel/s)	17.30	
No. of initial micrographs	5,356	
No. of initial particles	3,716,502	
No. of final particles	95,779	182,590
Symmetry	C1	C1
Resolution (Å)	3.17	3.14
FSC threshold	0.143	0.143
Map resolution range (Å)	2.80-30.00	2.67-30.00
Refinement		
Model resolution cut-off (Å)	3.17	3.14
Model composition		
No. of Protein residues	1,188	1,328
No. ligands	0	0
RMSD		
Bond lengths (Å)	0.006	0.006
Bond angles (°)	1.132	0.846
Validation		
MolProbity score	3.30	3.56
Ramachandran plot (%)		
Favored (%)	89.56	89.83
Allowed (%)	10.44	10.17
Disallowed (%)	0	0
CC[a] Mask	0.80	0.82
CC box	0.69	0.71
CC Vol	0.79	0.80

a, correlation coefficient

Table S3.2 Cryo-EM data collection, processing, and refinement statistics.

3.5 Materials and methods

3.5.1 $LptB_2FG$ expression and membranes isolation

The plasmids used for over-expression of *P. aeruginosa* $LptB_2FG$ and $LptB_2FGC$ were a kind gift from Yihua Huang (Chinese Academy of Sciences, Beijing, China). PCR. The plasmids were constructed as described in reference [25] and consisted of one pQLink-LptBFG plasmid and one pQLink-LptBFGC, whereby each gene (l*ptB*, *lptC*, *lptF* and *lptG*) had been amplified from *P. aeruginosa* PAO1-LAC genomic DNA (47085D-5 ATCC) using PCR and inserted in the pQlink plasmid by ligation-independent pathway. To allow affinity purification, a 6xHis-tag was attached to the C-terminus of LptB. All plasmids were amplified by transforming them into *E. coli* Stellar Competent Cells (Takara) and the DNA sequences were verified by Sanger sequencing.

Initial expression and purification of $LptB_2FG$ was performed adapting previously described protocol [25]. The amplified pQLink-LptBFG plasmids was transformed in *E. coli* C43(DE3) (Lucigen). Colonies were inoculated into 100 ml LB media and grown overnight at 37 °C supplemented with 100 μg/ml ampicillin. One litre of LB in 2 litre shaker flasks was inoculated with 10 ml of overnight culture and grown at 37 °C to a density of 0.6-0.7 at 600 nm at 37 °C, and the expression of the recombinant protein complex was induced by 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 20 °C for 16 hours. Cells were collected by centrifugation at 5,000xg for 10 min at 4 °C. Cell pellets were resuspended in buffer containing 150 mM NaCl, 20 mM Tris-HCl (pH 8.0) and stored at -80 °C.

Resuspended cells were thawed and supplemented with EDTA-free protease inhibitor cocktail (Roche) and 10 μg/ml DNaseI. The cell suspension was passed several times through an M-110 PS microfluidizer (Microfluidics) at 15,000 psi. Insoluble material was pelleted by centrifugation at 20,000xg for 20 min at 4 °C. The supernatant was and then ultracentrifuged at 200,000xg for 1 h to collect membrane fractions. Membranes were resuspended in ice-cold buffer containing 250 mM NaCl, 20 mM Tris-HCl (pH 8.0) and 20% glycerol. Resuspended membranes were used either directly or flash frozen in liquid nitrogen and stored at -80 °C.

3.5.2 $LptB_2FG$ purification

The protein was solubilized from the membrane fraction with 250 mM NaCl, 20 mM Tris-HCl (pH 8.0), 20% glycerol, 2% (w/v) DDM (Anatrace) for 1 hat 4°C. Insoluble material was removed by centrifugation at 20,000xg for 20 min at 4 °C. The supernatant was filtered before loading onto a 5 ml HisTrap-HP column (GE Healthcare, Piscataway, NJ) equilibrated in 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), 20 mM imidazole and 0.1% DDM. After the clarified supernatant was loaded, the column was initially washed with 50 ml (10 CVs) of 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), 20 mM imidazole and 0.1% DDM, and washed again with 50 ml of 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), 80 mM imidazole and 0.05% DDM. The bound protein was eluted with 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), 300 mM imidazole and 0.03% DDM. The protein was concentrated to 2.5 ml an Amicon Ultra-15 concentrator unit (Millipore) with a molecular cut-off of 100 kDa and buffer exchanged to 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10% glycerol and 0.03% DDM using PD-10 desalting column. The sample was further concentrated and loaded onto the Superdex 200 size exclusion chromatography (SEC) column in 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10% glycerol and 0.03% DDM. Protein concentration was measured using a DS-11 FX Spectrophotometer (DeNovix).

Protein purified in such a way yielded highly unresolved nMS spectra, thus we proceeded with optimisation of the purification protocol. We initially increased the solubilisation time from 1h to 16h and kept the remaining purification steps unchanged. This strategy resulted in purified $LptB_2FG$ bound to phospholipids and LPS as shown by the mass spectrum in Fig. 3.2a. To further optimise the nMS spectra we performed a delipidation protocol using 2% DDM + 2% OGNG as membrane solubilisation detergents, incubating the membranes for 16 h. In addition, to maximise the removal of bound lipids, we performed the wash step using 250ml (50 CVs) of a buffer with the following composition: 150 mM NaCl, 20 mM Tris-HCl (pH 8.0), 20 mM imidazole, 1% DDM, 1% OGNG. The protein then underwent the same desalting and SEC steps as described above. This strategy yielded a purified protein whose nMS spectrum is shown in Fig. 3.2b: while phospholipids were removed, LPS was still bound to the protein complex.

3.5.3 Protein expression and purification from ClearColi BL21(DE3)

To fully remove LPS adducts, we then expressed $LptB_2FG$ in ClearColi BL21(DE3) (Lucigen), an *E. coli* strain holding different gene knockouts that stop the biosynthetic pathway of LPS to a precursor, lipid IVA. The amplified pQLink-LptBFG plasmids were transformed in ClearColi BL21(DE3) through electroporation and the plates were incubated for 32-40 h at 37°C. Colonies were inoculated into 100 ml LB media and grown overnight at 37 °C supplemented with 100 µg/ml ampicillin. One litre of LB in 2 litre shaker flasks was inoculated with 40 ml of overnight culture and grown at 37 °C to a density of 0.6-0.7 at 600 nm at 37 °C, and the expression of the recombinant protein complex was induced by 0.5 mM IPTG at 20 °C for 16 hours. The steps for cell lysis and membrane isolation are the same as the ones described in 3.5.1 section.

The protein complex was initially purified following the "classic" protocol consisting in 16h membrane solubilisation in 2% DDM followed by a wash step during IMAC with 0.1% DDM. This procedure yielded the spectrum in Fig. 3.2c, whereby $LptB_2FG$ is bound to multiple lipid IVA molecules. The lipid IVA adducts were successfully removed using the previously described delipidation protocol: 2% DDM + 2% OGNG were used as membrane solubilisation agents and 50 CVs of wash buffer containing 1% DDM + 1% OGNG were used during IMAC. This protocol yielded a fully delipidated protein as shown in figure 3.2d.

$LptB_2FGC$ was expressed in ClearColi BL21(DE3), with the only exception that a pQLink-LptBFGC plasmid was used. Purification was performed following the delipidation protocol.

3.5.4 LptA expression and purification

LptA plasmid was obtained by inserting the gene fragment encoding the predicted mature part of the protein (amino acids 25-175) into a modified pET-28a expression vector (Novagen) between BamHI and XhoI cloning sites, using an In-Fusion cloning kit (Clonetech).

The vector contains a N-terminal 6xHis tag and TEV cleavage site. The plasmid was amplified by transforming them into *E. coli* Stellar Competent Cells (Takara) and the DNA sequences were verified by Sanger sequencing.

The amplified plasmid was transformed in *E. coli* BL21(DE3) (New England Biolabs). Colonies were inoculated into 100 ml LB media and grown overnight at 37 °C supplemented with 50 μg/ml kanamycin. One litre of LB in 2 litre shaker flasks was inoculated with 10 ml of overnight culture and grown at 37 °C to a density of 0.6-0.7 at 600 nm at 37 °C, and the expression of the recombinant protein complex was induced by 0.5 mM IPTG at 20 °C for 16 hours. Cells were collected by centrifugation at 5,000xg for 10 min at 4 °C. Cell pellets were resuspended in buffer containing 300 mM NaCl, 20 mM Tris-HCl (pH 7.5) and stored at -80 °C.

Resuspended cells were thawed and supplemented with EDTA-free protease inhibitor cocktail (Roche) and 10 μg/ml DNaseI. The cell suspension was passed several times through an M-110 PS microfluidizer (Microfluidics) at 15,000 psi. Insoluble material was pelleted by centrifugation at 20,000xg for 20 min at 4 °C. The supernatant was filtered before loading onto a 5 ml HisTrap-HP column equilibrated in 300 mM NaCl, 20 mM Tris-HCl (pH 7.5), and 20 mM imidazole. After the clarified supernatant was loaded, the column was initially washed with 50 ml of 300 mM NaCl, 20 mM Tris-HCl (pH 7.5), and 20 mM imidazole, and washed again with 300 mM NaCl, 20 mM Tris-HCl (pH 7.5), and 80 mM imidazole. The bound protein was eluted with 300 mM NaCl, 20 mM Tris-HCl (pH 7.5), and 300 mM imidazole. Peak fractions were pooled, incubated with TEV protease (produced in-house) and dialysed against 300 mM NaCl, 20 mM Tris-HCl (pH 7.5), and 10% glycerol. To remove the protein fragment containing the His-tag, the mixture was loaded on a 5 ml HisTrap-HP column equilibrated in 300 mM NaCl, 20 mM Tris-HCl (pH 7.5) and the flow through, containing the cleaved protein, was collected. The protein was then concentrated and loaded onto the Superdex 200 SEC column in 300 mM NaCl, 20 mM Tris-HCl (pH 7.5), and 10% glycerol.

3.5.5 Lipids, nucleotides and peptide preparation

Unless stated otherwise, all lipids [Kdo_2-Lipid A (Re-LPS), 1,2-dimyristoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (14:0 PG, DMPG), 1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine (14:0 PE, DMPE) 1',3'-bis[1,2-dimyristoyl-sn-glycero-3-phospho]-glycerol (14:0 cardiolipin, CDL)] used in this study were obtained from Avanti Polar Lipids powders and stock solutions were prepared following a previously established method [34]. ATP-Mg^{2+} and AMP-PNP (Sigma) were diluted in H_2O to a stock concentration of 10 mM. Thanatin (Biomatik) and murepavadin (DC Chemicals) were diluted in H_2O to a stock concentration of 2.5 mM.

3.5.6 Native MS experiments

Purified LptDE was buffer exchanged into MS Buffer (2x CMC of detergent of interest and 200 mM ammonium acetate) using a centrifugal buffer exchange device (Micro Bio-Spin 6, Bio-Rad) as previously described [34]. During the $LptB_2FG$ purification optimisation 0.5% C_8E_4 (Anatrace) was used as nMS detergent, then following a detergent screen (Fig. 3.2d and 3.6) 0.2% G1-OGD was chosen as it yielded more clear spectra. G1-OGD in 200 mM ammonium acetate was also used to analyse $LptB_2FGC$ and $LptB_2FGC$-A. LptA spectra were obtained using 200 mM ammonium acetate without any detergent as it is a soluble protein. The freshly buffer-exchanged proteins were kept on ice, with protein concentration measured as before.

The protein samples were diluted as desired in 200 mM ammonium acetate buffer with detergent as necessary and loaded into a gold-coated capillary Clark borosilicate capillary (Harvard Apparatus) prepared in the laboratory [46]. The experiments were performed using a Q-Exactive UHMR Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Typically, 2 µl of buffer exchanged protein solution was electrosprayed from gold-plated borosilicate capillaries prepared in house. The instrument parameters used for data collection were: capillary voltage 1.2 kV in positive mode and 0.9 kV in negative mode, S-lens RF 100%, quadrupole selection from 1,000 to 20,000 m/z range, HCD collision energy 150-250 V, source

fragmentation 0 V, in-source trapping -150 V to -250 V for analysis in positive mode and 150-250 V in negative mode. These parameters refer to membrane proteins. No energy was applied for LptA. The ion transfer optics in positive mode was set as follows: injection flatapole 5 V, inter-flatapole lens 4 V, bent flatapole 2 V, transfer multipole 0 V; opposite polarities were used for negative mode acquisition. The resolution of the instrument was 17,500 at m/z=200 (transient time of 64 ms), argon pressure in the HCD cell was maintained at approximately at 8×10^{-10} mbar and source temperature was kept at 100°C. The noise level was set at 3 rather than the default value of 4.64. Calibration of the instruments was performed using 10 mg/ml solution of caesium iodide in water. Where required, baseline subtraction was performed to achieve a better-quality mass spectrum. Where required, baseline subtraction was performed to achieve a better-quality mass spectrum. Data were analysed using the Xcalibur 3.0 (Thermo Scientific) and UniDec (www.unidec.chem.ox.ac.uk) [47] software packages. Theoretical and observed molecular masses of all species are described in Table S3.1.

Lipid and nucleotide binding experiments were performed by diluting the lipids in 200 mM ammonium acetate supplemented with 0.02% G1-OGD. Peak intensities were extracted using UniDec and the ratios of the intensity of the ligand bound peak versus the total intensity of all observed species were calculated [48]. The data were plotted using GraphPad Prism 8.0 and a Welch's T-test was used to evaluate statistical significant differences. All nMS experiments were performed three times and standard deviations are calculated from at least 3 datasets.

3.5.7 Structure and sequence alignment

Sequence alignments were performed using the NPS@ClustalW [31, 32] web service and rendered using the ESPript 3.0 web service [33]. Structure alignments were performed using the PyMOL (Schrödinger) *cealign* function.

3.5.8 Nanodisc preparation

To assemble $LptB_2FGC$ into nanodiscs, a mixture containing 10 μM $LptB_2FGC$, 10 μM membrane scaffold protein (MSP; 1E3D1), and 300 μM *E. coli* total extract lipid was incubated for 15 min at room temperature and 0.8 mg/ml prewashed Bio-Beads (Bio-Rad) were added. The resulting mixture was then incubated for 1 h on ice, followed by overnight incubation at 4°C. The Bio-Beads present in the protein-nanodisc solution were filtered out through 0.22 μm nitrocellulose filter tubes. To separate free nanodiscs from $LptB_2FGC$-loaded nanodiscs, the filtered protein-nanodisc solution was purified using a Superdex 200 column (GE Healthcare) equilibrated with 20 mM Tris-HCl (pH 7.5) and 100 mM NaCl. Fractions corresponding to the size of the $LptB_2FG$ complex were collected for cryo-EM.

3.5.9 Electron microscopy sample preparation

The $LptB_2FGC$ nanodisc sample was concentrated to 1 mg/ml and applied to glow-discharged holey carbon grids (Quantifoil R1.2/1.3, 300-mesh Cu), blotted for 2 to 3 s, and then plunge-frozen in liquid ethane using a Vitrobot (Thermo Fisher). The grids were transferred into cartridges and the images were recorded at 1- to 3.5- m defocus on a K2 summit direct electron detector (Gatan) with superresolution mode at nominal 130,000 magnification, corresponding to a sampling interval of 1.06 Å/pixel (superresolution, 0.53 Å/pixel). Each micrograph was exposed for 10 s with 4.48 e /s/physical pixel dose rate (total specimen dose, 40 e /Å2), and 40 frames were captured per specimen area using Latitude.

3.5.10 Data collection and processing

The image stack in super-resolution model was aligned using cryoSPARC [49]. The contrast transfer function (CTF) parameters of the micrographs were determined using Gctf [50]. After manual inspection and sorting to discard poor images, 1000 particles were manually picked to generate templates for automatic picking. Initially, 3,716,502 particles were selected after autopicking in cryoSPARC [49]. Several iterative rounds of two-dimensional (2D) classifications

were carried to remove false picks and classes with unclear features, ice contamination, or carbon. The resulting 654,084 particles were used to generate a reference free ab initio three-dimensional (3D) reconstruction. Two rounds of heterogeneous refinement were used, where two sets of particles were chosen for further processing with local motion correction using cryoSPARC [49] and local CTF reestimation by Gctf [50]. In particular, 229,810 particles corresponded to $LptB_2FG$ and 95,779 particles corresponded to $LptB_2FGC$. The nonuniform refinement using cryoSPARC resulted in 3.11 Å ($LptB_2FG$) and 3.17 Å ($LptB_2FGC$) global resolution maps based on the gold standard Fourier shell correlation (FSC).

3.5.11 Model building and refinement

Model building of $LptB_2FG$ and $LptB_2FGC$ were based on their 3.11 Å and 3.17 Å cryo-EM maps, respectively. The homology modelling structures generated by the FFAS server based on the atomic coordinates of previously published *pa*$LptB_2FG$ (PDB ID: 5X5Y) in the case of *pa*$LptB_2FG$ and *ecl*$LptB_2FGC$ (PDB ID: 6MIT) in the case of *pa*$LptB_2FGC$ were fit into the density map using Chimera [51]. The subsequent model rebuilding was performed using Coot [52]. Structure refinements were performed using the phenix.real_space_refine program [53] from the PHENIX suite [54]. The final atomic models were evaluated using MolProbity [55]. The statistics associated with data collection, 3D reconstruction, and model refinement are shown in Table S3.2.

3.6 References

1. Whitfield, C. and M.S. Trent, *Biosynthesis and export of bacterial lipopolysaccharides*. Annu. Rev. Biochem., 2014. **83**: p. 99-128.
2. Nikaido, H., *Molecular basis of bacterial outer membrane permeability revisited*. Microbiol. Mol. Biol. Rev., 2003. **67**(4): p. 593-656.
3. Ruiz, N., D. Kahne, and T.J. Silhavy, *Transport of lipopolysaccharide across the cell envelope: the long road of discovery*. Nat. Rev. Microbiol., 2009. **7**(9): p. 677-83.
4. Okuda, S., et al., *Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model*. Nat. Rev. Microbiol., 2016. **14**(6): p. 337-45.

5. Mi, W., et al., *Structural basis of MsbA-mediated lipopolysaccharide transport*. Nature, 2017. **549**(7671): p. 233-237.
6. Islam, S.T. and J.S. Lam, *Wzx flippase-mediated membrane translocation of sugar polymer precursors in bacteria*. Environ. Microbiol., 2013. **15**(4): p. 1001-15.
7. Sperandeo, P., et al., *Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of Escherichia coli*. J. Bacteriol., 2008. **190**(13): p. 4460-9.
8. Tran, A.X., C. Dong, and C. Whitfield, *Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in Escherichia coli*. J. Biol. Chem., 2010. **285**(43): p. 33529-33539.
9. Narita, S. and H. Tokuda, *Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides*. FEBS Lett., 2009. **583**(13): p. 2160-2164.
10. Tran, A.X., M.S. Trent, and C. Whitfield, *The LptA protein of Escherichia coli is a periplasmic lipid A,binding protein involved in the lipopolysaccharide export pathway*. J Biol Chem, 2008. **283**(29): p. 20342-9.
11. Suits, M.D., et al., *Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis*. J. Mol. Biol., 2008. **380**(3): p. 476-88.
12. Bollati, M., et al., *Crystal structure of LptH, the periplasmic component of the lipopolysaccharide transport machinery from Pseudomonas aeruginosa*. FEBS J, 2015. **282**(10): p. 1980-97.
13. Qiao, S., et al., *Structural basis for lipopolysaccharide insertion in the bacterial outer membrane*. Nature, 2014. **511**(7507): p. 108-111.
14. Dong, H., et al., *Structural basis for outer membrane lipopolysaccharide insertion*. Nature, 2014. **511**(7507): p. 52-56.
15. Botos, I., et al., *Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens*. Structure, 2016. **24**(6): p. 965-976.
16. Locher, K.P., A.T. Lee, and D.C. Rees, *The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism*. Science, 2002. **296**: p. 1091-1098.
17. Hollenstein, K., D.C. Frei, and K.P. Locher, *Structure of an ABC transporter in complex with its binding protein*. Nature, 2007. **446**: p. 213-216.
18. Oldham, M.L. and J. Chen, *Crystal structure of the maltose transporter in a pretranslocation intermediate state*. Science, 2011. **332**(6034): p. 1202-1205.
19. Locher, K.P., *Mechanistic diversity in ATP-binding cassette (ABC) transporters*. Nature Structural & Molecular Biology, 2016. **23**: p. 487-493.

20. Fiorentino, F., et al., *The Different Effects of Substrates and Nucleotides on the Complex Formation of ABC Transporters*. Structure, 2019. **27**(4): p. 651-659.
21. Owens, T.W., et al., *Structural basis of unidirectional export of lipopolysaccharide to the cell surface*. Nature, 2019. **567**(7749): p. 550-553.
22. Li, Y., B.J. Orlando, and M. Liao, *Structural basis of lipopolysaccharide extraction by the LptB2FGC complex*. Nature, 2019. **567**(7749): p. 486-490.
23. Tang, X., et al., *Cryo-EM structures of lipopolysaccharide transporter LptB2FGC in lipopolysaccharide or AMP-PNP-bound states reveal its transport mechanism*. Nat Commun, 2019. **10**(1): p. 4175.
24. Villa, R., et al., *The Escherichia coli Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains*. J. Bacteriol., 2013. **195**(5): p. 1100-8.
25. Luo, Q., et al., *Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG*. Nat. Struct. Mol. Biol., 2017. **24**(5): p. 469-474.
26. Willyard, C., *The drug-resistant bacteria that pose the greatest health threats*. Nature, 2017. **543**(7643): p. 15.
27. TH, S., et al., *Transgenic Cystic Fibrosis Mice Exhibit Reduced Early Clearance of Pseudomonas Aeruginosa From the Respiratory Tract*. J. Immunol., 2001. **166**(12): p. 7410-7418.
28. Lau, P.C.Y., et al., *Differential Lipopolysaccharide Core Capping Leads to Quantitative and Correlated Modifications of Mechanical and Structural Properties in Pseudomonas aeruginosa Biofilms*. J. Bacteriol., 2009. **191**(21): p. 6618-6631.
29. DK, A., et al., *Presence or Absence of Lipopolysaccharide O Antigens Affects Type III Secretion by Pseudomonas Aeruginosa*. J. Bacteriol., 2007. **189**(6): p. 2203-2209.
30. Dong, H., et al., *Structural and functional insights into the lipopolysaccharide ABC transporter LptB2FG*. Nat. Commun., 2017. **8**(1): p. 222.
31. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic Acids Res., 1994. **22**(22): p. 4673-4680.
32. Combet, C., et al., *NPS@: network protein sequence analysis*. Trends Biochem. Sci., 2000. **25**(3): p. 147-150.
33. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server*. Nucleic Acids Res., 2014. **42**(Web Server issue): p. W320-4.
34. Laganowsky, A., et al., *Mass spectrometry of intact membrane protein complexes*. Nat. Protoc., 2013. **8**(4): p. 639-651.

35. Reading, E., et al., *The Effect of Detergent, Temperature, and Lipid on the Oligomeric State of MscL Constructs: Insights From Mass Spectrometry*. Chem. Biol., 2015. **22**(5): p. 593-603.
36. Gupta, K., et al., *Identifying Key Membrane Protein Lipid Interactions Using Mass Spectrometry*. Nat. Protoc., 2018. **13**(5): p. 1106-1120.
37. Urner, L., et al., *Modular Detergents Tailor the Purification and Structural Analysis of Membrane Proteins Including G-protein Coupled Receptors*. Nat. Commun., 2020. **11**(564).
38. Chng, S.S., et al., *Characterization of the two-protein complex in Escherichia coli responsible for lipopolysaccharide assembly at the outer membrane*. Proc Natl Acad Sci U S A, 2010. **107**(12): p. 5363-8.
39. Xie, R., R.J. Taylor, and D. Kahne, *Outer Membrane Translocon Communicates with Inner Membrane ATPase To Stop Lipopolysaccharide Transport*. J Am Chem Soc, 2018. **140**(40): p. 12691-12694.
40. Andolina, G., et al., *A Peptidomimetic Antibiotic Interacts with the Periplasmic Domain of LptD from Pseudomonas aeruginosa*. ACS Chem Biol, 2018. **13**(3): p. 666-675.
41. Vetterli, S.U., et al., *Thanatin targets the intermembrane protein complex required for lipopolysaccharide transport in Escherichia coli*. Sci Adv, 2018. **4**(11): p. eaau2634.
42. Fehlbaum, P., et al., *Structure-activity analysis of thanatin, a 21-residue inducible insect defense peptide with sequence homology to frog skin antimicrobial peptides*. Proc Natl Acad Sci USA, 1996. **93**(3): p. 1221-1225.
43. Bolla, J.R., et al., *Direct observation of the influence of cardiolipin and antibiotics on lipid II binding to MurJ*. Nat. Chem., 2018. **10**(3): p. 363-371.
44. Musatov, A. and E. Sedlák, *Role of Cardiolipin in Stability of Integral Membrane Proteins*. Biochimie, 2017. **142**: p. 102-111.
45. Liko, I., et al., *Dimer Interface of Bovine Cytochrome C Oxidase Is Influenced by Local Posttranslational Modifications and Lipid Binding*. Proc. Natl. Acad. Sci. U.S.A., 2016. **113**(29): p. 8320-8325.
46. Hernandez, H. and C.V. Robinson, *Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry*. Nat. Protoc., 2007. **2**(3): p. 715-726.
47. Marty, M.T., et al., *Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles*. Anal. Chem., 2015. **87**(8): p. 4370-4376.
48. Cubrilovic, D., et al., *Determination of protein-ligand binding constants of a cooperatively regulated tetrameric enzyme using electrospray mass spectrometry*. ACS Chem. Biol., 2014. **9**(1): p. 218-226.

49. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination*. Nat. Methods, 2017. **14**(3): p. 290-296.

50. Zhang, K., *Gctf: Real-time CTF Determination and Correction*. J. Struct. Biol., 2016. **193**(1): p. 1-12.

51. Pettersen, E.F., et al., *UCSF Chimera--a Visualization System for Exploratory Research and Analysis*. J. Comput. Chem., 2004. **25**(13): p. 1605-1612.

52. Emsley, P. and K. Cowtan, *Coot: Model-Building Tools for Molecular Graphics*. Acta Crystallogr. D Biol. Crystallogr., 2004. **60**(Pt 12 Pt 1): p. 2126-2132.

53. Afonine, P.V., et al., *Real-space Refinement in PHENIX for cryo-EM and Crystallography*. Acta Crystallogr. D Struct. Biol., 2018. **74**(Pt 6): p. 531-544.

54. Adams, P.D., et al., *PHENIX: Building New Software for Automated Crystallographic Structure Determination*. Acta Crystallogr. D Biol. Crystallogr., 2002. **58**(Pt 11): p. 1948-1954.

55. Chen, V.B., et al., *MolProbity: All-Atom Structure Validation for Macromolecular Crystallography*. Acta Crystallogr. D Biol. Crystallogr., 2010. **66**(Pt 1): p. 12-21.

CHAPTER 4

Concerted dynamics of an LPS translocon in the presence of substrate and an antimicrobial peptide

Adapted from a paper accepted in *Nature Chemical Biology*

Francesco Fiorentino[1], Joshua B. Sauer[1,2], Xingyu Qiu[1], Robin A. Corey[2], C. Keith Cassidy[2], Benjamin Mynors-Wallis[2], Shahid Mehmood[1,#], Jani R. Bolla[1,*], Phillip J. Stansfeld[2,3,*], Carol V. Robinson[1,*]

[1]Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK

[2]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

[3]School of Life Sciences and Department of Chemistry, University of Warwick, Gibbet Hill Campus, Coventry, CV4 7AL, UK

[#]Current address: The Francis Crick Institute, 1 Midland Road, London NW1 1ST, UK

[*]Correspondence: jani.bolla@chem.ox.ac.uk; phillip.stansfeld@warwick.ac.uk; carol.robinson@chem.ox.ac.uk

Acknowledgements

Research in the C.V.R. laboratory is supported by a Medical Research Council program grant (MR/N020413/1), a European Research Council Advanced Grant ENABLE (695511), and a Wellcome Trust Senior Investigator Award (104633/Z/14/Z). Research in P.J.S.'s lab is funded by Wellcome (208361/Z/17/Z), the MRC (MR/S009213/1) and BBSRC (BB/P01948X/1, BB/R002517/1 and BB/S003339/1). This project made use of time on ARCHER and JADE granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim. P.J.S. acknowledges Athena at HPC Midlands+, funded by the EPSRC on grant EP/P020232/1, and the University of Warwick Scientific Computing Research Technology Platform for computational access. F.F. holds a SABS CDT studentship supported by the EPSRC and MRC (EP/L016044/1). J.B.S. is supported by the Oxford interdisciplinary DTP and the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/M011224/1). We also thank Dr. Sthitadhi Roy (University of Oxford) for his help with HDX-MS data analysis and rendering.

Authors contributions

F.F., J.R.B., P.J.S. and C.V.R. designed the research. F.F. expressed and purified the protein samples and performed all nMS measurements. F.F. and J.R.B. analysed nMS data. F. F. and X.Q. collected HDX-MS data. F.F. analysed HDX-MS data with the help of X.Q., J.R.B., and S.M. J.B.S. performed the MD simulations with the assistance of P.J.S, who modelled the initial substrate-bound LptDE states with B.M. J.B.S. analysed MD simulations data with the help of R.A.C., C.K.C., and P.J.S. F.F., J.B.S., R.A.C., C.K.C., J.R.B., P.J.S., and C.V.R. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests

C.V.R. is a co-founder and consultant at OMass Therapeutics.

Abstract

The transport of lipopolysaccharide (LPS) to the bacterial outer membrane (OM) is a crucial step in the biogenesis of the microbial surface defences. Although many structures and features of the translocation mechanism have been elucidated, molecular details of the LPS insertion process via the LptDE OM complex remain elusive. Here we integrate native mass spectrometry (nMS) with hydrogen/deuterium exchange mass spectrometry (HDX-MS) and molecular dynamics simulations (MD) to investigate the influence of substrate and antibiotic inhibitor binding on the conformational dynamics of LptDE. Our data reveal that LPS induces opening of the LptD periplasmic domain, coupled with conformational changes on key β-strands adjacent to the putative lateral exit gate. Conversely, an antimicrobial peptide, thanatin, stabilises the periplasmic domain of LptD, thus blocking these conformational fluctuations and preventing LPS transport. Our results illustrate that LPS insertion into the OM relies on concerted opening movements of both β-barrel and β-taco domains, and suggests a means for developing antimicrobial therapeutics that target this essential process in Gram-negative ESKAPE pathogens.

4.1 Introduction

The outer membrane (OM) of Gram-negative bacterial cells is an asymmetric lipid bilayer with lipopolysaccharide (LPS) molecules on the outer leaflet and phospholipids in the inner leaflet [1]. LPS is extracted from the inner membrane (IM) and inserted in the OM by seven essential LPS transport (Lpt) proteins LptA-G (Fig. 4.1a) [2, 3]. Three Lpt components, $LptB_2FG$, form a type VI ABC mechanotransducer which extracts LPS from the outer leaflet of the IM [4]. LptC is a single-pass membrane protein which inserts its helix between the transmembrane domains of LptF and LptG and regulates the transport of LPS from $LptB_2FG$ to LptA [5-7], a soluble oligomeric protein that mediates the passage of LPS across the periplasmic region [8]. Finally, the heterodimer formed by the integral β-barrel OM protein LptD and the lipid-anchored protein LptE is responsible for transporting LPS from the periplasm to the outer leaflet of the OM (Fig. 4.1b) [9, 10].

LptD possesses a 26-stranded β-barrel with a putative LPS lateral exit gate formed between strands β1 and β26. The N-terminal periplasmic region of LptD consists of a β-taco domain, a fold shared with LptA, LptC, LptF and LptG. Several crystal structures of the LptDE complex show that the lipoprotein LptE forms a plug within LptD, forming a plug-and-barrel complex (Fig. 4.1b) [11, 12]. Further structural, genetic and crosslinking studies have suggested that the lipid A portion of LPS interacts with the hydrophobic β-taco domain of LptD [13-15]. However, the molecular details of the insertion process remain elusive, particularly regarding the conformational changes induced by LPS binding, as well as the role of LptE in the translocation process.

Each LPS molecule is composed of three moieties: lipid A, core oligosaccharide and an O-antigen glycan which itself varies in its composition (Fig. 4.1c). The O-antigen contains up to 200 sugars, with many bearing phosphate groups which interact with bivalent cations, allowing the formation of a tight network between LPS molecules [2]. Consequently, while the lipidic portion of the OM is necessary to avoid the passage of polar molecules, the oligosaccharides of

LPS prevent toxic hydrophobic compounds from entering the cell [16], playing a major role in the resistance of Gram-negative bacteria to antibiotics [17].

Currently, two antibacterial peptides targeting LptD have been described: the peptidomimetic compound murepavadin and the insect-derived cyclic peptide thanatin. Murepavadin was reported to specifically bind *Pseudomonas aeruginosa* LptD [18, 19], and showed antimicrobial activity in a mouse septicaemia infection model [20]. On the other hand, thanatin (Fig. 4.1b, inset) demonstrated broad spectrum antimicrobial activity against *Escherichia coli and Klebsiella pneumoniae* [21] and has been shown to strongly interact with *E. coli* LptA and LptDE [22]. Its putative mechanism of action involves disrupting the protein-protein interaction networks in the periplasmic portion of the Lpt complex, thus impairing LPS transportation across the periplasm. While the key interactions between LptA and thanatin have been described [22], little is known about the effects of thanatin binding to LptDE and the mechanistic details of its influence on LPS translocation. Elucidating the molecular details on LptDE-thanatin interaction will therefore provide key information for the development of novel antimicrobial drugs.

Here, we combine native mass spectrometry (nMS) with hydrogen/deuterium exchange mass spectrometry (HDX-MS) and molecular dynamics simulations (MD) to examine the influence of substrate, lipids and inhibitor binding on the conformational dynamics of the LptDE complex. nMS afforded us an initial platform to investigate ligand binding properties [23-26] of LptDE with various native lipids as well as with the inhibitor thanatin. The conformational dynamics of these events were explored in detail using HDX-MS [27, 28]. MD was then used to aid in the mechanistic and molecular interpretation of nMS and HDX-MS experiments [29-31] using an *in silico* method for comparing predicted LptDE conformational states with the experimental data. Together, the results reveal that the soluble domain of LptDE binds to LPS, opening the lateral exit gate of the barrel in the process. We find that thanatin binds to the LptD at the base of the LptD β-taco, thus preventing the flow of LPS into the OM.

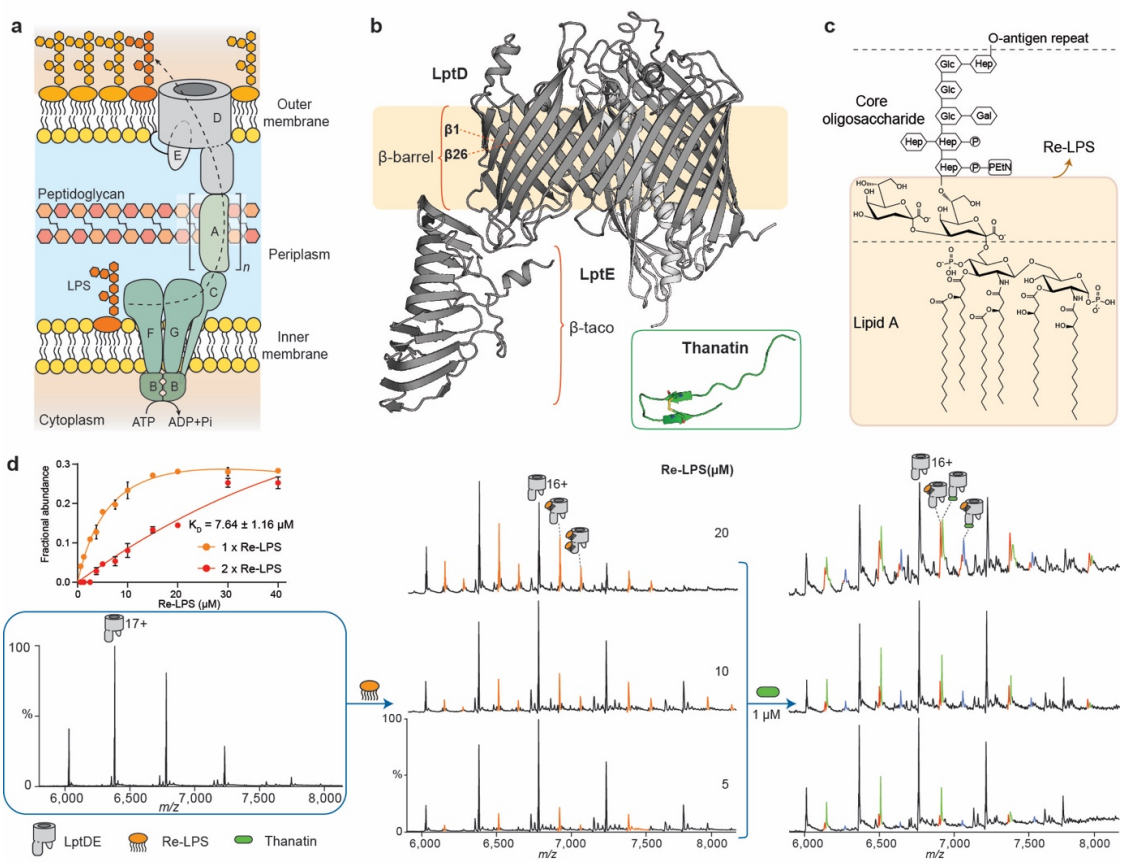


Figure 4.1 Schematic of the LPS transport system and nMS analysis of LptDE, Re-LPS and thanatin. (a) Schematic of the Lpt system: LPS is transported from the inner membrane (IM) to the outer membrane (OM) through a multiprotein complex comprising seven proteins. LPS is extracted from the IM via the ABC transporter $LptB_2FG$ which interacts with the single transmembrane protein LptC which forms a periplasmic bridge with the periplasmic LptA oligomer. The C-terminus of LptA interacts with the soluble N-terminal domain of LptD. LPS is then transported through the periplasmic space within the protein bridge and finally inserted in the outer leaflet of the OM by the heterodimer LptDE. (b) X-ray crystal structure of the OM plug-and-barrel complex LptDE (PDB ID: 5IV9). LptD presents a β-barrel transmembrane region and a β-taco soluble region that contributes to the formation of the periplasmic bridge. LptE is a lipoprotein inserted into the larger lobe formed of the LptD β-barrel. Inset: 3D-structure of thanatin (PDB ID 5XO4). (c) Molecular structure of LPS: lipid A portion, core oligosaccharide and the O-antigen repeat, consisting of a high MW polysaccharide with multiple branching. Re-LPS is a substructure of LPS consisting of lipid A bound to two Kdo moieties. (d) nMS analysis

of LptDE ligand binding. LptDE (5μM, left) in 0.5% C_8E_4 was incubated with increasing concentrations of Re-LPS (centre, orange peaks). The relative fractional intensity of lipid binding peaks over total peak intensity versus Re-LPS concentration, yielded a curve allowing for the calculation of the K_D for the first binding event. Each data point and standard deviation are calculated from the average of five charge states in three independent experiments. Error bars represent standard deviations (n = 3). In a further experiment, Re-LPS-bound LptDE was supplemented with 1 μM thanatin (right). New charge state series corresponding to LptDE bound to thanatin (green) and both Re-LPS and thanatin (purple) are observed. Detailed information on theoretical and experimental molecular masses is provided in Table S4.1.

4.2 Results

4.2.1 LPS and thanatin do not compete with each other for the same binding site on LptDE

We selected LptDE from the pathogenic *K. pneumoniae* as our model system. Following expression and purification of the protein complex (see methods section) using n-dodecyl-β-D-maltopyranoside (DDM) as detergent, we performed a detergent screen to optimise conditions for nMS analysis [32]. Tetraethylene glycol monooctyl ether (C_8E_4) yielded well-resolved mass spectra releasing the intact LptDE heterodimer into the gas phase (Fig. 4.1d, left and Table S4.1).

We then explored the substrate binding properties of the complex by incubating *K. pneumoniae* full length LPS mixture ("smooth" LPS) with LptDE. Mass spectra showed the existence of a heterogeneous array of adducts with a molecular mass ranging from ~4 to ~5 kDa indicative of LPS binding (Fig. S4.1 and Table S4.1). Therefore, given the heterogeneity of LPS, further nMS analysis at higher LPS concentrations would be technically challenging in terms of resolution particular when adding the inhibitor. To overcome this problem, we performed nMS experiments with Re-LPS, an LPS substructure consisting of the lipid A moiety and two ketodeoxyoctonic acids (Kdo) residues belonging to the core oligosaccharide (Fig. 4.1c). This LPS analogue has been found to have the same endotoxin activity as LPS [33] and has been used to investigate the functionality of the Lpt system in previous studies [9, 34]. We added increasing

concentrations of Re-LPS (5 μM, 10 μM and 20 μM) to solutions of LptDE (5 μM) and, using optimised conditions, we recoded mass spectra. We observed additional adduct peaks (orange) that increased in a concentration-dependent manner, as well as a second Re-LPS binding event at higher concentrations (Figure 4.1d, centre).

To assess the specificity of Re-LPS - LptDE interactions we used an MS-based method that outcompetes weaker annular lipid binding with detergents [35]. First, we analysed LptDE binding to cardiolipin (CDL) and phosphatidylglycerol (POPG) by adding a 2-fold excess of each lipid to LptDE. Using nMS we then compared the extent of binding with that observed with Re-LPS in 0.5% C_8E_4. We detected up to three lipid binding events for either CDL or POPG. Increasing concentrations of the detergent n-nonyl-β-D-glucopyranoside (NG), however displaced the second and third binding events for POPG and for CDL. By contrast the second binding event of Re-LPS to LptDE was only slightly reduced in 1% NG (Fig. S4.2). These results suggest that Re-LPS binds to LptDE in a specific manner, whereas CDL and POPG bind to LptDE in exposed regions that are readily displaced by detergents. To quantify the binding affinity of LptDE for Re-LPS, we recorded mass spectra of LptDE incubated with concentrations of Re-LPS from 0 to 40 μM. A dissociation constant (K_D) of 7.64 ± 1.16 μM was obtained for the first binding event (Fig. 4.1d).

We next examined the interaction between thanatin and LptDE. We observed binding even at sub-micromolar concentrations (Fig. S4.3a), in line with the reported binding affinity with *Ec*LptD [22]. To determine whether thanatin competes with Re-LPS we performed competition experiments in the presence of thanatin (1 μM) and increasing concentrations Re-LPS (5 to 20 μM). Interestingly, additional adduct peaks observed in the mass spectra showed simultaneous binding of Re-LPS and thanatin to LptDE (Fig. 4.1d and Table S4.1). In addition, the total amount of Re-LPS bound to LptDE was only slightly reduced in the presence of thanatin (Fig. S4.3b). These results suggest non-competitive binding of the thanatin peptide and Re-LPS substrate to LptDE.

4.2.2 LPS binding to LptDE enhances β-taco opening and decreases β-barrel dynamics

To understand the conformational dynamics of LptDE during LPS translocation, we used HDX-MS to probe structural changes as a function of time upon binding of LPS, Re-LPS, and thanatin. The peptide coverage for LptDE in DDM (68 peptides, 70.5% coverage of the LptD sequence and 17 peptides, 87.4% coverage of LptE, Fig. S4.4), enabled a relatively complete analysis of the conformational changes of LptDE. We performed HDX-MS experiments at 5 different time points (0.167 min to 420 min) to sample conformational changes happening at different time scales [36] (Fig. S4.5). We calculated the ΔHDX for all peptides at each time point and computed statistical significance as described (methods section).

First, we assessed the effects of binding LPS. We incubated LptDE with 15-fold excess of LPS. The LPS-bound state of LptD revealed a decrease in HDX for peptides covering β-strands β1-β2 (peptides 195-211, 195-218, 207-218) and β3-β4 (peptides 232-247, 232-251, Fig. 4.2a and Fig. S4.6a,b). These β-strands are adjacent to the putative LPS lateral exit gate, located at the edge between β1 and β26 [11-13, 37]. Conversely, we observed an increase in HDX in the β-taco domain (Fig. 4.2a), which has been proposed to accommodate LPS during the transport from the periplasmic component of the Lpt system to the OM [13-15].

Interestingly, we found that peptides covering residues 46-204 displayed bimodal isotopic distributions, characteristic of EX1 or mixed EX1/EX2 (EXX) kinetics (Fig. 4.2b, Fig. S4.7a-j). This would suggest that these peptides undergo concerted opening-closing motions at much lower rates compared to the exchange reaction [28, 38, 39]. Most of these peptides were amenable to quantitative analysis enabling us to extract kinetic constants, such as the rate of conformational opening (k_{op}) and the half-life of the closed state ($t_{1/2}$). The fits to the curves indicated that the β-taco k_{op} was increased in the presence of LPS (Fig. 4.2c, Fig. S4.8 and Table 4.1). The k_{op} values for the different peptides in the β-taco were within the same order of magnitude and neighbouring or overlapping peptides presented very similar values (Table 4.1 and Fig. S4.8). These results imply that the conformational events may be concerted, and that the presence of the substrate induces opening of the β-taco domain.

Notably, at the 420 min incubation time some peptides covering the β-barrel (375-386, 404-423, 404-425, 523-544, 657-669, 660-669, 676-689) showed decreased D uptake in the presence of LPS, a feature not displayed at shorter time points (Fig. S4.6a). This feature may be a consequence of structural changes following incubation of the protein at 20 °C for 420 min. Therefore, to assess the contribution from the formation of non-specific aggregates we ran analytical SEC (Fig. S4.9). This experiment indicated that LptDE is stable in the buffer used for protein purification (and HDX-MS experiments) over the incubation times used during HDX-MS; indeed, SEC profiles show no significant oligomerisation or aggregation. This indicates that the observed protection is not a consequence of structural destabilisation leading to protein aggregation, and confirms that the observed EX1 kinetics in the β-taco peptides is a specific feature of the protein complex. In order to investigate further, we calculated the ΔΔHDX for the β-barrel protected peptides between the 420 min time point and the shortest (0.167 min) time point. ΔΔHDX for all these peptides ranged between 0.45 and 0.94 Da, higher than the 0.28 Da cut-off (98% CI) at the 420 min incubation time. The same calculation applied to the peptides in the presence of thanatin indicated a ΔΔHDX ranging between 0.20 and 0.33 Da (~ 50% peptides below the 0.24 Da cut-off); in the presence of both LPS and thanatin ΔΔHDX ranged between 0.61 and 1.0 Da, higher than the 0.25 Da cut-off at 420 min. Since the highest protection was observed in the presence of LPS or both LPS and thanatin, rather than thanatin alone, this is likely to be a consequence of non-specific interactions between LPS and LptD. This phenomenon may be the result of prolonged non-specific contacts of LptD with LPS aggregates [40] over the incubation time; in addition, we cannot exclude that LptD structural changes have facilitated non-specific interactions.

To see if this opening of the β-taco and protection of the lateral gate could also be induced by Re-LPS we repeated our experiments using a 15-fold excess of the ligand. We observed analogous changes in dynamics in the β-taco and exit gate (Figs. S4.10 and S4.11a-c). As a control experiment we measured the HDX for the same peptides but this time in the presence of POPG, a lipid shown to bind in our nMS experiments above (Fig. S4.12). However, in this

case there was no difference in D uptake (+/- lipid), suggesting that the conformational changes observed are unique to LPS binding and cannot be recapitulated by lipid binding.

Considering LptE in the presence of LPS or Re-LPS, both the N- and C- terminal peptides were deprotected (Figs. S4.13a-b; S4.14a,d; S4.15a,d). From the crystal structure of *Kp*LptDE (PDB ID: 5IV9) we can see that peptide 15-35 forms a loop. The structure lacks the last 26 residues (170-196) which has not been observed in any X-ray crystal structure of LptDE from different species implying that it is a highly dynamic region. In line with this, the HDX of the peptides covering this region exchange rapidly, with more than 50% D uptake at the earliest time point.

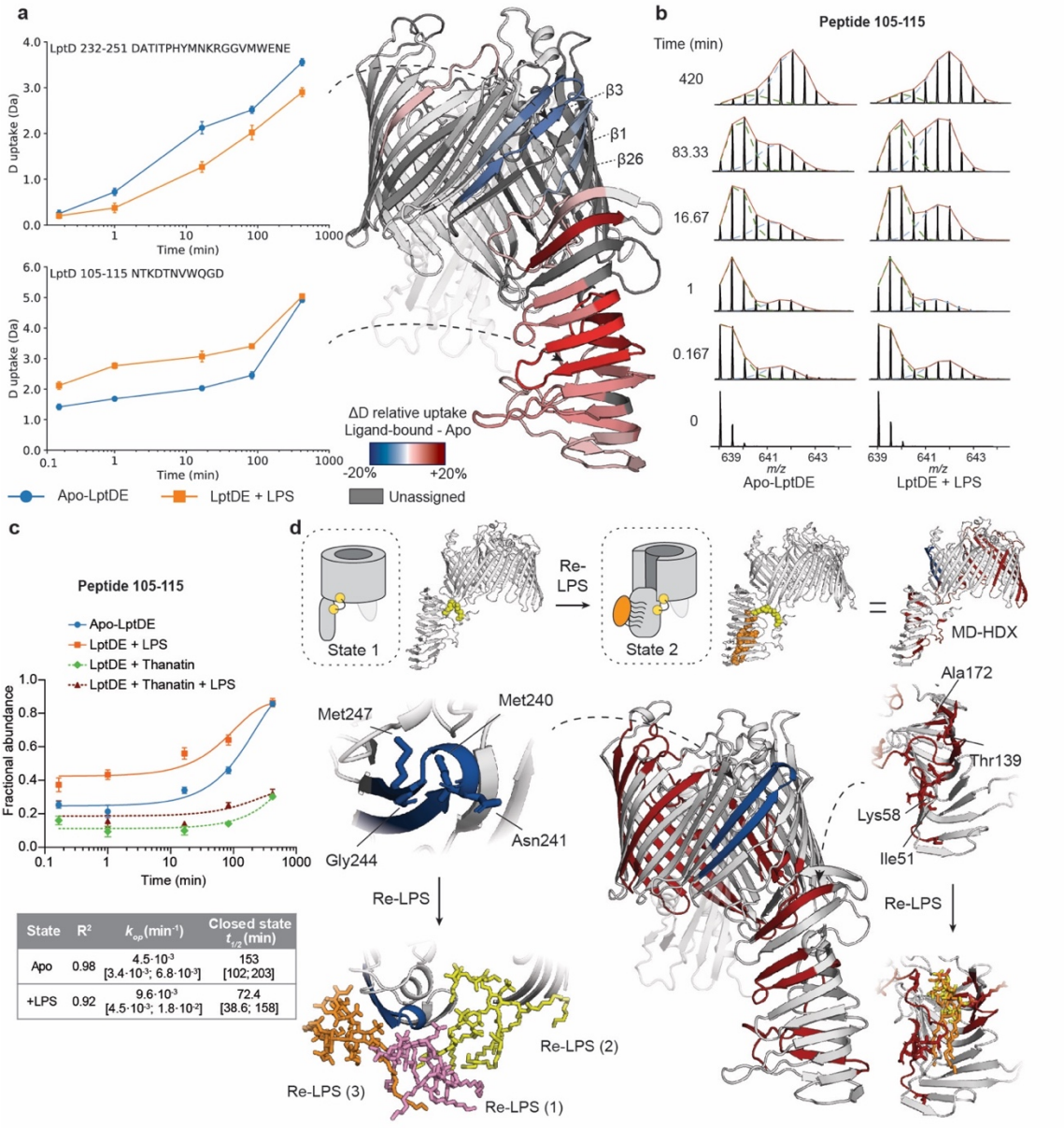


Figure 4.2 Conformational dynamics of LPS-bound LptDE. Experimental: (a) Left: the deuterium uptake of representative peptides (105-115, β-taco and 232-247, β-barrel) plotted as a function of labelling time (0.167-420 min) for apo-LptDE (blue) and LPS-bound LptDE (orange). Error bars indicate standard deviations. ($n_{biological}$ = 2; $n_{technical}$ = 3). Right: difference in relative deuterium uptake (scaled for the number of residues of each peptide) at the 16.67 min labelling time mapped on the crystal structure of LptD. Only peptides showing significant differences are coloured. Red and blue indicate increased and decreased deuterium uptake, respectively. Detailed information of HDX-MS data is provided in Table S4.2. Experimental: (b) Representative mass spectra for peptide 105-115 are shown for apo-LptDE and LPS-bound states. Two binomial isotopic envelopes produced the best fit for the spectra yielding low- (green) and high-mass (light blue) populations. The sums of the two binomial distributions are shown in red. Experimental: (c) Extracted relative abundances of the high- mass population of peptide 105-115 plotted as a function of labelling time and fitted to a monoexponential function to obtain the rate of translation from the low-mass population to the high-mass population (k_{op}). The rate was increased upon binding of LPS (orange) compared to the apo protein (blue). Computational: (d) Top: schematic comparing the solvent uptake for the apo-closed and lipid-bound open states and the resultant differential map. Bottom: highlighted portions of the differential solvent uptake map, where significant residues are labelled and shown as sticks for both the lateral exit gate peptide (232-251) and peptides on the β-taco (46-58, 130-141, 171-179). Re-LPS bound snapshots are illustrated for the β-taco and β-barrel (1-3) respectively.

4.2.3 Molecular dynamics simulations suggest an open LptDE conformation on substrate binding

To lend further structural insights to our HDX-MS results and elucidate the mechanisms by which LPS specifically affects LptDE dynamics, we performed a series of all-atom MD simulations of the LptDE complex in its open and closed states. To this end, we repurposed the alchembed methodology [41] to gradually grow the Re-LPS lipid inside the β-taco opening the

domain in the process. In order to model LPS insertion in the OM through the β1-β26 lateral exit gate, we templated *Kp*LptDE onto a structure of LptDE from *Salmonella typhimurium*, which the β-barrel has been previously modelled open [12]. In doing so, it became apparent that the process of opening may be concerted between the β-strands of the lateral exit gate and the β-taco. Our analysis identified that the *Kp*LptDE disulphide bonds, Cys7-Cys696 and Cys149-Cys697, where the latter is almost absolutely conserved, couple the C-terminal half of the β-barrel to the β-taco. We hypothesise, therefore, that binding of LPS and derivatives thereof (such as Re-LPS) to the β-taco not only drives the opening of this domain but also primes the opening of the lateral gate of the LptDE β-barrel, upon LPS translocation.

Altogether, five repeats of the apo-closed state and the lipid-bound open state were simulated for 300 ns. In these simulations, global solvent contacts were assessed to be converged after 150 ns (Fig. S4.16a). During the course of the open state simulations, the barrel tended towards closure in 3 replicates, though in 2 replicates we noticed that the gate remained in a stable open state at a distance of ~2.25 nm between four adjacent Cα atoms on β1 and β26 (Fig. S4.17a). Thus PLUMED [42] upper and lower walls were used to sample a consistent open conformation, allowing us to test how wide the lateral exit gate should be in order to observe protection seen in the experimental result (Figs. S4.16b-S4.17b).

We next established a pipeline to generate simulated HDX data. Here, we computed the backbone amine solvent contacts of each state, calculated the difference between states for each residue and then averaged over the peptide ranges obtained from the experimental sequence coverage (Fig. S4.4), as explained in the methods section. The differential solvent contact plot (Fig. 4.2d) resembled the experimental data in several regards, including reproducing the observed deprotected peptides in the β-taco (peptides 46-58, 130-141, 171-179) and protected peptides at the lateral exit gate (peptide 232-251). Furthermore, we determined the specific residues responsible for a significant change in solvent contacts (Fig. 4.2d and Fig. S4.18). Our model suggests that on binding to Re-LPS, the β-taco opens up and allows solvent access,

particularly for inward facing residues on the sides of the β-taco (Ile51, Lys58, Thr139 and Ala172).

Regarding the lateral exit gate, we found that the two significant residues on the outside of the barrel particularly interact with Re-LPS (Fig. 4.2d). One of these lipids (labelled Re-LPS(2)) is caught within the open barrel (Fig. 4.2d). Later within the simulation, these lipids slowly diffuse away from the lateral exit gate which correlates with the peptide 232-251 not being shielded from solvent and therefore not being protected (Fig. S4.19). We also observed that when the β-barrel opens up, the loop connecting β1 and β2 dips down and shields Met240 and Met247) within the barrel (Fig. S4.20). Thus, although local lipid effects contribute to solvent shielding, a subtle internal conformational change further facilitates protection of the amino acids at the lateral exit gate.

4.2.4 Thanatin binding stabilises the LptD β-taco

We next investigated the effects of thanatin on the dynamics of the transporter. We collected the HDX-MS data in presence of thanatin and compared the deuterium uptakes to the apo state (Figs. S4.5 and S4.6c,d). Regions belonging to the β-taco domain of LptD presented decreased dynamics, indicating stabilisation upon ligand binding (Fig. 4.3a). We also observed mild deprotection on the LptE C-terminus (Figs. S4.13a and S4.14b,d), probably because of the altered dynamics of the neighbouring LptD β-taco.

Peptides covering the central and terminal region of the β-taco showed significant protection at all measured time points. Indeed, incubation with thanatin stabilised the β-taco and decreased the EX1/EXX correlated exchange, indicating a slower conformational opening of this domain (Fig. 4.2b, Fig. S4.7a-j). Additionally, peptides covering the N-terminal region of LptD β-taco (27-42 and 27-45) also displayed decreased deuterium uptake in the presence of thanatin. However, these segments are to be distinguished from those covering the rest of the periplasmic domain, as they follow the EX2 HDX regime, leading us to the hypothesis that thanatin directly interacts with this region.

To assess the N-terminal thanatin binding mode, we performed sequence and structural alignments [43] between LptD and *Ec*LptA, which has been shown to bind thanatin on the N-terminus [22]. Key residues for LptA - thanatin interaction are retained at the LptD N-terminus (Fig. S4.21a), moreover we observed structural homology between the N-terminal regions of the two proteins (Fig. S4.21b). To validate these findings, we docked the peptide to either end of the β-taco in the closed state and assessed the stability of each docking pose. We observed that thanatin is more unstable when bound to the top (i.e. the membrane-facing end; RMSD = 0.34 ± 0.1 nm) of the β-taco than when bound to the N-terminal region (RMSD = 0.22 ± 0.05 nm) (Fig. S4.22). In addition, we observed that the N-terminal of β-taco itself deviates more from its original position is the crystal structure (distance = 1.13 ± 0.5 nm) when thanatin is docked at the top than when it is bound to the N-terminal (distance = 0.65 ± 0.2 nm) (Fig. S4.22). This suggests that the β-taco would have to shift extensively its conformation to accommodate thanatin at the top, making it less plausible for it to bind there. Additionally, differential solvent contact plot of thanatin bound to the bottom position in the closed state of the protein (Fig. 4.3b) reveals that the residues which cause the peptide to be protected are those that predominantly form H-bonds with thanatin (Val28, Ile30, Ala32, Asp33). Taken together the MD results in relation to the HDX experiments suggest that thanatin is more stable when docked to the N-terminal end of the β-taco compared to the C-terminus and directly interacts with this region causing peptides to be protected.

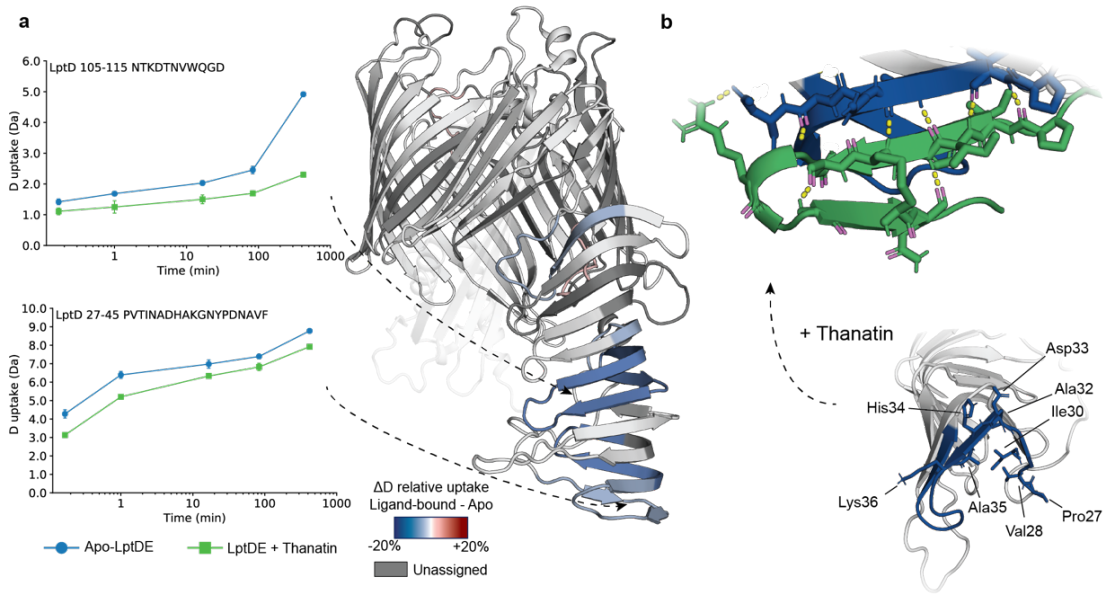


Figure 4.3 Conformational dynamics of thanatin-bound LptDE. Experimental**:** (a) Left: the deuterium uptake of representative peptides (27-45, β-taco N-terminus; 105-115, β-taco central region) plotted as a function of labelling time for apo-LptDE (blue) and thanatin-bound LptDE (green). Error bars indicate standard deviations. ($n_{biological}$ = 2; $n_{technical}$ = 3). Right: difference in relative deuterium uptake (scaled for the number of residues of each peptide) at the 16.67 min labelling time mapped on the crystal structure of LptD. Only peptides showing significant difference are coloured. Red and blue indicate increased and decreased deuterium uptake, respectively. Detailed information of HDX-MS data is provided in Table S4.2. Computational: (b) Top: final snapshot of thanatin bound to the β-taco in the simulation. The H-bond network is shown as yellow lines between residues represented as sticks on the β-taco and thanatin respectively. Bottom: differential solvent contact plot between the apo-closed state and state in which thanatin is docked to the bottom of the β-taco with significant residues shown as sticks.

4.2.5 Thanatin interferes with LPS-induced opening of LptD β-taco

Incubation of LptDE with LPS (or Re-LPS) and subsequent addition of thanatin yielded an HDX profile of the β-taco similar to what we observed in the presence of thanatin only (Fig. 4.4a and Figs. S4.5; S4.6e,f and S4.11d-f), suggesting that the stabilising effect of thanatin is stronger than substrate-induced opening. Notably, we also detected protection on β1-β2 and β3-

β4 segments, similar to what was observed in the presence of LPS only. The decreased conformational dynamics of the LptD β-barrel suggests that substrate may still bind to LptD despite the presence of thanatin, in line with our nMS results (Figure 4.1d) and MD simulations suggesting that thanatin interacts with the N-terminus of the β-taco (Fig. 4.3).

Next, we generated a differential solvent contact plot to establish how thanatin affects solvent uptake in the open Re-LPS bound state (Fig. 4.4b). The plot showed the N-terminus of the β-taco being protected as when LptDE was in its closed state (Fig. 4.3b). In addition, a residue-by-residue analysis revealed that protection is present throughout the length of the β-taco (Fig. 4.4b). In line with these results, when comparing the Re-LPS + thanatin open state with LptD closed state (Fig. 4.4c) we noticed decreased solvent contacts in both β-taco (although Re-LPS still binds in this region) and β-barrel lateral gate, in agreement with the experimental data (Fig. 4.4a).

Assessment of the number of H-bonds in the β-taco (Fig. 4.4d and Fig. S4.23) indicated that Re-LPS binding and resultant β-taco opening causes a decrease in the number of backbone H-bonds pertaining to the β-taco itself, while thanatin docking restores these H-bonds (Fig. 4.4d, left upper plot). In line with this, the number of H-bonds made between the β-taco and solvent (Fig. 4.4d, left lower plot) increases from the closed state to the open (Re-LPS bound) state indicating that opening the β-taco weakens the internal H-bond network affording a greater propensity for solvent interactions. Not only the total number of H-bonds, but also their relative occupancy within the β-taco changes in the different states. Indeed, in the presence of thanatin most of the H-bonds are between β-strands, while H-bonds between loops were prevalent when thanatin was absent (Fig. 4.4d, right). These results indicate that thanatin has a stabilising influence through a rearrangement of the H-bond network, propagating along several strands of the β-taco domain, explaining the global protection of the domain in both the closed (Fig. 4.3a) and open (Fig. 4.4a) states of the protein.

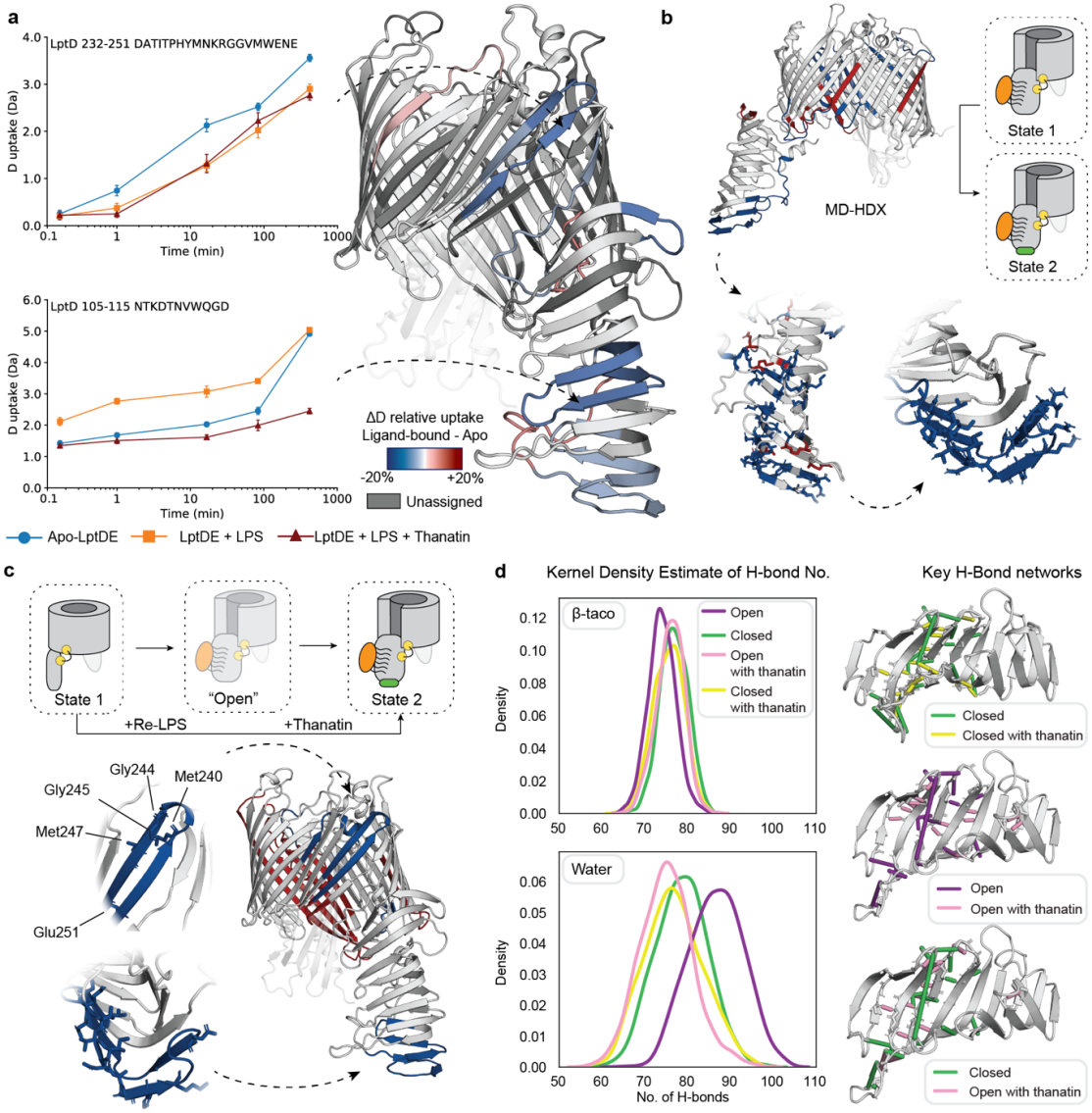


Figure 4.4 Conformational dynamics of LptDE in the presence of both LPS and thanatin.

Experimental: (a) Left: the deuterium uptake of representative peptides (105-115, β-taco and 232-247, β-barrel) plotted as a function of labelling time for apo-LptDE (blue) and LptDE in the presence of both LPS and thanatin (maroon). Error bars indicate standard deviations. ($n_{biological}$ = 2; $n_{technical}$ = 3). Right: difference in relative deuterium uptake (scaled for the number of residues of each peptide) at the 16.67 min labelling time mapped on the crystal structure of LptD. Only peptides showing significant difference are coloured. Red and blue indicate increased and decreased deuterium uptake, respectively. Detailed information of HDX-MS data is provided in Table S4.2. Computational: (b) Top: differential solvent contact map between the open Re-LPS bound state with and without thanatin docked to the bottom of the β-taco illustrated in the cartoon

schematic. Bottom: comparison of the differential solvent map of the β-taco shown residue by residue (left) and averaged over experimentally determined peptides (right) where all significant residues are highlighted in a stick representation. Computational: (c) Top: cartoon schematic illustrating the proposed changes in LptDE conformation after incubation with Re-LPS followed by thanatin; resulting in the comparison between the apo-closed state (1) and the open Re-LPS bound state (2) with thanatin docked. Bottom: HDX differential plot between states 1 and 2, with protected peptides highlighted on β-taco and β-barrel with significant residues labelled and shown as sticks. Computational: (d) Left: Kernel Density Estimate (KDE) distributions of H-bond frequency per frame of simulation for each state of the β-taco, considering the H-bond network on the β-taco itself (upper) and that which it makes with the solvent (lower). Right: H-bonds which are occupied more than 25% in one state compared to the other. Comparisons are made between thanatin bound and non-bound states in the closed (upper), open (middle) conformations as well as between the closed state and open thanatin bound states (lower).

4.3 Discussion

The insertion of LPS in the outer leaflet of the OM is an essential step for Gram-negative bacteria membrane biogenesis. Mechanistic details of insertion process by LptDE and associated conformational dynamics upon substrate binding are largely unknown. Using nMS we showed that Re-LPS specifically binds to LptDE, with a K_D in the low μM range. We further demonstrated that binding of the antimicrobial peptide thanatin does not impair the LptDE-LPS interaction, and we observed a small population of the ternary complex formed by the protein and the two ligands (Figure 4.1d). These results indicate that thanatin does not compete for the same binding site of LPS, but may act with another mechanism, such as allosteric modulation.

To broaden the investigation on the conformational dynamics of LptDE we employed HDX-MS and integrated the results with MD simulations. HDX-MS revealed an extended region in the periplasmic domain where peptides undergo correlated exchange. Our observations can be interpreted as the H-bonds between the β-strands defining the β-taco breaking in a slow concerted

manner (Fig. 4.2a-c). In the presence of LPS or Re-LPS, we observed increased an increased rate of cooperative fluctuations of the peptides under the EX1 kinetic regime. Thus, the presence of substrate promotes the opening of the cavity of the β-taco, consequently breaking of the H-bond network between the β strands. MD simulations showed that Re-LPS binding inside the β-taco yields an open conformation which promotes solvent exposure, particularly with residues that are directed towards the lipid (Fig. 4.2d). In light of these observations, we suggest that the specific binding site detected by nMS (Figure 4.1d) is found in the hydrophobic pocket of the β-taco itself. Such a specific binding site would be largely unaffected by competing detergent, as also confirmed by delipidation experiments on the Re-LPS bound complex (Fig. S4.2).

Furthermore, opening of the β-taco whilst maintaining the cysteine disulphide network on the β-barrel would appear to also open the lateral exit gate in the process. Indeed, both HDX-MS and MD simulations indicate protection on strands β1-β2 and β3-β4, adjacent to the putative LPS lateral exit gate (Fig. 4.2a). The observed decreased dynamics is related to both direct interaction with substrate (Asn241 and Gly245) as well as small conformational fluctuations on the β-barrel itself, providing evidence these rearrangements are necessary to accommodate the substrate. Interestingly, an equivalent degree of lateral opening in an OM protein was recently reported for the BAM complex in response to substrate binding [44].

We observed that thanatin stabilises the periplasmic region of LptD, decreasing both the dynamics and opening rate of the β-taco. According to our data, the thanatin binding site resides on the N-terminal segment of the β-taco, acting both as protein-protein interaction disruptor [22] and non-competitive inhibitor of LPS transportation. The predominating effect of thanatin over LPS on the dynamics of the β-taco observed in the presence of both ligand, along with the protection on β1-β2 and β3-β4 strands, suggest simultaneous binding of both ligands (Fig. 4.4). We can therefore infer that, given the stabilising effect of thanatin, the periplasmic domain is essentially in its closed state while the β-barrel undergoes the conformational changes characteristic of LPS interaction (Fig. 4.4). These findings support the idea that thanatin acts non-competitively to impair LPS transport through stabilisation of LptD periplasmic domain.

Although this binding does not allow substrate sliding through the β-taco, LPS may still interact with LptD.

Our results indicate a situation whereby the complex binds multiple LPS molecules, in a similar fashion to the proposed PEZ model, which suggests that the whole Lpt system is filled with LPS molecules at any given time [2]. The interaction of different protein domains with LPS leads to an increase of the opening rate of the β-taco to accommodate the substrate and decreased solvent contacts in the lateral exit gate region (Fig. 4.5, state ii). The observed β-barrel protection is given by both direct interactions with the substrate and conformational changes necessary to allow the passage of LPS from the lumen of LptD to the outer leaflet of the OM. Thanatin-bound LptD exhibits a clear difference in HDX, even in the presence of substrate, suggesting an additional mode of action relying on the stabilisation of the β-taco domain and consequent impairment of the LPS translocation (Fig. 4.5). Our findings provide a deeper understanding of the LPS insertion process in the OM, highlighting crucial residues and specific conformational fluctuations for domains that are critical for LPS transport.

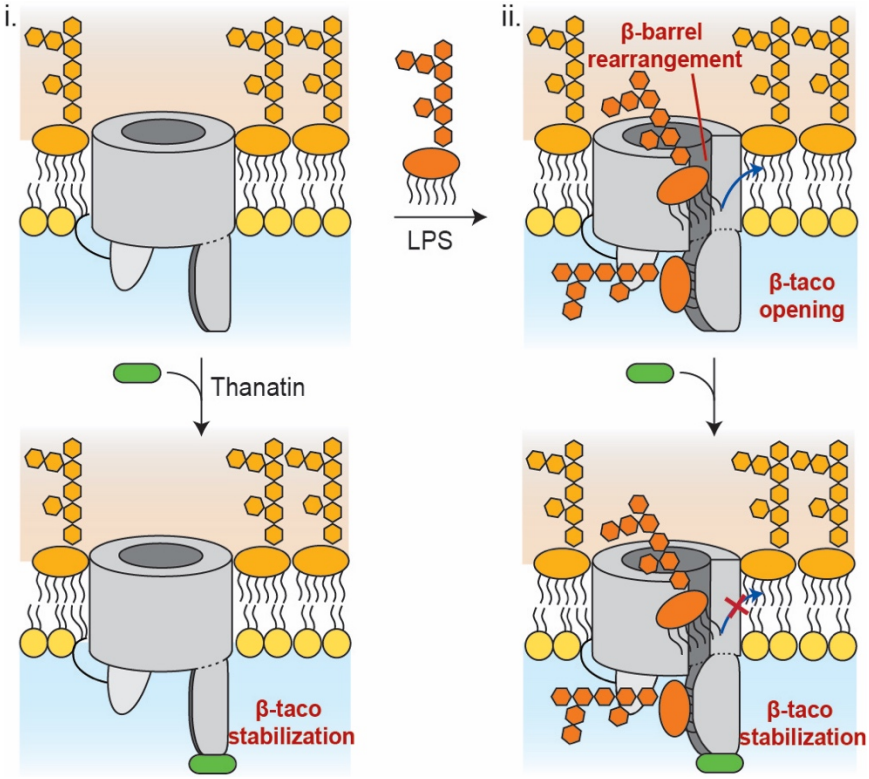


Figure 4.5 Schematic of LptDE-mediated LPS insertion in the OM. Starting from a putative apo form of LptDE (i), the interaction between LPS and LptDE determines β-taco opening to accommodate the substrate and consequent loosening of the β strands between β-26 and β-4. This conformational change creates enough space to accommodate an LPS molecule within the β-barrel (ii), allowing the insertion in the OM through the lateral exit gate. Thanatin may bind LptD at any point of the insertion process, thus obstructing LPS flow through the stabilisation of the β-taco.

Peptide	State	R^2	k_{op} (min^{-1})	Closed state $t_{1/2}$ (min)
66-92	Apo	0.98	$2.6 \cdot 10^{-3}$ [$2.6 \cdot 10^{-3}$; $4.7 \cdot 10^{-3}$]	267 [147; 273]
	+LPS	0.96	$9.6 \cdot 10^{-3}$ [$6.2 \cdot 10^{-3}$; $1.4 \cdot 10^{-2}$]	72.2 [50.0; 111]
93-104	Apo	0.91	$2.9 \cdot 10^{-3}$ [$2.4 \cdot 10^{-3}$; $4.2 \cdot 10^{-3}$]	238 [166; 290]
	+LPS	0.89	$6.2 \cdot 10^{-3}$ [$2.8 \cdot 10^{-3}$; $1.0 \cdot 10^{-2}$]	112.4 [68.4; 251]
105-115	Apo	0.98	$4.5 \cdot 10^{-3}$ [$3.4 \cdot 10^{-3}$; $6.8 \cdot 10^{-3}$]	153 [103; 202]
	+LPS	0.92	$9.6 \cdot 10^{-3}$ [$4.4 \cdot 10^{-3}$; $1.8 \cdot 10^{-2}$]	72.4 [38.6; 158]
105-118	Apo	0.98	$3.1 \cdot 10^{-3}$ [$2.8 \cdot 10^{-3}$; $4.0 \cdot 10^{-3}$]	226 [176; 251]
	+LPS	0.97	$7.4 \cdot 10^{-3}$ [$4.4 \cdot 10^{-3}$; $1.1 \cdot 10^{-2}$]	72.4 [63.7; 158]
116-129	Apo	0.98	$3.6 \cdot 10^{-3}$ [$2.2 \cdot 10^{-3}$; $5.9 \cdot 10^{-3}$]	193 [116; 306]
	+LPS	0.93	$1.8 \cdot 10^{-2}$ [$1.1 \cdot 10^{-2}$; $3.4 \cdot 10^{-2}$]	39.6 [19.9; 67.5]
119-129	Apo	0.98	$3.3 \cdot 10^{-3}$ [$2.6 \cdot 10^{-3}$; $5.8 \cdot 10^{-3}$]	213 [120; 268]
	+LPS	0.96	$1.1 \cdot 10^{-2}$ [$7.5 \cdot 10^{-3}$; $1.7 \cdot 10^{-2}$]	61.1 [41.0; 92.3]
130-141	Apo	0.98	$3.4 \cdot 10^{-3}$ [$2.6 \cdot 10^{-3}$; $5.3 \cdot 10^{-3}$]	205 [130; 263]
	+LPS	0.96	$2.2 \cdot 10^{-2}$ [$1.4 \cdot 10^{-2}$; $3.9 \cdot 10^{-2}$]	31.7 [18.9; 49.2]
171-179	Apo	0.97	$1.4 \cdot 10^{-3}$ [$1.2 \cdot 10^{-3}$; $3.2 \cdot 10^{-3}$]	497 [219; 570]
	+LPS	0.89	$8.0 \cdot 10^{-3}$ [$3.4 \cdot 10^{-3}$; $1.4 \cdot 10^{-2}$]	86.8 [48.0; 202]

Table 4.1 Kinetic parameters for LptD segments showing EX1 or EXX kinetics.

4.4 Supplementary information

4.4.1 Supplementary Figures

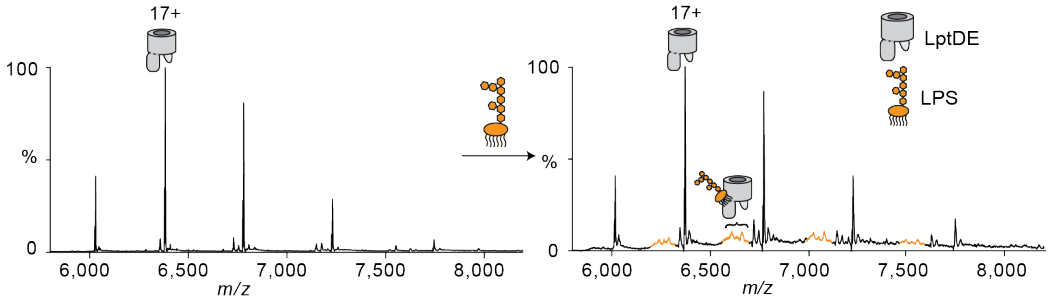


Figure S4.1 nMS analysis of LptDE-LPS interaction. LptDE (5 µM, left) in 0.5% C_8E_4 was mixed with 10 µM LPS. A series of different adducts of molecular mass between 4 and 5 kDa was observed upon LPS addition (orange).

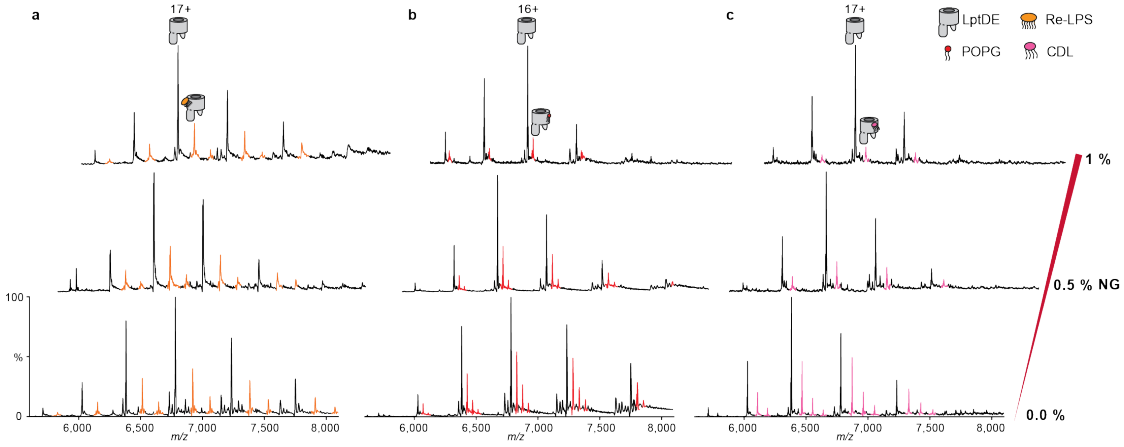


Figure S4.2 Detergent competition experiments to assess LptDE lipid binding. LptDE (5 µM, left) in 0.5% C_8E_4 was mixed with 10 µM Re-LPS (a, orange adducts), POPG (b, red adducts) or CDL (c, magenta adducts). The protein-lipid mixture was then supplemented with increasing concentrations of n-nonyl-β-D-glucopyranoside (NG). Detergent addition decreases POPG and CDL binding, but has no pronounced effect on Re-LPS.

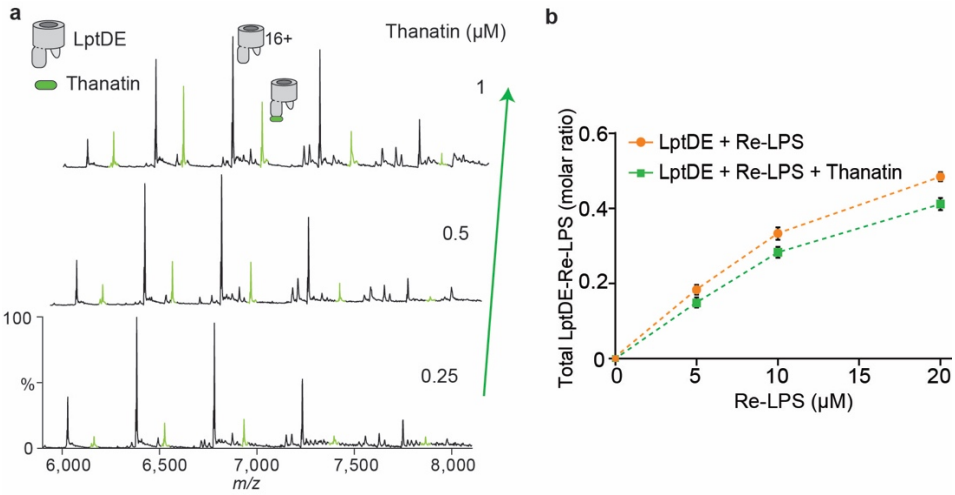


Figure S4.3 Thanatin binding efficiency and effect on Re-LPS binding. (a) Mass spectra recorded for solutions of LptDE (5 μM) with increasing concentrations of thanatin (green adducts). (b) Quantification of the total amount of Re-LPS bound to LptDE in the absence or in the presence of thanatin (related to Fig. 4.1d). Error bars represent standard deviation (n = 3).

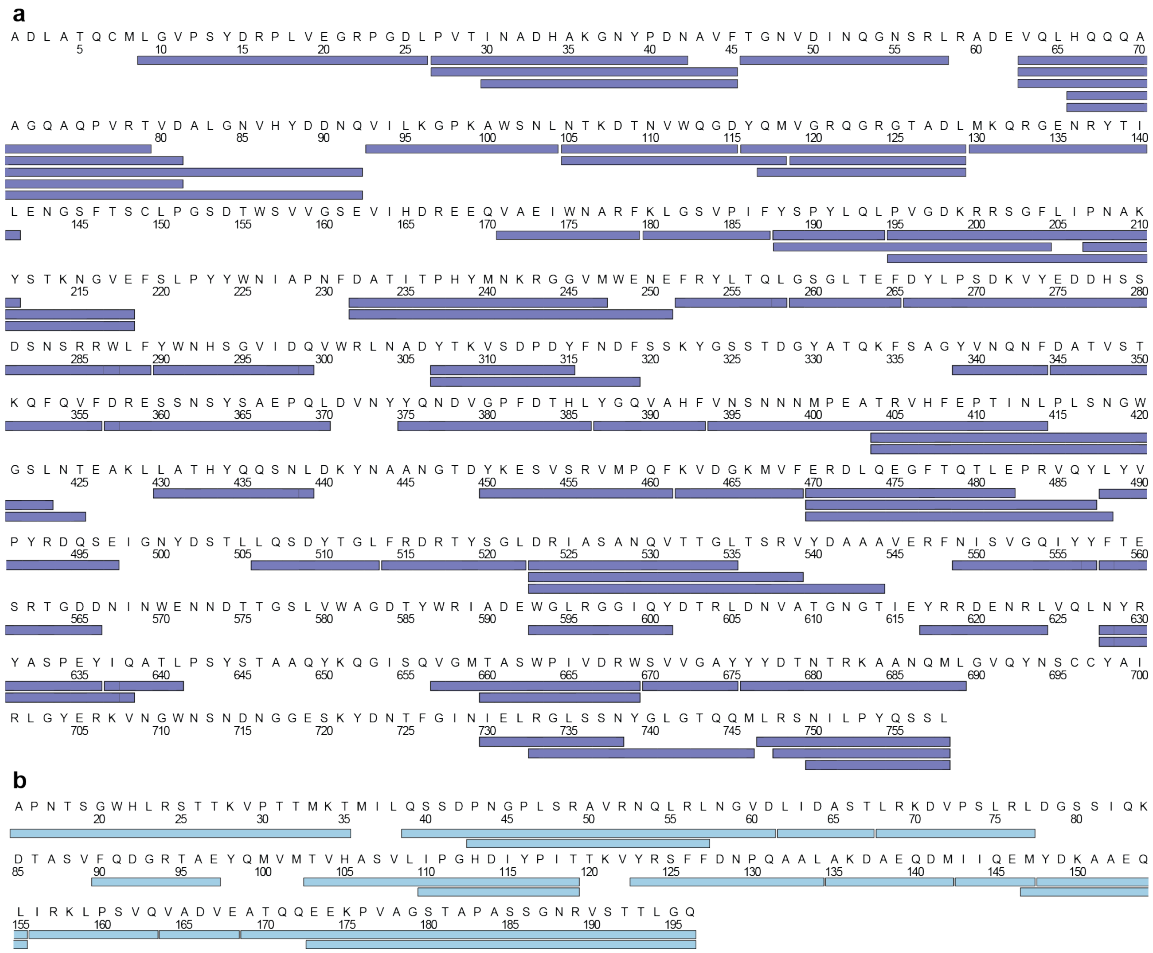
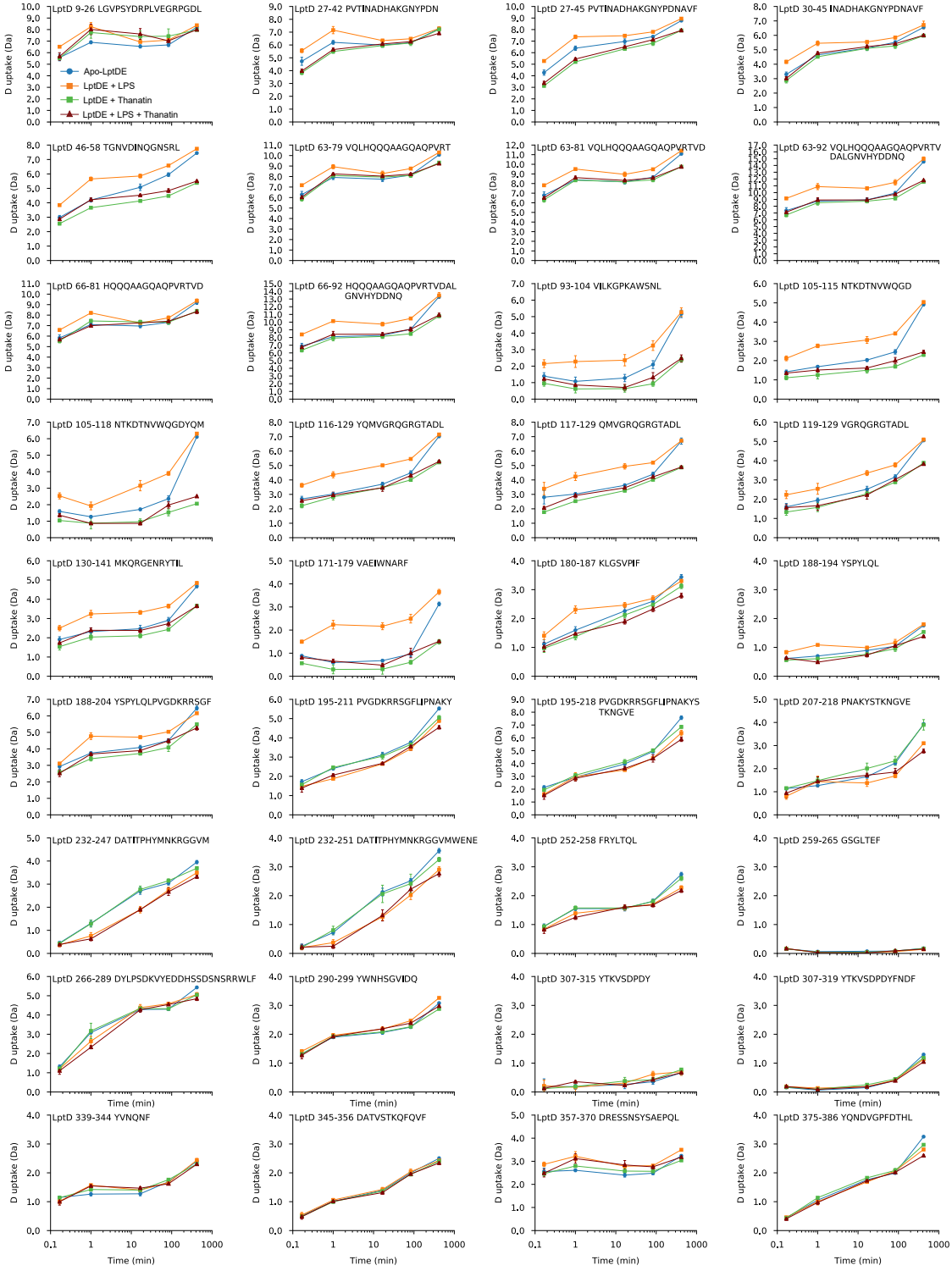


Figure S4.4 Sequence coverage of LptDE. 68 peptides covering 70.5% of LptD (b) sequence and 17 peptides covering 87.4% of LptE (b) were identified following digestion with immobilised pepsin.


LptD 9-26 LGVPSYDRPLVEGRPGDL
Apo-LptDE
LptDE + LPS
LptDE + Thanatin
LptDE + LPS + Thanatin
LptD 27-42 PVTINADHAKGNYPDN
LptD 27-45 PVTINADHAKGNYPDNAVF
LptD 30-45 INADHAKGNYPDNAVF
LptD 46-58 TGNVDINQGNSRL
LptD 63-79 VQLHQQQAAGQAQPVRT
LptD 63-81 VQLHQQQAAGQAQPVRTVD
LptD 63-92 VQLHQQQAAGQAQPVRTV DALGNVHYDDNQ
LptD 66-81 HQQQAAGQAQPVRTVD
LptD 66-92 HQQQAAGQAQPVRTVDAL GNVHYDDNQ
LptD 93-104 VILKGPKAWSNL
LptD 105-115 NTKDTNVWQGD
LptD 105-118 NTKDTNVWQGDYQM
LptD 116-129 YQMVGRQGRGTADL
LptD 117-129 QMVGRQGRGTADL
LptD 119-129 VGRQGRGTADL
LptD 130-141 MKQRGENRYTIL
LptD 171-179 VAEIWNARF
LptD 180-187 KLGSVPIF
LptD 188-194 YSPYLQL
LptD 188-204 YSPYLQLPVGDKRRSGF
LptD 195-211 PVGDKRRSGFLIPNAKY
LptD 195-218 PVGDKRRSGFLIPNAKYS TKNGVE
LptD 207-218 PNAKYSTKNGVE
LptD 232-247 DATITPHYMNKRGGVM
LptD 232-251 DATITPHYMNKRGGVMWENE
LptD 252-258 FRYLTQL
LptD 259-265 GSGLTEF
LptD 266-289 DYLPSDKVYEDDHSSDSNSRRWLF
LptD 290-299 YWNHSGVIDQ
LptD 307-315 YTKVSDPDY
LptD 307-319 YTKVSDPDYFNDF
LptD 339-344 YVNQNF
LptD 345-356 DATVSTKQFQVF
LptD 357-370 DRESSNSYSAEPQL
LptD 375-386 YQNDVGPFDTHL
D uptake (Da)
Time (min)

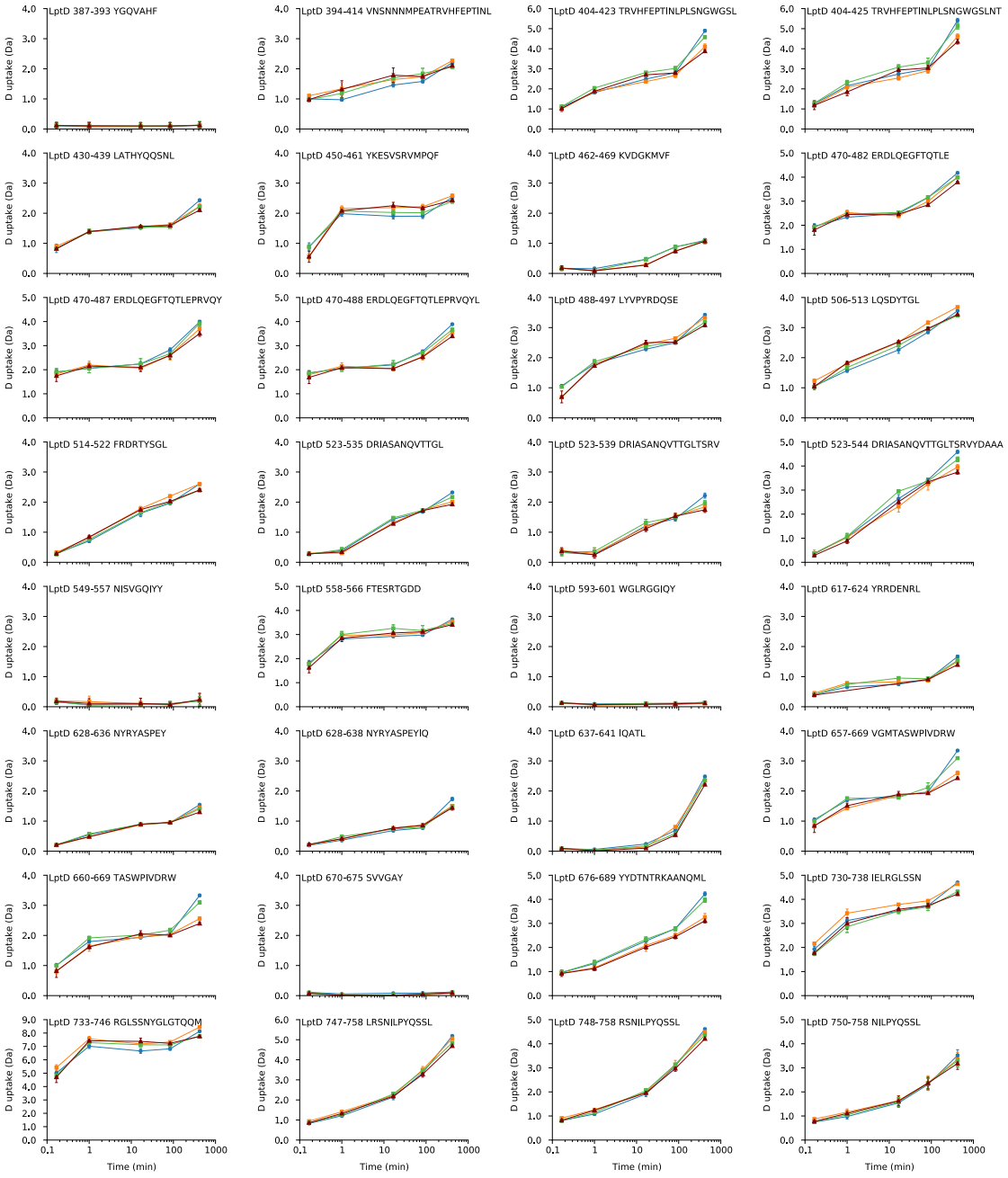


Figure S4.5 Deuterium uptake plots for LptD in four states. Deuterium uptake is plotted as a function of labelling time (0.167 – 420 min) for all identified peptides. The apo state is indicated in blue, LPS-bound state in orange, thanatin-bound state in green, and LPS + thanatin state in maroon. Standard deviations are plotted as error bars ($n_{biological}$ = 2; $n_{technical}$ = 3) but are in some instances too small to be visible.

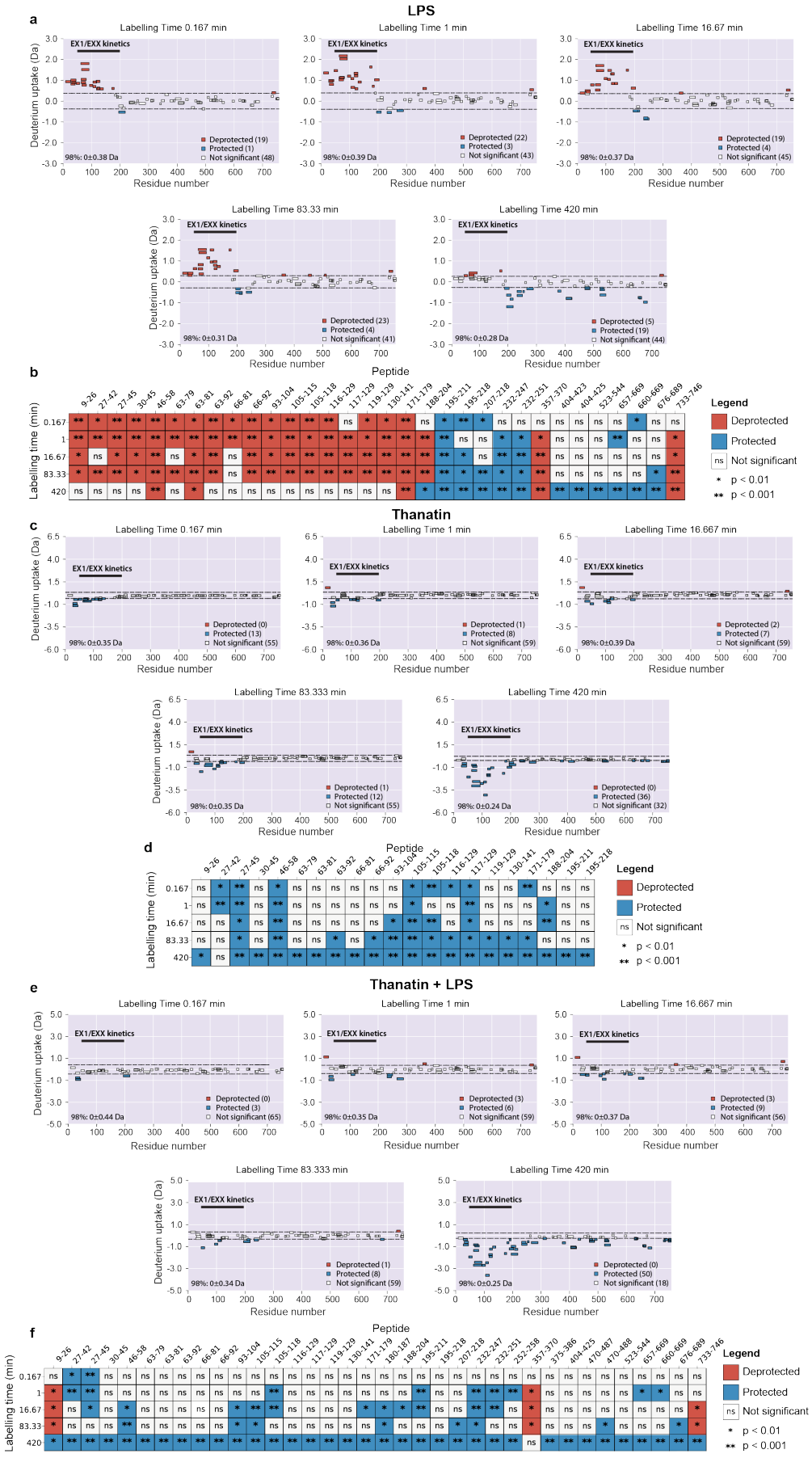
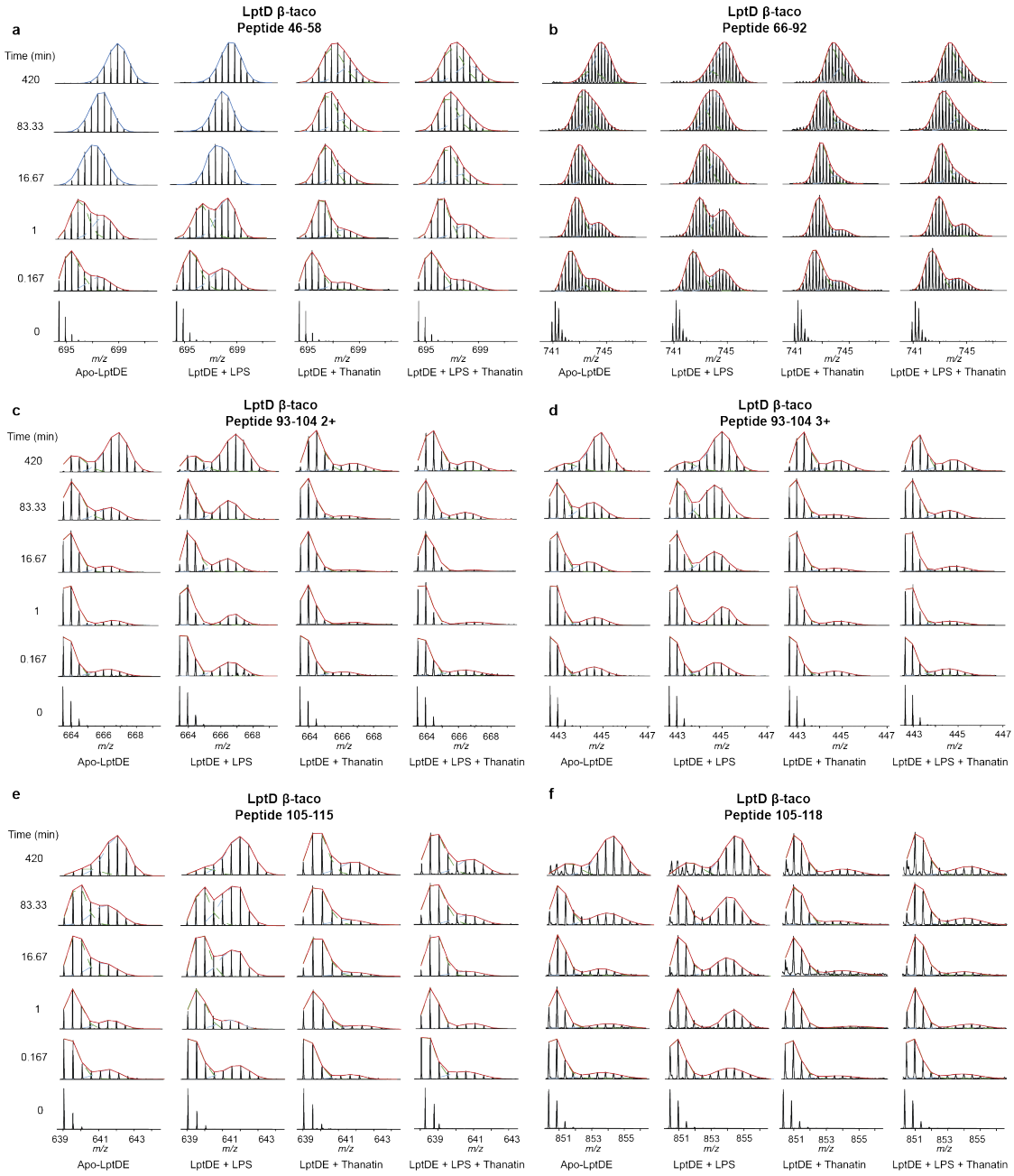

LPS
a
Labelling Time 0.167 min
Labelling Time 1 min
Labelling Time 16.67 min
Labelling Time 83.33 min
Labelling Time 420 min
EX1/EXX kinetics
Deuterium uptake (Da)
Residue number
Deprotected (19)
Protected (1)
Not significant (48)
98%: 0±0.38 Da
Deprotected (22)
Protected (3)
Not significant (43)
98%: 0±0.39 Da
Deprotected (19)
Protected (4)
Not significant (45)
98%: 0±0.37 Da
Deprotected (23)
Protected (4)
Not significant (41)
98%: 0±0.31 Da
Deprotected (5)
Protected (19)
Not significant (44)
98%: 0±0.28 Da
b
Peptide
Labelling time (min)
9-26
27-42
27-45
30-45
46-58
63-79
63-81
63-92
66-81
66-92
93-104
105-115
105-118
116-129
117-129
119-129
130-141
171-179
188-204
195-211
195-218
207-218
232-247
232-251
357-370
404-423
523-544
657-669
660-669
676-689
733-746
Legend
Deprotected
Protected
Not significant
p < 0.01
p < 0.001
Thanatin
c
Labelling Time 0.167 min
Labelling Time 1 min
Labelling Time 16.667 min
Labelling Time 83.333 min
Labelling Time 420 min
Deprotected (0)
Protected (13)
Not significant (55)
98%: 0±0.35 Da
Deprotected (1)
Protected (8)
Not significant (59)
98%: 0±0.36 Da
Deprotected (2)
Protected (7)
Not significant (59)
98%: 0±0.39 Da
Deprotected (1)
Protected (12)
Not significant (55)
98%: 0±0.35 Da
Deprotected (0)
Protected (36)
Not significant (32)
98%: 0±0.24 Da
d
Peptide
Labelling time (min)
9-26
27-42
27-45
30-45
46-58
63-79
63-81
63-92
66-81
66-92
93-104
105-115
105-118
116-129
117-129
119-129
130-141
171-179
188-204
195-211
195-218
Legend
Deprotected
Protected
Not significant
p < 0.01
p < 0.001
Thanatin + LPS
e
Labelling Time 0.167 min
Labelling Time 1 min
Labelling Time 16.667 min
Labelling Time 83.333 min
Labelling Time 420 min
Deprotected (0)
Protected (3)
Not significant (65)
98%: 0±0.44 Da
Deprotected (3)
Protected (6)
Not significant (59)
98%: 0±0.35 Da
Deprotected (3)
Protected (9)
Not significant (56)
98%: 0±0.37 Da
Deprotected (1)
Protected (8)
Not significant (59)
98%: 0±0.34 Da
Deprotected (0)
Protected (50)
Not significant (18)
98%: 0±0.25 Da
f
Peptide
Labelling time (min)
9-26
27-42
27-45
30-45
46-58
63-79
63-81
63-92
66-81
66-92
93-104
105-115
105-118
116-129
117-129
119-129
130-141
171-179
180-187
188-204
195-211
195-218
207-218
232-247
232-251
252-258
357-370
375-386
404-425
470-487
470-488
523-544
657-669
660-669
676-689
733-746
Legend
Deprotected
Protected
Not significant
p < 0.01
p < 0.001

Figure S4.6 Woods plots and peptide table for LptD in the presence of LPS, Thanatin or LPS+Thanatin. (a) Woods plots showing the difference in deuterium uptake for each identified peptide in the presence of LPS. The length of the lines represents the length of the peptides. The confidence interval (CI) used to analyse each labelling time is 98%. The region showing EX1 kinetics is indicated in each Woods plot. (b) Peptides having a difference in deuterium uptake greater than the calculated CI according to the described methods are included in the table and a Student's t-test (two-sided, unpaired, $p < 0.01$) was performed to finally include only statistically significant differences. (c) Woods plots showing the difference in deuterium uptake for each identified peptide in the presence of thanatin. (d) Peptides having a difference in deuterium uptake greater than the calculated CI according to the described methods are included in the table and a Student's t-test (two-sided, unpaired, $p < 0.01$) was performed to finally include only statistically significant differences. (e) Woods plots showing the difference in deuterium uptake for each identified peptide in the presence of LPS and thanatin. (f) Peptides having a difference in deuterium uptake greater than the calculated CI according to the described methods are included in the table and Student's t-test (two-sided, unpaired, $p < 0.01$) was performed to finally include only statistically significant differences. Red colour indicates deprotected peptides, blue indicates protected; white and 'ns' indicate non-significant difference; * $p < 0.01$; ** $p < 0.001$. Detailed information of HDX-MS data is provided in Table S4.2.


a
LptD β-taco
Peptide 46-58
Time (min)
420
83.33
16.67
1
0.167
0
695
699
m/z
Apo-LptDE
LptDE + LPS
LptDE + Thanatin
LptDE + LPS + Thanatin
b
LptD β-taco
Peptide 66-92
741
745
c
LptD β-taco
Peptide 93-104 2+
664
666
668
d
LptD β-taco
Peptide 93-104 3+
443
445
447
e
LptD β-taco
Peptide 105-115
639
641
643
f
LptD β-taco
Peptide 105-118
851
853
855

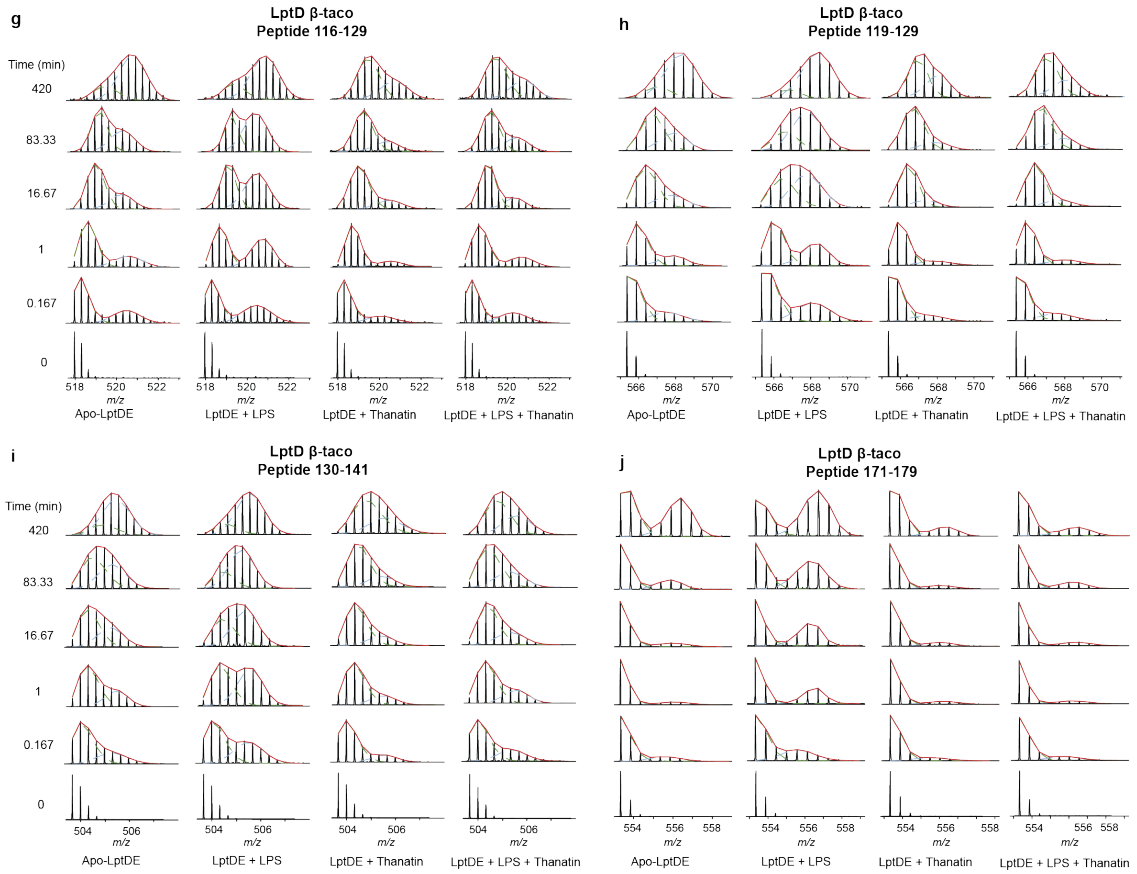


Figure S4.7 Representative mass spectra for peptides showing EX1/EXX kinetics. Mass spectra are shown for apo-LptDE, LPS, thanatin or LPS + thanatin states. Two binomial isotopic envelopes produced the best fit for the spectra yielding low- (green) and high-mass (light blue) populations. The sums of the two binomial distributions are shown in red. In the case of peptide 46-58 (a) it was not always possible to unambiguously determine the two populations; therefore, the blue line represents a single binomial fit. (b) Peptide 66-92. (c) Peptide 93-104 (charge state 2+). (d) Peptide 93-104 (charge state 3+). (e) Peptide 105-115. (f) Peptide 105-118. (g) Peptide 116-19. (h) Peptide 119-129. (i) Peptide 130-141. (j) Peptide 171-179.

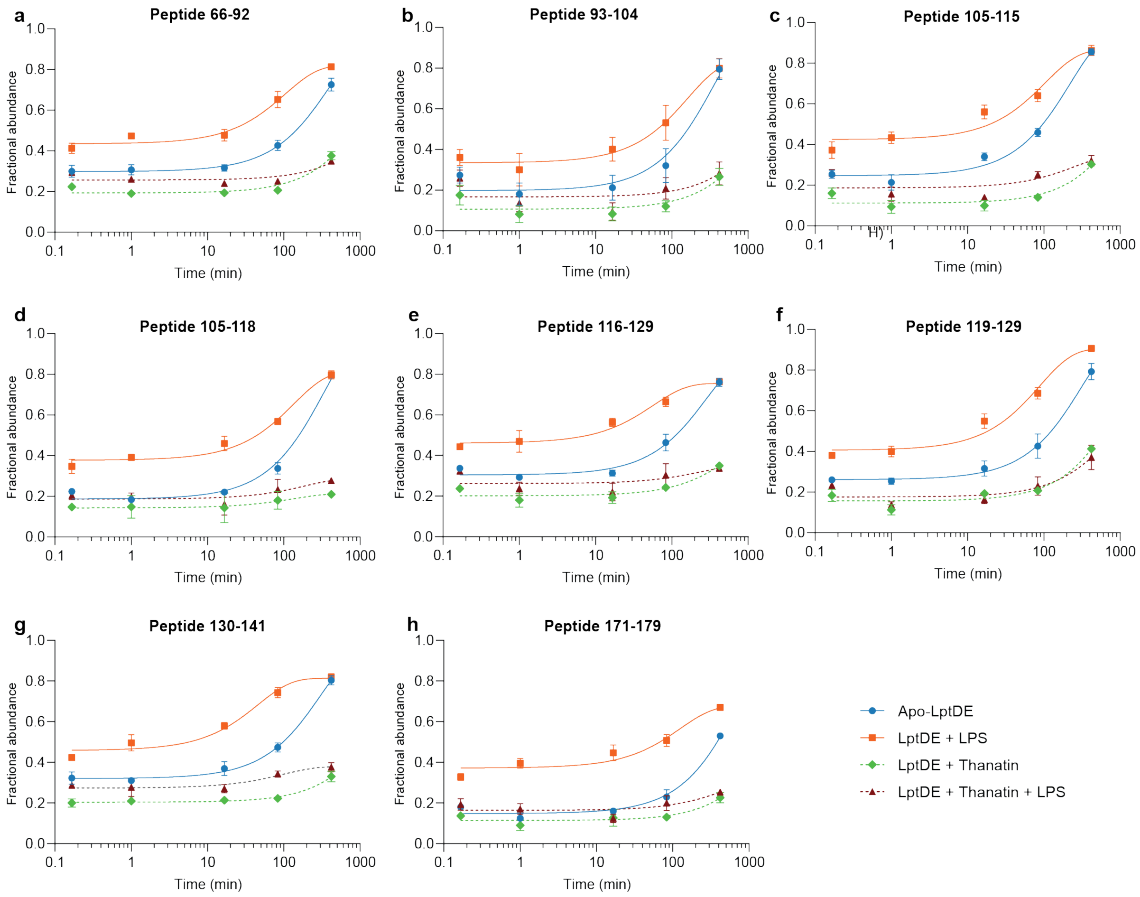


Figure S4.8 Monoexponential fitting of the high-mass population in peptides showing EX1 kinetics. Extracted relative abundances of high- mass populations plotted as a function of labelling time and fitted to a single exponential function with a variable intercept and plateau (to account for the lack of saturation seen in most peptides) to obtain the rate of translation from the low-mass population to the high-mass population (k_{op}) and the half-life of the low mass population ($t_{1/2}$). The apo state is indicated in blue, LPS-bound state in orange, thanatin-bound state in green, and LPS + thanatin state in maroon. Standard deviations are plotted as error bars ($n_{biological} = 2$; $n_{technical} = 3$) but are in some instances too small to be visible. In the case of thanatin-bound state and LPS + thanatin state kinetic values could not be extracted because of poor fitting (the increase of high-mass population is incremental given the slow kinetics). In these cases, the dotted line is included only for visual guidance.

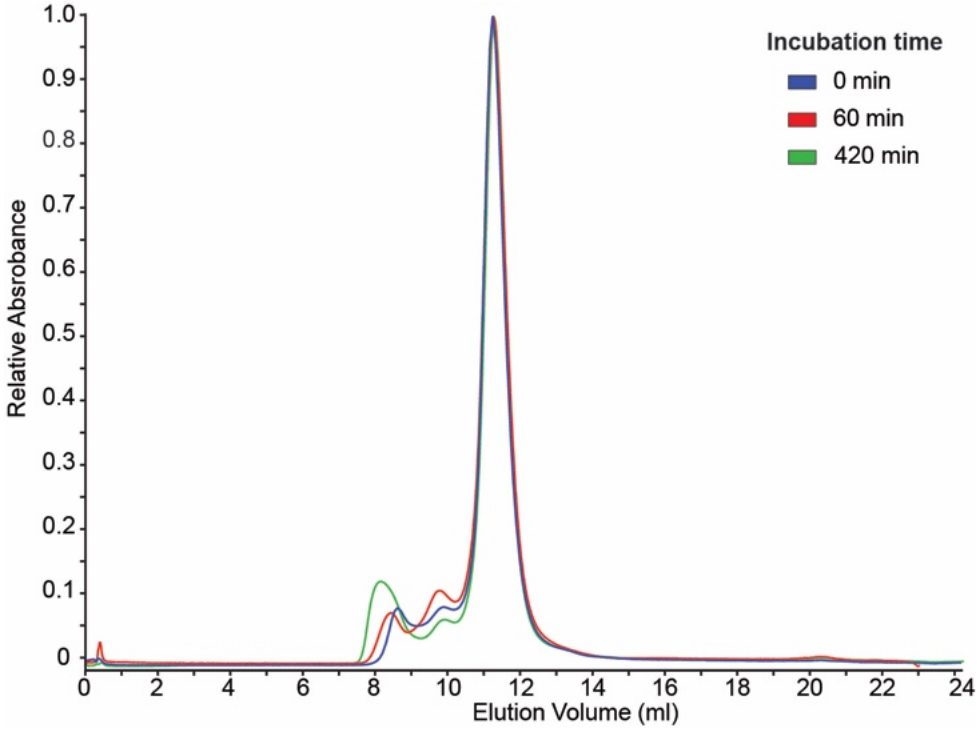
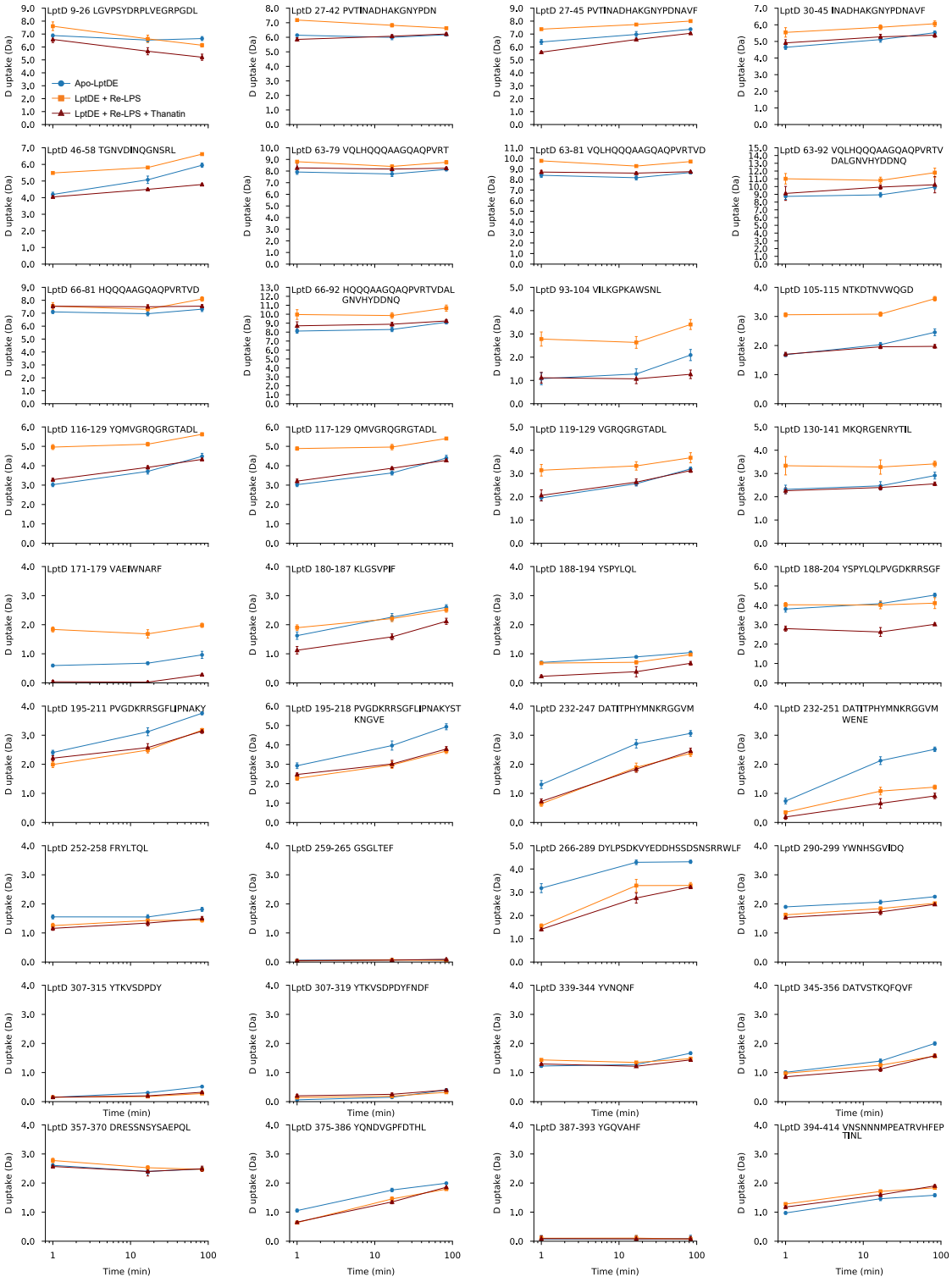


Figure S4.9 Analytical SEC. Analytical SEC of LptDE performed in the buffer used for protein purification and HDX-MS experiments. The protein has been incubated at 20 °C for 0, 60, and 420 min before injection in the SEC column.


LptD 9-26 LGVPSYDRPLVEGRPGDL
Apo-LptDE
LptDE + Re-LPS
LptDE + Re-LPS + Thanatin
LptD 27-42 PVTINADHAKGNYPDN
LptD 27-45 PVTINADHAKGNYPDNAVF
LptD 30-45 INADHAKGNYPDNAVF
LptD 46-58 TGNVDINQGNSRL
LptD 63-79 VQLHQQQAAGQAQPVRT
LptD 63-81 VQLHQQQAAGQAQPVRTVD
LptD 63-92 VQLHQQQAAGQAQPVRTV DALGNVHYDDNQ
LptD 66-81 HQQQAAGQAQPVRTVD
LptD 66-92 HQQQAAGQAQPVRTVDAL GNVHYDDNQ
LptD 93-104 VILKGPKAWSNL
LptD 105-115 NTKDTNVWQGD
LptD 116-129 YQMVGRQGRGTADL
LptD 117-129 QMVGRQGRGTADL
LptD 119-129 VGRQGRGTADL
LptD 130-141 MKQRGENRYTIL
LptD 171-179 VAEIWNARF
LptD 180-187 KLGSVPIF
LptD 188-194 YSPYLQL
LptD 188-204 YSPYLQLPVGDKRRSGF
LptD 195-211 PVGDKRRSGFLIPNAKY
LptD 195-218 PVGDKRRSGFLIPNAKYST KNGVE
LptD 232-247 DATITPHYMNKRGGVM
LptD 232-251 DATITPHYMNKRGGVM WENE
LptD 252-258 FRYLTQL
LptD 259-265 GSGLTEF
LptD 266-289 DYLPSDKVYEDDHSSDSNSRRWLF
LptD 290-299 YWNHSGVIDQ
LptD 307-315 YTKVSDPDY
LptD 307-319 YTKVSDPDYFNDF
LptD 339-344 YVNQNF
LptD 345-356 DATVSTKQFQVF
LptD 357-370 DRESSNSYSAEPQL
LptD 375-386 YQNDVGPFDTHL
LptD 387-393 YGQVAHF
LptD 394-414 VNSNNNMPEATRVHFEP TINL
D uptake (Da)
Time (min)

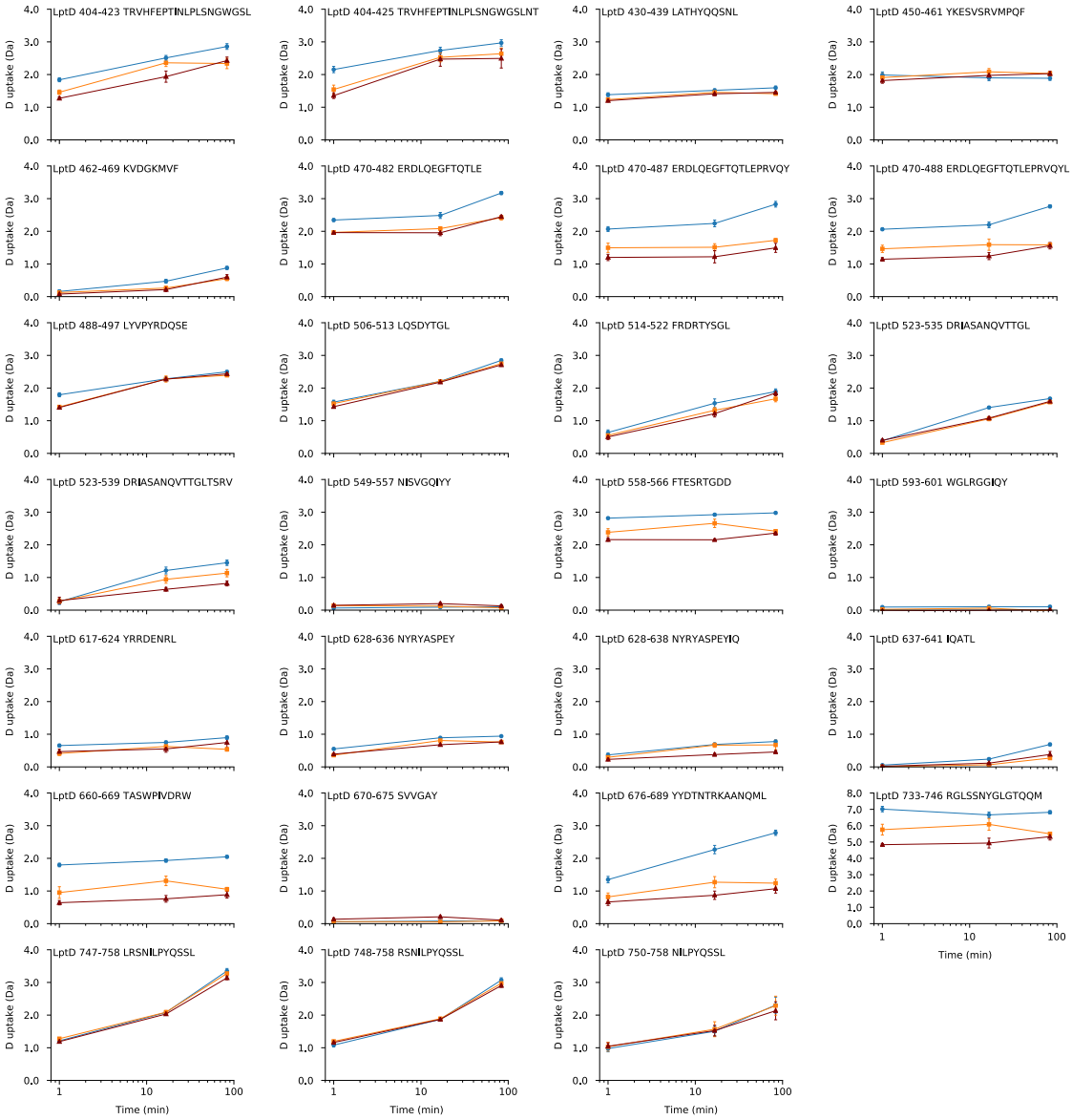


Figure S4.10 Deuterium uptake plots for LptD in three states. Deuterium uptake is plotted as a function of labelling time (1 – 83.33 min) for all identified peptides. The apo state is indicated in blue, Re-LPS-bound state in orange, and Re-LPS + thanatin state in maroon. Standard deviations are plotted as error bars ($n_{biological}$ = 1; $n_{technical}$ = 3) but are in some instances too small to be visible.

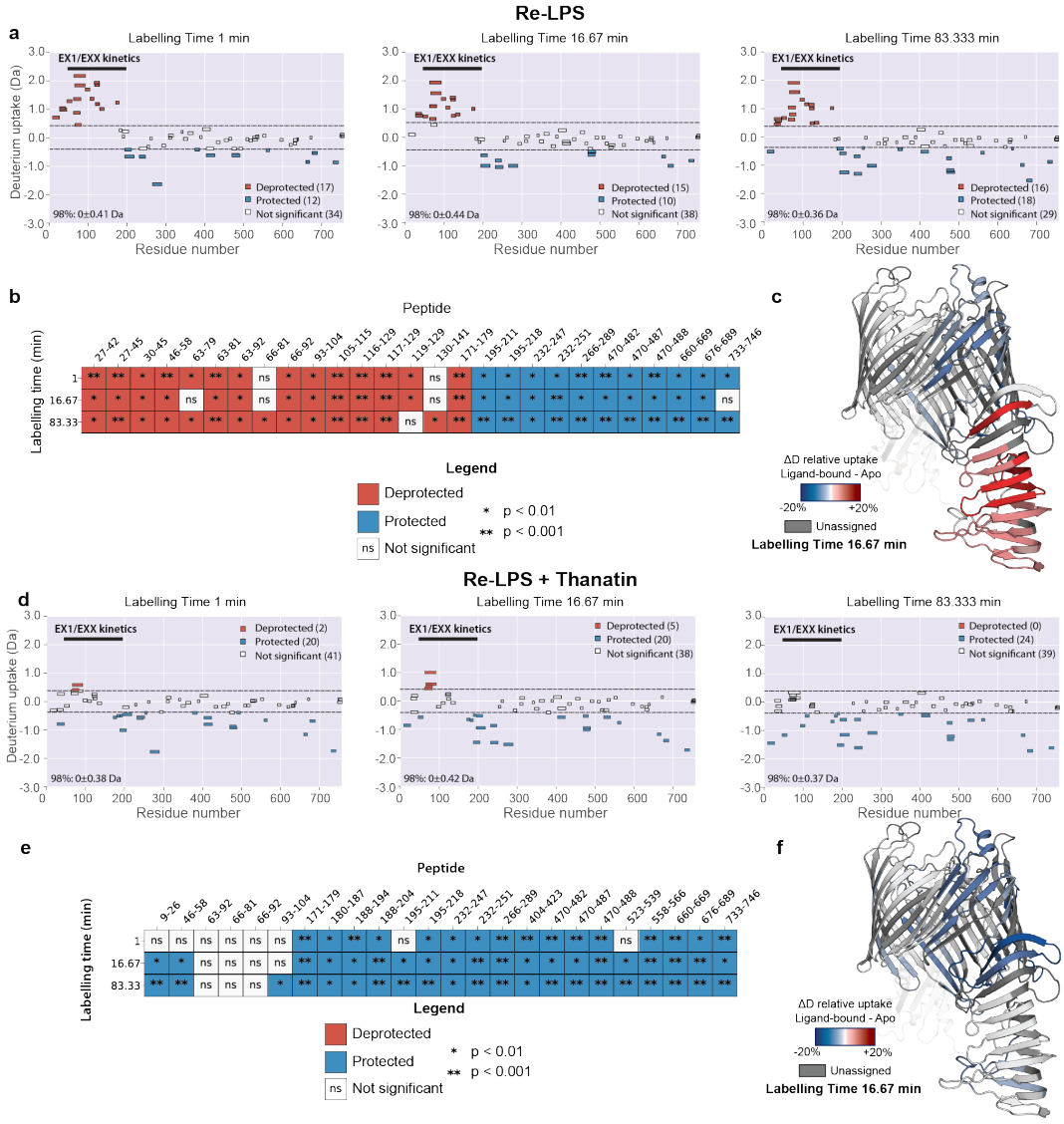
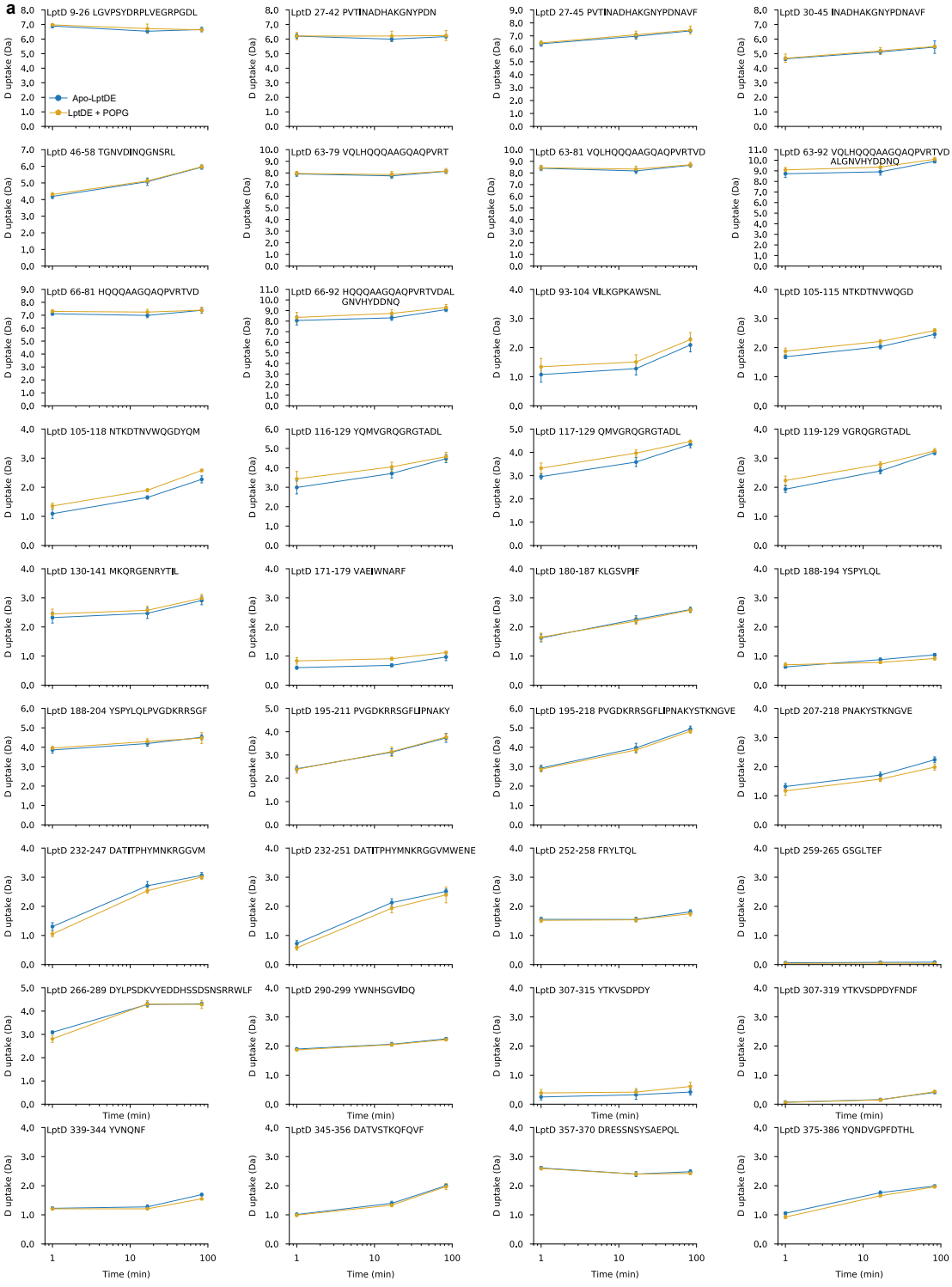


Figure S4.11 Woods plots and peptide table for LptD in the presence of Re-LPS or Re-LPS+Thanain. (a) Woods plots showing the difference in deuterium uptake for each identified peptide in the presence of Re-LPS. The length of the lines represents the length of the peptides. The CI used to analyse each labelling time is 98%. The region showing EX1 kinetics is indicated in each Woods plot. (b) Peptides having a difference in deuterium uptake greater than the calculated CI according to the described methods are included in the table and Student's t-test (two-sided, unpaired, $p < 0.01$) was performed to finally include only statistically significant differences. (c) Difference in relative deuterium uptake (scaled for the number of residues of each peptide) at the 16.67 min labelling time mapped on the crystal structure of LptD (PDB ID: 5IV9). (d) Woods plots showing the difference in deuterium uptake for each identified peptide

in the presence of Re-LPS and thanatin. (e) Peptides having a difference in deuterium uptake greater than the calculated CI according to the described methods are included in the table and Student's t-test (two-sided, unpaired, $p < 0.01$) was performed to finally include only statistically significant differences. (f) Difference in relative deuterium uptake (scaled for the number of residues of each peptide) at the 16.67 min labelling time mapped on the crystal structure of LptD (PDB ID: 5IV9). Only peptides showing significant difference are coloured. Red colour indicates deprotected peptides, blue indicates protected; white and 'ns' indicate non-significant difference; * $p < 0.01$; ** $p < 0.001$. Detailed information of HDX-MS data is provided in Table S4.3.


a
LptD 9-26 LGVPSYDRPLVEGRPGDL
LptD 27-42 PVTINADHAKGNYPDN
LptD 27-45 PVTINADHAKGNYPDNAVF
LptD 30-45 INADHAKGNYPDNAVF
LptD 46-58 TGNVDINQGNSRL
LptD 63-79 VQLHQQQAAGQAQPVRT
LptD 63-81 VQLHQQQAAGQAQPVRTVD
LptD 63-92 VQLHQQQAAGQAQPVRTVD ALGNVHYDDNQ
LptD 66-81 HQQQAAGQAQPVRTVD
LptD 66-92 HQQQAAGQAQPVRTVDAL GNVHYDDNQ
LptD 93-104 VILKGPKAWSNL
LptD 105-115 NTKDTNVWQGD
LptD 105-118 NTKDTNVWQGDYQM
LptD 116-129 YQMVGRQGRGTADL
LptD 117-129 QMVGRQGRGTADL
LptD 119-129 VGRQGRGTADL
LptD 130-141 MKQRGENRYTIL
LptD 171-179 VAEIWNARF
LptD 180-187 KLGSVPIF
LptD 188-194 YSPYLQL
LptD 188-204 YSPYLQLPVGDKRRSGF
LptD 195-211 PVGDKRRSGFLIPNAKY
LptD 195-218 PVGDKRRSGFLIPNAKYSTKNGVE
LptD 207-218 PNAKYSTKNGVE
LptD 232-247 DATITPHYMNKRGGVM
LptD 232-251 DATITPHYMNKRGGVMWENE
LptD 252-258 FRYLTQL
LptD 259-265 GSGLTEF
LptD 266-289 DYLPSDKVYEDDHSSDSNSRRWLF
LptD 290-299 YWNHSGVIDQ
LptD 307-315 YTKVSDPDY
LptD 307-319 YTKVSDPDYFNDF
LptD 339-344 YVNQNF
LptD 345-356 DATVSTKQFQVF
LptD 357-370 DRESSNSYSAEPQL
LptD 375-386 YQNDVGPFDTHL
D uptake (Da)
Time (min)
Apo-LptDE
LptDE + POPG

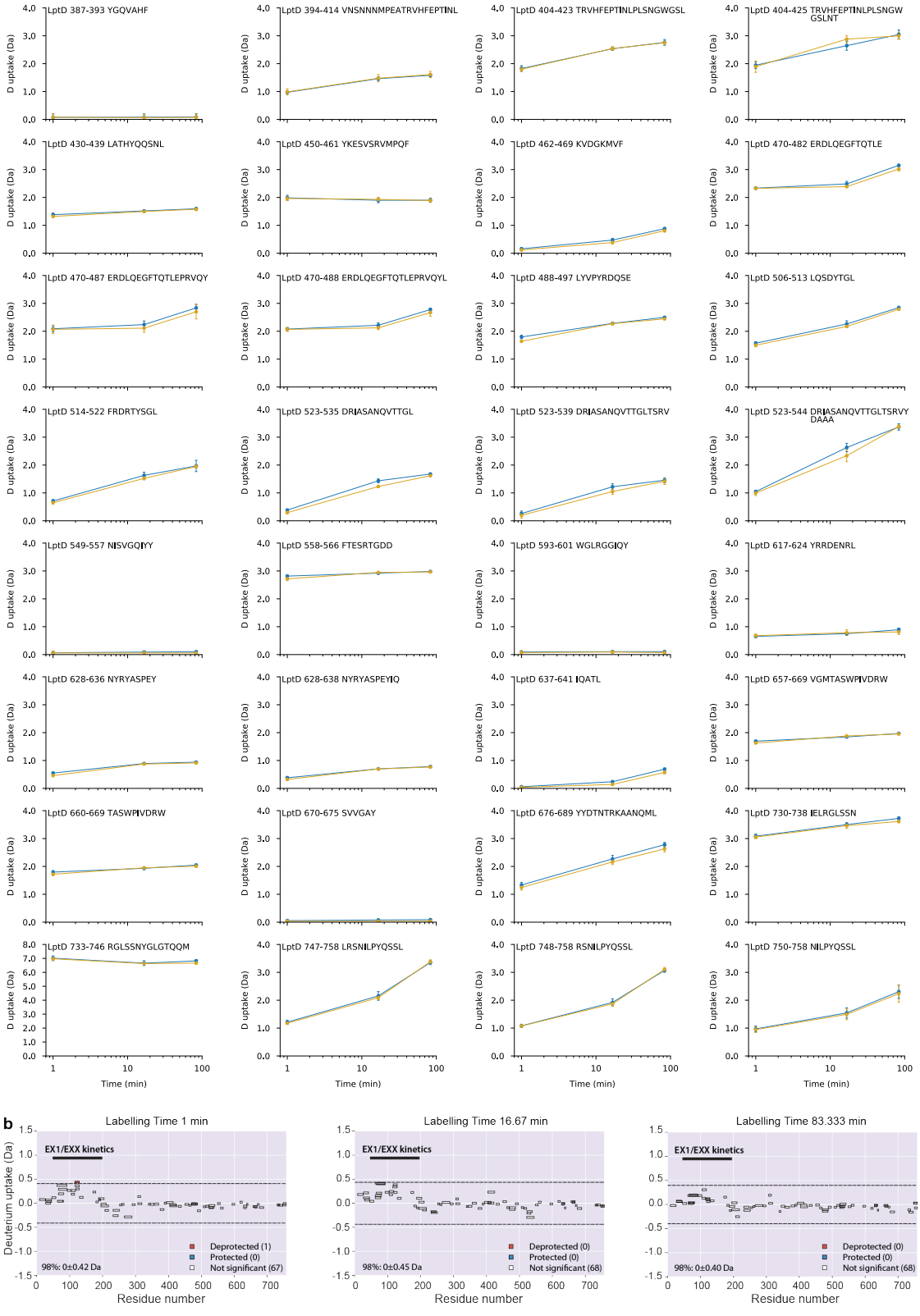
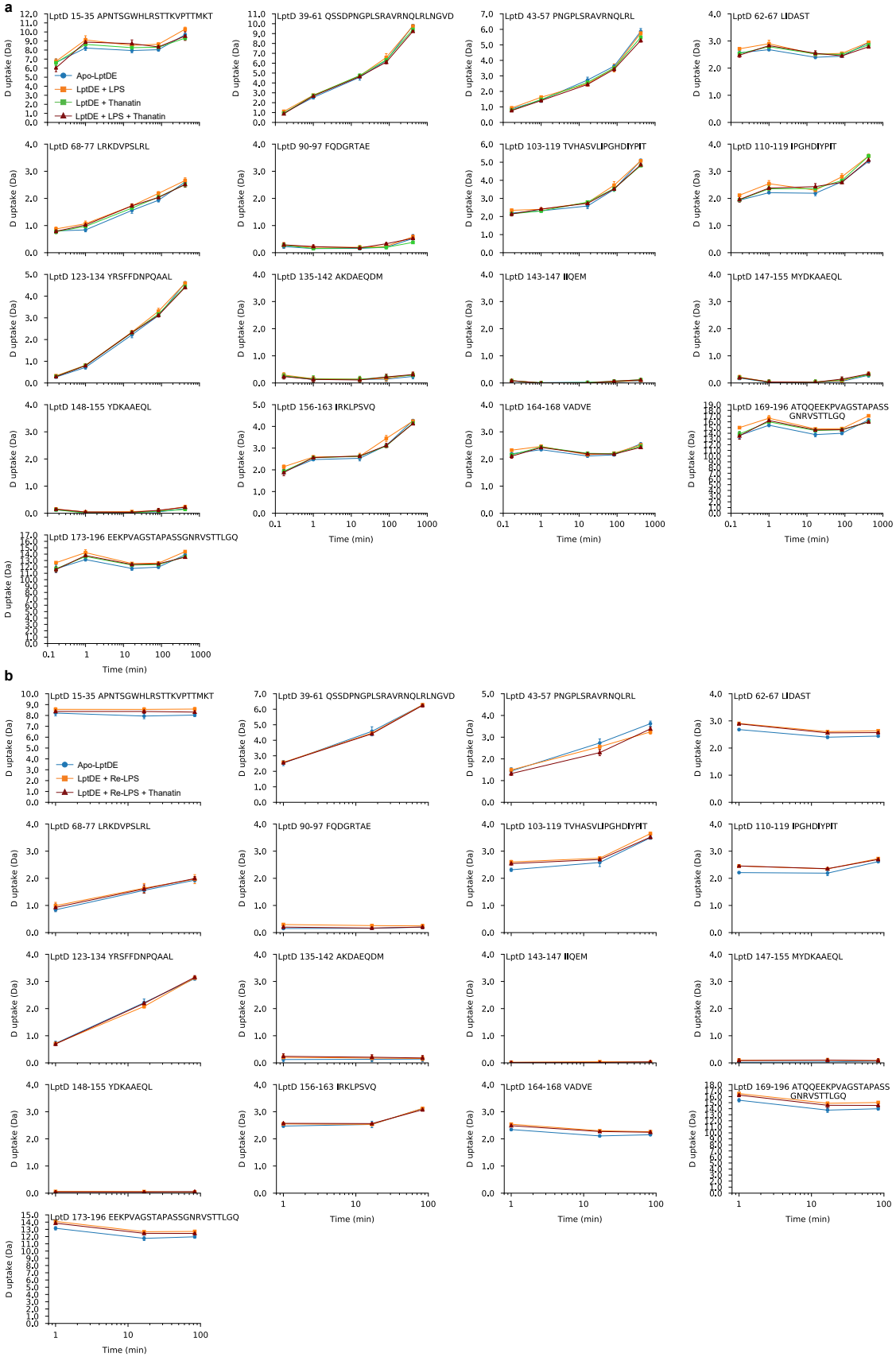


Figure S4.12 Deuterium uptake and woods plots for LptD in the presence of POPG. (a) Deuterium uptake is plotted as a function of labelling time (1 – 83.33 min) for all identified peptides. The apo state is indicated in blue, and POPG state in gold. Standard deviations are

plotted as error bars ($n_{biological}$ = 1; $n_{technical}$ = 3) but are in some instances too small to be visible. (b) Woods plots showing the difference in deuterium uptake for each identified peptide in the presence of Re-LPS. The length of the lines represents the length of the peptides. The CI used to analyse each labelling time is 98%. The region showing EX1 kinetics is indicated in each Woods plot. Red colour indicates deprotected peptides, blue indicates protected, white indicate non-significant difference. No statistically significant peptides were identified. Detailed information of HDX-MS data is provided in Table S4.3.


a
LptD 15-35 APNTSGWHLRSTTKVPTTMKT
LptD 39-61 QSSDPNGPLSRAVRNQLRLNGVD
LptD 43-57 PNGPLSRAVRNQLRL
LptD 62-67 LIDAST
LptD 68-77 LRKDVPSLRL
LptD 90-97 FQDGRTAE
LptD 103-119 TVHASVLIPGHDIYPIT
LptD 110-119 IPGHDIYPIT
LptD 123-134 YRSFFDNPQAAL
LptD 135-142 AKDAEQDM
LptD 143-147 IIQEM
LptD 147-155 MYDKAAEQL
LptD 148-155 YDKAAEQL
LptD 156-163 IRKLPSVQ
LptD 164-168 VADVE
LptD 169-196 ATQQEEKPVAGSTAPASS GNRVSTTLGQ
LptD 173-196 EEKPVAGSTAPASSGNRVSTTLGQ
Apo-LptDE
LptDE + LPS
LptDE + Thanatin
LptDE + LPS + Thanatin
D uptake (Da)
Time (min)
b
LptD 15-35 APNTSGWHLRSTTKVPTTMKT
LptD 39-61 QSSDPNGPLSRAVRNQLRLNGVD
LptD 43-57 PNGPLSRAVRNQLRL
LptD 62-67 LIDAST
LptD 68-77 LRKDVPSLRL
LptD 90-97 FQDGRTAE
LptD 103-119 TVHASVLIPGHDIYPIT
LptD 110-119 IPGHDIYPIT
LptD 123-134 YRSFFDNPQAAL
LptD 135-142 AKDAEQDM
LptD 143-147 IIQEM
LptD 147-155 MYDKAAEQL
LptD 148-155 YDKAAEQL
LptD 156-163 IRKLPSVQ
LptD 164-168 VADVE
LptD 169-196 ATQQEEKPVAGSTAPASS GNRVSTTLGQ
LptD 173-196 EEKPVAGSTAPASSGNRVSTTLGQ
Apo-LptDE
LptDE + Re-LPS
LptDE + Re-LPS + Thanatin
D uptake (Da)
Time (min)

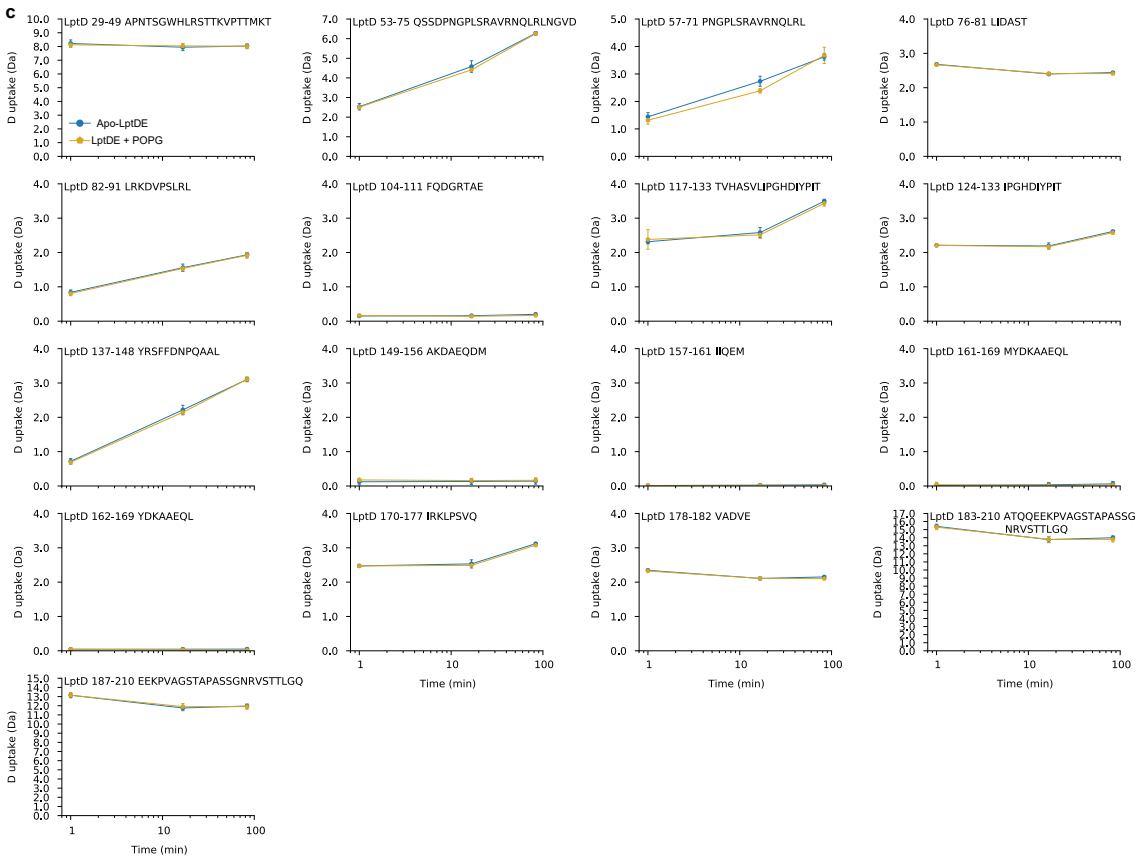


Figure S4.13 Deuterium uptake plots for LptE in all states. (a) Deuterium uptake is plotted as a function of labelling time (0.167 – 420 min) for all identified peptides in the apo, LPS, thanatin and thanatin + LPS states. The apo state is indicated in blue, LPS-bound state in orange, thanatin-bound state in green, and LPS + thanatin state in maroon. Error bars indicate standard deviations ($n_{biological}$ = 1; $n_{technical}$ = 3). (b) Deuterium uptake is plotted as a function of labelling time (1 – 83.33 min) for all identified peptides in apo, Re-LPS, and Re-LPS + thanatin states. The apo state is indicated in blue, Re-LPS-bound state in orange, and Re-LPS + thanatin state in maroon. Error bars indicate standard deviations ($n_{biological}$ = 1; $n_{technical}$ = 3). (c) Deuterium uptake is plotted as a function of labelling time (1 – 83.33 min) for all identified peptides in apo and POPG state. The apo state is indicated in blue, and POPG state in gold. Error bars indicate standard deviations ($n_{biological}$ = 1; $n_{technical}$ = 3). In some instances, error bars are too small to be visible.

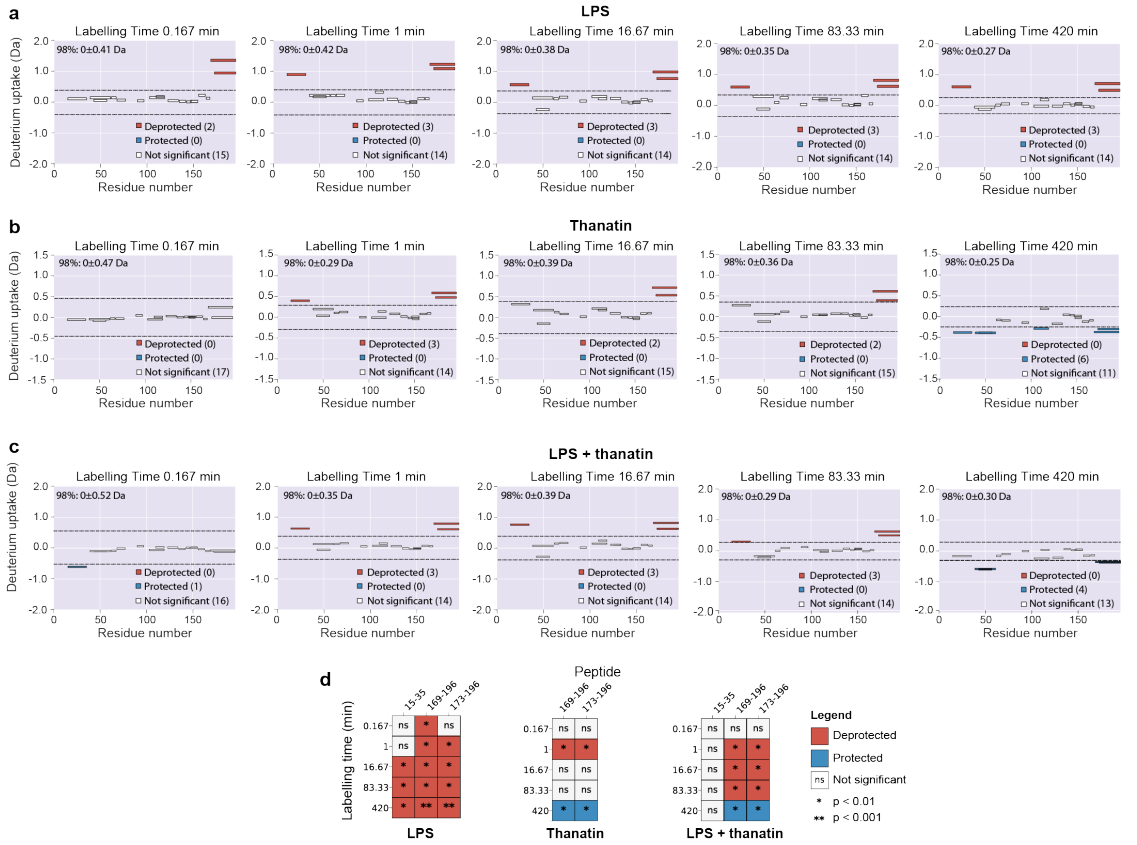


Figure S4.14 Woods plots and peptide table for LptE in the presence of LPS, thanatin and LPS + thanatin. Woods plots showing the difference in deuterium uptake for each identified peptide in the presence of LPS (a), thanatin (b), and thanatin + LPS (c). The length of the lines represents the length of the peptides. The CI used to analyse each labelling time is 98%. (d) Peptides having a difference in deuterium uptake greater than the calculated CI according to the described methods are included in the table and Student's t-test (two-sided, unpaired, $p < 0.01$) was performed to finally include only statistically significant differences. Red colour indicates deprotected peptides, blue indicates protected; white and 'ns' indicate non-significant difference; * $p < 0.01$; ** $p < 0.001$. Detailed information of HDX-MS data is provided in Table S4.4

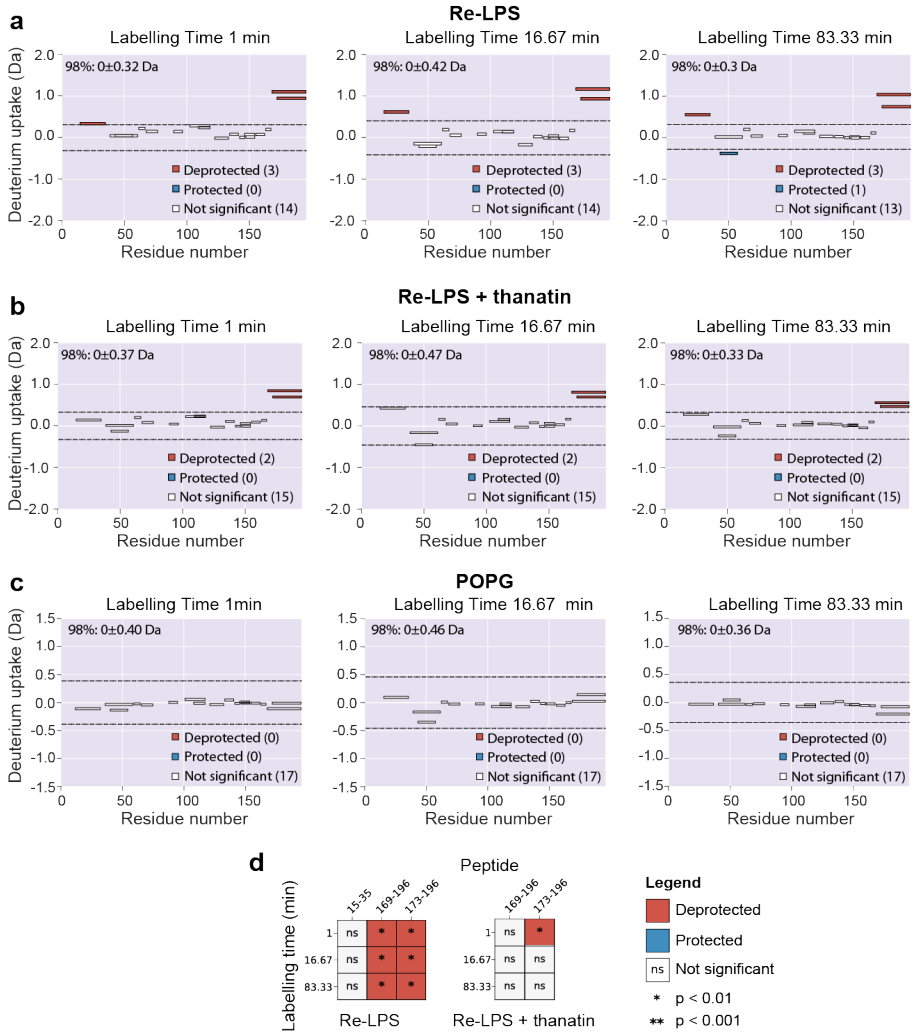


Figure S4.15 Woods plots and peptide table for LptE in the presence of Re-LPS, Re-LPS + thanatin and POPG. Woods plots showing the difference in deuterium uptake for each identified peptide in the presence of Re-LPS (a), Re-LPS + thanatin (b), POPG (c). The length of the lines represents the length of the peptides. The CI used to analyse each labelling time is 98%. (d) Peptides having a difference in deuterium uptake greater than the calculated CI according to the described methods are included in the table and Student's t-test (two-sided, unpaired, $p < 0.01$) was performed to finally include only statistically significant differences. Red colour indicates deprotected peptides, blue indicates protected; white and 'ns' indicate non-significant difference; * $p < 0.01$; ** $p < 0.001$. Detailed information of HDX-MS data is provided in Table S4.5.

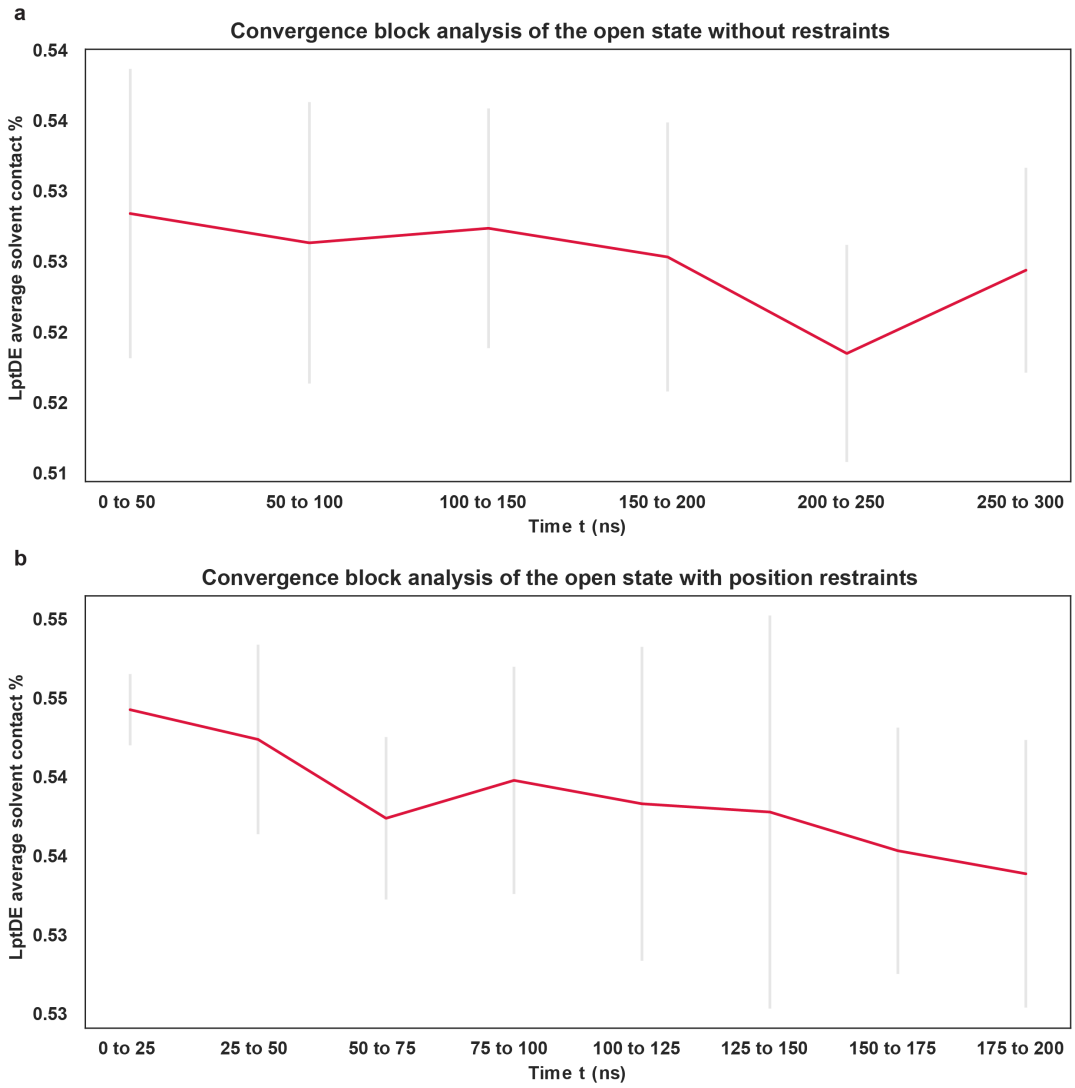


Figure S4.16 Block convergence analysis. Convergence plot of the open Re-LPS bound state (a) without and with (b) PLUMED restraints, monitoring average solvent contact over all residues in LptDE. Error bars correspond to mean and standard deviation of the five repeats.

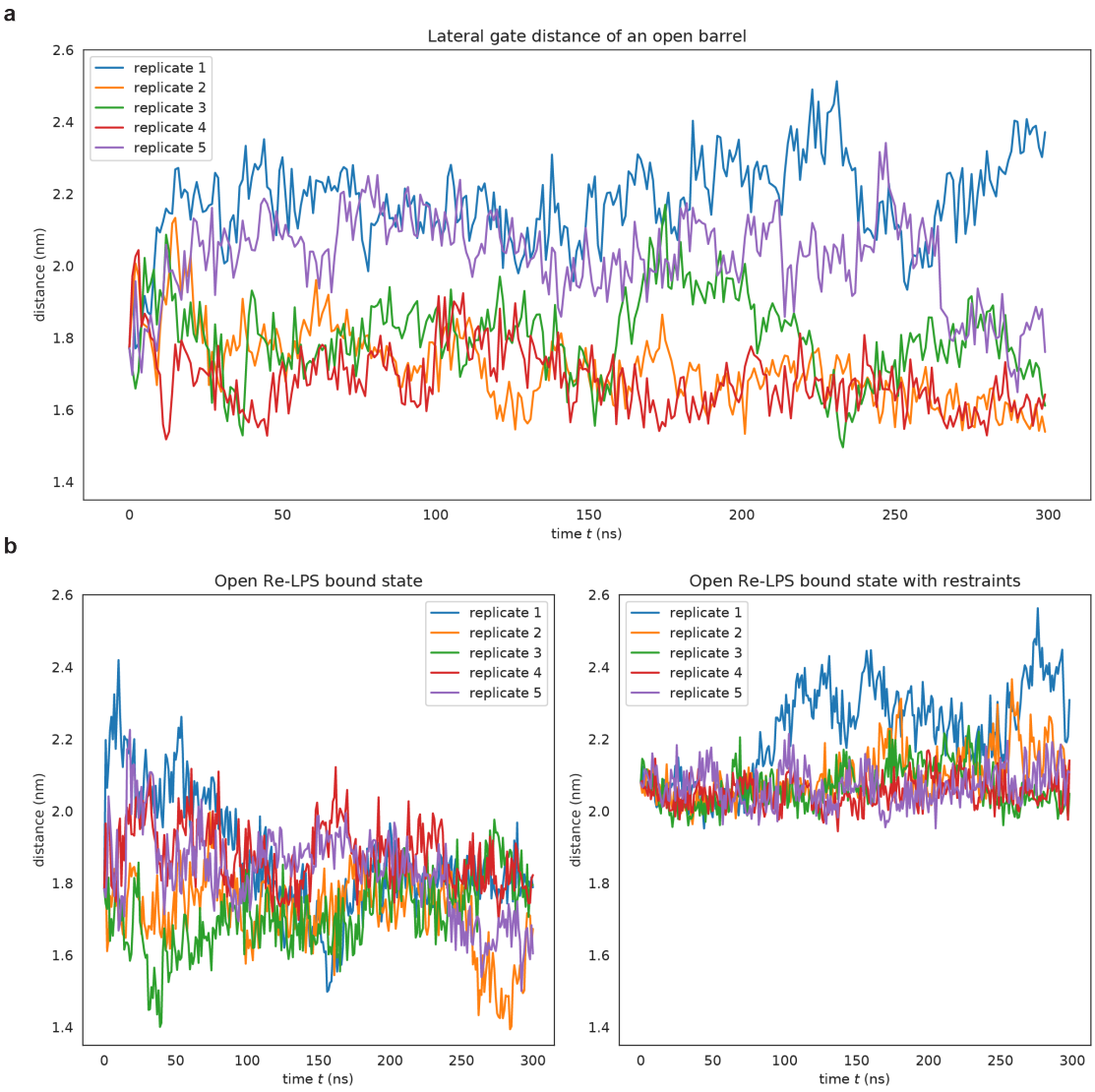


Figure S4.17 Lateral gate distance. (a) Distance of putative gate on LptDE β-barrel when opened without Re-LPS bound to the β-jellyroll. Replicates 1 and 5 were found to be stable over the course of the simulation after sampling a distance above 2.25 nm. (b) Comparison of the lateral gate distance without (left) and with (right) PLUMED restraints.

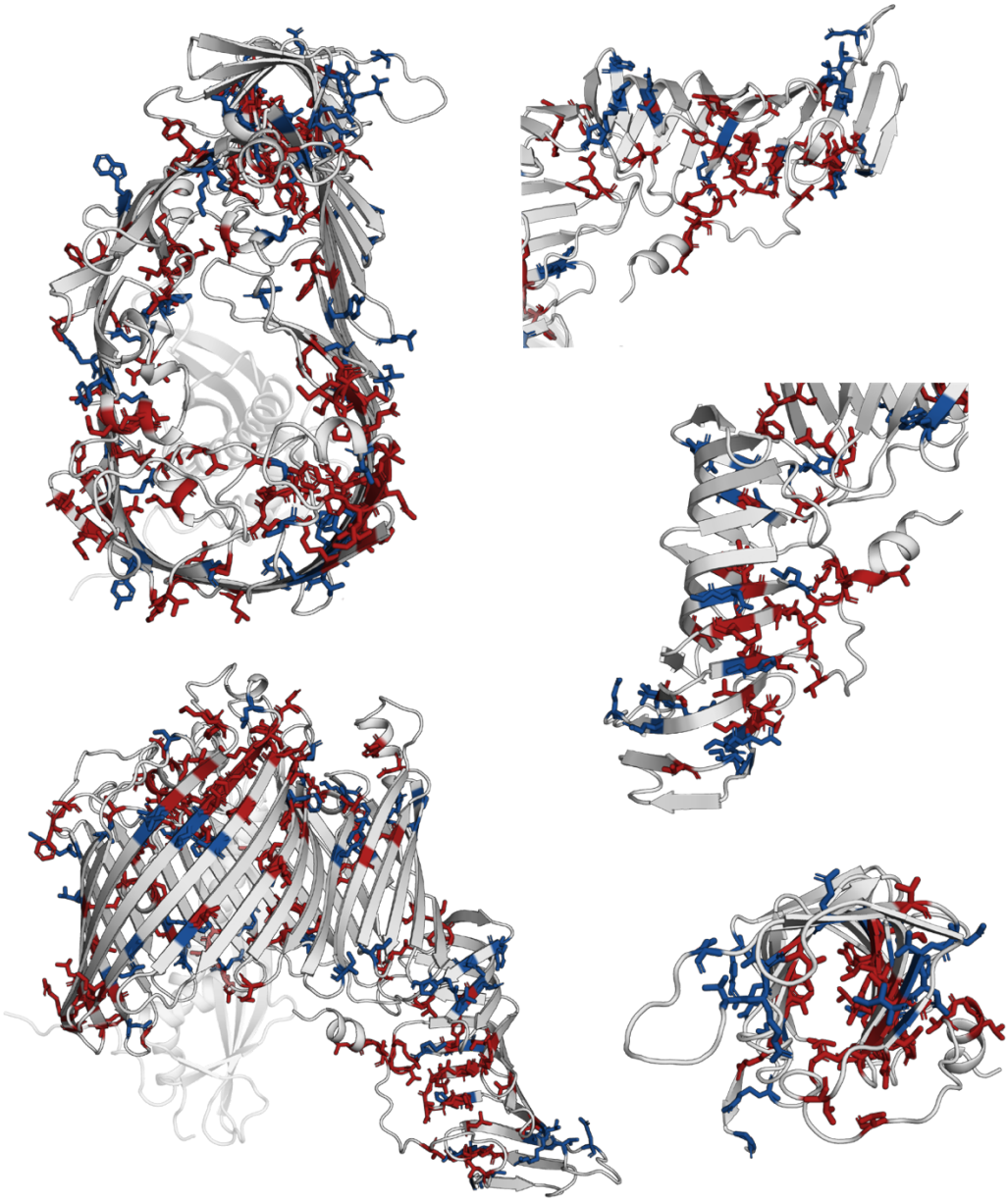


Figure S4.18 LptD open (Re-LPS bound) state vs closed state differential solvent contact map. Different views of the residue by residue solvent contact map where every significant residue is shown in the sticks representation. Loops pointing towards the extracellular side are highly solvated in our simulations, likely being more solvent exposed in a bilayer than in detergent, thus causing some peptides (e. g. 394-425 and 462-497) to be overly deprotected. Differences within the central peptides of the β-taco (e.g 105-129) were found not to be statistically significant, which is likely attributed to the relatively short timescales of the simulation. Nevertheless, the binding of Re-LPS causes a substantial amount of deprotection in the β-taco by virtue of an open conformation.

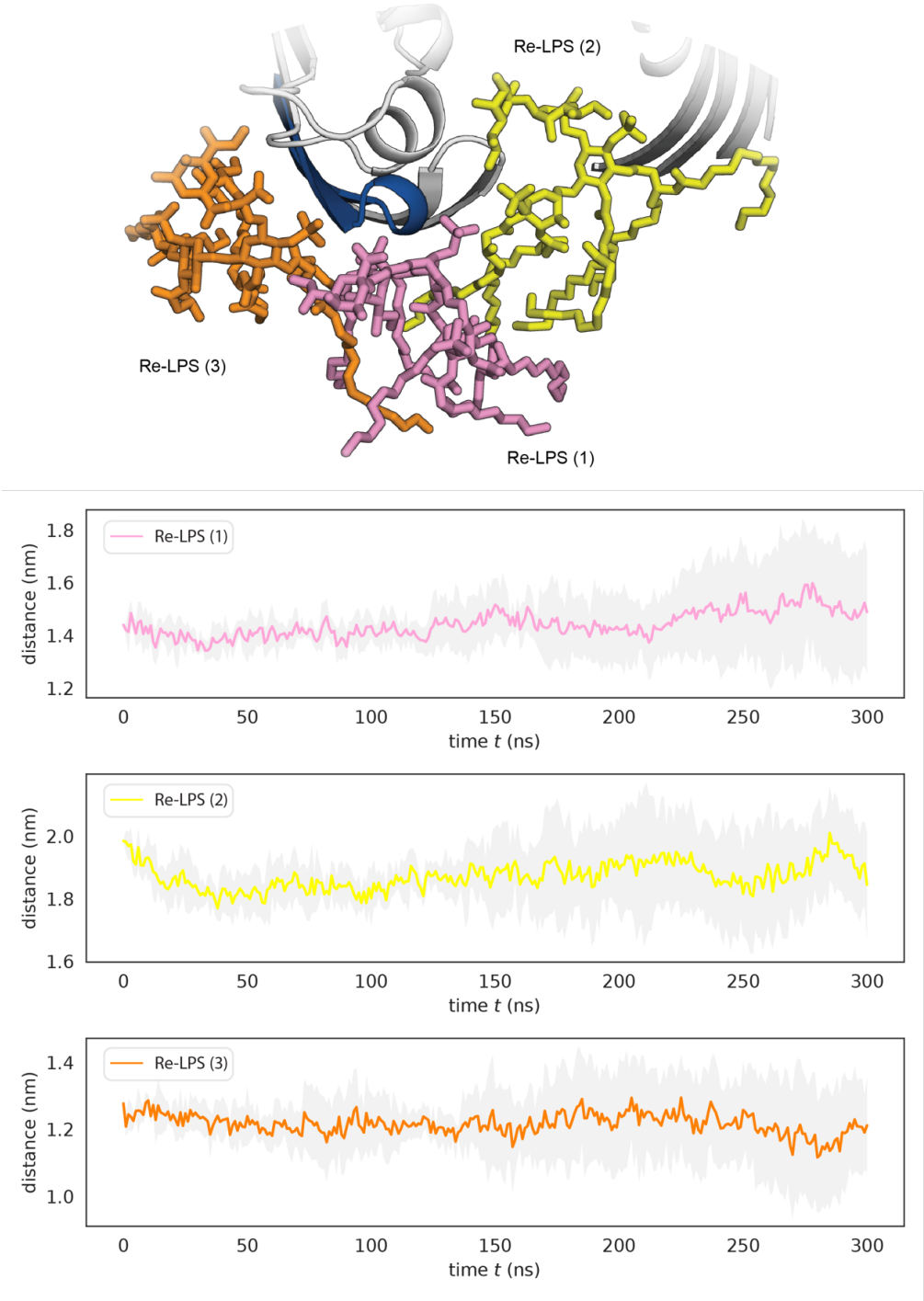


Figure S4.19 Re-LPS distance plots. Top: snapshot of the key lipid interactions between Re-LPS and the lateral gate peptide (232-251). Bottom: Plots measuring the distance from the geometric centre of the peptide (232-251) to the geometric centre of the lipids. Data plotted as the average of five repeats throughout the course of the simulation, standard deviation shaded grey.

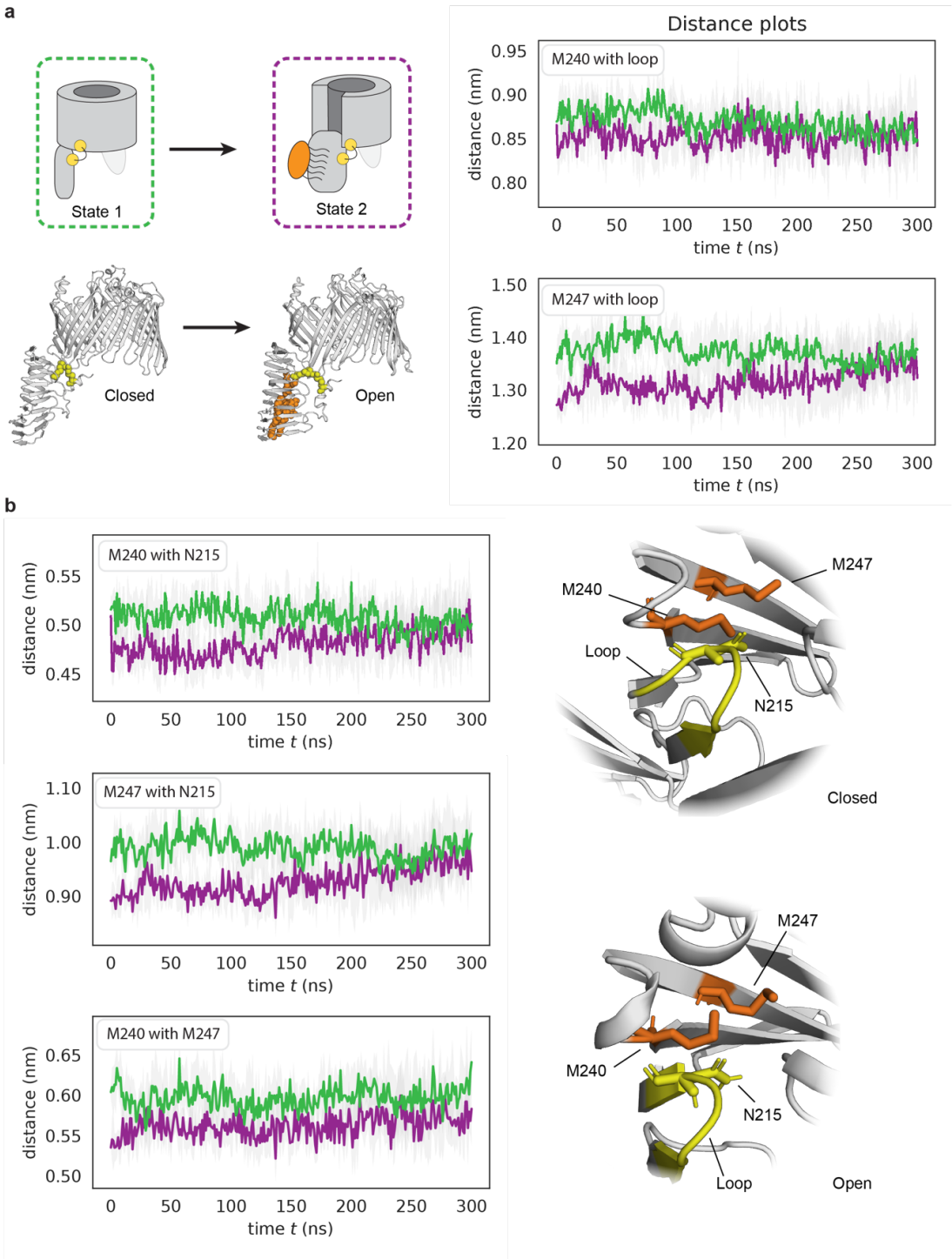


Figure S4.20 Residue distance plots. (a) Left: cartoon schematic illustrating the difference between the apo closed state (green) and the Re-LPS bound state (purple). Right: distance plots between Met240 and Met247 from the loop located between β1 and β2 of the β-barrel. (b) Left: Residue distance plots between Met240 and Met247 as well as N215. Right: Schematic of the open and closed state conformations; methionines coloured in orange with the asparagine and loop coloured yellow.

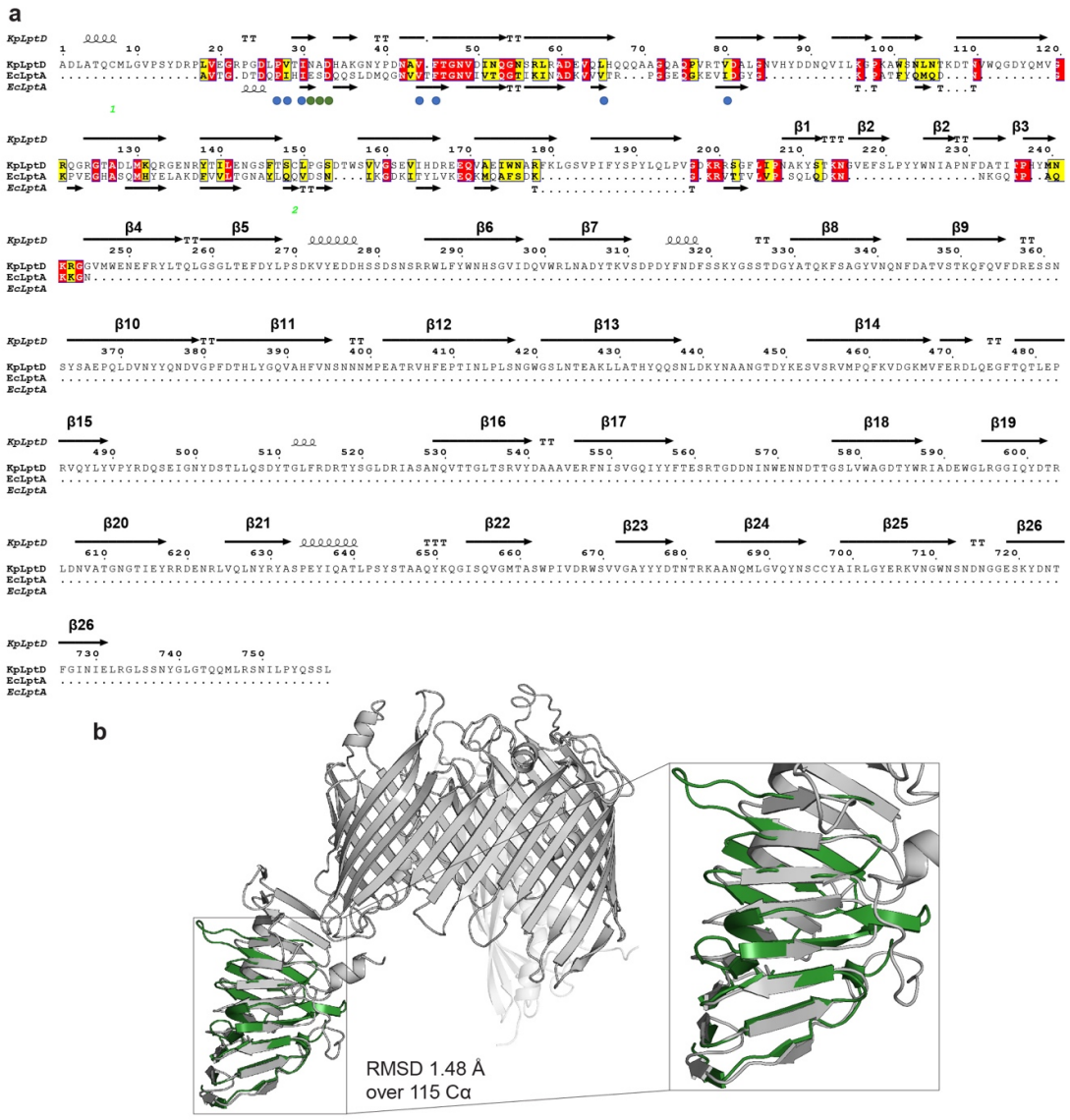


Figure S4.21 Sequence and structure alignments of *Kp*LptD and *Ec*LptA. (a) Sequence alignment between *Kp*LptD and *Ec*LptA. Red colour indicates conserved residues, yellow indicates highly similar residues. Circles indicate crucial residues for LptA-thanatin interaction. Blue colour represents van-der-Waals interactions while green represents ionic bridges and H-bonds. (b) Structure alignment of LptD (PDB ID: 5IV9) with apo-LptA (PDB ID: 2R1A). The RMSD is equal to 1.48 Å over 115 Cα.

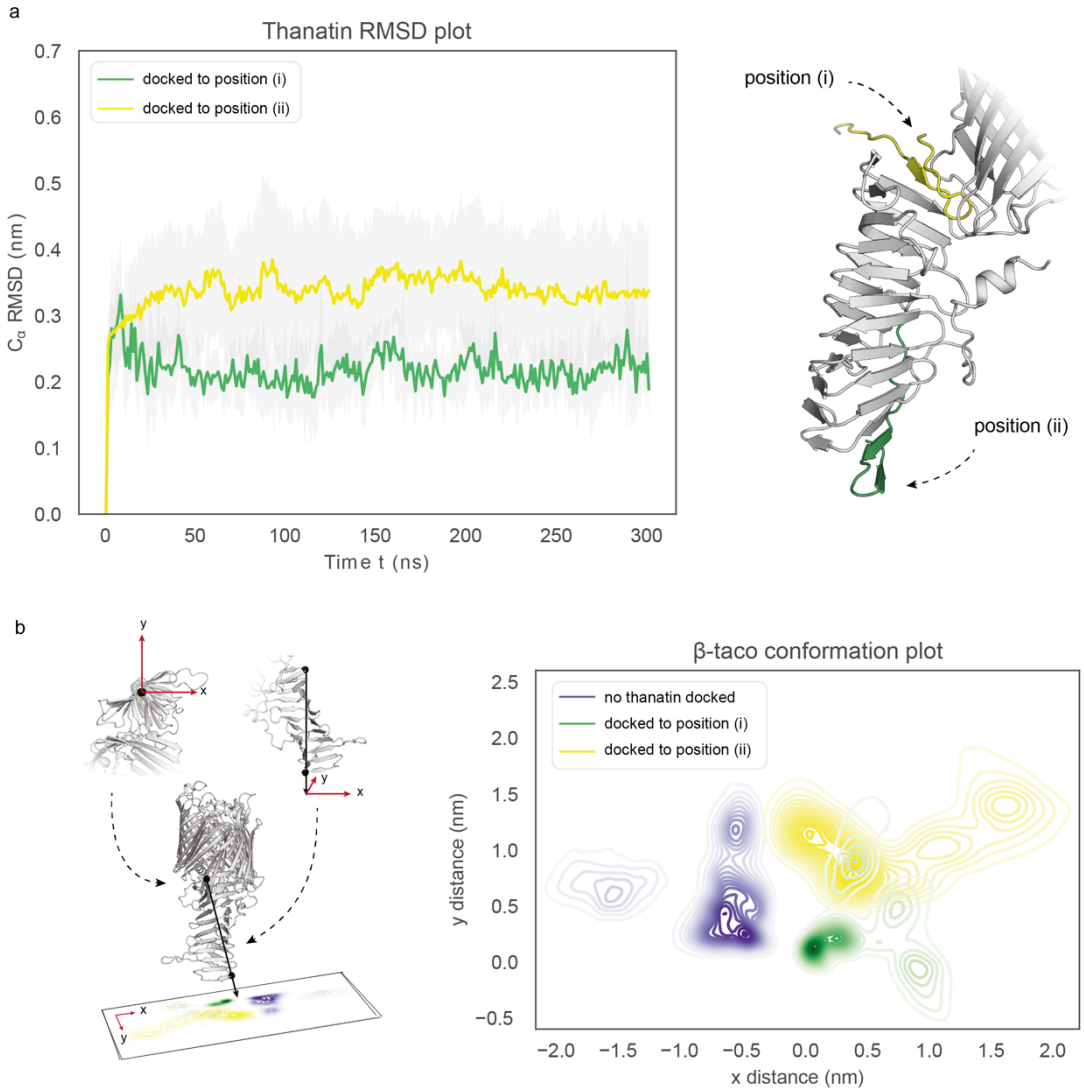
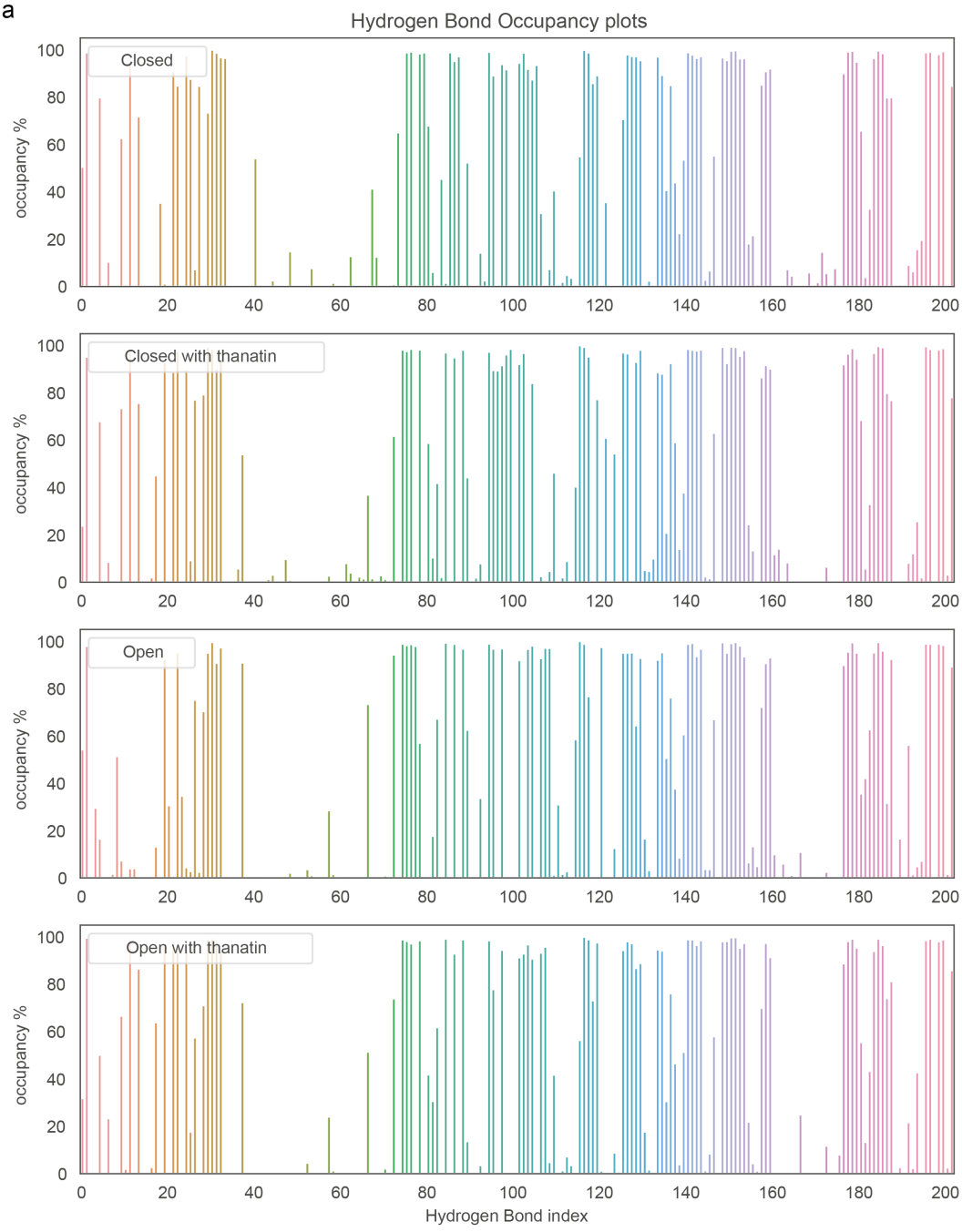


Figure S4.22 Thanatin stability. (a) RMSD of thanatin docked to the bottom (i) of the β-jellyroll in its closed conformation compared to being docked to the top (ii) where standard deviation is coloured grey. (b) Right: schematic showing the orientation of the β-taco vector drawn through C_α of residues Asp33 and Pro190, where x, y displacement is shown as red arrows. Right: x,y displacement of the β-taco vector mapped relative to the origin 0.0, 0.0 (defined as the original position of β-taco in the crystal structure) comparing the conformation space sampled of the apo-β-taco against when thanatin is docked to the (i) bottom and (ii) top of the β-taco.

a


Hydrogen Bond Occupancy plots
Closed
Closed with thanatin
Open
Open with thanatin
occupancy %
Hydrogen Bond index

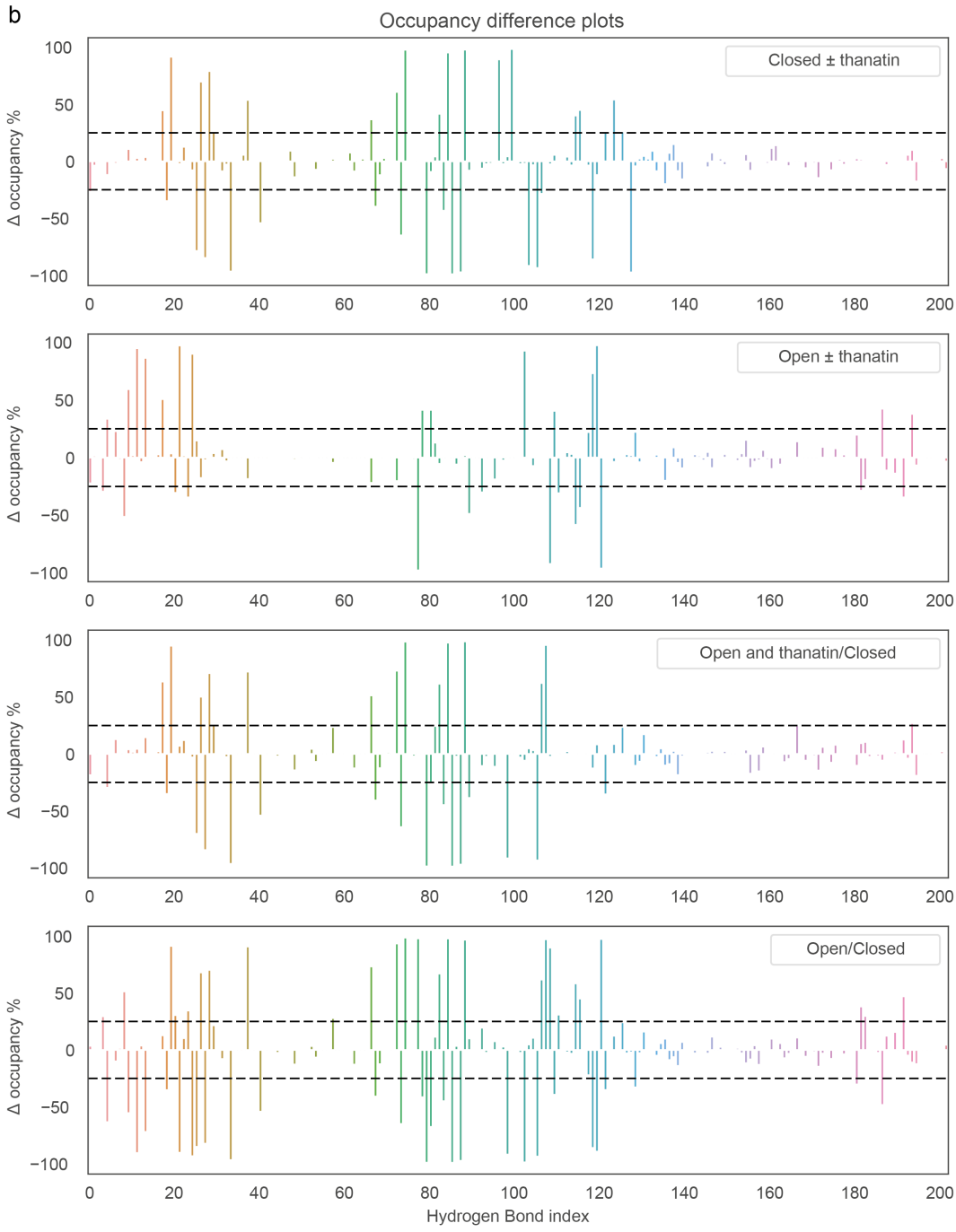


Figure S4.23 Hydrogen bond occupancy. (a) Percentage occupancy of hydrogen bonds on different states of the β-taco, considering only the backbone hydrogen bonds between residues 28 to 182 (excluding the first strand which has direct contact with thanatin). Occupancy of the hydrogen bonds was calculated over all five replicates of the four different states of the β-taco: closed, closed with thanatin docked, open (with Re-LPS) and open (with Re-LPS) and thanatin (plots listed from top to bottom respectively). (b) Difference plots of hydrogen bond occupancy

between states. From top to bottom, difference plot between: closed with thanatin and closed states, open (Re-LPS bound) with thanatin and open (Re-LPS bound) states, open (Re-LPS bound) with thanatin and closed states, open (Re-LPS bound) and closed states. In each chart, respectively arbitrary cut-offs at 25% and -25% occupancy difference are plotted.

4.4.2 Supplementary Tables

Complex	Experimental molecular mass ± s.d. (Da)	Theoretical molecular mass (Da)
LptDE	108,472 ± 1.0	108,473
LptDE + 1x Re-LPS	110,712 ± 1.2	110,711
LptDE + 2x Re-LPS	112,949 ± 3.7	112,950
LptDE + thanatin	110,907 ± 4.1	110,909
LptDE + Re-LPS + thanatin	113,145 ± 0.7	113,147
LptDE + LPS (peak 1)	112,489 ± 13	N/A
LptDE + LPS (peak 2)	113,373 ± 3.9	N/A
LptDE + 1x 16:0/18:1 PG	109,221 ± 2.2	109,222
LptDE + 2x 16:0/18:1 PG	109,972 ± 4.3	109,971
LptDE + 3x 16:0/18:1 PG	110,718 ± 2.2	110,720
LptDE + 1x 18:1 CDL	109,930 ± 0.5	109,931
LptDE + 2x 18:1 CDL	111,388 ± 1.6	111,389
LptDE + 3x 18:1 CDL	112,845 ± 4.5	112,847

Table S4.1 Molecular masses determined by nMS.

Data set	Apo-LptDE	LptDE+ LPS	LptDE + thanatin	LptDE+ LPS+thanatin
HDX reaction details	200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol, 0.03% DDM. Percent deuterium: 91.7%. Temperature: 20°C			
HDX time course	0.167, 1, 16.67, 83.33, 420 min			
Back-exchange (mean/IQR)	ND			
Number of peptides	68			
Sequence coverage	70.5%			
Average peptide length/redundancy	Average peptide length = 13.4 residues Redundancy = 1.66			
Replicates	$n_{biological} = 2$; $n_{technical} = 3$			
Repeatability (average SD)	0.079	0.087	0.083	0.088
Significant differences in HDX (16.67 min time point)	Reference state	CI 98% = 0.37	CI 98% = 0.39	CI 98% = 0.37

Table S4.2 HDX-MS data summary table for LptD in apo state and LPS, thanatin or LPS + thanatin states.

Data set	LptDE + Re-LPS	LptDE+ Re-LPS + thanatin	LptDE + POPG
HDX reaction details	200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol, 0.03% DDM. Percent deuterium: 91.7%. Temperature: 20°C		
HDX time course	1, 16.67, 83.33 min		
Back-exchange (mean/IQR)		ND	
Number of peptides	63	63	68
Sequence coverage	69.1%	69.1%	70.5%
Average peptide length/redundancy	13.3/1.56	13.3/1.56	13.4/1.66
Replicates	$n_{biological} = 1$ $n_{technical} = 3$	$n_{biological} = 1$ $n_{technical} = 3$	$n_{biological} = 1$ $n_{technical} = 3$
Repeatability (average SD)	0.082	0.076	0.080
Significant differences in HDX (16.67 min time point)	CI 98% = 0.44	CI 98% = 0.42	CI 98% = 0.45

Table S4.3 HDX-MS data summary table for LptD in Re-LPS, Re-LPS + thanatin, and POPG states.

Data set	Apo-LptDE	LptDE+ LPS	LptDE + thanatin	LptDE+ LPS+thanatin
HDX reaction details	200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol, 0.03% DDM. Percent deuterium: 91.7%. Temperature: 20°C			
HDX time course	0.167, 1, 16.67, 83.33, 420 min			
Back-exchange (mean/IQR)	ND			
Number of peptides	17			
Sequence coverage	87.4%			
Average peptide length/redundancy	12.8/1.36			
Replicates	$n_{biological} = 2$; $n_{technical} = 3$			
Repeatability (average SD)	0.090	0.083	0.079	0.088
Significant differences in HDX (16.67 min time point)	Reference state	CI 98% = 0.38	CI 98% = 0.39	CI 98% = 0.39

Table S4.4 HDX-MS data summary table for LptE in apo state and LPS, thanatin or LPS + thanatin states.

Data set	LptDE + Re-LPS	LptDE+ Re-LPS + thanatin	LptDE + POPG
HDX reaction details	200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol, 0.03% DDM. Percent deuterium: 91.7%. Temperature: 20°C		
HDX time course	1, 16.67, 83.33 min		
Back-exchange (mean/IQR)	ND		
Number of peptides	17		
Sequence coverage	87.4%		
Average peptide length/redundancy	12.8/1.36		
Replicates	$n_{biological} = 1$ $n_{technical} = 3$	$n_{biological} = 1$ $n_{technical} = 3$	$n_{biological} = 1$ $n_{technical} = 3$
Repeatability (average SD)	0.052	0.063	0.073
Significant differences in HDX (16.67 min time point)	CI 98% = 0.42	CI 98% = 0.47	CI 98% = 0.46

Table S4.5 HDX-MS data summary table for LptE in Re-LPS, Re-LPS + thanatin, and POPG states.

4.5 Materials and methods

4.5.1 Protein expression and purification

The plasmids used for over-expression of *K. pneumoniae* LptD and LptE were a kind gift from Susan Buchanan (NIH Bethesda). The mature LptD sequence (25-782) was inserted into a pET9 vector containing LIC cloning sites, a N-terminal PelB signal sequence, and a 10xHis affinity tag followed by a TEV protease cleavage site. The mature LptE sequence (20-196) was inserted into a pCDF-1b vector containing LIC cloning sites, and PelB signal sequence. All plasmids were amplified by transforming them into *E. coli* Stellar Competent Cells (Takara) and the DNA sequences were verified by Sanger sequencing.

Protein was expressed and purified adapting previously described protocols [13]. Amplified plasmids were then co-transformed in *E. coli* BL21(DE3) (New England Biolabs) in

order to co-express LptD and LptE. Colonies were inoculated into 100 ml LB media and grown overnight at 21 °C supplemented with 50 μg/ml kanamycin and 25 μg/ml streptomycin. One litre of LB in 2 litre shaker flasks was inoculated with 10 ml of overnight culture and grown at 21 °C without IPTG induction until the culture reached stationary phase. Cells were collected by centrifugation at 5,000xg for 10 min at 4 °C. Cell pellets were resuspended in buffer containing 200 mM NaCl, 25 mM Tris-HCl (pH 7.5) and stored at -80 °C.

Resuspended cells were thawed and supplemented with EDTA-free protease inhibitor cocktail (Roche) and 10 μg/ml DNaseI. The cell suspension was passed several times through an M-110 PS microfluidizer (Microfluidics) at 15,000 psi. Insoluble material was pelleted by centrifugation at 20,000xg for 20 min at 4 °C. The supernatant was was incubated with 2% (v/v) Triton X-100 for 30 min at 4°C and then ultracentrifuged at 200,000xg for 1 h to collect membrane fractions. Membranes were resuspended in ice-cold buffer containing 250 mM NaCl, 50 mM Tris-HCl (pH 7.5) and 20% glycerol. Resuspended membranes were used either directly or flash frozen in liquid nitrogen and stored at -80 °C.

The protein was solubilized from the membrane fraction with 250 mM NaCl, 25 mM Tris-HCl (pH 7.5), 20% glycerol, 5% (w/v) Elugent (Millipore Merck) for 16 h at 4°C. Insoluble material was removed by centrifugation at 20,000xg for 20 min at 4 °C. Supernatant was filtered before loading onto a 5 ml HisTrap-HP column (GE Healthcare, Piscataway, NJ) equilibrated in 200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol, 20 mM imidazole and 0.1% (w/v) dodecyl maltoside (DDM) (Anatrace). After the clarified supernatant was loaded, the column was initially washed with 50 ml of 200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol, 20 mM imidazole and 0.05% DDM, and washed again with 50 ml of 200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol, 80 mM imidazole and 0.05% DDM. The bound protein was eluted with 200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol, 300 mM imidazole and 0.03% DDM. The protein was concentrated to 2.5 ml an Amicon Ultra-15 concentrator unit (Millipore) with a molecular cut-off of 100 kDa and buffer exchanged to 200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol and 0.03% DDM using PD-10 desalting column. The sample was further

concentrated and loaded onto the Superdex 200 size exclusion chromatography (SEC) column in 200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol and 0.03% DDM. Protein concentration was measured using a DS-11 FX Spectrophotometer (DeNovix).

4.5.2 Lipids and peptides preparation

Unless stated otherwise, all lipids [Kdo_2-Lipid A (Re-LPS), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-1'-rac-glycerol (16:0/18:1 PG, POPG), 1',3'-bis[1,2-dioleoyl-sn-glycero-3-phospho]-glycerol (18:1 cardiolipin, CDL)] used in this study were obtained from Avanti Polar Lipids powders and stock solutions were prepared following a previously established method [32]. LPS from *K. pneumoniae* (Sigma-Aldrich) was diluted in H_2O to a stock concentration of 5 mg/ml. Thanatin was purchased from Biomatik and diluted in H_2O to a stock concentration of 2.5 mM. Lipids and peptides were diluted in 200 mM AmAc supplemented with C_8E_4 for nMS experiments or 200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol and 0.03% DDM for HDX-MS experiments.

4.5.3 Native MS experiments

Purified LptDE was buffer exchanged into MS Buffer (2x CMC of detergent of interest and 200 mM ammonium acetate) using a centrifugal buffer exchange device (Micro Bio-Spin 6, Bio-Rad) as previously described [32]. The best quality mass spectra were obtained using 0.5% C_8E_4 (Anatrace) as detergent. The freshly buffer-exchanged proteins were kept on ice, with protein concentration measured as before.

The protein samples were diluted as desired in 200 mM ammonium acetate buffer with detergent as necessary and loaded into a gold-coated capillary Clark borosilicate capillary (Harvard Apparatus) prepared in the laboratory [45]. The experiments were performed using a Q-Exactive UHMR Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Typically, 2 µl of buffer exchanged protein solution was electrosprayed from gold-plated borosilicate capillaries prepared in house. The instrument parameters for MS are: 1.2kV capillary

voltage, S-lens RF 100%, quadrupole selection from 1,000 to 20,000 *m/z* range, collisional activation in the HCD cell 200 V, nitrogen UHV pressure $8.0x10^{-10}$ mbar and 100°C capillary temperature. The ion transfer optics in positive mode was set as follows: injection flatapole 5 V, inter-flatapole lens 4 V, bent flatapole 2 V, transfer multipole 0 V. The resolution of the instrument was 17,500 at *m/z*=200 (transient time of 64 ms). The noise level was set at 3 rather than the default value of 4.64. Calibration of the instruments was performed using 10 mg/ml solution of caesium iodide in water. Where required, baseline subtraction was performed to achieve a better-quality mass spectrum. Data were analysed using the Xcalibur 3.0 (Thermo Scientific) and UniDec (www.unidec.chem.ox.ac.uk) [46] software packages. Theoretical and observed molecular masses of all species are described in Table 4.1.

Lipid binding experiments. LPS and Re-LPS binding experiments were performed by diluting the lipids in 200 mM ammonium acetate supplemented with 0.5% (w/v) C_8E_4. Figure S4.2 refers to addition of 10 μM LPS (considering average MW ~ 4.5 kDa) to 5 μM LptDE. In order to obtain the binding constant for the interaction between Re-LPS and LptDE, Re-LPS was added in increasing amounts while keeping the protein concentration constant. Peak intensities were extracted and the ratios of the intensity of the Re-LPS bound peak versus the total intensity of all observed species were calculated [47]. Average and standard deviation of these ratios from at least 5 charge states from three independent experiments were plotted against Re-LPS concentration. The data were fitted globally using GraphPad Prism 8.0 with the equation $Y=B_{max}*X/(X+K_D)$ where Bmax is the maximum specific binding in the same units as Y. Using our methods, we could not deduce the low-affinity K_D for the second Re-LPS binding as higher concentrations of Re-LPS binding to the protein were impractical to measure using our mass spectrometry method due to associated decreases in spectral quality, and the potential for favouring the capturing of non-specific binding events through the electrospray process.

Protein delipidation experiments were performed by adding 10 μM of the relevant lipid (Re-LPS, POPG, CDL) to 5 μM protein in 0.5% (w/v) C_8E_4 and supplementing the mixture with increasing

concentrations of n-nonyl-β-D-glucopyranoside (NG) (0, 0.5 and 1% final NG concentration) as described previously [48].

Thanatin binding experiments. Peptide binding experiments were performed by adding increasing concentration of thanatin to 5 μM LptDE. Effects of thanatin binding on Re-LPS binding were monitored by acquiring the mass spectra by adding a fixed amount of thanatin (1 μM) to the LptDE-Re-LPS complex obtained by adding increasing amount of Re-LPS (5 – 10 – 20 μM). The relative abundance of LptDE-Re-LPS species in the absence and in the presence of thanatin were plotted as a function of Re-LPS concentration.

All nMS experiments were performed three times and standard deviations are calculated from at least four charge states in these 3 datasets.

4.5.4 Hydrogen–deuterium exchange mass spectrometry (HDX-MS)

Prior to HDX labelling, purified LptDE was purified at a concentration of 20 μM and all ligands were diluted in equilibration buffer (200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% glycerol and 0.03% DDM). In the case of LPS binding experiments it was incubated with LPS in 15-fold excess (considering an average MW of 4.5 kDa). The same molar ratio was used for the experiments in the presence of Re-LPS and POPG. In the case of thanatin, the peptide was added in a 1:1 molar ratio. All ligands were incubated for 1h at 4°C to allow equilibration before HDX labelling. In order to study thanatin effect on substrate binding to LptDE, LPS or Re-LPS were added in 15-fold excess, the protein-substrate mixture was equilibrated for 1h followed by the addition of thanatin (1:1 LptDE:thanatin ratio) and further 1h incubation at 4 °C.

The HDX reaction was initiated by a 12× dilution into deuterated buffer (0.03% DDM, 200 mM NaCl, 25 mM Tris adjusted to pD 8 using DCl at 20 °C) resulting in a labelling solution ~92 % D_2O. The labelling reaction was incubated for a time course of 10 s (0.1667 min), 60 s (1 min), 1000 s (16.67 min), 5000 s (83.33 min) and 7 hrs (420 min) at 20 °C and quenched with equal volume of ice cold quench buffer (200 mM NaCl, 25 mM Tris-HCl (pH 1.9), 0.03% DDM, 15 mM TCEP).

The quenched sample was injected into a nanoACQUITY UPLC System (Waters) and online digested by Enzymate™ BEH Pepsin Column (2.1 x 30 mm, Waters) at 20 °C using an isocratic 0.1 % v/v formic acid-H_2O solution (200 µL/min).

Peptide products were collected on a BEH C18 trap column (1.7 µm, 2.1 x 5 mm, Waters) at 0 °C for 3 minutes for de-salting. Peptides were then eluted from the trap column on to a BEH C18 analytical column (1.7 µm, 1 x 100 mm, Waters) for separation using a reverse-phase gradient. The mobile phase consisted of a H_2O/MeCN (+0.1% formic acid v/v) linear gradient from 3% to 35% MeCN at a flow rate of 40 µL/min. Wash buffer (1.5 M guanidinium chloride pH 2.8, 4 % acetonitrile, 0.8% formic acid) was used to clean the pepsin column. To prevent peptide carry-over a wash and a blank injection runs were performed between each labelling experiment.

Following chromatographic separation, the peptides were injected in a hybrid ESI-Q-ToF mass spectrometer (Synapt G2-Si, Waters) for mass spectrometric analysis. Mass analysis was performed in in positive ion mode. MS/MS spectra were obtained using data-independent MS^E mode. MS conditions were optimised to avoid H/D exchange in the step-wave source region with water vapour leading to false EX1 kinetics [49]. MS conditions were as follows: capillary 2.8 kV, sample cone 30 V, source offset 30 V, trap activation 4 V, transfer activation 2 V. The source temperature was set to 80 °C and cone gas flow 80 L/hr, the desolvation temperature was 150 °C and the desolvation gas flow of 250 L/hr. Leucine Enkephalin was used as internal calibrant for mass accuracy correction and NaI was used for MS calibration. A summary for HDX experimental conditions is provided in tables S4.2, S4.3, S4.4 according to published guidelines [36].

4.5.5 HDX-MS data analysis

Peptides from un-deuterated samples were analysed and identified by Protein Lynx Global Server 3.0 (Waters). Peptide identifications were filtered using DynamX (Waters), according to the following parameters: minimum and maximum peptide sequence length of 5

and 30, respectively, minimum fragmentation products per amino acid of 0.15, minimum MS/MS products of 3, maximum $[MH]^+$ error of 5 ppm. Additionally, the peptides had to be identified in 50% of the acquired MS/MS files. Moreover, all the spectra were visually examined and only those with high signal to noise ratios were used for HDX-MS analysis.

The amount of relative deuterium uptake for each peptide was determined using DynamX 3.0 (Waters) and were not corrected for back exchange since analyses compared different states of the protein and there was no benefit from normalising the data. Confidence intervals (CI) and Woods plots were generated using Deuteros software [50]. In-house Python scripts were used to generate the uptake plots and to apply further statistical filters as described below.

In the comparative HDX analyses on different ligand-binding states, a peptide was considered to have a significant difference in HDX at a certain time point if this difference was greater than the calculated CI for two consecutive time points and statistically significant based on a Student's *t*-test (two-sided, unpaired, $p < 0.01$). As an exemption to this rule, a difference was also considered significant if the last time point (420 min) displayed a difference in HDX at least two times larger than the CI, since this may indicate an increasing difference not completely sampled by the HDX time course. HDX-MS experiments were performed in n = 2 biological replicate measurements obtained and n = 2-3 technical replicates. HDX results were mapped onto the LptDE crystal structure (PDB ID: 5IV9) using PyMOL (Schrödinger LLC). Only significant differences were mapped onto the protein structure.

Spectra displaying bimodal distributions following the EX1 kinetics were quantified using HX-Express 2.0 [51, 52]. In EX1 kinetics, residues undergo concerted opening-closing motions at much lower rates compared to the exchange reaction. The outcome is a bimodal isotopic distribution: the low-mass envelope represents those residues that have not yet undergone the conformational change required for exchange, while the high-mass population represents multiple backbone hydrogens that have already undergone the H-D exchange following the correlated conformational fluctuations [53, 54]. Therefore, some of these peptides were amenable

for quantitative analysis to extract kinetic constants, such as the rate of conformational opening (k_{op}) and the half-life of the closed state ($t_{1/2}$). The relative abundances of high- and low-mass populations were calculated as described previously [51, 52] and the high-mass population relative abundance was fitted to an exponential equation:

$$y = y_0 + (P - y_0)[1 - e^{(-k_{op}\, t)}]$$

where y is the fractional abundance of the high-mass population and t is the labelling time. The initial y intercept (y_0) accounts for the fact that all peptides showing EX1 or EXX kinetics start with an initial open population; the plateau parameter (P) is set to be 1 at maximum. In the case of thanatin and thanatin + LPS states, we could not extract kinetic values since the exponential function could not fit the data. This may be explained by the fact that the growth in the relative abundance of the high-mass population between early and late time points is incremental.

GraphPad Prism 8 software (GraphPad, San Diego) was used to extract the best-fit values for the kinetic parameters (k_{op} and the $t_{1/2}$ of the closed state).

4.5.6 Analytical SEC

LptDE samples were incubated at 20 °C in equilibration buffer (200 mM NaCl, 25 mM Tris-HCl (pH 7.5), 10% (v/v) glycerol and 0.03% (w/v) DDM). After 0, 60 or 420 min the samples were subjected to size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated with the equilibration buffer. UNICORN 7.0 (GE Healthcare) was used to process analytical SEC data.

4.5.7 Structure and sequence alignment

Sequence alignments were performed using the NPS@ClustalW [55, 56] web service and rendered using the ESPript 3.0 web service [43]. Structure alignment between LptD (PDB

ID: 5IV9) with apo-LptA (PDB ID: 2R1A) was performed using PyMOL and the TopMatch web service [57].

4.5.8 MD Modelling

The open Re-LPS state of LptDE. As there is not currently a structure that describes the open state for either the β-taco and the β-barrel of LptD we sought to produce an open-state model of *Kp*LptDE based on the current structure of its closed state (PDB ID: 5IV9). The hydrophobic groove of the β-taco domain of *Kp*LptD is too constricted to allow the binding of Re-LPS, however other structures (e.g. PDB ID: 4Q35 from *Shigella flexneri*) have shown that detergents, *N*,*N*-dimethyldodecylamine N-oxide (LDAO) and tetraethylene glycol monooctyl ether (C_8E_4) are able to bind to this site. We therefore used the detergents to guide the position of a Re-LPS molecule within the groove, with the six acyl tails binding internally, and the KDO sugars directed towards the solvent. The β-taco was opened by alchemically inserting one Re-LPS lipid within the hydrophobic groove by inverting the alchembed methodology to grow the lipid within the protein, rather than the protein within a lipid membrane [41]. This permitted the β-taco to gradually open while retaining the overall tertiary and secondary structures, until sufficiently wide (~2.0 nm across) to accommodate a Re-LPS molecule without steric clashes.

For the β-barrel domain, we initially considered previous simulations of the open-state of the β-barrel of LptDE from *Salmonella typhimurium*, where the lateral gate was opened by applying a negative pressure of -100 bar within the membrane plane [12]. This provides a range of states that describe the open lateral exit gate. We used these models to create a series of *Kp*LptDE models with different distances between the β1 and β26 strands. For our preliminary simulations of the open state, we chose a distance of 2.0 nm that would be sufficient to enable a single Re-LPS molecule to be laterally inserted through the exit gate. This is equivalent to the observed separation of β1 and β16 strands observed for the BAM machinery for OMP insertion.

Through structural alignments of the open state models for both β-taco and β-barrel with the closed state structure, it was noted that the extent of domain opening was similar. This is a

consequence of synchronized motions between domains given by the disulfide bonds (Cys7-Cys696 and Cys149-Cys697), in addition to the unzipping of the β1- β26 strands on interaction of the Re-LPS with the luminal gate [15].

4.5.9 Atomistic simulations

All simulations not requiring PLUMED were run using GROMACS 2019.1 [58]. Initially, all states of the LptDE complex were first converted to a coarse-grain (CG) representation using the Martinize python script [59] and individually embedded in a bilayer using INSANE [60]. The outer leaflet contained pure Re-LPS and the inner leaflet was composed of cardiolipin, POPG and POPE to a 1:2:7 ratio. Elastic bonds were applied between protein beads, between 5 and 10 Å apart, with a force of 1,000 kJ mol^{-1} nm^{-2}. The systems were solvated in a box with martini water and sodium and chloride ions for charge neutrality at a concentration of 0.15 M. Following this, systems were energy minimised using the steepest descents method to a target force of 1,000 kJ mol^{-1} nm^{-1}. CG simulations were run with the Martini 2.2 force field [59] for 100 ns using 20 fs time steps. The final snapshots of the 100 ns equilibration were converted back to atomistic (AT) resolution with the AT protein structure aligned with the CG coordinates within the equilibrated bilayer [61]. Upon AT conversion, systems were further equilibrated for 1 ns, with position restraints applied to the protein and described by the CHARMM36 forcefield [62]. Finally, production simulations were run for 300 ns without restraints (with 2 fs time steps) in explicit TIP3P water, using the V-rescale thermostat at 310 K and semi-isotropic Parrinello-Rahman pressure coupling at 1 atm [63, 64]. For each state of the LptDE complex five replicates with different initial seeds were run.

All simulations using PLUMED [42] were run with GROMACS 2018.6 [58]. For those states of the protein complex requiring an open β-barrel PLUMED upper and lower walls were applied to the putative lateral gate at a distance of 2.25 nm and 2.75 nm respectively, applied on the post 150 ns equilibrated LptDE open systems. These restraints were added to already equilibrated simulations, requiring only a further 50ns equilibration for convergence. These

constraints were imposed between the centre of mass of grouped Cα atoms from residues Asn208, Ala209, Lys210 and Tyr211 on β1 and Thr725, Phe726, Gly727 and Ile728 on β26. The wall restraint was set to a force constant of 10,000 kJ mol^{-1} with the default exponent, offset and rescaling factors.

4.5.10 Differential solvent uptake analysis

Solvent contacts based on a cutoff distance of 0.4 nm were calculated between the oxygen of the water and the nitrogen of the protein backbone using a modified MDAnalysis contact script, as a proxy for HDX [65-67]. The solvent contact data from apo and ligand-bound states of the LptDE complex were averaged for each residue between replicates for both states, with the averaged ligand-bound state subtracted from the averaged apo state. This produces an average differential solvent contact value for each residue. The confidence intervals are calculated using a two tailed test using a 95% significance level in much the same way as the Deuteros software [50]. Significant residues are extracted and mapped onto the structure as B-factors, then subsequently visualised in PyMOL. A positive difference is coloured red to represent 'deprotected' residues and a negative difference coloured blue to represent 'protected' residues.

Once the average differential solvent contact value for each residue was calculated, these values were averaged over the same peptide coverage map as obtained by the HDX experiments described above. For simplicity, 'cleavage sites' are recognised only when there is no overlap and areas which are not resolved by experiment are not considered for the analysis. Calculating confidence intervals and mapping peptides considered to be statistically significant is conducted in the same manner as the above, where a positive difference represents 'deprotected' peptides and a negative difference is used to model for 'protected' peptides accordingly.

4.6 References

1. Whitfield, C. and M.S. Trent, *Biosynthesis and export of bacterial lipopolysaccharides*. Annu. Rev. Biochem., 2014. **83**: p. 99-128.
2. Okuda, S., et al., *Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model*. Nat. Rev. Microbiol., 2016. **14**(6): p. 337-45.
3. Sperandeo, P., et al., *Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of Escherichia coli*. J. Bacteriol., 2008. **190**(13): p. 4460-9.
4. Luo, Q., et al., *Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG*. Nat. Struct. Mol. Biol., 2017. **24**(5): p. 469-474.
5. Li, Y., B.J. Orlando, and M. Liao, *Structural basis of lipopolysaccharide extraction by the LptB2FGC complex*. Nature, 2019. **567**(7749): p. 486-490.
6. Tang, X., et al., *Cryo-EM structures of lipopolysaccharide transporter LptB2FGC in lipopolysaccharide or AMP-PNP-bound states reveal its transport mechanism*. Nat Commun, 2019. **10**(1): p. 4175.
7. Owens, T.W., et al., *Structural basis of unidirectional export of lipopolysaccharide to the cell surface*. Nature, 2019. **567**(7749): p. 550-553.
8. Villa, R., et al., *The Escherichia coli Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains*. J. Bacteriol., 2013. **195**(5): p. 1100-8.
9. Chng, S.S., et al., *Characterization of the two-protein complex in Escherichia coli responsible for lipopolysaccharide assembly at the outer membrane*. Proc Natl Acad Sci U S A, 2010. **107**(12): p. 5363-8.
10. Freinkman, E., S.S. Chng, and D. Kahne, *The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel*. Proc Natl Acad Sci U S A, 2011. **108**(6): p. 2486-91.
11. Qiao, S., et al., *Structural basis for lipopolysaccharide insertion in the bacterial outer membrane*. Nature, 2014. **511**(7507): p. 108-111.
12. Dong, H., et al., *Structural basis for outer membrane lipopolysaccharide insertion*. Nature, 2014. **511**(7507): p. 52-56.
13. Botos, I., et al., *Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens*. Structure, 2016. **24**(6): p. 965-976.
14. Li, X., et al., *Trapped lipopolysaccharide and LptD intermediates reveal lipopolysaccharide translocation steps across the Escherichia coli outer membrane*. Sci Rep, 2015. **5**: p. 11883.

15. Gu, Y., et al., *Lipopolysaccharide is inserted into the outer membrane through an intramembrane hole, a lumen gate, and the lateral opening of LptD*. Structure, 2015. **23**(3): p. 496-504.
16. Nikaido, H., *Molecular basis of bacterial outer membrane permeability revisited*. Microbiol. Mol. Biol. Rev., 2003. **67**(4): p. 593-656.
17. Ruiz, N., D. Kahne, and T.J. Silhavy, *Transport of lipopolysaccharide across the cell envelope: the long road of discovery*. Nat. Rev. Microbiol., 2009. **7**(9): p. 677-83.
18. Srinivas, N., et al., *Peptidomimetic Antibiotics Target Outer-Membrane Biogenesis in Pseudomonas aeruginosa*. Science, 2010. **327**(5968): p. 1010-1013.
19. Andolina, G., et al., *A Peptidomimetic Antibiotic Interacts with the Periplasmic Domain of LptD from Pseudomonas aeruginosa*. ACS Chem Biol, 2018. **13**(3): p. 666-675.
20. Werneburg, M., et al., *Inhibition of lipopolysaccharide transport to the outer membrane in Pseudomonas aeruginosa by peptidomimetic antibiotics*. Chembiochem, 2012. **13**(12): p. 1767-1775.
21. Fehlbaum, P., et al., *Structure-activity analysis of thanatin, a 21-residue inducible insect defense peptide with sequence homology to frog skin antimicrobial peptides*. Proc Natl Acad Sci USA, 1996. **93**(3): p. 1221-1225.
22. Vetterli, S.U., et al., *Thanatin targets the intermembrane protein complex required for lipopolysaccharide transport in Escherichia coli*. Sci Adv, 2018. **4**(11): p. eaau2634.
23. Gault, J., et al., *High-resolution mass spectrometry of small molecules bound to membrane proteins*. Nat. Methods, 2016. **13**(4): p. 333-336.
24. Bolla, J.R., et al., *Direct observation of the influence of cardiolipin and antibiotics on lipid II binding to MurJ*. Nat. Chem., 2018. **10**(3): p. 363-371.
25. Fiorentino, F., et al., *The Different Effects of Substrates and Nucleotides on the Complex Formation of ABC Transporters*. Structure, 2019. **27**(4): p. 651-659.
26. Yen, H.Y., et al., *PtdIns(4,5)P2 stabilizes active states of GPCRs and enhances selectivity of G-protein coupling*. Nature, 2018. **559**(7714): p. 423-427.
27. Martens, C., et al., *Direct protein-lipid interactions shape the conformational landscape of secondary transporters*. Nat Commun, 2018. **9**(1): p. 4151.
28. Nielsen, A.K., et al., *Substrate-induced conformational dynamics of the dopamine transporter*. Nat Commun, 2019. **10**(1): p. 2714.
29. Landreh, M., et al., *Integrating mass spectrometry with MD simulations reveals the role of lipids in Na(+)/H(+) antiporters*. Nature communications, 2017. **8**: p. 13993-13993.
30. Skinner, J.J., et al., *Benchmarking all-atom simulations using hydrogen exchange*. Proc Natl Acad Sci U S A, 2014. **111**(45): p. 15975-80.

31. Persson, F. and B. Halle, *How amide hydrogens exchange in native proteins*. Proceedings of the National Academy of Sciences, 2015. **112**(33): p. 10383.
32. Laganowsky, A., et al., *Mass spectrometry of intact membrane protein complexes*. Nat. Protoc., 2013. **8**(4): p. 639-651.
33. Raetz, C.R., et al., *Kdo2-Lipid A of Escherichia coli, a defined endotoxin that activates macrophages via TLR-4*. J. Lipid. Res., 2006. **47**(5): p. 1097-1111.
34. Xie, R., R.J. Taylor, and D. Kahne, *Outer Membrane Translocon Communicates with Inner Membrane ATPase To Stop Lipopolysaccharide Transport*. J Am Chem Soc, 2018. **140**(40): p. 12691-12694.
35. Bolla, J.R., et al., *A mass spectrometry-based approach to distinguish annular and specific lipid binding to membrane proteins*. Angew. Chem. Int. Ed. Engl., 2019.
36. Masson, G.R., et al., *Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments*. Nat. Methods, 2019. **16**(7): p. 595-602.
37. Lundquist, K.P. and J.C. Gumbart, *Presence of substrate aids lateral gate separation in LptD*. Biochim Biophys Acta Biomembr, 2020. **1862**(1): p. 183025.
38. Merkle, P.S., et al., *Substrate-modulated unwinding of transmembrane helices in the NSS transporter LeuT*. Sci. Adv., 2018. **4**(5): p. eaar6179.
39. Zhou, J., et al., *Conformational dynamics of 1-deoxy-d-xylulose 5-phosphate synthase on ligand binding revealed by H/D exchange MS*. Proc Natl Acad Sci USA, 2017. **114**(35): p. 9355-9360.
40. Wang, C., et al., *Understanding lipopolysaccharide aggregation and its influence on activation of Factor C*. J Colloid Interf Sci, 2019. **552**: p. 540-553
41. Jefferys, E., et al., *Alchembed: A Computational Method for Incorporating Multiple Proteins into Complex Lipid Geometries*. J Chem Theory Comput, 2015. **11**(6): p. 2743-2754.
42. Bonomi, M., et al., *PLUMED: A portable plugin for free-energy calculations with molecular dynamics*. Computer Physics Communications, 2009. **180**(10): p. 1961-1972.
43. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server*. Nucleic Acids Res., 2014. **42**(Web Server issue): p. W320-324.
44. Tomasek, D., et al., *Structure of a nascent membrane protein as it folds on the BAM complex*. Nature, 2020. **583**(7816): p. 473–478
45. Hernandez, H. and C.V. Robinson, *Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry*. Nat. Protoc., 2007. **2**(3): p. 715-726.

46. Marty, M.T., et al., *Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles*. Anal. Chem., 2015. **87**(8): p. 4370-4376.
47. Cubrilovic, D., et al., *Determination of protein-ligand binding constants of a cooperatively regulated tetrameric enzyme using electrospray mass spectrometry*. ACS Chem. Biol., 2014. **9**(1): p. 218-226.
48. Bolla, J.R., et al., *A Mass-Spectrometry-Based Approach to Distinguish Annular and Specific Lipid Binding to Membrane Proteins*. Angew. Chem. Int. Ed. Engl., 2020. **59**(9): p. 3523-3528.
49. Guttman, M., et al., *Tuning a High Transmission Ion Guide to Prevent Gas-Phase Proton Exchange During H/D Exchange MS Analysis*. J. Am. Soc. Mass Spectrom., 2016. **27**(4): p. 662-668.
50. Lau, A.M.C., et al., *Deuteros: software for rapid analysis and visualization of data from differential hydrogen deuterium exchange-mass spectrometry*. Bioinformatics, 2019. **35**(17): p. 3171-3173.
51. Weis, D.D., J.R. Engen, and I.J. Kass, *Semi-automated data processing of hydrogen exchange mass spectra using HX-Express*. J. Am. Soc. Mass. Spectrom., 2006. **17**(12): p. 1700-1703.
52. Guttman, M., et al., *Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra*. J. Am. Soc. Mass Spectrom., 2013. **24**(12): p. 1906-1912.
53. Weis, D.D., et al., *Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis*. J Am Soc Mass Spectrom, 2006. **17**(11): p. 1498-1509.
54. Konermann, L., J. Pan, and Y.H. Liu, *Hydrogen exchange mass spectrometry for studying protein structure and dynamics*. Chem Soc Rev, 2011. **40**(3): p. 1224-34.
55. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic Acids Res., 1994. **22**(22): p. 4673-4680.
56. Combet, C., et al., *NPS@: network protein sequence analysis*. Trends Biochem. Sci., 2000. **25**(3): p. 147-150.
57. Sippl, M.J. and M. Wiederstein, *Detection of spatial correlations in protein structures and molecular complexes*. Structure, 2012. **20**(4): p. 718-728.
58. Abraham, M.J., et al., *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*. SoftwareX, 2015. **1-2**: p. 19-25.
59. Monticelli, L., et al., *The MARTINI Coarse-Grained Force Field: Extension to Proteins*. Journal of Chemical Theory and Computation, 2008. **4**(5): p. 819-834.

60. Wassenaar, T.A., et al., *Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations*. Journal of Chemical Theory and Computation, 2015. **11**(5): p. 2144-2155.

61. Stansfeld, P.J. and M.S.P. Sansom, *From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations*. Journal of Chemical Theory and Computation, 2011. **7**(4): p. 1157-1166.

62. Best, R.B., et al., *Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles*. Journal of chemical theory and computation, 2012. **8**(9): p. 3257-3273.

63. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling*. The Journal of Chemical Physics, 2007. **126**(1): p. 014101.

64. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*. Journal of Applied Physics, 1981. **52**(12): p. 7182-7190.

65. Newport, T.D., M.S P. Sansom, and P.J. Stansfeld, *The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions*. Nucleic Acids Research, 2018. **47**(D1): p. D390-D397.

66. Michaud-Agrawal, N., et al., *MDAnalysis: A toolkit for the analysis of molecular dynamics simulations*. Journal of Computational Chemistry, 2011. **32**(10): p. 2319-2327.

67. R. J. Gowers, M.L., J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, and O. Beckstein., *MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations*. Proceedings of the 15th Python in Science Conference, 2016: p. 98-105.

CHAPTER 5

Structural basis of the ER membrane-associated steps in tail anchored protein insertion by the TRC40 (Get3)-insertase complex

Adapted from a paper accepted in *Molecular Cell*

Melanie A. McDowell*[1], Michael Heimes[1], Francesco Fiorentino[2], Shahid Mehmood[2#], Ákos Farkas[3], Javier Coy-Vergara[3], Di Wu[2], Jani Reddy Bolla[2], Volker Schmid[3], Roger Heinze[1], Klemens Wild[1], Dirk Flemming[1], Stefan Pfeffer[4], Blanche Schwappach[3], Carol V. Robinson[2], Irmgard Sinning*[1,5]

[1]Heidelberg University Biochemistry Centre (BZH), Im Neuenheimer Feld 328, 69120 Heidelberg, Germany

[2]Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK

[3]Department of Molecular Biology, University Medical Center Göttingen, 37073 Göttingen, Germany

[4]Centre for Molecular Biology of Heidelberg University (ZMBH), Im Neuenheimer Feld 282, 69120 Heidelberg, Germany

[5]Lead contact

*Corresponding authors: irmi.sinning@bzh.uni-heidelberg.de and melanie.mcdowell@bzh.uni-heidelberg.de

[#]Present address: The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

Acknowledgements

We thank G. Stier for the TRC40 construct, M. Thoms and E. Hurt for the pMT929 vector and V. Denic for the *S. cerevisiae Δget3* strain. We are grateful to J. Sloan and K. Lapouge for careful reading and useful discussion of the manuscript. Cryo-EM data were collected at the University of Heidelberg with support from G. Hofhaus and at the European Molecular Biology Laboratory (EMBL), Heidelberg with F. Weiss. We acknowledge the data storage service SDS@hd and bwHPC supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) and the Deutsche Forschungsgemeinschaft (DFG) through grants INST 35/1314-1 FUGG and INST 35/1134-1 FUGG. This work was supported by an EMBO Long Term Fellowship (ALTF 1230-2013) to M.A.M. and by the DFG through the Leibniz Programme (SI 586/6-1) and TRR83 (TP22) to I.S, who is an investigator of the Cluster of Excellence:CellNetworks. C.V.R holds a Medical Research Council program grant (MR/N020413/1), an ERC Advanced Grant ENABLE (695511) and a Wellcome Trust Investigator Award (104633/Z/14/Z). F.F. holds a SABS CDT Scholarship supported by the Engineering and Physical Sciences Research Council and Medical Research Council grant (EP/L016044/1). A.F. and B.S. were funded by the DFG through SFB 1190 (P04) and under Germany's Excellence Strategy (EXC 2067/1-390729940).

Author contributions

M.A.M., M.H. and R.H. purified protein samples. M.A.M., D.F. and S.P. collected EM data. M.A.M., M.H., K.W. and S.P. processed EM data and built the structural models. F.F., S.M., J.R.B and C.V.R performed native mass spectrometry and interpreted the data. F.F. and D.W. performed and analysed the lipidomics data. A. F., J. C-V., V. S. and B. S. performed and analysed functional assays in yeast. M.A.M., M.H., K.W. and I.S. interpreted the data. M.A.M., F.F. and I.S. wrote the manuscript with contributions from all authors. M.A.M. and I.S. conceived the study and designed the research.

Declaration of Interests

The authors declare no competing interests

Abstract

Membrane protein biogenesis faces the challenge of chaperoning hydrophobic transmembrane helices for faithful membrane insertion. The guided entry of tail-anchored proteins (GET) pathway targets and inserts tail-anchored (TA) proteins into the ER membrane with an insertase (yeast Get1/Get2 and mammalian WRB/CAML) somehow capturing the TA from the Get3 or TRC40 cytoplasmic chaperone. Here we present cryo-electron microscopy reconstructions, native mass spectrometry and structure-based mutagenesis of human WRB/CAML/TRC40 and yeast Get1/Get2/Get3 complexes. Get3 binding to the membrane insertase supports heterotetramer formation, and phosphatidylinositol binding at the heterotetramer interface stabilises the insertase for efficient TA insertion *in vivo*. Structural homology with YidC and the EMC implicates an evolutionarily conserved insertion mechanism for divergent substrates involving a hydrophilic groove. We identify an unanticipated, conserved Get2/CAML cytoplasmic helix that forms a 'gating' interaction with Get3/TRC40 and assists in TA insertion. Thus, we provide a detailed structural and mechanistic framework for TA insertion.

5.1 Introduction

Around a third of proteins translated by eukaryotic ribosomes are integral membrane proteins. These all contain hydrophobic transmembrane domains (TMDs) that must be shielded from the cytoplasm and inserted into the lipid bilayer. An important class is the tail-anchored (TA) proteins, which mediate a diverse range of essential activities in almost every cellular membrane, including vesicular trafficking (N¬-ethylmaleimide-sensitive factor attachment receptors (SNARES)), apoptosis (Bcl-2 family members) and protein translocation (Sec61βγ) [1]. TA proteins are characterised by a single TMD at their C-terminus, which only becomes accessible to the cytoplasm after translation termination, necessitating post-translational recognition and targeting pathways for their membrane integration [2]. The most extensively studied pathway for targeting and insertion of endoplasmic reticulum (ER)-destined TA proteins is the yeast guided entry of tail-anchored proteins (GET) pathway (transmembrane recognition complex (TRC) pathway in mammals) [3, 4]. More recently it was found that this widely conserved pathway works in parallel with the SRP-independent (SND) and ER membrane protein complex (EMC) pathways to mediate TA protein biogenesis at the ER [5, 6].

Within the GET/TRC pathway, the central targeting factor is the yeast Get3 or mammalian TRC40 homodimeric ATPase, which binds and releases the TA protein through nucleotide-dependent conformational changes [7-11]. In the cytoplasm, the TA protein interacts with the closed ADP-Pi bound form of Get3, which shields the substrate TMD within a hydrophobic groove [12]. This complex is subsequently targeted to the ER membrane through interactions with the cytoplasmic domains (CDs) of the membrane complex, Get1/Get2 in yeast [13, 14] and WRB/CAML in mammals [15, 16]. In the current membrane targeting model, the ~100 amino acid, flexible CD of Get2 initially captures the Get3/TA protein complex from the cytoplasm. Subsequent Get1-CD binding is then expected to drive TA release by bringing the targeting complex closer to the membrane and stabilising the open form of the Get3 dimer, leading to nucleotide release and disruption of the substrate binding site [10, 14, 17, 18].

Get1/Get2 and WRB/CAML form the minimal machinery required for TA protein membrane insertion [14, 16, 19, 20]. Whilst their TMDs provide a membrane-integrated docking site for the TA substrate [21], it is unknown how Get1/Get2 acts as a membrane insertase in the absence of a structure of the complex. As members of the Oxa1 superfamily, Get1 and WRB are predicted to share a three-TMD core structure with other proteins involved in membrane protein biogenesis, including the EMC3 subunit of the EMC and the bacterial YidC insertase [22]. Indeed, Get1-CD forms a coiled-coil between the first two TMDs analogous to that present between the equivalent TMDs in structures of EMC3 and YidC [17, 18, 23-26]. However, whether Get1/Get2 recapitulates the insertion mechanisms of these distant homologues is yet to be ascertained. Furthermore, although predicted to share the same three-TMD overall topology, Get2 and CAML have no sequence homology to each other or other proteins [16], making structural predictions challenging.

Whilst Get1/Get2 forms a stoichiometric complex [14, 21], it is unclear whether this complex acts as a 1:1 heterodimer or 2:2 heterotetramer *in vivo*. Given that the Get3 homodimer presents two binding sites for each Get1 or Get2 CD, the quaternary structure of the insertase dictates the stoichiometries that are possible for the interaction between the CDs and the targeting complex, leading to two published models for TA targeting and release [17, 18, 20]. In the 'heterodimer asymmetric binding' model, Get1/Get2 CDs bind simultaneously to opposite halves of a single Get3 dimer (Fig. 5.1a), as implicated by single-molecule FRET efficiencies within Get1/Get2 heterodimers [20]. This study further corroborated their model by showing that Get1/Get2 reconstituted into lipid bilayers *in vitro* remained heterodimeric even in the presence of Get3 and that this minimal complex was sufficient for TA insertion. Alternatively, in the 'heterotetramer symmetric binding' model Get3 sequentially binds to a pair of Get2 then Get1 CDs (Fig. 5.1b). This scenario was directly visualised in high-resolution structures showing two copies of either Get1- or Get2-CD in isolation bound to their symmetry-related sites within the Get3 dimer [17, 18, 27]. Intrinsic to this model is the formation of a Get1/Get2 membrane

heterotetramer, however this has not been observed either constitutively or dynamically upon Get3 binding.

A molecular-level characterisation of Get1/Get2 and WRB/CAML complexes is required to understand the mechanisms associated with Get3 binding, TA release and membrane insertion. Here we present cryo-electron microscopy (cryo-EM) reconstructions and native mass spectrometry (nMS) of human WRB/CAML/TRC40 and *S. cerevisiae* Get1/Get2/Get3 complexes. Our data show that the membrane insertase is a heterotetramer, which is formed upon Get3 binding. We further identify positively charged residues within CAML/Get2 TMD1 required for phosphatidylinositol binding at the oligomer interface and show they are important for heterotetramer formation and insertase function *in vivo*. Our WRB/CAML reconstruction reveals homology with the EMC and YidC insertases, including a hydrophilic groove that likely allows the TA substrate to traverse the bilayer. We also identify a functionally important cytoplasmic helix α3' in CAML involved in a novel 'gating' interaction with TRC40 and propose that helix α3' guides the TA substrate towards the membrane for insertion. We therefore provide a new model for TA targeting and insertion by Get1/Get2 homologues.

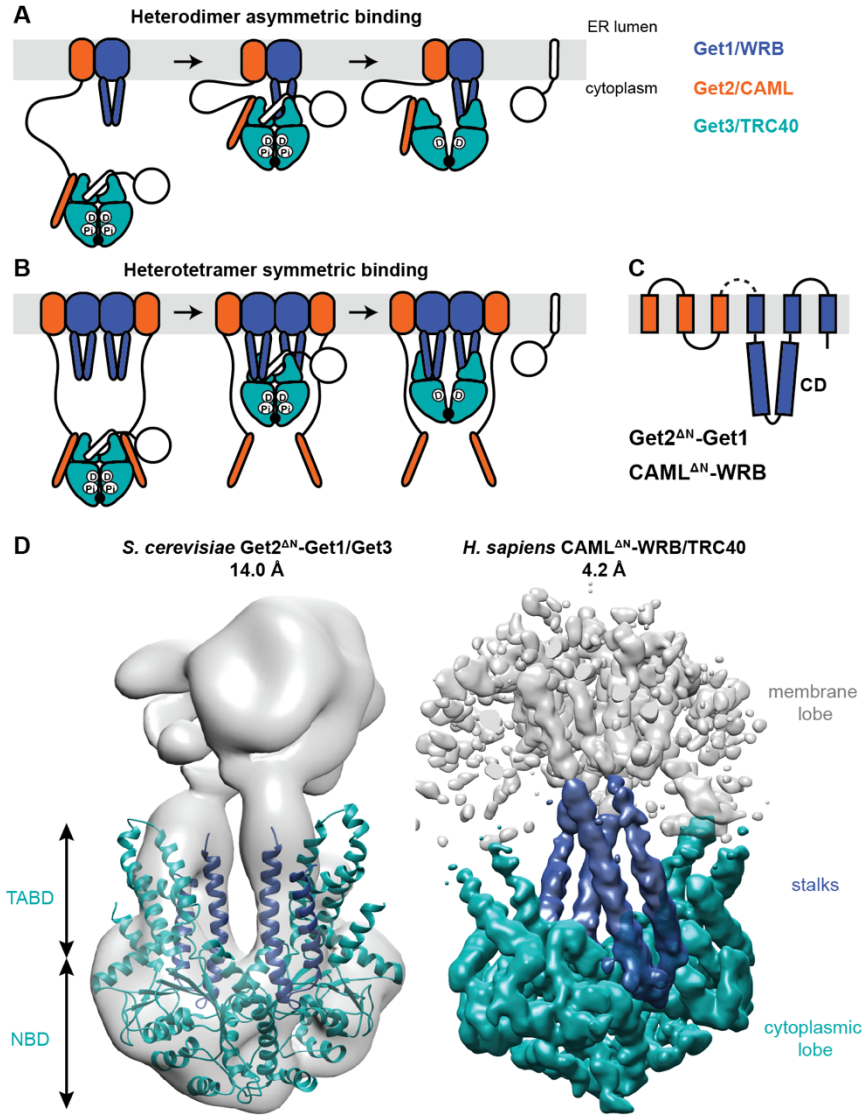


Figure 5.1 The Get1/Get2 insertase forms an evolutionarily conserved heterotetramer. (A, B) Possible models for Get3 or TRC40 binding and TA protein insertion by Get1/Get2 or WRB/CAML oligomers. (c) Topology of Get2ΔN-Get1 and CAMLΔN-WRB single-chain constructs used in this study. Dashed lines represent the non-native linker sequence and the Get1 CD is shown. (d) Cryo-EM reconstructions for *S. cerevisiae* Get2ΔN-Get1/Get3 and *H. sapiens* CAMLΔN-WRB/TRC40 complexes. The *sc*Get1 (blue)/Get3 (cyan) open state structure (PDB accession 3SJA) is docked in the *S. cerevisiae* density (CC=0.92) and the equivalent regions of the *H. sapiens* density are coloured. The approximate positions of the tail anchor binding domains (TABD) and nucleotide binding domains (NBD) within the Get3 dimer are shown. See also Figs. S5.1 and S5.2 and Table S5.2.

5.2 Results

5.2.1 The Get1/Get2 insertase forms an evolutionarily conserved heterotetramer

Starting with a functional *Saccharomyces cerevisiae* Get2-Get1 fusion [21], we engineered a single-chain construct (*sc*Get2$^{\Delta N}$-Get1) lacking the flexible N-terminal CD of Get2 to make it more amenable to *in vitro* structural characterisation (Fig. 5.1c). This fusion (henceforth referred to as a heterodimer) was overexpressed in *S. cerevisiae* and co-purified with apo *sc*Get3 expressed in *Escherichia coli*, whilst an analogous *Homo sapiens* CAML$^{\Delta N}$-WRB construct was purified from *Spodoptera frugiperda* Sf9 membranes and complexed with *E. coli* expressed TRC40 (Fig. S5.1A). We analysed the structures of these *sc*Get2$^{\Delta N}$-Get1/Get3 and CAML$^{\Delta N}$-WRB/TRC40 complexes in various amphipols using single-particle cryo-EM (Figs. 5.1d, S5.1b-d and S5.2). The overall resolution as defined by Fourier shell correlation (FSC) was much higher for the 150 kDa *H. sapiens* complex in PMAL-C8 amphipol (4.2 Å) than for the 174 kDa *S. cerevisiae* complex in A835 (14.0 Å; Fig. S5.1e). However, both reconstructions had a similar overall morphology, showing major cytoplasmic and membrane lobes connected by stalks of density (Fig. 5.1d). Docking of a high-resolution structure of *sc*Get3/Get1-CD [18] into the *S. cerevisiae* density revealed that the cytoplasmic lobe represents Get3 in the fully open conformation, indicating that the captured state represents the complex immediately following TA release and membrane insertion. Most interestingly, a pair of Get1-CDs accounts for the connecting stalks in these reconstructions, showing that both *sc*Get1/Get2 and WRB/CAML are present as a heterotetramer.

5.2.2 Get3 symmetric binding promotes heterotetramerisation of the Get1/Get2 insertase

Our cryo-EM structures show the symmetric binding of Get3 homologues by two Get1-CDs that was previously observed in high-resolution structures [17, 18, 27] (Fig. 5.1d). Indeed, nMS confirmed that *sc*Get2$^{\Delta N}$-Get1 or CAML$^{\Delta N}$-WRB purified in different mild detergents bind predominantly as heterotetramers to one *sc*Get3 or TRC40 homodimer respectively (Figs. 5.2a, 5.2b, S5.3a and S5.3b). Similarly, full-length *sc*Get1/Get2 (Fig. S5.3c) with N72A/R73A

mutations in *sc*Get1-CD that abolish *sc*Get3 binding [18] (*sc*Get1$^{N72A/R73A}$/Get2), exclusively formed complexes with a 2:2 stoichiometry for Get2:Get3 (Fig. S5.3d). Notably, these Get3-associated Get1/Get2 variants formed heterotetramers in detergents where the isolated membrane insertase was heterodimeric (Figs. S5.3f). Therefore, interaction of either the Get1- or Get2-CDs with dimeric Get3 induces stable formation of Get1/Get2 heterotetramers *in vitro*.

Whilst these data confirm the 'heterotetramer symmetric binding' model (Fig. 5.1b) for Get3 recruitment in the presence of only Get1- or Get2-CDs, 'heterodimer asymmetric binding' (Fig. 5.1a) could still be preferable when both CDs are presented simultaneously by the full-length proteins. However, we found that wild-type *sc*Get1/Get2 also heterotetramerised upon addition of *sc*Get3 (Fig. 5.2c and S5.3e). This indicates that CDs from distinct *sc*Get1/Get2 heterodimers contribute to Get3 binding, as would only be implicit to the 'heterotetramer symmetric binding' model. Interestingly, *sc*Get1/Get2 heterotetramers complexed with Get3 exclusively had a 2:2:4 stoichiometry (Figs. 5.2c and S5.3e), indicating for the first time that in the context of the full-length insertase the CDs are appropriately positioned to sterically accommodate two *sc*Get3 dimers simultaneously.

The observed Get3-induced heterotetramerisation is at odds with reports that *sc*Get1/Get2 in a reconstituted lipid bilayer remains heterodimeric even in the presence of *sc*Get3 [20]. Indeed, we found that *sc*Get2$^{\Delta N}$-Get1 was predominantly heterodimeric when purified in its native lipid environment within styrene maleic acid lipid particles (SMALPs; Fig. 5.2d), indicating that interactions between *sc*Get1/Get2 heterodimers, which result in heterotetramers, are transient in a lipid bilayer. To investigate the stoichiometry of endogenous insertase complexes, ER-enriched membranes (microsomes) were isolated from *S. cerevisiae*. Subsequent blue native polyacrylamide gel electrophoresis (BN-PAGE) and western blot detection of *sc*Get3 demonstrated that this cytosolic protein is mostly found within two membrane-associated complexes (Fig. 5.2e). Whilst the smaller complex is presumably the 78 kDa *sc*Get3 homodimer, we hypothesised that the ~200 kDa population is likely to represent a complex between *sc*Get3 and *sc*Get1/Get2 (Fig. 5.2e 'T'). Indeed, when microsomes were isolated from a *Δget1* strain

expressing a 4PC-tagged version of *sc*Get1 this complex had an increased molecular weight and was specifically detected by an anti-PC antibody (Fig. 5.2e 'T^{4PC}'), confirming that *sc*Get1 is an integral component. Given the relative sizes of a *sc*Get3-associated heterodimeric (136 kDa) versus heterotetrameric (194 kDa) insertase, it is most likely the isolated complex contains *sc*Get1/Get2/Get3 in a 2:2:2 stoichiometry. Therefore, *sc*Get3 associates with heterotetramers of the *sc*Get1/Get2 insertase in their native membrane environment. Notably, the predominant 2:2:4 species (272 kDa) observed in nMS experiments was not isolated from microsomes, suggesting either that the additional *sc*Get3 dimer had dissociated during sample preparation (possibly accounting for the membrane-associated population of *sc*Get3 homodimer in Fig. 5.2e) or that this complex does not form *in vivo*.

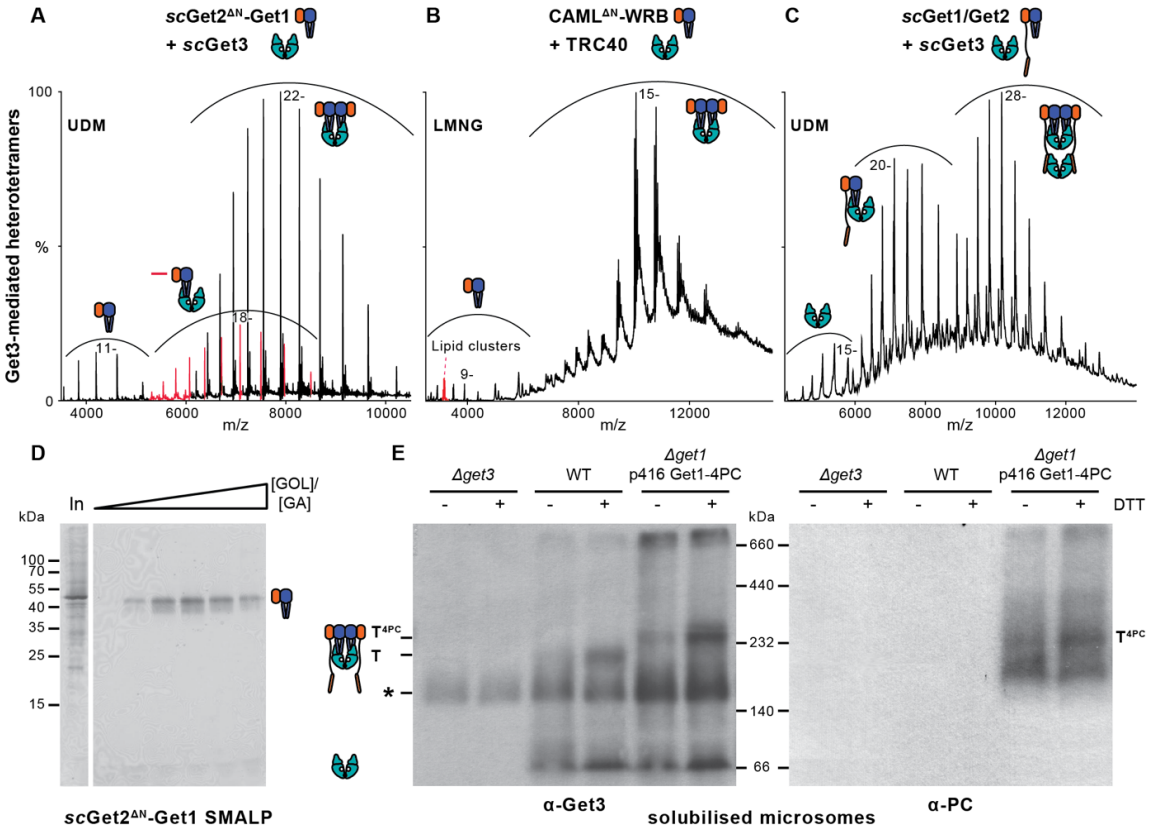


Figure 5.2 Get3 symmetric binding promotes heterotetramerisation of the Get1/Get2 insertase. (A-C) Mass spectra of stable *sc*Get2$^{\Delta N}$-Get1/Get3, CAML$^{\Delta N}$-WRB/TRC40 and *sc*Get1/Get2/Get3 oligomers liberated from detergent micelles collected under non-denaturing conditions. For each assigned complex the charge state for the most intense peak is shown, with the sign denoting data collection in either positive or negative mode. Liberation from the

specified detergents results in stable heterotetramers in the gas phase only upon complex formation with *sc*Get3/TRC40. Heterodimers observed in the absence of *sc*Get3/TRC40 are shown in Fig. S5.3f. (d) GraFix cross-linking of *sc*Get2$^{\Delta N}$-Get1 isolated directly from *S. cerevisiae* membranes within SMALPs. 12.5% silver-stained SDS-PAGE gel shows the input (In) and fractions taken from an increasing gradient of glycerol (GOL) and glutaraldehyde (GA). (e) Anti-Get3 (α-Get3) and anti-protein C tag (α-PC) western blots after separation of microsomal proteins on a 5-13% BN-PAGE gel, as previously shown [28]. Microsomes were isolated from *S. cerevisiae Δget3*, wild-type (WT) and *Δget1* expressing Get1-4PC from a plasmid, and solubilised with 1% digitonin ± 100 mM DTT prior to loading. Bands corresponding to the putative 2:2:2 complexes of Get1/Get2/Get3 (T) and Get1-4PC/Get2/Get3 (T^{4PC}) are indicated, along with a species non-specifically recognised by α-Get3 (*). See also Figs. S5.3 and Table S5.1.

5.2.3 The structure of CAML$^{\Delta N}$-WRB/TRC40 reveals a YidC-like fold in the membrane

To gain insights into the structure of these heterotetramers, the CAML$^{\Delta N}$-WRB/TRC40 reconstruction was further analysed. The membrane region shows clear density for 12 TMDs, adopting a concave arrangement towards the cytoplasm (Figure 5.3a). These can be divided into two groups of six TMDs related by C2 pseudo-symmetry, each corresponding to one copy of CAML$^{\Delta N}$-WRB. Homology models for a TRC40 dimer and the WRB-CD based on the crystal structure of *sc*Get3/Get1-CD could be rigid-body docked into the cytoplasmic density, allowing TMD1/TMD2 of WRB to be modelled based on helical extension of the cytoplasmic coiled-coil (Fig. 5.3b). Although both TMDs form a continuous helix with the WRB-CD, the connection to TMD2 is straight, whilst TMD1 is connected at the cytoplasmic membrane face by a ~20 Å amphipathic helix (AH) tilted at 45° relative to the membrane normal (Fig. 5.3b and S5.4a). In order to identify the remaining TMDs, we first compared our TMD arrangement to that of structurally characterised members of the Oxa1 superfamily. Superimposition of WRB TMD1/TMD2 with *E. coli* YidC [24] TMD2/TMD3 revealed a helix in our cryo-EM density

coincident with TMD6, which we designated as TMD3 of WRB in agreement with previous bioinformatics [22] (Figs. 5.3b and 5.3c). CAML therefore accounts for the outstanding TMDs, which are arranged as a three-helical bundle. Density for the connecting luminal and cytoplasmic loops was more clearly present in a 6.5 Å low-pass filtered reconstruction (Fig. S5.2a), allowing us to assign CAML TMD1, 2 and 3 from their predicted topology (Fig. 5.3b). Using these assignments, we were able to build relevant models for all TMDs of CAML$^{\Delta N}$-WRB based on density features and assisted by sequence alignment, TMD prediction and secondary structure prediction (Figs. S5.4b-d).

These data show that WRB/CAML together form a structure topologically comparable to the hexahelical membrane domain of *E. coli* YidC and the EMC3/EMC6 subcomplex of the EMC insertase (Fig. 5.3c). The TMD arrangement of WRB is highly conserved with EMC3 and the core three-TMD structure of YidC, providing further evidence for a common membrane architecture within the Oxa1 superfamily. Notably, a unique feature of WRB appears to be the rigid helical connection between the CD and TMDs, as the cytoplasmic coiled-coils of both YidC and EMC3 are connected by flexible linkers. The CAML TMDs form a three-helical bundle that interacts with WRB TMD3 in a similar manner to TMDs 1, 4 and 6 of *E. coli* YidC [24, 29] and the EMC6 interaction partner of EMC3 (Fig. 5.3c) [25]. However, the TMD arrangement of CAML has virtually no structural homology with these YidC helices and, despite being crudely similar to EMC6, shows notable differences in TMD length and tilt. Therefore, Oxa1 superfamily members have evolved to interact with a highly divergent set of three-TMD proteins.

A distinguishing feature of CAML is TMD3, which forms a long 55 Å helix that is tilted at 45° relative to the membrane normal and protrudes into the cytoplasm (Fig. 5.3b). This helix acts as a central linchpin within the complex, bridging the WRB TMDs with the helix-turn-helix formed by CAML TMD1/TMD2. Indeed, Get2 TMD3 is the only TMD within *sc*Get1/Get2 that cannot be substituted without destabilising the complex [21]. Given the helical register of the TMDs, F278 of CAML TMD3 is likely to stabilise the interaction with WRB TMD3 via aromatic stacking with the absolutely conserved W158 (Fig. S5.4e). In addition, our model places the two

cysteine residues within CAML TMD3 in proximity to cysteines from WRB TMD3 and CAML TMD1 (Figs. 5.3b and S5.4f), indicating they could form disulphides. Whilst the reconstructions show clear density for the C208/C284 intramolecular disulphide within CAML, there is only weakly observable density for the putative disulphide connecting CAML C275 and WRB C157, suggesting it may only form under certain conditions.

On the luminal side of the membrane, connectivities within the density were ambiguous and therefore residues corresponding to the ER loop of WRB and the artificial linker fusing CAML to WRB were not included in the final model. However, a horseshoe-shaped density caps one CAML$^{\Delta N}$-WRB heterodimer along the ER membrane interface and likely belongs to the 24-residue proline-rich loop connecting WRB TMD2/TMD3 (ER cap), which has a single α-helix predicted at its C-terminal end (Fig. S5.4c, S5.4g). As this ER cap forms intimate interactions with the luminal ends of WRB TMD1 and CAML TMD2/TMD3, it probably stabilises the TMD arrangement and also restricts access between the bilayer and ER lumen. Interestingly, the equivalent loop within EMC3 also contains a luminal helix that is predicted to fulfil a similar function [25].

Notably, CAML TMD3 and all the WRB TMDs are loosely packed on the cytoplasmic side of the membrane, forming a groove that is sealed on the luminal side by these TMDs and the ER cap (Fig. 5.3d). Furthermore, this groove is lined by many hydrophilic or charged residues from WRB (N22, R25, Q112, N163, K164; Fig. 5.3d) and thus creates a hydrophilic environment within the membrane. Such a lipid-exposed hydrophilic groove is also formed by the equivalent helices in both YidC and EMC3/EMC6, where it is required for membrane insertion [23, 25]. Indeed, cross-links have previously been observed between the TA substrate and the equivalent TMDs in *sc*Get1/Get2 [21], suggesting the hydrophilic groove of WRB/CAML provides a route for TA protein insertion.

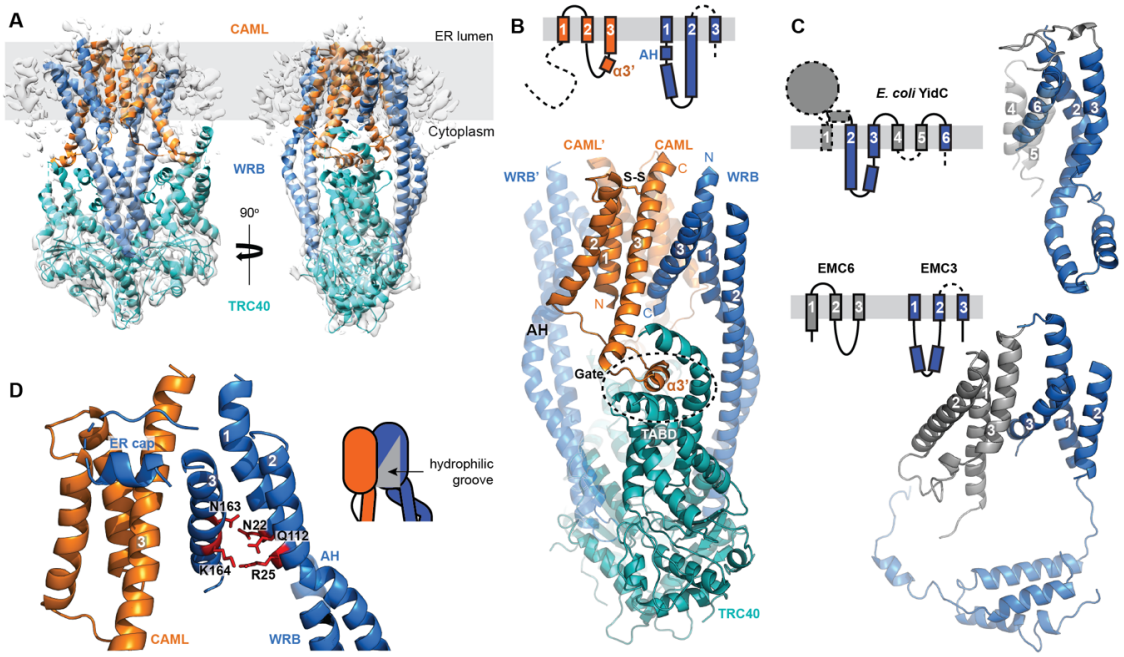


Figure 5.3 The structure of human CAML$^{\Delta N}$-WRB/TRC40 reveals a YidC-like fold in the membrane. (a) Model for CAML$^{\Delta N}$-WRB/TRC40 superposed to the cryo-EM density. Membrane plane defined perpendicular to the TRC40 symmetry axis with boundaries inferred from the location of the TMDs. Unassigned density corresponds to the amphipol belt (b) Structure of the CAML$^{\Delta N}$-WRB heterotetramer. A dashed circle indicates the α3'-TABD gate. Inset, topology of WRB and CAML, with dashed lines representing features not present in our model. (c) Side-by-side comparison of Oxa1 superfamily members after superimposition with the CAML$^{\Delta N}$-WRB TMDs in the view shown in (b). Regions homologous to WRB in E. coli YidC (root mean square deviation (RMSD) of 3.18 Å over 56 Cα atoms; PDB accession 3WVF) and EMC3/EMC6 from the EMC (root mean square deviation (RMSD) for EMC3 of 2.93 Å over 57 Cα atoms; PDB accession 6WW7) are coloured blue. Dashed lines in the topology schematics represent regions not shown in the structural models for clarity. (d) The hydrophilic groove formed by the WRB TMDs and CAML TMD3, lined with the hydrophilic residues shown in red. A possible model for the WRB ER cap is also shown. See also Figs. S5.1, S5.2, S5.4 and S5.8 and Table S5.2.

5.2.4 The heterotetramer is stabilised by phosphatidylinositol binding

The heterotetramer interface between CAML$^{\Delta N}$-WRB involves CAML TMD1/TMD3 and WRB TMD3 and buries a surface area of ~700 Å^2 (Fig. 5.4a), suggesting it is weak in comparison to other membrane protein oligomers and thus is likely to be maintained by interfacial lipids [30]. To further characterise the interactions stabilising these membrane heterotetramers, the oligomeric state of *sc*Get2$^{\Delta N}$-Get1 and CAML$^{\Delta N}$-WRB was analysed in a variety of mild detergents. NMS (Fig. 5.4b), size exclusion chromatography (SEC; Fig. S5.5a, S5b) and GraFix cross-linking [31] (Fig. S5.5c) showed that *sc*Get2$^{\Delta N}$-Get1 and CAML$^{\Delta N}$-WRB formed stable heterotetramers only when solubilised in nonaethylene glycol monododecyl ether (C12E9) and CYMAL-5 neopentyl glycol (NG) respectively. As the collisional activation applied during nMS removes the detergent micelle from the membrane protein complex [32] and we observed no detergent molecules specifically bound to the complexes (Figs. 5.4b and S5.3), the interactions maintaining these heterotetramers are likely to be physiologically relevant.

nMS of *sc*Get2$^{\Delta N}$-Get1 also revealed adduct peaks that were only associated with the heterotetrameric species, which lipidomics analyses showed to be a mixture of phospholipids present in the *S. cerevisiae* ER membrane (Fig. 5.4c). The mass of the adduct observed in nMS was ~839 Da, suggesting the bound phospholipid is predominantly phosphatidylinositol (PI; 790-870 Da, Fig. 5.4c). In addition, the analyses confirmed that ~1390 Da adduct peaks were also present for cardiolipin (CL; 1290-1490 Da, Fig. 5.4c), a lipid predominantly found within the inner mitochondrial membrane in eukaryotes. Given their low natural abundance in the ER membrane [33], the acidic phospholipids PI and CDL are likely to interact specifically with heterotetrameric *sc*Get1/Get2.

These lipids were also found associated with CAML$^{\Delta N}$-WRB in different detergents (Figs. 5.4d and S5.6a). Firstly, CDL (~1345 Da) was bound to CAML$^{\Delta N}$-WRB heterodimers (Fig. S5.6a), indicating Get1/Get2 homologues have a propensity for this mitochondrial lipid. Although there have been some reports that this ER-resident complex may be connected to mitochondrial function [34, 35], it is more probable the observed CDL binding is an artefact of protein

purification from total membranes, with lipid binding sites likely having a high affinity for this doubly negatively charged lipid. More interestingly, only heterotetrameric CAML$^{\Delta N}$-WRB displays an adduct of ~826 Da for bound phospholipids, again indicating predominant association with PI (Fig. 5.4d). Therefore, we can generally conclude that *sc*Get1/Get2 and WRB/CAML heterotetramers are only stable in detergents that allow co-purification with a single molecule of PI. This strongly suggests that conserved lipid-protein interactions are playing a role in stabilising this higher oligomeric state and could explain why Get3-mediated heterotetramerisation of *sc*Get1/Get2 was previously not observed in proteoliposomes lacking PI [20].

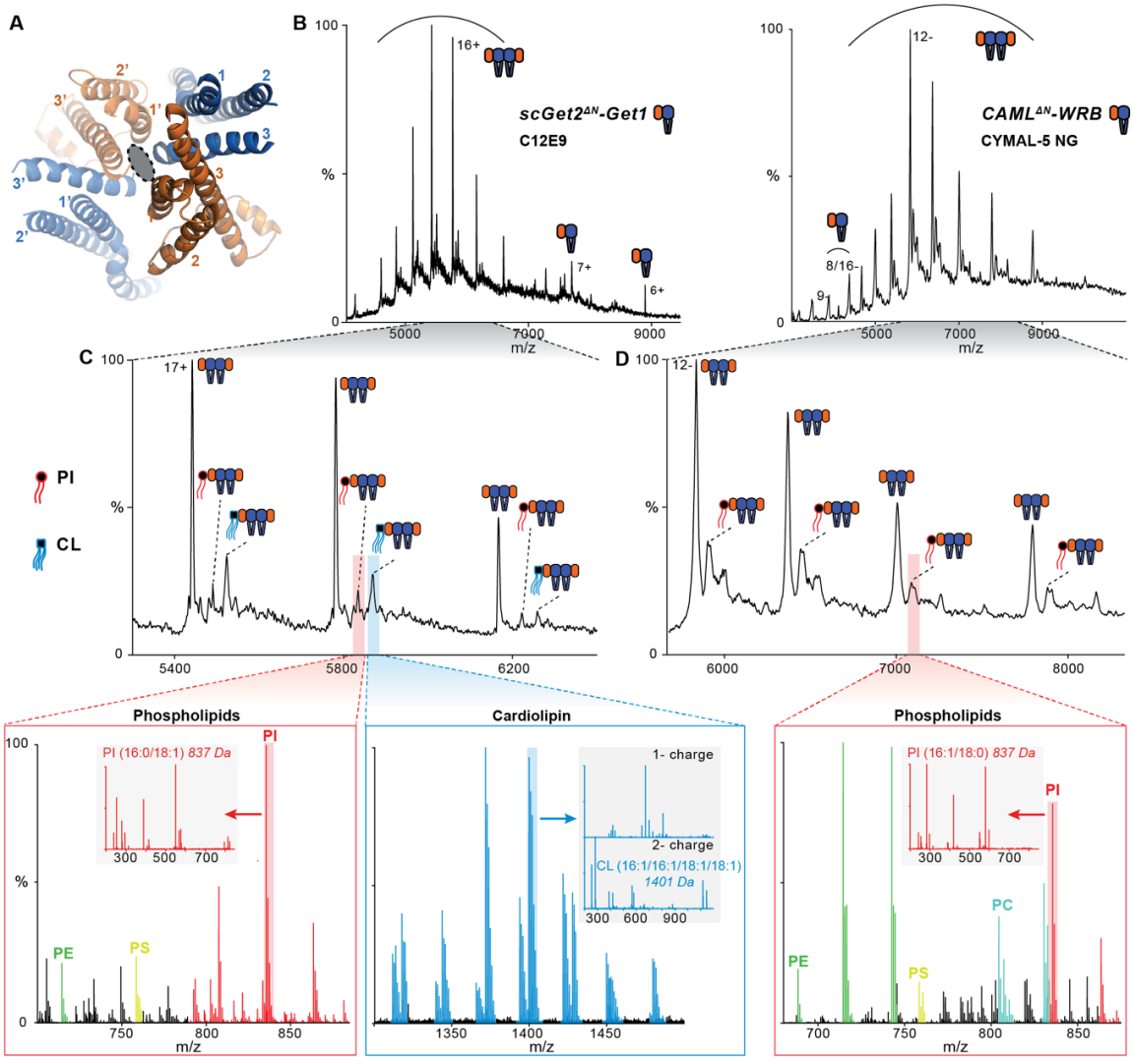


Figure 5.4 *sc*Get2$^{\Delta N}$-Get1 and CAML$^{\Delta N}$-WRB heterotetramers bind lipids. (a) Structure of the CAML$^{\Delta N}$-WRB heterotetramer viewed from the ER lumen. A grey oval highlights the 700 Å^2 heterotetramer interface. (b) Mass spectra of stable *sc*Get2$^{\Delta N}$-Get1 and CAML$^{\Delta N}$-WRB

heterotetramers liberated from the specified detergent micelles collected under non-denaturing conditions. For each assigned complex, the charge state for the most intense peak is shown, with the sign denoting data collection in either positive or negative mode. (C, D) Zoomed regions of the mass spectra in (b) showing the main charge states for the major species. Adduct peaks with a defined mass greater than the amino acid sequence mass correspond to bound phospholipid or CDL. MS profiles for phospholipid and CDL species identified through in-solution lipidomics are shown beneath each spectrum. Phospholipid adducts represent a lipid mixture that is likely dominated by PI. Insets, MS/MS of the highlighted ions identified the major PI and CDL species associated with the protein. See also Fig. S5.6 and Table S5.1.

5.2.5 PI binds to functionally important interfacial residues

Interestingly, in our model R204 and K209 at the luminal end of CAML TMD1 point towards the equivalent residues from the opposite dimer (Fig. S5.4h), which is energetically unfavourable for both the interaction and the membrane environment. Similarly, *sc*Get2 TMD1 contains the positively charged residues K150 and K157 on one helical face (Fig. S5.4d), although their exact orientation with respect to the heterotetramer interface cannot be predicted due to poor sequence conservation. Given that basic residues commonly interact with phospholipid headgroups [36], the membrane heterotetramer could be stabilised by PI binding to CAML/Get2 TMD1. To test this hypothesis, alanine mutants of the equivalent R202 and K207 residues in a mouse CAML variant (97% sequence identity to the human CAML TMDs) and of K150 and K157 in *sc*Get2 were coexpressed with human WRB or *sc*Get1 respectively in a *S. cerevisiae Δget1/get2* strain. When expressed at the same levels as wild type sequences (Fig. 5.5a), WRB/mCAML^{R202A}, WRB/mCAML$^{R202A/K207A}$ and *sc*Get1/Get2$^{K150A/K157A}$ were unable to rescue the growth phenotypes of the mutant yeast cells (Fig. 5.5b and S5.7). TA protein insertion was specifically impaired in the presence of these point mutants, as the GFP-tagged TA protein Sed5 retained a diffuse cytosolic distribution and showed a highly reduced formation of the fluorescent punctate structures indicative of correct Golgi localisation (Fig. 5.5d and S5.7) [13,

19]. Whilst these mCAML mutants were still able to recruit a genomically GFP-tagged version of Get3 to the ER, Get3 localisation was impaired in the presence of the *sc*Get1/Get2$^{K150A/K157A}$ double mutant (Fig. 5.5c), indicating that overall insertase stability requires K150 and K157 together but not individually.

In order to delineate the exact role of these residues in lipid binding and oligomerisation, CAML$^{\Delta N/R204A}$-WRB and *sc*Get2$^{\Delta N/K150A}$-Get1 single mutants were purified in detergents where the wild type proteins are predominantly heterotetrameric. SEC and nMS analysis of CAML$^{\Delta N/R202A}$-WRB demonstrated a shift in oligomeric state towards the heterodimer (Fig. S5.6b), whilst nMS showed complete loss of the *sc*Get2$^{\Delta N/K150A}$-Get1 heterotetramer (Fig. S5.6c). More strikingly, no PI was associated with the *sc*Get2$^{\Delta N/K150A}$-Get1 heterodimer in C12E9 (Fig. S5.6c), indicating that PI binding is dependent on the protein sequence rather than the detergent present and that it is likely bound at the heterotetramer interface. Given that this mutant still forms a functional insertase *in vivo* (Fig. 5.5b-d and S5.7), Get3 may still be able to drive heterotetramerisation of *sc*Get1/Get2^{K150A} in the membrane. Taken together, these results show that PI binding and heterotetramer formation by the insertase complex are both intrinsically linked and dependent on positively charged residues within the interfacial TMD1 of CAML/*sc*Get2. Furthermore, the functional importance of these residues *in vivo* shows that the Get1/Get2 heterotetramer is required for efficient TA insertion.

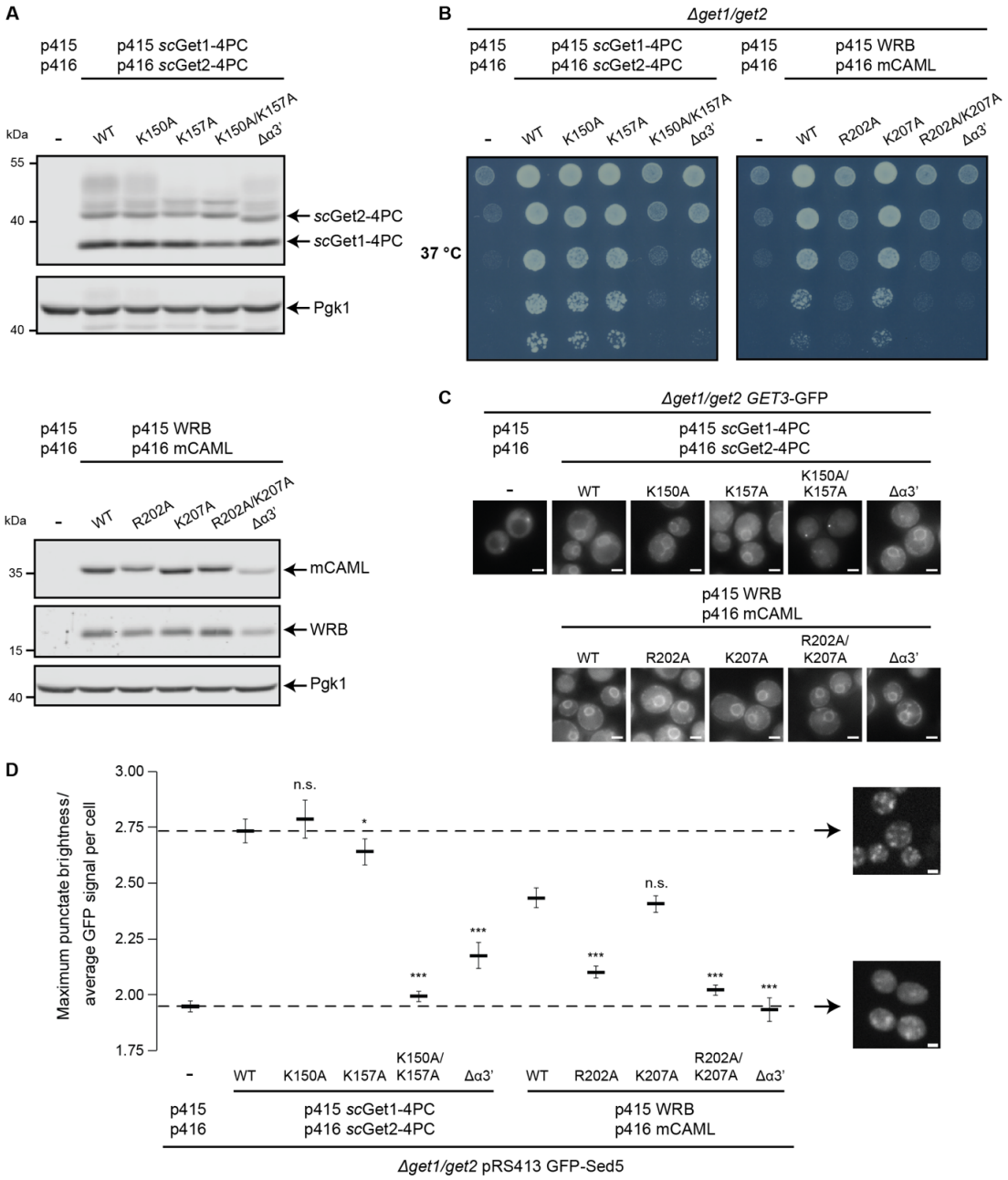


Figure 5.5 Helix α3' and positively charged residues in the heterotetramer interface are important for *sc*Get1/Get2 and WRB/mCAML function *in vivo*. (a)-(d) *sc*Get1-4PC/Get2-4PC and WRB/mCAML were co-expressed from the indicated plasmids in *Δget1/get2* yeast strains for wild type (WT) and mutant *sc*Get2/mCAML sequences. Transformants with the empty vector were taken as a negative control (-). (a) Western blot of *sc*Get1/Get2 and WRB/mCAML expression in *Δget1/get2* yeast cells. An anti-protein C tag antibody was used to detect Get1-4PC and Get2-4PC. Pgk1 was analysed as a loading control. The immunoblots are representative of n=3. (b) Serial dilutions of *Δget1/get2* yeast cells expressing *sc*Get1/Get2 or WRB/mCAML

constructs spotted on synthetic dropout plates at 37ºC. The spotting assays are representative of n=3. (c) Representative images from fluorescence microscopy showing the subcellular localisation of *sc*Get3-GFP in *Δget1/get2* yeast cells expressing genomically tagged Get3 and *sc*Get1/Get2 or WRB/mCAML constructs. Scale bar: 2 μM. (d) Quantification of GFP-Sed5 distribution from fluorescence microscopy images of *Δget1/get2* yeast cells expressing GFP-Sed5 together with *sc*Get1/Get2 or WRB/mCAML constructs. The dashed line at the bottom and representative microscopy image shows a lower maximum punctate fluorescence relative to average cellular GFP fluorescence for cells carrying empty plasmid, whilst the dashed line and representative microscopy image at the top shows a higher ratio for cells expressing WT *sc*Get1/Get2. Several hundreds of cells were quantified for each replicate and error bars indicate standard error calculated from six independent experiments. Statistically significant differences between the WT and mutant *sc*Get2 or mCAML variants were determined using the Welch's *t*-test: ***: $p<0.0001$; *: $p<0.05$; n.s.: not significant. Scale bar: 2 μM. See also Figs. S5.6 and S5.7.

5.2.6 A conserved CAML helix α3' forms a functionally important gate with TRC40

Our structure shows a cavity between the cytoplasmic and membrane regions that would be accessible to the TA protein after its release from TRC40. The WRB-CD and the missing CAML N-terminus likely act as barriers between this cavity and the cytoplasm, limiting the movement of the substrate TMD and its adjoining cytoplasmic domain. In addition, the cytoplasmic loop region between CAML TMD2/TMD3 has a high proportion of basic residues (Fig. S5.4d), suggesting it is likely to interact with the membrane surface and reduce the volume of the cavity accessible to the TA protein. An unexpected feature of this otherwise disordered loop is a short helix (α3'; residues 243-250) directly preceding CAML TMD3. Interestingly, helix α3' forms a contact with helices α5, α7 and α9 of the TABD of TRC40, largely through hydrophobic interactions between aliphatic residues (Fig. 5.6a and S5.4i). Therefore, helix α3' effectively forms a closed 'gate' with the TRC40 TABD (Fig. 5.3b), providing a further barrier

to TA protein movement. Notably, binding of helix α3' to TRC40 is accompanied by large conformational changes in its TABD compared to open *sc*Get3, namely a 15 Å displacement of the long α7-α9 helices, a 180º rotation of α5 and reorientation of α4 and α6 (Fig. 5.6b). Given that helices α4-α9 of closed *sc*Get3 together form the binding site for the TA TMD [12], the observed rearrangement within the TRC40 TABD is likely to determine how the TA substrate is released first to the cavity and then to the membrane.

To assess the functional importance of the helix α3'-TABD gate, mCAML with a polyglycine substitution of helix α3' (mCAML$^{\Delta\alpha3'}$) was coexpressed with WRB in the *S. cerevisiae* *Δget1/get2* strain. It was observed that WRB/mCAML$^{\Delta\alpha3'}$ was unable to rescue the growth phenotypes of these cells (Fig. 5.5b and S5.7a), although the proteins were expressed at slightly lower levels than their wild type counterparts (Fig. 5.5a). Crucially however, these mutant insertase levels were still able to restore ER localisation of Get3 (Fig. 5.5c), indicating that helix α3' is not required for Get3 recruitment per se. In contrast, GFP-Sed5 was mis-localised in the presence of mCAML$^{\Delta\alpha3'}$ (Fig. 5.5d and S5.7b), demonstrating that helix α3' is important for substrate insertion by the WRB/CAML insertase complex. Furthermore, secondary structure predictions imply that helix α3' is a conserved feature preceding TMD3 in CAML homologues; whilst an analogous *sc*Get2$^{\Delta\alpha3'}$ variant was expressed at the same level as wild type *sc*Get2 (Fig. 5.5a) and mostly restored the growth phenotypes of *Δget1/get2* cells (Fig. 5.5b and S5.7a), GFP-Sed5 Golgi localisation was severely impaired (Fig. 5.5d and S5.7b). Therefore, TA insertion from lower to higher eukaryotes requires helix α3' of Get2 homologues.

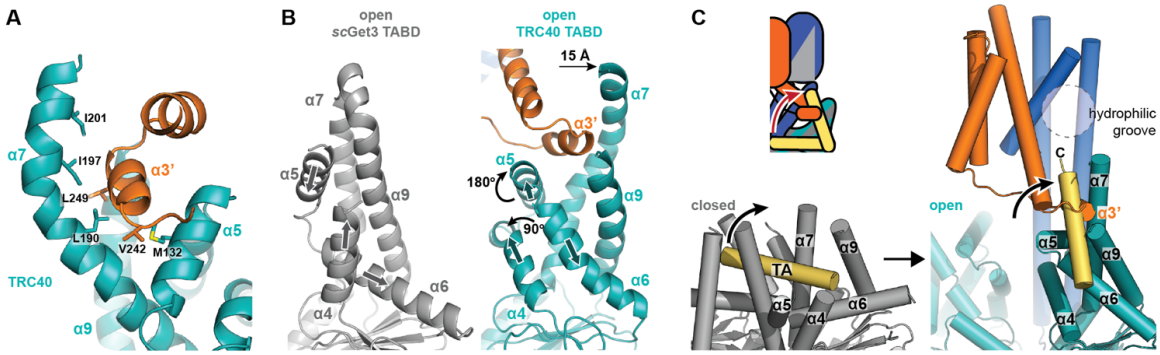


Figure 5.6 CAML helix α3'forms a functionally important gate with TRC40. (a) The interaction between CAML helix α3' and the TABD of TRC40. (b) Side-by-side comparison of the TABDs from open *sc*Get3 (grey; PDB accession 3SJA) and TRC40 (cyan) after superimposition of the dimers (RMSD 2.41 Å over 447 Cα atoms). The arrows show the directionality of the α helices. (c) Comparison of the TABDs from closed *sc*Get3 (grey) bound to Pep12 (yellow; PDB accession 4XTR) and open TRC40 TABD (cyan) bound to CAML helix α3'. Equivalent helices are labelled and only one CAML$^{\Delta N}$-WRB heterodimer is shown for clarity. The Pep12 TMD is manually positioned within the corresponding binding site in TRC40.

5.3 Discussion

5.3.1 A YidC-like mechanism for TA insertion involving a hydrophilic groove

Our structure of CAML$^{\Delta N}$-WRB reveals features within the TMDs and adjoining loops that cannot be predicted by bioinformatics alone. For example, topology predictions for CAML fail to reach a consensus on the position or even the existence of TMD2 [37], whilst the functionally important helix α3' was not anticipated to be a distinct feature from TMD3. Another notable feature uncovered by the CAML$^{\Delta N}$-WRB structure is the membrane-embedded hydrophilic groove that is accessible from the cytoplasm. As this region was previously shown to form a membrane-integrated docking site for the TA substrate [21], we propose that the C-terminus of the TA protein is initially drawn into the lipid bilayer via interactions with the hydrophilic groove, analogous to the mechanisms discussed for YidC and the EMC [23, 25]. The C-termini of ER-destined TA proteins display a wide variation in their charge properties [2], which may explain why the hydrophilic groove does not contain an absolutely conserved charged

residue as in YidC and EMC3 (Fig. S5.4c). Given the proximity of the WRB AH to the hydrophilic groove (Fig. 5.3d) and that amphipathic helices are known to affect membrane structure [38, 39], it is also tempting to speculate that this helix causes a deformation of the bilayer that assists TA insertion. Furthermore, the observed concave arrangement of WRB/CAML TMDs would also cause membrane curvature, indicating future studies with membrane-embedded proteins are important for determining whether Get1/Get2 homologues can induce membrane deformation in the same manner as YidC and EMC3/EMC6 [25, 40].

5.3.2 The membrane heterotetramer increases the efficiency of TA insertion

Notably, the hydrophilic groove is formed entirely by one CAML$^{\Delta N}$-WRB heterodimer within the membrane heterotetramer, explaining why the heterodimer alone is sufficient for TA insertion [20]. This uncovers further similarities between Oxa1 superfamily members, as YidC also has the propensity to form dimers in both membranes and detergents [41, 42], despite both protomers being functional as independent insertases [43]. Furthermore, the minimal functional units of structurally unrelated bacterial (SecYEG) [44] and mitochondrial (TOM complex) [45] protein translocases also undergo lipid-mediated higher order oligomerisation, demonstrating that the Get1/Get2 insertase follows a general theme in the field of protein transport. It is also not surprising that Get1/Get2 heterotetramers have so far eluded experimental observation; the weak interface stabilised by lipid binding within our heterotetramer structure is a common trait shared by dynamic and controversial membrane protein oligomers such as class A G protein-coupled receptors [30, 46, 47]. Indeed, Get3 binding is required to trigger stable heterotetramerisation in a membrane environment, akin to ligand-induced oligomerisation of membrane receptors involved in cell signalling [48].

Nevertheless, our data show that disruption of the heterotetramer interface within Get1/Get2 homologues reduces TA insertion, implicating a role for this oligomeric state *in vivo*. We can hypothesise that the heterotetramer serves to improve the efficiency of TA targeting and insertion, thereby fine-tuning and increasing the robustness of the GET pathway. For example,

the provision of distinct hydrophilic grooves by the Get1/Get2 heterotetramer would allow the released TA protein to choose the optimal path for insertion. This is particularly pertinent as the Get3 homodimer binds the substrate TMD asymmetrically [12] and thus the C-terminus of the TA protein would be more proximal to one of the two opposing hydrophilic grooves when presented to the heterotetrameric insertase complex. In addition, the symmetric binding of Get3 by the CDs of the Get1/Get2 heterotetramer increases the avidity of the interaction and thus the targeting efficacy. Furthermore, our nMS data suggest that Get1/Get2 has the propensity to simultaneously bind two Get3 homodimers. In functional terms this could mean that whilst the Get1-CDs are occupied with a Get3 dimer, a second Get3 dimer could already be captured by the Get2-CDs and be 'primed' for efficient handover to Get1.

5.3.3 A dynamic model for TA insertion

Existing models for TA insertion predict that the arrangement of Get1/Get2 heterodimers is dynamically coupled to Get3 conformation; after handover from the Get2-CDs, Get1-CDs engage with a closed TA-bound Get3 dimer through interaction with one monomer and remains bound during the subsequent transition to the open state of Get3, in which it makes additional contacts with the second monomer [18, 27, 49]. To understand how the interactions within the WRB/CAML heterotetramer changes when bound to different TRC40 conformers, we superimposed monomers of TRC40 from our model on *sc*Get3 in closed [12] and semi-open conformations [18], making the assumption that TRC40 exists in the same states as Get3 homologues. The resulting models are analogous to those constructed previously for the *sc*Get1-CDs in isolation [18, 49] and corroborate the assertion that opening of the Get3 dimer correlates with closure of the heterodimeric insertase halves (Fig. S5.8a). As a result, WRB/CAML heterodimers bound to closed TRC40 are positioned too far apart to interact (Fig. S5.8a). This suggests that initial *sc*Get2-CD mediated 'heterotetramerisation' likely occurs solely via interactions between Get2 and Get3 and serves to simply localise a pair of Get1/Get2 heterodimers within the membrane.

It is striking from this dynamic model that during the transition from closed to open TRC40, CAML$^{\Delta N}$-WRB has to concomitantly undergo conformational changes in the membrane (Fig. S5.8a). Given that the WRB-CDs penetrate the ATP binding site of TRC40 at one end [17, 18] and are rigidly linked to the TMDs at the other, it is likely they communicate nucleotide-driven conformational changes in TRC40 to the membrane region of WRB/CAML, analogous to the role played by the coupling helices in ATP-binding cassette (ABC) transporters [49, 50]. Interestingly, superimposition of a high-resolution structure of *sc*Get3/Get1-CD [18] with TRC40 within our model shows that the WRB-CD is rotated 15° relative to *sc*Get1-CD (Fig. S5.8b). In addition, the two WRB-CDs within our heterotetramer structure display an 8° variation in their position relative to TRC40 (Fig. S5.8c), suggesting there is plasticity in the WRB/TRC40 interaction. Furthermore, this minor change in WRB-CD orientation is propagated to the adjoining TMDs in a lever-like manner, resulting in notably different positions of the WRB/CAML heterodimeric halves (Fig. S5.8c). Therefore, we speculate that opening of TRC40 is coupled to conformational changes in the WRB-CDs that allow the WRB/CAML heterotetramer to dynamically sample different conformations during the insertion cycle.

As TRC40 opens and the TABD rearranges, the C-terminus of the TA substrate is likely to be increasingly tilted towards the hydrophilic groove (Fig. 5.6c), assuming the TMD remains within the equivalent TABD binding site to closed Get3 [9] and follows the TRC40 protomer adjacent to the N-terminal soluble domain. Therefore, opening of the TRC40 dimer guides the TA substrate towards the site for membrane insertion. Interestingly, in this model the TA substrate is pushed into the observed binding site for helix α3' (Fig. 5.6c). This would imply that the hydrophobic helix α3' mimics the substrate TMD and acts as a placeholder in its interaction with TRC40, much like the plug helix within the SecYEG channel [51]. In a similar manner, the helix α3'-TABD gate effectively blocks the route of the TA substrate from its TRC40 binding site to the hydrophilic groove. However, in contrast to plug deletion mutants stimulating the protein transport activity of SecYEG [52], our data show that deletion of helix α3' severely impairs TA insertion by *sc*Get1/Get2 and WRB/CAML. This suggests that helix α3' assists

rather than impedes insertion via the hydrophilic groove. It is tempting to hypothesise that, during binding to TRC40, helix α3' actively pushes the TA substrate along the TABD towards the membrane. The α3'-TABD gate could also drive insertion by simply preventing backwards movement of the substrate.

5.3.4 Perspectives

For many years, TA insertion by Get1/Get2 homologues has remained the least understood aspect of the GET pathway, as structural information for these small and dynamic integral membrane proteins eluded the field. The structural, nMS and functional complementation data presented here reveal the architecture of the WRB/CAML membrane insertase and provide a new model for Get3 or TRC40 recognition and TA insertion (Fig. 5.7). Our data also highlight new evolutionary relationships between membrane insertases, with the the YidC-like fold and hydrophilic groove of WRB indicating that both the architecture and mechanistic principles of membrane protein insertion are conserved from prokaryotes to higher eukaryotes. Furthermore, the structural homology between WRB/CAML and the membrane insertase within the EMC [25] demonstrates that the distinct pathways for TA biogenesis adopt common mechanistic principles. Furthermore, other members of the Oxa1 superfamily use the same protein fold to facilitate the insertion, folding and assembly of a much wider variety of mono- and polytopic membrane proteins [22], therefore the structure of the Get1/Get2 insertase has implications not only for TA insertion, but as a module utilised for more general aspects of membrane protein biogenesis. Indeed, it was recently shown that WRB acts independently of its role as a TA protein insertase to catalytically promote the correct integration of CAML TMDs [37]. The first structures of the archetypal Sec, Tat and YidC protein transporters [23, 51, 53] have kickstarted detailed molecular studies into how hydrophobic sequences are able to permeate the lipid bilayer. In this manner, our structure of the WRB/CAML insertase paves the way for future studies into the mechanistic details of substrate insertion and the role of specific lipids within the ER membrane. Further, functional similarities with Oxa1 superfamily members and

other protein transporters can now be ascertained, allowing the full integration of the GET pathway into the landscape of membrane protein transport.

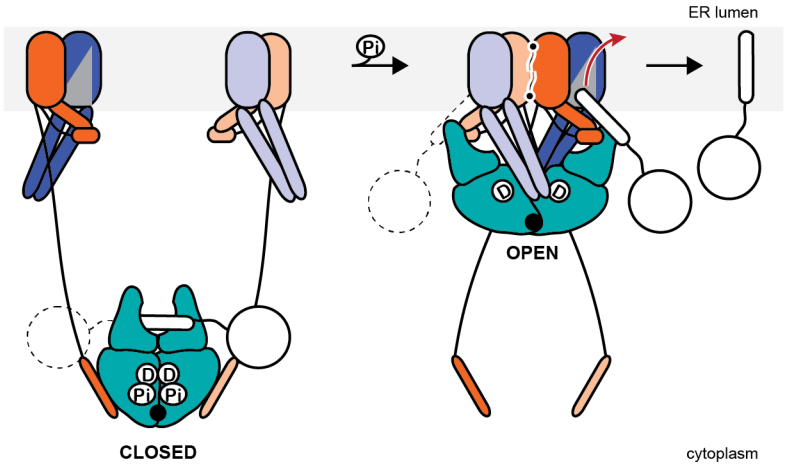


Figure 5.7 A model for TA insertion by Get1/Get2 homologues. TA-bound Get3/TRC40 in the closed state is initially captured by Get2/CAML-CDs from a pair of heterodimers. Handover to Get1/WRB-CDs triggers opening of Get3/TRC40 and formation of an intimate membrane heterotetramer stabilised by lipid-protein interactions. Get2/CAML α3' guides the substrate TMD along the TABD towards the bilayer. The C-terminus of the TA substrate interacts with the hydrophilic groove of Get1/WRB, allowing it to traverse the membrane. The heterotetramer provides two distinct insertion pathways to account for delivery of the TA substrate in one of two possible orientations.

5.4 Supplementary Information

5.4.1 Supplementary Figures

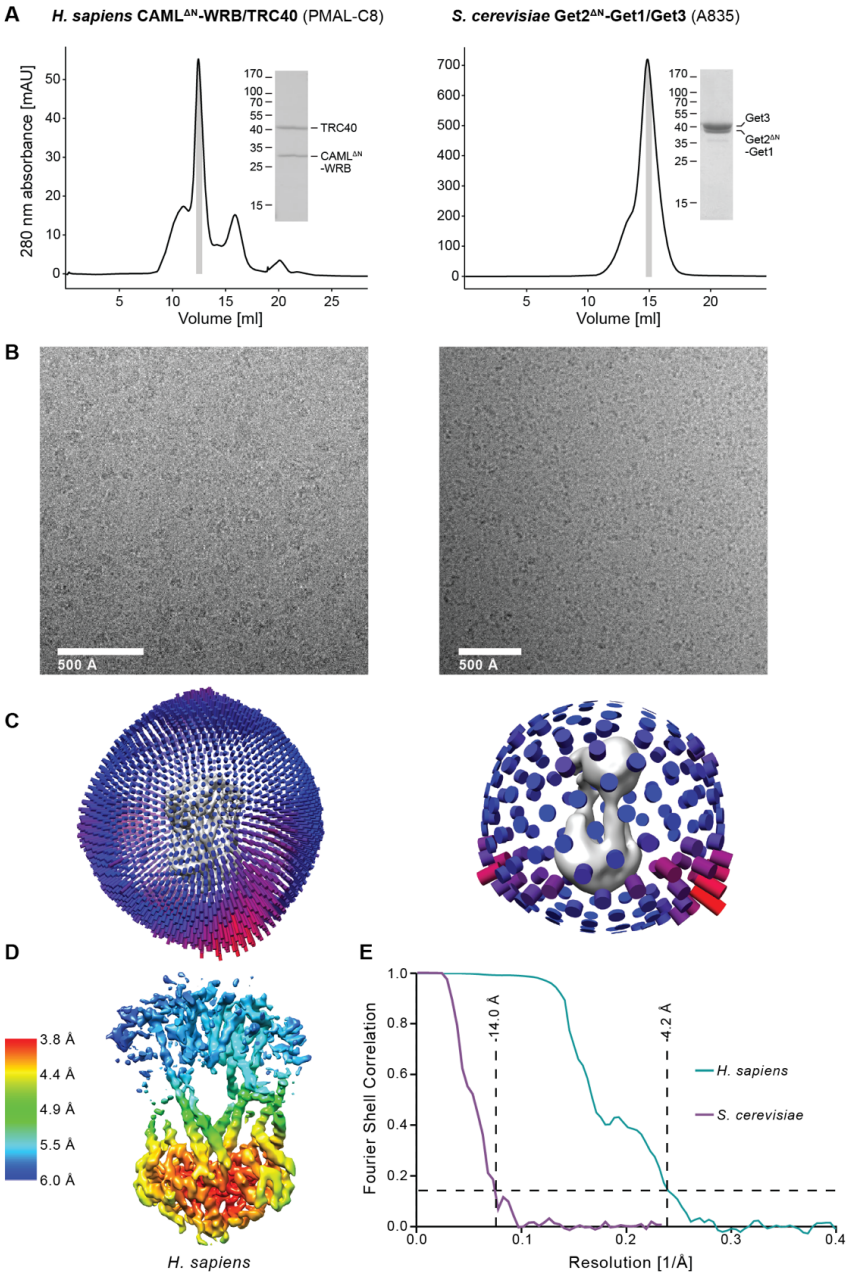


Figure S5.1 Sample preparation and experimental quality for cryo-EM. Related to Figures 5.1 and 5.3. (a) Size exclusion chromatography profiles for the complexes in amphipol, with the single fraction taken for cryo-EM analysis highlighted in grey. Insets, Coomassie-stained 12.5% SDS-PAGE gels showing the purity of the final complexes. (b) Representative micrographs for each complex. (c) Angular distribution plots following 3D classification (*H. sapiens*) or 3D auto-refinement (*S. cerevisiae*). (d) Local resolution estimates of the final reconstruction obtained

from the *H. sapiens* complex. (e) Gold-standard FSC curves of the final reconstructions calculated using a soft-edge mask.

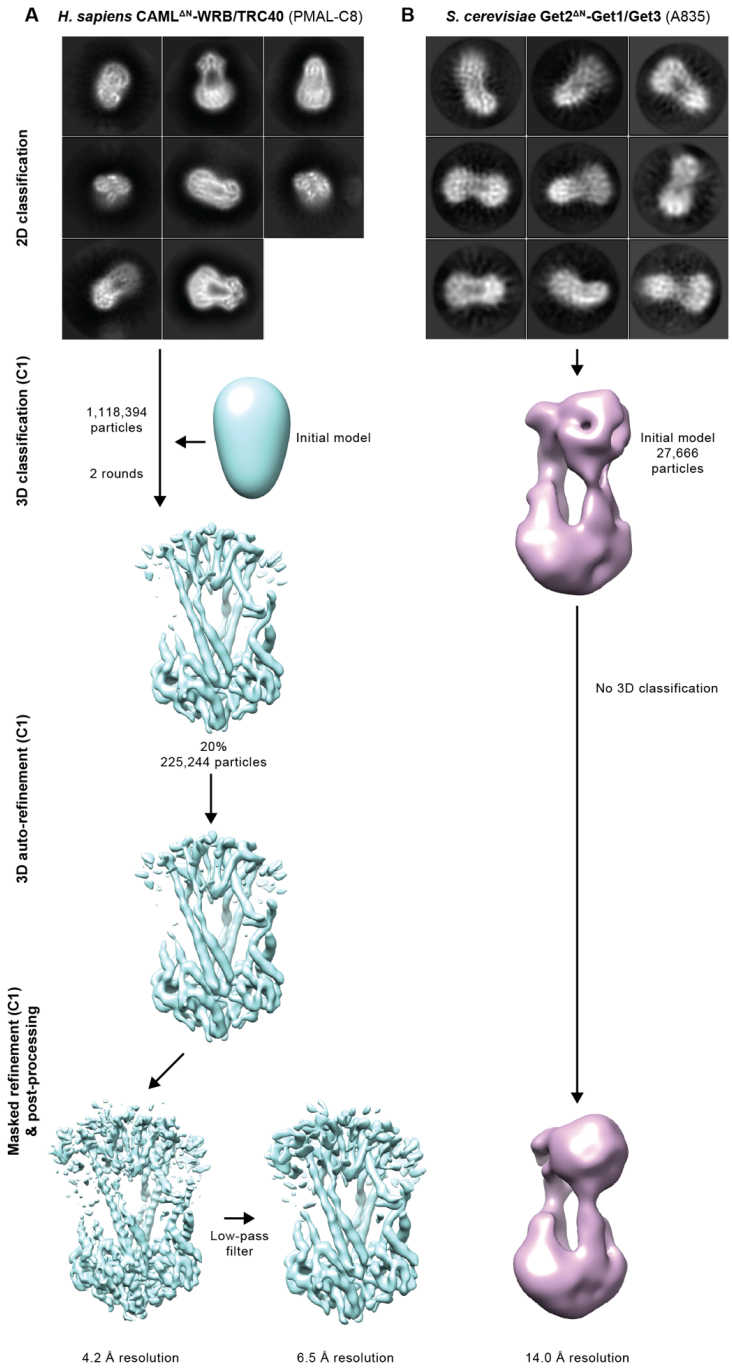


Fig. S5.2 Workflow for cryo-EM single-particle data processing of CAML$^{\Delta N}$-WRB/TRC40 and *sc*Get2$^{\Delta N}$-Get1/Get3 complexes. Related to Figures 5.1 and 5.3. (a) Autopicked particles for *H. sapiens* CAML$^{\Delta N}$-WRB/TRC40 in PMAL-C8 were subjected to 2D classification

followed by 3D classification against a low-resolution *de novo* model. Particle images were then refined using a soft mask, resulting in a reconstruction with an average resolution of 4.2 Å. More complete density for low resolution regions was present in a reconstruction low-pass filtered to 6.5 Å. (c) For *S. cerevisiae* Get2$^{\Delta N}$-Get1/Get3 in A835, 3D classification did not produce sensible particle subsets. Therefore, the initial model was refined directly, resulting in a reconstruction at 14.0 Å resolution.

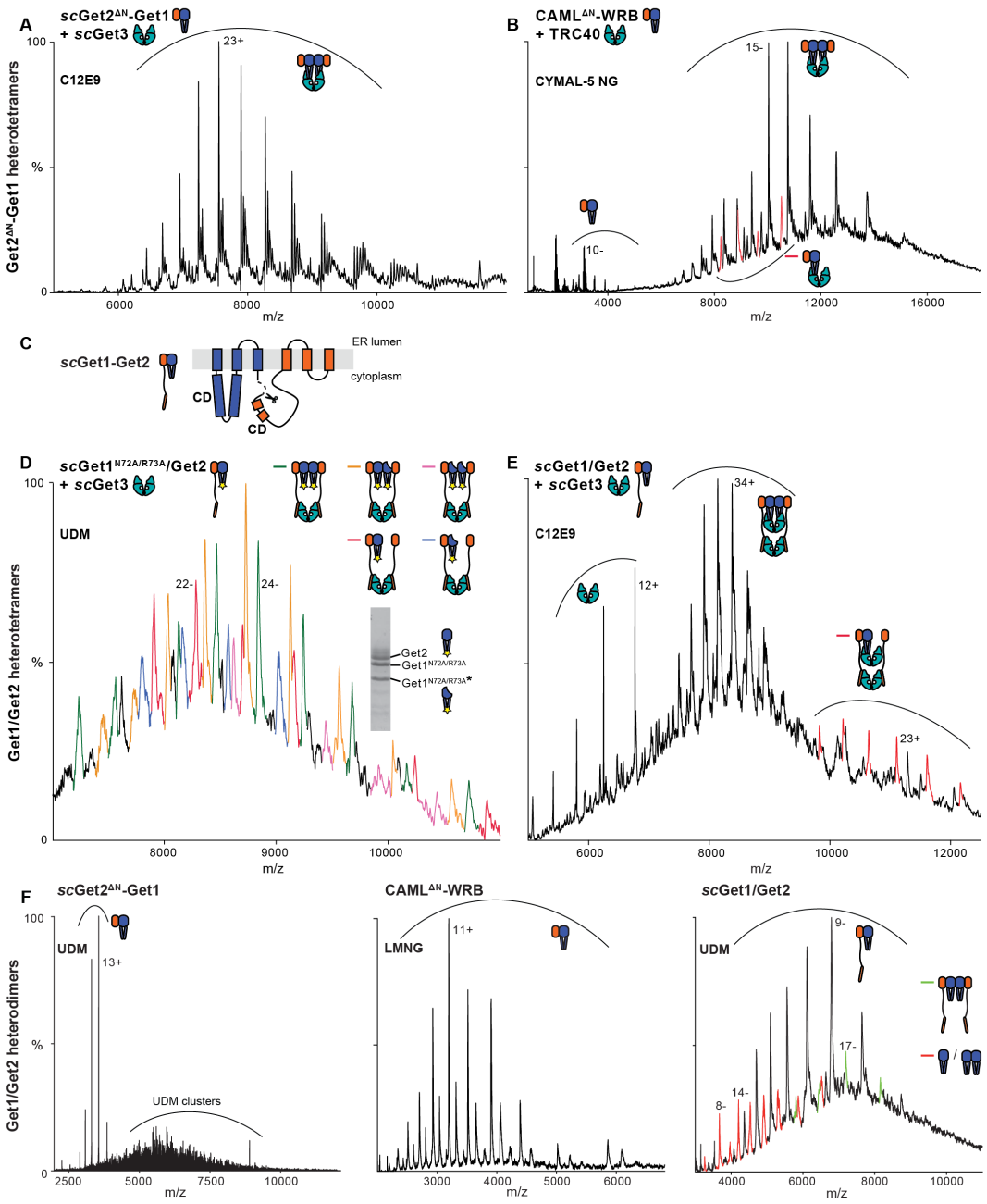


Fig. S5.3 Characterisation of *sc*Get1/Get2/Get3 and CAML$^{\Delta N}$-WRB/TRC40 by nMS. Related to Fig. 5.2. Mass spectra of (a) *sc*Get2$^{\Delta N}$-Get1/Get3 and (b) CAML$^{\Delta N}$-WRB/TRC40 complexes liberated from detergent micelles collected under non-denaturing conditions. (c) Topology of the full-length *sc*Get1-Get2 single-chain construct used in this study showing the Get1 and Get2 CDs. The dashed line represents the non-native linker sequence that is cleaved by specific proteases during purification to yield *sc*Get1/Get2. Mass spectra of full-length *sc*Get1/Get2, (d) with and (e) without N72A/R73A mutations in the Get1-CD, liberated from detergent micelles collected under non-denaturing conditions. (d) Inset, Coomassie-stained

12.5% SDS-PAGE gel showing the degradation product of *sc*Get1$^{N72A/R73A}$ (*sc*Get1$^{N72A/R73A}$*) observed during nMS. (f) Mass spectra of *sc*Get2$^{\Delta N}$-Get1, CAML$^{\Delta N}$-WRB and *sc*Get1/Get2 heterodimers liberated from detergent micelles collected under non-denaturing conditions. For each assigned complex in all mass spectra the charge state for the most intense peak is shown, with the sign denoting data collection in either positive or negative mode.

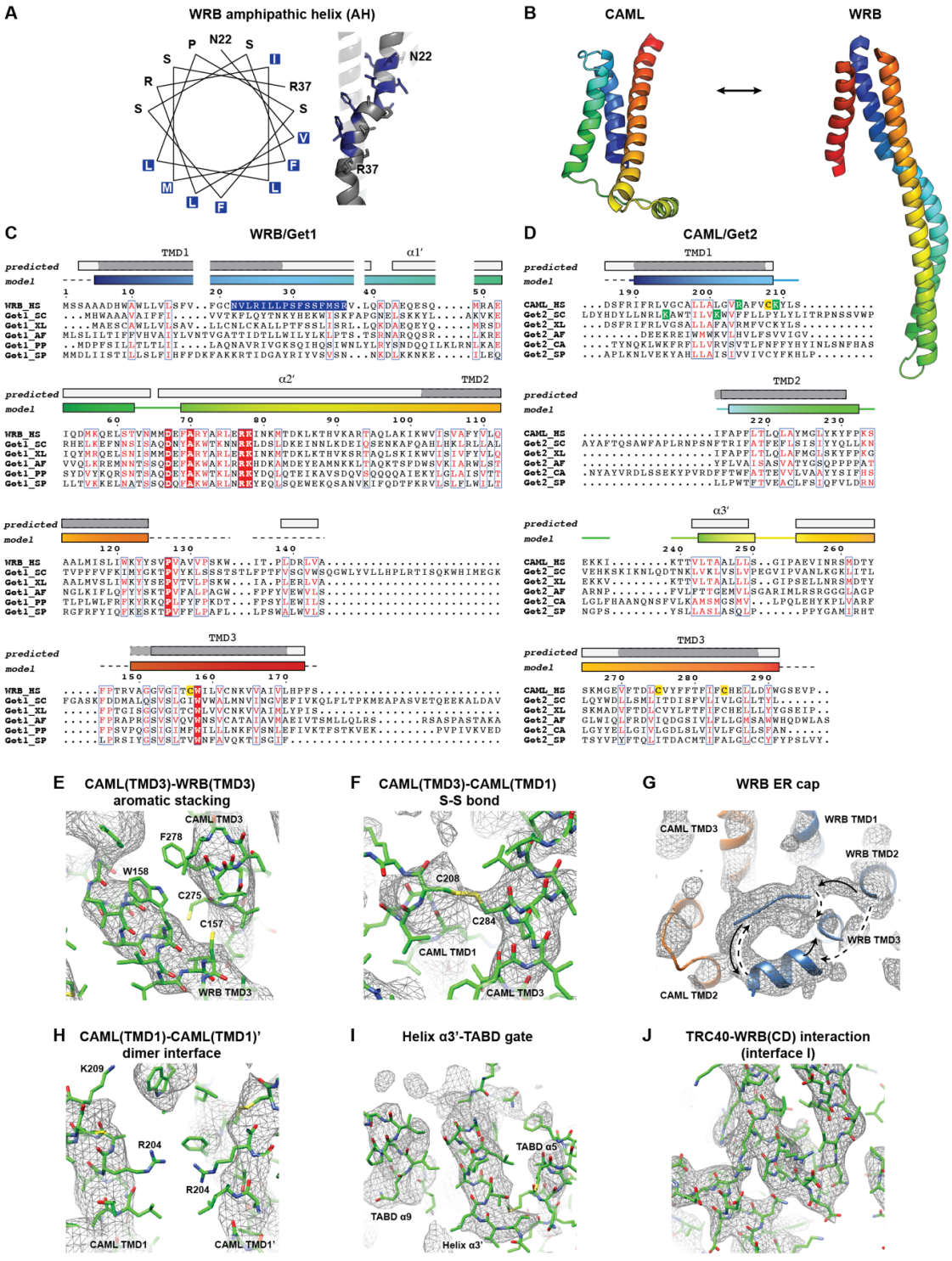


Fig. S5.4 Model building of CAML$^{\Delta N}$-WRB/TRC40 in the cryo-EM density. Related to Fig. 5.3. a) The amphipathic helix of WRB (residues 22-37) is shown in both a helical wheel and structural representation. Hydrophobic residues are coloured blue. b) The structures of CAML and WRB (artificially separated in the membrane plane for clarity) coloured from the N- (blue) to the C- (red) terminus. c,d) Sequence alignments for (c) WRB/Get1 and (d) CAML/Get2

generated using simultaneous transmembrane prediction and secondary structure alignment in PRALINE [54] and visualised using Espript [55]. Sequences are shown for *H. sapiens* (HS), *S. cerevisiae* (SC), *Xenopus laevis* (XL), *Aspergillus fumigatus* (AF), *Candida albicans* (CA), *Schizosaccharomyces pombe* (SP) and *Pichia pastoris* (PP). Conserved and similar residues are shown in red, CAML/*sc*Get2 TMD1 positively charged residues that may point into the heterotetramer interface are highlighted in green and CAML/WRB cysteines that may be involved in disulphide bond formation are highlighted in yellow. The amphipathic helix present in the structure of WRB is highlighted in blue. Residues missing in our CAML$^{\Delta N}$-WRB model are indicated by dashed lines. Helices predicted for CAML/WRB by PSIPRED [56] are shown in grey, with dark shading showing transmembrane domains predicted by Phobius [57]. Helices present within our model for CAML$^{\Delta N}$-WRB are coloured according to their representation in (b). e-j) Representative density for CAML$^{\Delta N}$-WRB/TRC40 showing key structural features. For the WRB ER cap (g), two possible connectivities are indicated with arrows on the cartoon representation, with the solid arrows indicating the most likely arrangement based on density shape and size.

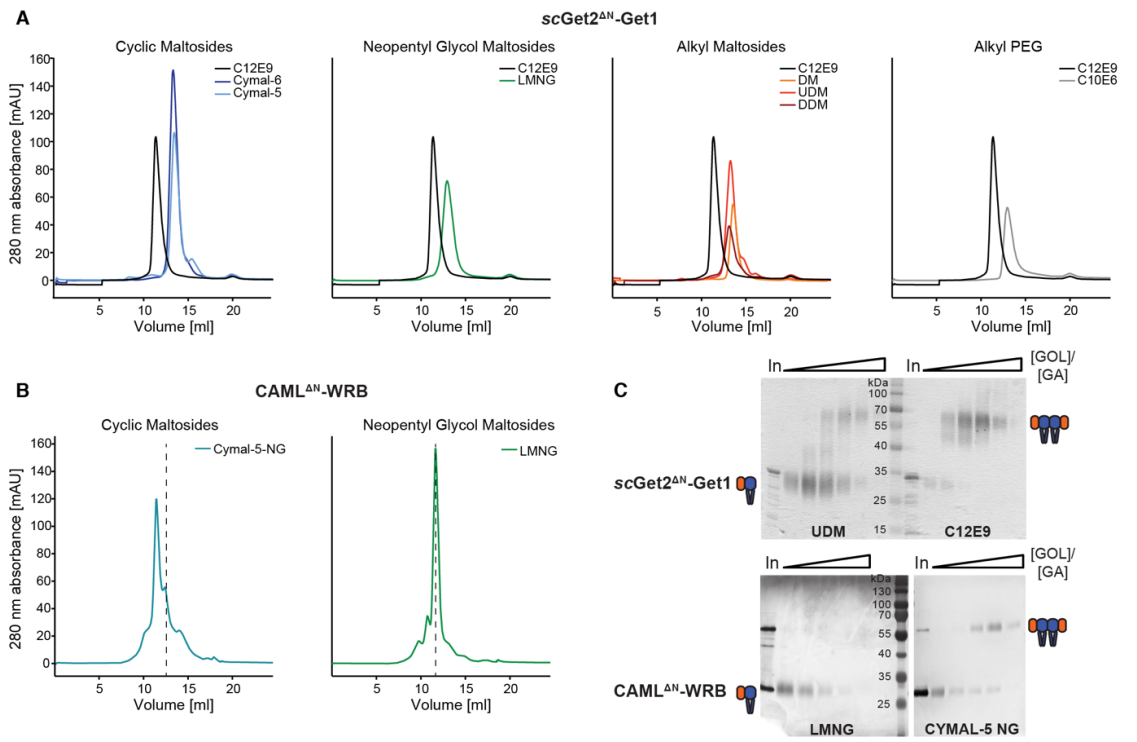


Fig. S5.5 *sc*Get2$^{\Delta N}$-Get1 and CAML$^{\Delta N}$-WRB form heterotetramers in C12E9 and Cymal5-NG respectively. Related to Fig. 5.4.

(a) Size exclusion chromatography for *sc*Get2$^{\Delta N}$-Get1 heterodimers in different detergents using a Superdex 200 increase 10/30 GL column shown relative to the elution profile for the *sc*Get2$^{\Delta N}$-Get1 heterotetramer in C12E9 (black). (b) Size exclusion chromatography for CAML$^{\Delta N}$-WRB in different detergents using a Superdex 200 increase 10/30 GL column. The dashed line denotes the elution volume for the heterodimeric fusion. (c) GraFix cross-linking of *sc*Get2$^{\Delta N}$-Get1 and CAML$^{\Delta N}$-WRB oligomers in the given detergents. Silver-stained 12.5% SDS-PAGE gels show the input (In) and fractions taken from an increasing gradient of glycerol (GOL) and glutaraldehyde (GA).

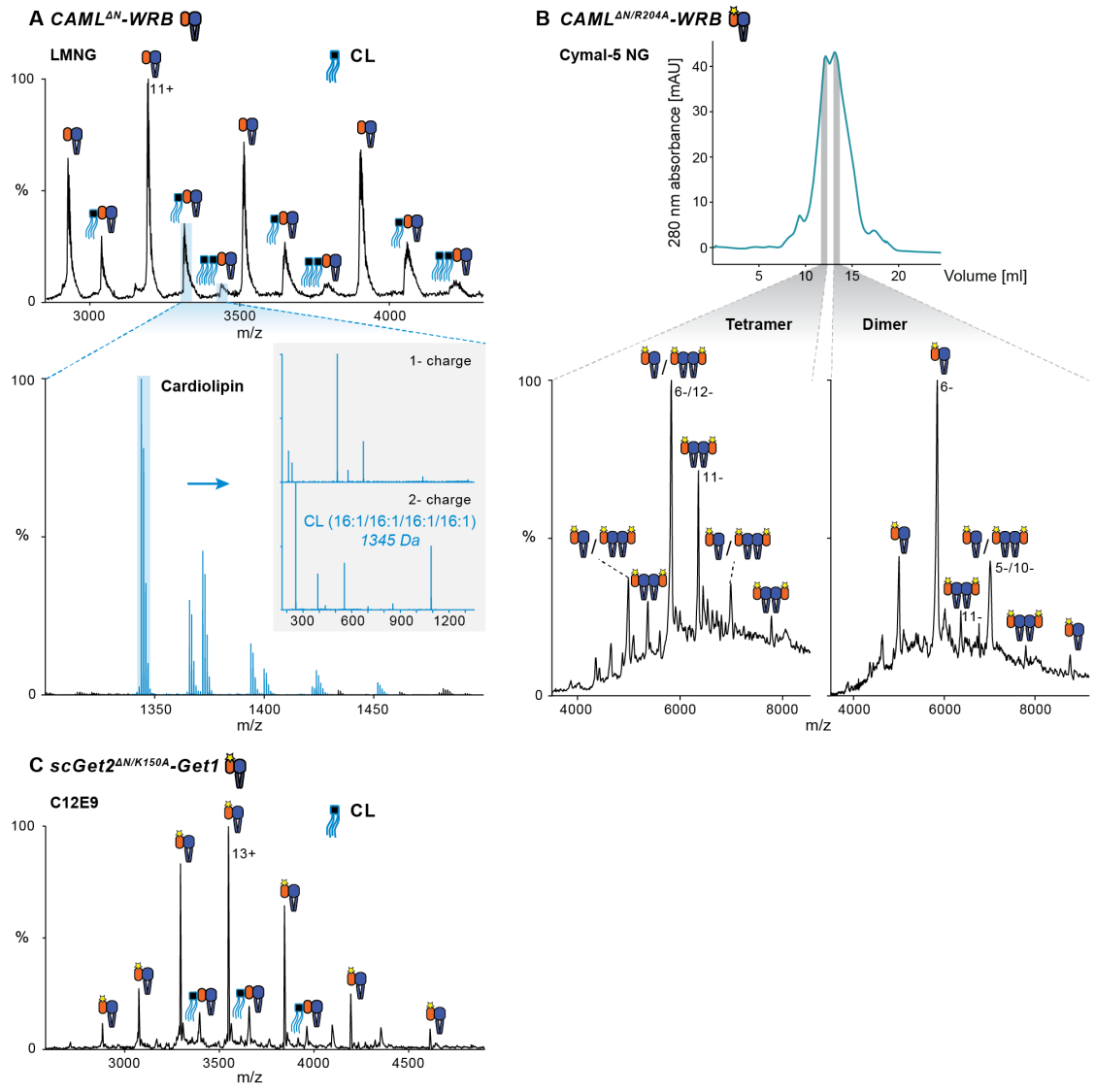


Fig. S5.6 CAML$^{\Delta N}$-WRB and *sc*Get2$^{\Delta N/K150A}$-Get1 heterodimers do not bind PI. Related to Fig. 5.4 and 5.5. (a) Zoomed region of the CAML$^{\Delta N}$-WRB mass spectrum from Fig. S5.3f showing the main charge state for the major species. Adduct peaks with a defined mass greater than the amino acid sequence mass correspond to bound CDL. MS profile for CDL species identified through in-solution lipidomics is shown beneath the spectrum. Inset, MS/MS of the highlighted ions identified the major CDL species associated with the protein. (b) Size exclusion chromatography for CAML$^{\Delta N/R204A}$-WRB in Cymal-5 NG using a Superdex 200 increase 10/30 GL column. Shown beneath are mass spectra collected under non-denaturing conditions for the indicated fractions (corresponding to tetrameric and dimeric species) liberated from detergent micelles. (c) Mass spectrum of *sc*Get2$^{\Delta N/K150A}$-Get1 liberated from C12E9 micelles under non-

denaturing conditions. (B, C) For each assigned complex the charge state for the most intense peak is shown, with the sign denoting data collection in either positive or negative mode.

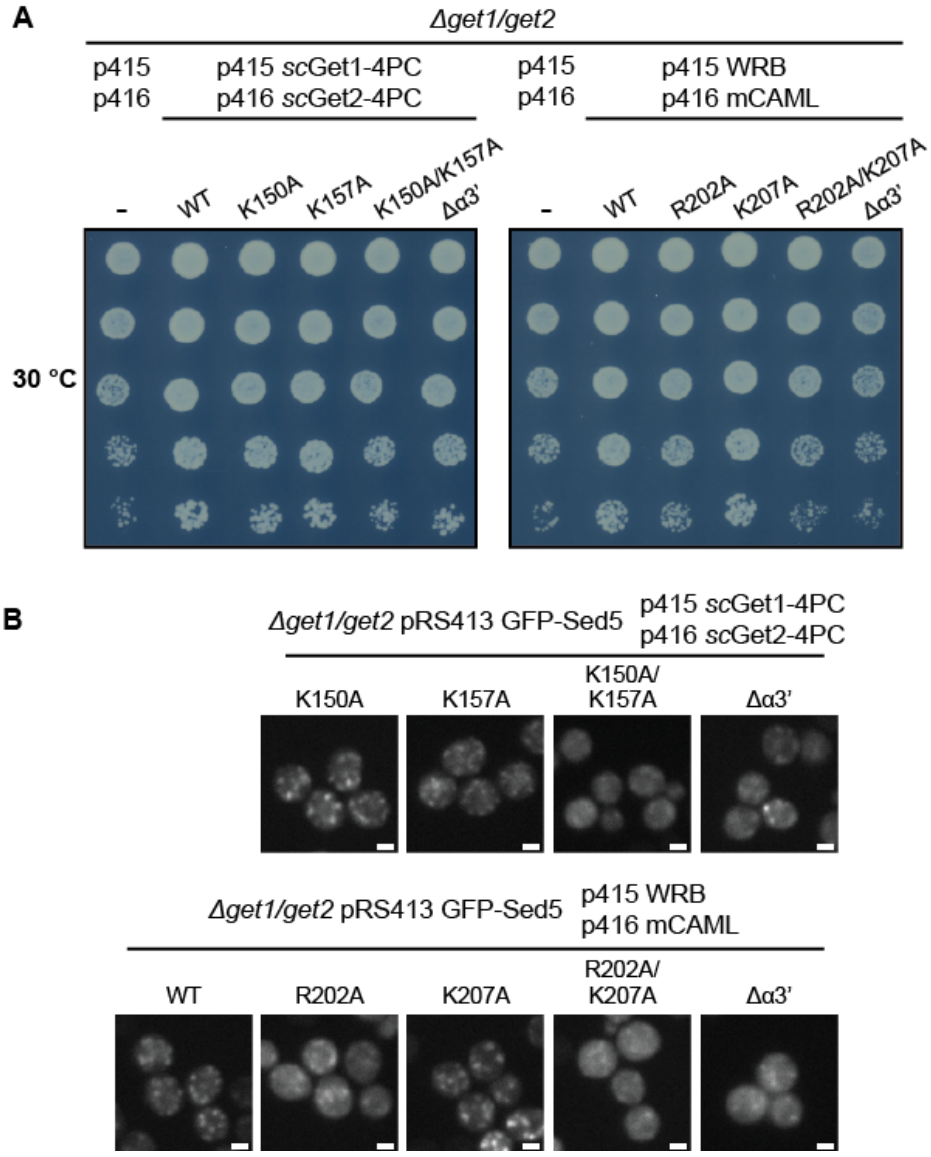


Fig. S5.7. Helix α3' and positively charged residues in the heterotetramer interface are important for *sc*Get1/Get2 and WRB/mCAML function *in vivo*. Related to Fig. 5.4, 5 and 6. (a), (b) *sc*Get1-4PC/Get2-4PC and WRB/mCAML were co-expressed from the indicated plasmids in *Δget1/get2* yeast strains for wild type (WT) and mutant *sc*Get2/mCAML sequences. Transformants with the empty vector were taken as a negative control (-). (a) Control plates from the spotting assay shown in Fig. 5.5b incubated at 30ºC, demonstrating comparable growth of the strains under normal growth conditions. (b) Representative fluorescence microscopy images

of *Δget1/get2* cells expressing GFP-Sed5 and *sc*Get1-4PC/Get2-4PC or WRB/mCAML constructs from separate plasmids. Scale bar: 2 μM.

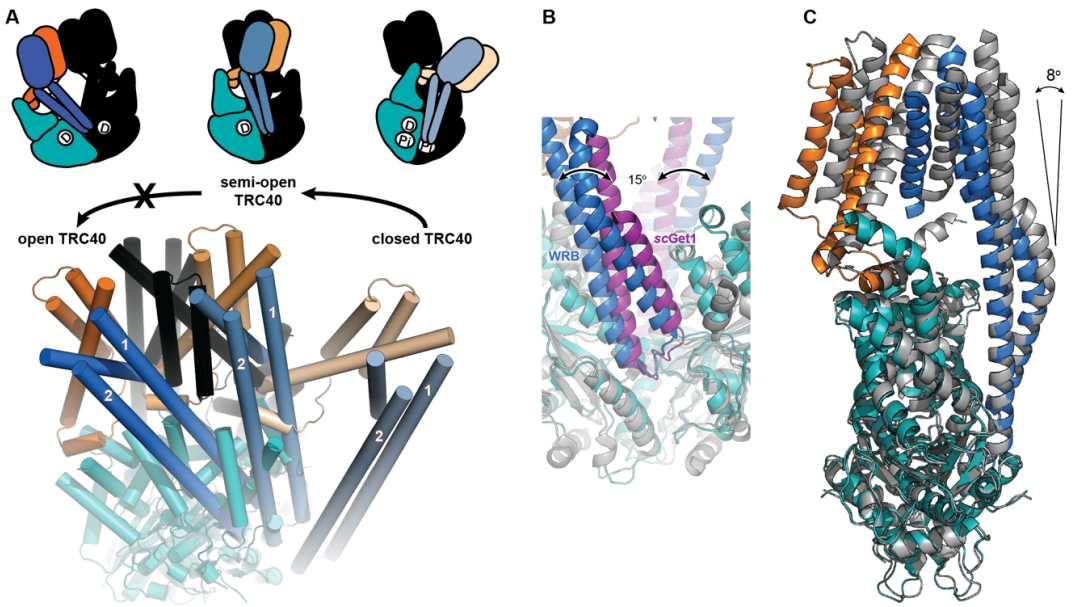


Fig. S5.8. Structural analysis of the WRB-CD interaction with TRC40. Related to Figures 5.3 and 5.7. (a) Model for CAML$^{\Delta N}$-WRB coupled to the closed to semi-open to open transition of TRC40. One CAML$^{\Delta N}$-WRB/TRC40 (black) is kept in a fixed position by first aligning semi-open and closed *sc*Get3 templates (PDB accession codes 3SJC and 4XTR) via chain A (RMSD 1.32 Å over 234 Cα atoms), then superimposing chain A of TRC40 (RMSD 1.72 Å over 230 Cα atoms) with the semi-open state. Chain B of TRC40 (RMSD 1.60 Å over 226 Cα atoms) from our model is then superimposed on semi-open and closed *sc*Get3. The adjoining CAML$^{\Delta N}$-WRB maintains interactions via interface I [18]. Arrows show the expected movement of the mobile CAML$^{\Delta N}$-WRB half. (b) Superimposition of the open *sc*Get3 dimer (grey; PDB accession 3SJB) with TRC40 (cyan; RMSD 2.31 Å over 456 Cα atoms), showing the tilt of bound *sc*Get1-CD relative to WRB. (c) Superimposition of one TRC40 on the other within the CAML$^{\Delta N}$-WRB/TRC40 complex (RMSD 1.01 Å over 271 Cα atoms). The superimposed TRC40 monomer and the CAML$^{\Delta N}$-WRB heterodimer interacting via interface I [18] are shown in grey for clarity. The given angle represents the difference in tilt of the WRB-CD between each dimer half.

5.4.2 Supplementary Tables

Figure	Complex	Experimental molecular mass ± s.d. (Da)	Theoretical molecular mass (Da)
	***sc*Get2$^{\Delta N}$-Get1**		
Figure 5.4b	(*sc*Get2$^{\Delta N}$-Get1)$_2$	92,493 ± 2	92,328
Figure 5.4c	(*sc*Get2$^{\Delta N}$-Get1)$_2$+ phospholipid	93,332 ± 11	N/A
Figure 5.5.4c	(*sc*Get2$^{\Delta N}$-Get1)$_2$ + cardiolipin	93,883 ± 8	N/A
Figure S5.3f	*sc*Get2$^{\Delta N}$-Get1	46,163 ± 9	46,164
Figure S5.3a	(*sc*Get2$^{\Delta N}$-Get1)$_2$/Get3$_2$	173,745 ± 10	173,600
Figure 5.2a	(*sc*Get2$^{\Delta N}$-Get1)$_2$/Get3$_2$	173,702 ± 3	173,600
	*sc*Get2$^{\Delta N/K150A}$-Get1		
Figure S5.6c	*sc*Get2$^{\Delta N/K150A}$-Get1	46,105 ± 1	46,164
Figure S5.6c	*sc*Get2$^{\Delta N/K150A}$-Get1+ cardiolipin	47,521 ± 18	N/A
	*sc*Get1/Get2		
Figure S5.3f	*sc*Get1/Get2	61,152 ± 12	61,136
Figure S5.3e	(*sc*Get1/Get2)$_2$/Get3$_2$	284,851 ± 37	284,816
Figure 5.2c	(*sc*Get1/Get2)$_2$/Get3$_2$	284,835 ± 12	284,816
	*sc*Get1/Get2/Get3$_2$	142,346 ± 6	142,408
	*sc*Get1$^{N72A/R73A}$/Get2		
Figure S5.3d	(*sc*Get1$^{N72A/R73A}$/Get2)$_2$/Get3$_2$	174,006 ± 20	173,917
	(*sc*Get1$^{N72A/R73A}$/Get2)$_2$/Get3$_2$	203,302 ± 48	203,287
	CAML$^{\Delta N}$-WRB		
Figure 5.4b	(CAML$^{\Delta N}$-WRB)$_2$	70,067 ± 26	70,095
Figure 5.4d	(CAML$^{\Delta N}$-WRB)$_2$+ phospholipid	70,893 ± 16	N/A
Figure S5.3f	CAML$^{\Delta N}$-WRB	35,090 ± 0.5	35,047
Figure S5.6a	CAML$^{\Delta N}$-WRB + cardiolipin	36,435 ± 0.2	N/A
Figure S5.3b	(CAML$^{\Delta N}$-WRB)$_2$/TRC40$_2$	150,548 ± 18	150,305
Figure 5.2b	(CAML$^{\Delta N}$-WRB)$_2$/TRC40$_2$	150,375 ± 46	150,305
	CAML$^{\Delta N/R204A}$-WRB		

Figure S5.6b	$(CAML^{\Delta N/R204A}\text{-}WRB)_2$	69,952	± 77	70,095
Figure S5.6b	$CAML^{\Delta N/R204A}\text{-}WRB$	35,077	± 33	35,047

Table S5.1 Molecular masses determined by native mass spectrometry. Related to Figures 5.2 and 5.4.

	H. sapiens $CAML^{\Delta N}$-WRB/ TRC40 in PMAL-C8 (EMDB-10266) (PDB 6SO5)	*S. cerevisiae* $Get2^{\Delta N}$-Get1/ Get3 in A835
Data collection and processing		
Detector	K2	K2
Magnification	165,000	128,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	46	25
Defocus range (μm)	0.8-2.0	1.5-2.5
Pixel size (Å)	0.81	1.08
Symmetry imposed	C1	C1
Initial particle images (no.)	1,698,096	150,857
Final particle images (no.)	225,244	27,666
Map resolution (Å)	4.2	14.0
FSC threshold	0.143	0.143
Map resolution range (Å)	3.8-6.0	10.2-29.3
Refinement		
Initial model used (PDB code)	3SJA	
Model resolution (Å)	4.2	
FSC threshold	0.143	
Model resolution range (Å)	3.8-6.0	
Map sharpening *B* factor (Å^2)	-219	
Model composition		

Non-hydrogen atoms	8737
Protein residues	1097
Ligands	ZN:1
B factors (Å^2)	
Protein	136
Ligand	128
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.809
Validation	
MolProbity score	2.41
Clashscore	23.32
Poor rotamers (%)	0.00
Ramachandran plot	
Favored (%)	89.79
Allowed (%)	10.21
Disallowed (%)	0.00

H. sapiens complex, model to map fit CC_mask = 0.79

Table S5.2 Cryo-EM data collection, refinement and validation statistics. Related to Figures 5.1 and 5.3.

5.5 Materials and Methods

5.5.1 Construct design

For recombinant protein expression, wild type *S. cerevisiae* Get1 and Get2 sequences were cloned into a 2μ vector harbouring a bi-directional GAL1/GAL10 promoter, TRP1, LEU2 and Amp resistance markers (pMT929; M. Thoms, Heidelberg) via the multiple cloning site upstream of GAL1. Wild type *H. sapiens* WRB and CAML sequences were cloned between BamHI/XhoI of the pFastBac1 vector (Thermo Fisher), which was then transformed into *E. coli* DH10Bac cells (Invitrogen) to generate a recombinant bacmid. For the *sc*Get2$^{\Delta N}$-Get1 construct, residues 139-285 of *sc*Get2 were linked to *sc*Get1 via a $(GS)_7$ sequence and cloned with a C-terminal 3C-His_8 tag. The CAML$^{\Delta N}$-WRB construct was designed similarly, except residues 186-296 of CAML were linked to WRB by a $(GS)_2$-TEV-$(GS)_2$ linker. For the expression of full-length proteins, *sc*Get1 was joined to *sc*Get2 with a thrombin-His_6-TEV linker. Wild type *sc*Get3 was cloned between NcoI/XhoI sites of the pET24d vector (Novagen), modified to incorporate a C-terminal Strep-tag II into the protein sequence. Wild type *H. sapiens* TRC40 was cloned into a pET24d vector modified to incorporate an N-terminal His_6-ZZ-TEV tag and a C-terminal Strep-tag II (G. Stier, Heidelberg).

For *S. cerevisiae* growth and functional assays, wild type *S. cerevisiae* Get1 and Get2 sequences were cloned into p415 and p416 MET25 promoter plasmids [58] respectively. The protein coding sequences were C-terminally fused to four copies of the protein C epitope (PC; Roche) to allow evaluation of relative expression levels. Wild type *H. sapiens* WRB and *Mus musculus* CAML sequences were also cloned into p415 and p416 vectors respectively without additional tags. p416 *sc*Get2/mCAML mutants were subsequently created using the Quikchange Lightning site-directed mutagenesis kit (Stratagene). In p416 mCAML$^{\Delta\alpha3'}$ residues 240-248 are replaced by a G_4 linker, whilst in p416 *sc*Get2$^{\Delta\alpha3'}$ residues 233-243 are replaced by a G_5 linker.

5.5.2 Strain growth conditions

For recombinant protein expression, pMT929 Get1/Get2 constructs were transformed into a *S. cerevisiae Δget3* strain and plated on synthetic complete (SD) -Leu plates for 72 h. A stationary SD -Leu pre-culture containing raffinose as the carbon source was diluted 12.5-fold into YP media containing 1% galactose and incubated at 30ºC for 24 h. Sf9 insect cells were cultured in Sf900-II SFM (Thermo Fisher) supplemented with 5% (v/v) EX-CELL TiterHigh (Sigma-Aldrich) and P1 virus was generated from cells transfected with the CAML$^{\Delta N}$-WRB-containing bacmid and Escort IV Transfection Reagent (Sigma-Aldrich) according to manufacturer's instructions. P2 virus was generated by infecting 50 ml log phase Sf9 cells and incubating them at 27ºC for 72 h. The virus containing solution was subsequently used to infect 300 ml cells at 1.5 × 10^6 cells/mL, which were incubated at 27ºC for 72 h to express CAML$^{\Delta N}$-WRB. *sc*Get3 was expressed in *E. coli* BL21 (DE3) (Novagen) overnight at 18ºC in 2x YT medium supplemented with 1 mM IPTG. TRC40 was expressed in *E. coli* Rosetta2 (DE3) (Novagen) by autoinduction [59] overnight at 22ºC. All cells were harvested by centrifugation and stored at -20ºC until further use.

For *S. cerevisiae* growth and functional assays, p415 Get1-4PC/p416 Get2-4PC or p415 WRB/p416 mCAML variants were co-transformed into a *Δget1/get2* strain ± the pRS413 GFP-Sed5 plasmid [60] or a *Δget1/get2 GET3*-GFP strain [13]. *Δget1/get2* cells were grown at 30ºC in SC -Ura -Leu and cells additionally expressing GFP-Sed5 or Get3-GFP in SC -Ura -Leu -His. For microsome preparation, *S. cerevisiae* wild type (BY4741), *Δget3* and *Δget1* + p416 Get1-4PC strains were grown in 1 l YPAD media until OD_{600nm} = 2.

5.5.3 Protein purification

For the purification of His_8-tagged Get1/Get2 homologues, *S. cerevisiae* cells were resuspended in lysis buffer A (50 mM HEPES (pH 8.0), 500 mM NaCl, 10% (v/v) glycerol, 1 mM EDTA) and lysed by beating with glass beads at 500 rpm for 3x 4 min. Sf9 cells were resuspended in lysis buffer B (lysis buffer A at pH 7.5 supplemented with 0.01% (v/v) 1-

thioglycerol) and lysed using a M-110L Microfluidiser (Microfluidics). All lysates were cleared by centrifugation at 1000 xg for 30 min and the cell membranes harvested and washed once by centrifugation at 160,000 xg for 30 min. The membranes were then solubilised for 30 min in membrane buffer (50 mM HEPES (pH 7.5), 500 mM NaCl) supplemented with the required detergent (0.5% C12E9, 1% Cymal-5, 1% Cymal-6, 1% DM, 1% UDM, 0.5% DDM, 0.5% LMNG, 0.5% CYMAL-5 NG or 0.5% C10E6; Anatrace) and, for CAML$^{\Delta N}$-WRB only, 0.01% (v/v) 1-thioglycerol. This detergent was present in all subsequent buffers at a concentration closer to its critical micelle concentration (CMC; 0.02% C12E9, 0.3% Cymal-5, 0.08% Cymal-6, 0.2% DM, 0.08% UDM, 0.03% DDM, 0.01% LMNG, 0.02% CYMAL-5 NG or 0.1% C10E6). The solubilised membranes were subsequently centrifuged at 160,000 xg for 30 min and the supernatant was incubated with TALON metal affinity resin (Clontech) for 30 min at 4ºC. This resin was washed with 30 column volumes (CV) Co^{2+} wash buffer (membrane buffer with 20 mM imidazole). The full-length Get1-Get2 fusion was eluted from the resin with 5 CV Co^{2+} elution buffer (membrane buffer with 150 mM imidazole). The eluate was then supplemented with 1 mg TEV protease, 0.3 mM oxidised glutathione and 3 mM reduced glutathione, applied to 10000 MWCO SnakeSkin Dialysis Tubing (Thermo Fisher) and dialysed overnight at 4ºC against 500 ml SEC buffer (20 mM HEPES pH 7.5, 200 mM NaCl, detergent). The *sc*Get2$^{\Delta N}$-Get1 and CAML$^{\Delta N}$-WRB constructs were eluted from the TALON resin by on-column cleavage via incubation with 5 mg His_6-tagged HRV-3C Protease in 5 CV Co^{2+} wash buffer for 1 h at 4ºC. All cleaved membrane proteins were then subjected to SEC with a Superdex S200 increase 10/30 GL column (GE Healthcare) equilibrated with SEC buffer, supplemented with 1 mM TCEP for CAML$^{\Delta N}$-WRB. Protein-containing fractions were concentrated to 10 μM for nMS with a VivaSpin Protein Concentrator MWCO 50000 (GE Healthcare).

For the purification of Get3 homologues, *E. coli* cells were resuspended in lysis buffer C (50 mM HEPES (pH 7.5), 500 mM NaCl) and lysed as for Sf9 cells. Clarified *sc*Get3-Strep lysate was applied to 2x 5 ml StrepTrap HP columns (GE Healthcare). The resin was washed with 20 CV lysis buffer C and eluted in 1 CV lysis buffer C supplemented with 2.5 mM

desthiobiotin. Clarified His_6-ZZ-TRC40-Strep lysate was applied to a 5 ml HisTrap HP column (GE Healthcare). The resin was washed with 20 CV Ni^{2+} wash buffer (lysis buffer C with 50 mM imidazole) and eluted with 1 CV Ni^{2+} elution buffer (lysis buffer C with 250 mM imidazole). The His_6-ZZ tag was cleaved from TRC40 by dialysis with TEV protease (see above) and removed by a reverse HisTrap HP column. All constructs were then subjected to SEC with a HiLoad 16/60 Superdex 200 pg (GE Healthcare) column equilibrated in 20 mM HEPES (pH 7.5), 150 mM NaCl and protein-containing fractions were concentrated using a 50 kDa Amicon-Ultra centrifugal filter (Millipore).

For purification of complexes, the TALON resin bound to *sc*Get2$^{\Delta N}$-Get1 or CAML$^{\Delta N}$-WRB was incubated with 5 mg of the corresponding purified Strep-tagged Get3 homologue in 5 CV Co^{2+} wash buffer for 1 h at 4ºC and then washed with 15 CV Co^{2+} wash buffer, before on-column cleavage. The flow-through was then applied to a 1 ml StrepTrap HP column (GE Healthcare), which was washed with 30 CV membrane buffer prior to elution with 10 CV membrane buffer supplemented with 2.5 mM desthiobiotin. For nMS, the eluate was then purified by SEC as previously and protein-containing fractions concentrated to 10 μM with a VivaSpin Protein Concentrator MWCO 50000 (GE Healthcare). For preparation of cryo-EM samples, eluates for *sc*Get2$^{\Delta N}$-Get1 in C12E9 and CAML$^{\Delta N}$-WRB in LMNG were incubated with A835 (Jena Bioscience) or PMAL-C8 (Anatrace) amphipol respectively at a 3x weight excess over the membrane protein fusion for 2 h at 4ºC, followed by 1% (w/v) α-cyclodextrin (Sigma Aldrich) overnight at 4ºC. SEC was performed with a Superose 6 10/30 GL (GE Healthcare) equilibrated in 20 mM HEPES pH7.5, 200 mM NaCl. For native MS of full-length *sc*Get1/Get2/Get3 complexes, purified Get1/Get2 was supplemented directly with Strep-tagged *sc*Get3 at a 1-2 molar excess.

5.5.4 SMALP preparation

Membranes containing His_8-tagged *sc*Get2$^{\Delta N}$-Get1 were resuspended in membrane buffer and solubilised for 30 min at 4ºC with 2.5% (v/v) SMA 2.3:1 (XIRAN SL30010-S30,

Polyscope Polymers), which had been dialysed overnight against 1 l membrane buffer. The solubilised membranes were centrifuged at 160,000 xg for 30 min and the supernatant incubated with Ni-NTA Agarose (Qiagen) for 30 min at 4ºC. The resin was washed with 30 CV membrane buffer supplemented with 30 mM imidazole and eluted with 5 CV membrane buffer supplemented with 250 mM imidazole. The eluate was applied to 10000 MWCO SnakeSkin Dialysis Tubing (Thermo Fisher) and dialysed overnight at 4ºC against 1 l 20 mM HEPES pH7.5, 200 mM NaCl.

5.5.5 GraFix cross-linking

GraFix crosslinking was carried out as previously described [31]. Briefly, 4 ml gradients of 5-20% glycerol and 0-0.075% glutaraldehyde in a uniform background of the final purification buffer containing the required detergent were prepared using a Gradient Master (Biocomp). 200 µl each of a cushion of 2.5% glycerol in buffer and the purified protein sample were then applied to the top of the gradient. Gradients were then subjected to centrifugation at 215,000 xg for 18 h. 200 µl fractions were extracted from the top of the gradient, quenched with 50 mM Tris (pH 7.5) and subjected to a Pierce Coomassie Protein Assay (Thermo Fisher) to identify protein-containing fractions.

5.5.6 Microsome preparation and BN PAGE analysis

Cells were washed briefly with 100 mM Tris-HCl (pH 9.4) and incubated with spheroplast buffer (10 mM Tris-HCl (pH 7.4), 0.7 M sorbitol, 1.5% (w/v) bacto peptone, 0.75% (w/v) yeast extract, 0.5% (w/v) glucose) supplemented with 600 U zymolylase for 1.5 h at 37ºC under constant shaking. The resulting spheroplasts were applied to a sucrose cushion and centrifuged at 5,500 xg for 20 min. The pellet was thoroughly resuspended in 1/3 volume cold homogenisation buffer (20 mM HEPES (pH 7.4), 0.1 M sorbitol, 50 mM potassium acetate, 2 mM EDTA, 1 mM PMSF, 25 mM N-ethylmaleimide (NEM)), the supernatant extracted after re-pelleting at 3,000 xg and the process repeated. The pooled supernatants were centrifuged at

27,000 xg for 10 min and the resulting membranes resuspended in 3 ml homogenisation buffer before application to a 9 ml two-layer 1.2 M/1.5 M sucrose gradient in a uniform background of 20 mM HEPES (pH 7.4). After centrifugation at 100,000 xg for 1 h, microsomes were collected from the two-layer interface, centrifuged at 27,000 xg for 10 min, washed twice with 20 mM HEPES (pH 6.8), 0.25 M sorbitol, 150 mM potassium acetate, 5 mM magnesium acetate, 25 mM NEM and finally resuspended in 300 µl 50 mM HEPES (pH 7.5), 30 mM NaCl, 3.75 mM $MgCl_2$, 2.5 mM EGTA, 25% (v/v) glycerol, 1.25 M 6-aminohexanoic acid, 25 mM NEM. To prepare samples for BN-PAGE, microsomes were solubilised on ice with 1% digitonin for 30 min and 50 µg/lane microsomal protein was applied to gels cast with the Mighty Small II Mini Vertical Electrophoresis Unit (Hoefer). BN-PAGE was performed using 5-13% acrylamide gradient gels at 4ºC as described previously [61] and transferred on to a Hybond PVDF membrane (Amersham) activated in methanol. Membranes were incubated with anti-Get3 [62] or anti-PC (Roche) primary antibodies, then with HRP-conjugated secondary antibodies, and detected using the ECL system (Amersham).

5.5.7 Western blotting of yeast lysates

For preparation of whole cell lysates a previously described NaOH lysis protocol was adapted [63]. Briefly, 750 μl cells in the logarithmic phase were pelleted, then resuspended in 1 mL 250 mM NaOH. Samples were incubated on ice for 10 min, 100 μL of the solution was used to determine their optical density. After incubation, samples were centrifuged for 1 min at 5 000 g, and normalised by resuspending the pellets in one-tenth mL SDS-PAGE sample buffer of their optical density. After incubating the samples for 5 min at room temperature, they were centrifuged for 1 min at 5 000 g. 8 μL of the supernatant was separated by SDS-PAGE and transferred onto a Protran nitrocellulose membrane (Amershan). Membranes were incubated with the indicated primary antibodies, then with fluorescent secondary antibodies, and detected in a LI-COR Odyssey scanner. Images of immunoblots were visualised in ImageStudioLite 5.2 (LI-COR).

5.5.8 Plate growth assays

Strains were grown to logarithmic phase, then normalised to OD_{600} 0.8 and resuspended in water. 5 μl of a 1:5 dilution series was spotted onto synthetic dropout plates lacking uracil and leucine, which were incubated at 30 °C or 37 °C.

5.5.9 Microscopy of fluorescent strains

For GFP-Sed5 expression cells, exponentially growing cells were diluted into 384-well glass-bottom microtiter plates (Matriplate, Brooks Life Science Systems) containing low fluorescence minimal media (FORMEDIUM Ltd.) supplemented with methionine. The plates were automatically imaged on an Imaging Machine 03-dual (Acquifer) widefield high-content screening microscope, equipped with a white LED array for bright field imaging, an LED fluorescence excitation light source, a sCMOS camera (2048 x 2048 pixels) and a stationary plate holder in combination with movable optics. Images were acquired with 470 nm filter cubes (Ex 469/35, Em 525/39, dichroic 497) or without filter cube for brightfield images using 3 z-slices (dz = 1 µ m) and a 40x CFI Super Plan Fluor ELWD N.A. 0.60 (Nikon) objective. The focal plane was automatically detected in the brightfield channel using an autofocus algorithm.
Cells expressing Get3-GFP from overnight cultures were diluted and imaged at room temperature with a Nikon Ti2 2-E inverted microscope equipped with a computer-controlled stage, a Lumencor Spectra X light source, a pco.edge 5.5 M-AIR-CL-PCO sCMOS camera (2560 x 2156 pixels). Focal plane was detected with the Perfect Focus System (Nikon). Images were acquired with a 100x CFI Plan Apo Lambda 100x/1.45 oil objective using settings appropriate for GFP (Ex 470/24, Em 520/35, Dichroic 488).

5.5.10 Native MS

Protein samples were buffer exchanged into 200 mM ammonium acetate (pH 8.0) supplemented with 2x CMC of the respective detergent using Micro Bio-Spin 6 columns (BioRad) prior to mass spectrometry analyses. These samples were directly introduced into the

mass spectrometer using gold-coated capillary needles prepared in-house [64]. Data were collected on a modified Q-Exactive mass spectrometer [65] and a Q-Exactive UHMR Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher). The instrument parameters used for data collection were: capillary voltage 1.2 kV in positive mode and 0.9 kV in negative mode, S-lens RF 100%, quadrupole selection from 1,000 to 20,000 *m/z* range, HCD collision energy 150-250 V, source fragmentation 0-100 V, in-source trapping -100 V to -250 V for analysis in positive mode and 100-250 V in negative mode. The ion transfer optics in positive mode was set as follows: injection flatapole 5 V, inter-flatapole lens 4 V, bent flatapole 2 V, transfer multipole 0 V; opposite polarities were used for negative mode acquisition. The resolution of the instrument was 17,500 at *m/z*=200 (transient time of 64 ms), argon pressure in the HCD cell was maintained at approximately at 8×10^{-10} mbar and source temperature was kept at 100°C. The noise level was set at 3 rather than the default value of 4.64. Calibration of the instruments was performed using 10 mg/ml solution of caesium iodide in water. Where required, baseline subtraction was performed to achieve a better-quality mass spectrum.

5.5.11 Lipidomics analysis

Co-purified lipids have been identified in a similar manner as described previously [66]. Briefly, proteins were digested with trypsin (1:20 units) for 15 h at 37°C in a thermomixer (Eppendorf). The digest was lyophilised and re-dissolved in 68% solution A (ACN:H2O 60:40, 10 mM ammonium formate and 0.1% formic acid) and 32% solution B (IPA:ACN 90:10, 10 mM ammonium formate and 0.1% formic acid) for analysis by reverse phase liquid chromatography tandem MS (RP LC-MS/MS). LC-MS/MS was performed using a Dionex UltiMate 3000 RSLC Nano system coupled to an LTQ Orbitrap XL hybrid mass spectrometer (Thermo Fisher). The tryptic digest mixture was loaded onto a pre-equilibrated C18 trap (Acclaim PepMap μ-Precolumn 300 μm, 5 mm) followed by a C18 column (Acclaim PepMap 100, C18, 75 μm × 15 cm; Thermo Fisher) and separated by using the following step gradient of solution A and B at 300 nl/min and 40ºC: 0-10 min with 32% solvent B, solvent B was ramped to 65% over 1 min,

then concentration of solvent B was increased to 80% buffer over 6 min and held for 10 min, then ramped to 99% B over 6 min and this concentration was held for 7 min. The nanoflow reversed-phase liquid chromatography was directly coupled to the mass spectrometer via a dynamic nanospray source. Typical MS conditions were spray voltage of 1.6 kV and capillary temperature of 275°C. The LTQ-Orbitrap XL was set-up in negative ion mode and in data-dependent acquisition mode to perform five MS/MS scans per MS scan. Survey full-scan MS spectra were acquired in the orbitrap (*m/z* 350 – 2,000) with a resolution of 60,000. Collision-induced dissociation (CID) fragmentation in the linear ion trap was performed for the five most intense ions at an automatic gain control target of 30,000 and a normalised collision energy of 38% at an activation of q = 0.25 and an activation time of 30 ms. Data were analysed using Xcalibur 3.0 (Thermo Fisher).

5.5.12 Cryo-EM grid preparation and data collection

Purified *sc*Get2$^{\Delta N}$-Get1/Get3 or CAML$^{\Delta N}$-WRB/TRC40 complexes in A835 or PMAL-C8 amphipol respectively were concentrated to 1-2 mg/ml with a VivaSpin Protein Concentrator MWCO 100000 (GE Healthcare). 3 µl sample was applied to a glow-discharged holey carbon-coated grid (Quantifoil 300 mesh, Cu R2/1), adsorbed for 20s, blotted for 5 s at 95% humidity and 6ºC before plunge-freezing in liquid ethane using a Vitrobot Mark IV (Thermo Fisher). All data were collected in counting mode on a 300 kV Titan Krios (Thermo Fisher). For *sc*Get2$^{\Delta N}$-Get1/Get3, 678 movies were collected on a K2 detector (Gatan), at 1.08 Å/pixel and 0.7 e^-/pixel/frame over 40 frames and 4s total exposure. 9470 images were collected for the CAML$^{\Delta N}$-WRB/TRC40 dataset at 0.81 Å/pixel and 0.6 e^-/pixel/frame over 40 frames and 12 s total exposure.

5.5.13 Quantification and statistical analysis

Quantification of GFP-Sed5 images. A script was written in KNIME (https://www.knime.com) to automatically identify cells and quantify the distribution of the fluorescence within them similarly to a method described previously [60]. In short, using an ImageJ [67] macro in KNIME, the difference of two brightfield images (± 1 μM from focal plane) was used to identify segments (cells). Background areas containing no irregularities were identified based on areas of the image with low variance. Cells were excluded if they overlapped with the calculated background area. Cells were also filtered for circularity, variance in the bright field signal within the cell, the range of bright field signal values within cells in order to remove cells from analysis that contained irregularities or were stacked on each other. Outliers in size, i.e. too big or small cells, were also excluded. Fluorescent channel background was estimated based on the fluorescent signal observed in the background areas identified with the brightfield images, then subtracted from the fluorescent channel image. Cells identified in the brightfield channels, as described above, were used to calculate the pixel distribution for each cell in the fluorescent channel. Outliers with too little variance in the fluorescent signal or too high signal intensity were removed in order to exclude dead cells. Cells with low signal intensity were also removed to exclude cells not expressing GFP-Sed5. The ratio of the maximum and the average pixel intensity of each cell was determined and averaged for all cells in the image. Six images were quantified as described above for each strain, each containing several hundreds of cells. Error bars represent average ± standard error of the mean. Statistical significance was determined using the Welch's *t*-test. The KNIME workflow containing the script is available upon request.

Native MS analysis. Data were analysed using the Xcalibur 3.0 (Thermo Fisher) and UniDec [68] (unidec.chem.ox.ac.uk) software packages. For each assigned complex, a comparison of experimental and theoretical molecular mass is given in Table S5.1.

Image processing. *sc*Get2$^{\Delta N}$-Get1/Get3 images were motion-corrected using MotionCorr2 [69] and contrast transfer function (CTF) was estimated using GCTF [70]. 6020 particles were manually selected in EMAN2.2 [71] and their 2D class averages generated in Relion2.2 [72] were used for automatic picking with GAUTOMATCH (K. Zhang,

http://www.mrc-lmb.cam.ac.uk/kzhang/). All further processing was carried out in Relion2.2 [72]. 150,857 particles were extracted with a 200 x 200 Å box size and subjected to 2D classification to select 27,666 good particles. An *ab initio* initial model was directly used in 3D auto-refinement and postprocessing to produce the final reconstruction at 14 Å resolution, as estimated by the gold-standard FSC cutoff value of 0.143. CAML$^{\Delta N}$-WRB/TRC40 images were processed using Relion3.0 [73]. Motion correction was carried out with dose-weighting using the Relion3.0 implementation of MotionCorr2 [69] and CTF estimation was performed via GCTF [70]. 1,698,096 particles were automatically picked during data collection using Warp [74], then filtered to 1,118,394 good particles by 2D classification. These were subjected to 3D classification using an oblate spheroid as an initial model and the best class after two rounds of classification (225,244 particles) was subjected to 3D auto-refinement and postprocessing to give a 4.2 Å final reconstruction. Subsequent low-pass filtering of this reconstruction to 6.5 Å using Relion 3.0 yielded more continuous density for flexible, low resolution regions. Local resolution was estimated using Relion's local postprocessing implementation. The processing workflow for both reconstructions is illustrated in Fig. S5.2.

Structural modelling, refinement and analysis. The 6.5 Å low-pass filtered reconstruction for CAML$^{\Delta N}$-WRB/TRC40 was first used to assign the secondary structure elements of the TRC40 dimer and CAML$^{\Delta N}$-WRB heterotetramer, then the 4.2 Å reconstruction was used for model building and refinement. Initially, high-resolution structures of *sc*Get3 and *sc*Get1-CD (PDB accession 3SJA) were jiggle-fit into the density within Coot [75] then manually mutated to the homologous TRC40 and WRB-CD sequences respectively based on published sequence alignments [18]. TRC40 loop and helical deviations were manually repositioned to fit the density, most notably for α4-α9 within the TABD. WRB TMD1/TMD2 were built by helical extension of the CD. The remaining WRB and CAML helices were modelled by placing ideal helices in the density and assigning the sequence based on sequence alignment with PRALINE [54], TMD prediction with Phobius [57] and secondary structure prediction with PSIPRED [56].

Cysteines within CAML TMD1/TMD3 and WRB TMD3 aligned with putative density for disulphide bonds and the unassigned residues of CAML fully accounted for the remaining luminal and cytoplasmic loop density, providing confidence in helical register.

Real-space refinement in Phenix [76] with secondary structure and geometry restraints produced the final model (Table S5.2). Although a membrane protein fusion was used, the amino acid residues are numbered relative to the start of the native WRB/CAML sequences. Structural figures were prepared in UCSF Chimera [77] and PyMOL (Schrödinger, LLC). Superimpositions were performed with Superpose [78].

Data and code availability. The cryo-EM volumes have been deposited in the Electron Microscopy Data Bank under accession codes EMD-10266 (*H. sapiens*) and the coordinates have been deposited in the Protein Data Bank under accession number 6SO5.

5.6 References

1. Kalbfleisch, T., A. Cambon, and B.W. Wattenberg, *A Bioinformatics Approach to Identifying Tail-Anchored Proteins in the Human Genome*. Traffic, 2007. **8**(12): p. 1687-1694.
2. Borgese, N., et al., *The Ways of Tails: the GET Pathway and more*. Protein J., 2019. **38**(3): p. 289-305.
3. Mateja, A. and R.J. Keenan, *A structural perspective on tail-anchored protein biogenesis by the GET pathway*. Current Opinion in Structural Biology, 2018. **51**: p. 195-202.
4. Shan, S.-o., *Guiding tail-anchored membrane proteins to the endoplasmic reticulum in a chaperone cascade*. Journal of Biological Chemistry, 2019. **294**(45): p. 16577-16586.
5. Aviram, N., et al., *The SND Proteins Constitute an Alternative Targeting Route to the Endoplasmic Reticulum*. Nature, 2016. **540**(7631): p. 134-138.
6. Guna, A., et al., *The ER Membrane Protein Complex Is a Transmembrane Domain Insertase*. Science, 2018. **359**(6374): p. 470-473.
7. Bozkurt, G., et al., *Structural Insights Into Tail-Anchored Protein Binding and Membrane Insertion by Get3*. Proc. Natl. Acad. Sci. U.S.A., 2009. **106**(50): p. 21131-21136.
8. Hu, J., et al., *The Crystal Structures of Yeast Get3 Suggest a Mechanism for Tail-Anchored Protein Membrane Insertion*. PLoS ONE, 2009. **4**(11): p. e8061.

9. Mateja, A., et al., *The structural basis of tail-anchored membrane protein recognition by Get3*. Nature, 2009. **461**(7262): p. 361-366.
10. Rome, M.E., et al., *Differential gradients of interaction affinities drive efficient targeting and recycling in the GET pathway*. Proceedings of the National Academy of Sciences, 2014. **111**(46): p. E4929-E4935.
11. Suloway, C.J.M., et al., *Model for eukaryotic tail-anchored protein binding based on the structure of Get3*. Proceedings of the National Academy of Sciences, 2009. **106**(35): p. 14849-14854.
12. Mateja, A., et al., *Structure of the Get3 targeting factor in complex with its membrane protein cargo*. Science, 2015. **347**(6226): p. 1152-1155.
13. Schuldiner, M., et al., *The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane*. Cell, 2008. **134**(4): p. 634-645.
14. Wang, F., et al., *The Mechanism of Tail-Anchored Protein Insertion into the ER Membrane*. Molecular Cell, 2011. **43**(5): p. 738-750.
15. Vilardi, F., H. Lorenz, and B. Dobberstein, *WRB is the receptor for TRC40/Asna1-mediated insertion of tail-anchored proteins into the ER membrane*. Journal of Cell Science, 2011. **124**(8): p. 1301-1307.
16. Yamamoto, Y. and T. Sakisaka, *Molecular Machinery for Insertion of Tail-Anchored Membrane Proteins into the Endoplasmic Reticulum Membrane in Mammalian Cells*. Molecular Cell, 2012. **48**(3): p. 387-397.
17. Mariappan, M., et al., *The mechanism of membrane-associated steps in tail-anchored protein insertion*. Nature, 2011. **477**(7362): p. 61-66.
18. Stefer, S., et al., *Structural Basis for Tail-Anchored Membrane Protein Biogenesis by the Get3-Receptor Complex*. Science, 2011. **333**(6043): p. 758-762.
19. Vilardi, F., et al., *WRB and CAML Are Necessary and Sufficient to Mediate Tail-Anchored Protein Targeting to the ER Membrane*. PLoS ONE, 2014. **9**(1): p. e85033.
20. Zalisko, B.E., et al., *Tail-Anchored Protein Insertion by a Single Get1/2 Heterodimer*. Cell Reports, 2017. **20**(10): p. 2287-2293.
21. Wang, F., et al., *The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase*. Nature, 2014. **512**(7515): p. 441-444.
22. Anghel, S.A., et al., *Identification of Oxa1 Homologs Operating in the Eukaryotic Endoplasmic Reticulum*. Cell Reports, 2017. **21**(13): p. 3708-3716.
23. Kumazaki, K., et al., *Structural basis of Sec-independent membrane protein insertion by YidC*. Nature, 2014. **509**(7501): p. 516-520.
24. Kumazaki, K., et al., *Crystal structure of Escherichia coli YidC, a membrane protein chaperone and insertase*. Scientific Reports, 2015. **4**(1): p. 7299.

25. Pleiner, T., et al., *Structural basis for membrane insertion by the human ER membrane protein complex*. Science, 2020: p. eabb5008.
26. Xin, Y., et al., *Structure of YidC from Thermotoga maritima and its implications for YidC-mediated membrane protein insertion*. The FASEB Journal, 2018. **32**(5): p. 2411-2421.
27. Kubota, K., et al., *Get1 Stabilizes an Open Dimer Conformation of Get3 ATPase by Binding Two Distinct Interfaces*. Journal of Molecular Biology, 2012. **422**(3): p. 366-375.
28. Schmid, V., *The function of the GET pathway in tail-anchored protein biogenesis*. 2009, University of Heidelberg: HEIDI, https://katalog.ub.uni-heidelberg.de/titel/66884722
29. Wickles, S., et al., *A structural model of the active ribosome-bound membrane protein insertase YidC*. eLife, 2014. **3**: p. e03035.
30. Gupta, K., et al., *The Role of Interfacial Lipids in Stabilizing Membrane Protein Oligomers*. Nature, 2017. **541**(7637): p. 421-424.
31. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy*. Nature Methods, 2008. **5**(1): p. 53-55.
32. Laganowsky, A., et al., *Mass spectrometry of intact membrane protein complexes*. Nat. Protoc., 2013. **8**(4): p. 639-651.
33. Zinser, E., et al., *Phospholipid synthesis and lipid composition of subcellular membranes in the unicellular eukaryote Saccharomyces cerevisiae*. Journal of Bacteriology, 1991. **173**(6): p. 2026-2034.
34. Joshi, A.S., N. Fei, and M.L. Greenberg, *Get1p and Get2p are required for maintenance of mitochondrial morphology and normal cardiolipin levels*. FEMS Yeast Research, 2016: p. fow019.
35. Onishi, M., et al., *The ER membrane insertase Get1/2 is required for efficient mitophagy in yeast*. Biochemical and Biophysical Research Communications, 2018. **503**(1): p. 14-20.
36. Yeagle, P.L., *Non-covalent binding of membrane lipids to membrane proteins*. Biochimica et Biophysica Acta (BBA) - Biomembranes, 2014. **1838**(6): p. 1548-1559.
37. Carvalho, H.J.F., et al., *The WRB Subunit of the Get3 Receptor is Required for the Correct Integration of its Partner CAML into the ER*. Scientific Reports, 2019. **9**(1): p. 11887.
38. Giménez-Andrés, M., A. Čopič, and B. Antonny, *The Many Faces of Amphipathic Helices*. Biomolecules, 2018. **8**(3): p. 45.
39. Opaliński, Ł., et al., *Membrane curvature during peroxisome fission requires Pex11: Pex11-induced membrane curvature*. The EMBO Journal, 2011. **30**(1): p. 5-16.

40. Chen, Y., et al., *YidC Insertase of Escherichia coli: Water Accessibility and Membrane Shaping*. Structure, 2017. **25**(9): p. 1403-1414.e3.
41. Boy, D. and H.G. Koch, *Visualization of Distinct Entities of the SecYEG Translocon During Translocation and Integration of Bacterial Proteins*. Mol. Biol. Cell, 2009. **20**(6): p. 1804-1815.
42. Kohler, R., et al., *YidC and Oxa1 Form Dimeric Insertion Pores on the Translating Ribosome*. Molecular Cell, 2009. **34**(3): p. 344-353.
43. Spann, D., et al., *Each protomer of a dimeric YidC functions as a single membrane insertase*. Scientific Reports, 2018. **8**(1): p. 589.
44. Gold, V.A.M., et al., *The action of cardiolipin on the bacterial translocon*. Proceedings of the National Academy of Sciences, 2010. **107**(22): p. 10044-10049.
45. Tucker, K. and E. Park, *Cryo-EM structure of the mitochondrial protein-import channel TOM complex at near-atomic resolution*. Nature Structural & Molecular Biology, 2019. **26**(12): p. 1158-1166.
46. Dorsch, S., et al., *Analysis of receptor oligomerization by FRAP microscopy*. Nature Methods, 2009. **6**(3): p. 225-230.
47. Milligan, G., R.J. Ward, and S. Marsango, *GPCR homo-oligomerization*. Current Opinion in Cell Biology, 2019. **57**: p. 40-47.
48. Arkhipov, A., et al., *Architecture and Membrane Interactions of the EGF Receptor*. Cell, 2013. **152**(3): p. 557-569.
49. Denic, V., V. Dotsch, and I. Sinning, *Endoplasmic Reticulum Targeting and Insertion of Tail-Anchored Membrane Proteins by the GET Pathway*. Cold Spring Harbor Perspectives in Biology, 2013. **5**(8): p. a013334-a013334.
50. Locher, K.P., *Mechanistic diversity in ATP-binding cassette (ABC) transporters*. Nature Structural & Molecular Biology, 2016. **23**: p. 487-493.
51. Van den Berg, B., et al., *X-ray Structure of a Protein-Conducting Channel*. Nature, 2004. **427**(6969): p. 36-44.
52. Maillard, A.P., et al., *Deregulation of the SecYEG Translocation Channel upon Removal of the Plug Domain*. Journal of Biological Chemistry, 2007. **282**(2): p. 1281-1287.
53. Rollauer, S.E., et al., *Structure of the TatC core of the twin-arginine protein transport system*. Nature, 2012. **492**(7428): p. 210-214.
54. Pirovano, W., K.A. Feenstra, and J. Heringa, *PRALINETM: a strategy for improved multiple alignment of transmembrane proteins*. Bioinformatics, 2008. **24**(4): p. 492-497.
55. Gouet, P., X. Robert, and E. Courcelle, *ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins*. . Nucleic Acids Research 2003. **31**: p. 3320-3323.

56. Buchan, D.W.A. and D.T. Jones, *The PSIPRED Protein Analysis Workbench: 20 years on*. Nucleic Acids Research, 2019. **47**(W1): p. W402-W407.

57. Kall, L., A. Krogh, and E.L.L. Sonnhammer, *Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server*. Nucleic Acids Research, 2007. **35**(Web Server): p. W429-W432.

58. Mumberg, D., R. Muller, and M. Funk, *Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression*. Nucleic Acids Research, 1994. **22**(25): p. 5767-5768.

59. Studier, F.W., *Protein production by auto-induction in high-density shaking cultures*. Protein Expression and Purification, 2005. **41**(1): p. 207-234.

60. Jonikas, M.C., et al., *Comprehensive Characterization of Genes Required for Protein Folding in the Endoplasmic Reticulum*. Science, 2009. **323**(5922): p. 1693-1697.

61. Wittig, I., H.-P. Braun, and H. Schägger, *Blue native PAGE*. Nature Protocols, 2006. **1**(1): p. 418-428.

62. Metz, J., et al., *The Yeast Arr4p ATPase Binds the Chloride Transporter Gef1p When Copper Is Available in the Cytosol*. Journal of Biological Chemistry, 2006. **281**(1): p. 410-417.

63. Kushnirov, V.V., *Rapid and reliable protein extraction from yeast*. Yeast, 2000. **16**: p. 857-860.

64. Hernandez, H. and C.V. Robinson, *Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry*. Nat. Protoc., 2007. **2**(3): p. 715-726.

65. Gault, J., et al., *High-resolution mass spectrometry of small molecules bound to membrane proteins*. Nat. Methods, 2016. **13**(4): p. 333-336.

66. Bolla, J.R., et al., *Direct observation of the influence of cardiolipin and antibiotics on lipid II binding to MurJ*. Nat. Chem., 2018. **10**(3): p. 363-371.

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

68. Marty, M.T., et al., *Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles*. Anal. Chem., 2015. **87**(8): p. 4370-4376.

69. Zheng, S.Q., et al., *MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy*. Nature Methods, 2017. **14**(4): p. 331-332.

70. Zhang, K., *Gctf: Real-time CTF Determination and Correction*. J. Struct. Biol., 2016. **193**(1): p. 1-12.

71. Tang, G., et al., *EMAN2: An extensible image processing suite for electron microscopy*. Journal of Structural Biology, 2007. **157**(1): p. 38-46.

72. Kimanius, D., et al., *Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2*. eLife, 2016. **5**: p. e18722.

73. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3*. 2018, eLife. p. e42166.

74. Tegunov, D. and P. Cramer, *Real-time cryo-electron microscopy data preprocessing with Warp*. Nature Methods, 2019. **16**(11): p. 1146-1152.

75. Brown, A., et al., *Tools for Macromolecular Model Building and Refinement Into Electron Cryo-Microscopy Reconstructions*. Acta Crystallogr. D Biol. Crystallogr., 2015. **71**(Pt 1): p. 136-153.

76. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta Crystallographica Section D Biological Crystallography, 2010. **66**(2): p. 213-221.

77. Pettersen, E.F., et al., *UCSF Chimera--a Visualization System for Exploratory Research and Analysis*. J. Comput. Chem., 2004. **25**(13): p. 1605-1612.

78. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions*. Acta Crystallographica Section D Biological Crystallography, 2004. **60**(12): p. 2256-2268.

CHAPTER 6

Concluding remarks

In this dissertation, we described how the integration of nMS with other biophysical techniques led to the elucidation of different aspects of the assembly and modulation of membrane transporters and insertases. The investigation started with an analysis of the mechanism of transport of ABC importers (Chapter 2), followed by the examination of an atypical ABC transporter, the LPS transport protein $LptB_2FG$, which interacts with the single pass membrane protein LptC (Chapter 3). Following the interest in the LPS transport system, we then focussed the study on the OM LPS translocon LptDE (Chapter 4). While the LptDE heterodimer is a lipid insertase, the protein complex described in Chapter 5 is a protein insertase, devoted to the insertion of TA proteins in the ER.

In Chapter 2, we proposed that type I and type II ABC importers have a fundamentally different mechanism of substrate transport. Using nMS, we demonstrated that the interaction between the periplasmic substrate-binding protein (SBP) ModA and the type I importer ModBC relies on substrate binding. Conversely, in the case of type II importer BtuCD, the substrate, vitamin B_{12}, destabilises the interaction between the transporter and its cognate SBP, BtuF. Furthermore, while ATP binds cooperatively to BtuCD-F, and acts synergistically with vitamin B_{12} to disrupt the BtuCD-F complex, we did not observe any effect on the stability of ModBC-A following ATP binding. These observations highlight the ability of nMS to capture these transporter-SBP complexes and add missing details to the proposed transport mechanisms. We trapped different states of the transport cycle for both transporters and suggest a transport cycle for type I importer ModBC whereby the initiating step is the docking of substrate-loaded ModA onto ModBC. Then, ATP binding and hydrolysis trigger conformational rearrangements leading to substrate release in the cytoplasm and following dissociation of ModA. In the case of type II importer BtuCD, since binding of both substrate and nucleotide impair the interaction with BtuF, the resting state is represented by the SBP-bound transporter. Binding of ATP destabilises this

interaction, allowing the capture of the substrate by BtuF and subsequent release in concert with ATP hydrolysis. Hence, the driving force of the transport mechanism is the attainment of the minimum energy state represented by substrate-free BtuCD-F.

In Chapter 3, we optimised the expression and purification conditions for the observation of the LPS transporter $LptB_2FG$ from *P. aeruginosa* using nMS. Using an approach similar to the one described in Chapter 2, we showed that nucleotide binding increases $LptB_2FG$ affinity for its substrate, while ATP hydrolysis has the opposite effect. We also demonstrated the importance of the sugar moieties of the core oligosaccharide part of LPS protein interaction and discovered that cardiolipin binding causes a decrease in substrate binding, probably acting as a competitor for the pocket located between the TMDs. We then extended our analysis to $LptB_2FGC$ and integrated nMS with cryo-EM, to obtain structural insights. We observed that our $LptB_2FG$ has a different conformation from previously published $LptB_2FG$ from *P. aeruginosa*, probably because our protein is LPS-free. Similarly, comparison of our $LptB_2FGC$ structure with other Gram-negative bacterial orthologues provided unique molecular details on the conformational fluctuations of the TMDs between substrate-free (represented by our $LptB_2FGC$) and substrate-bound state, but also yielded insights on species-specific differences of the β-taco orientation. We also expanded our investigation to LptA and showed that it is mainly dimeric in solution and interact with $LptB_2FGC$ in 1:1 ratio in the absence of substrate.

The interest in the LPS transport system led us to explore the OM heterodimer LptDE in Chapter 4. Here we combined nMS with HDX-MS and MD simulations to dissect the conformational changes of the transporter and clarify the mechanism of insertion of LPS in the OM. We also studied the effect of binding of a naturally occurring antimicrobial peptide, thanatin. Using nMS, we found that LPS binds LptDE specifically compared to other membrane lipids. We also showed that the thanatin strongly binds to LptDE in a non-competitive manner towards LPS. Upon substrate binding, we discovered that the periplasmic domain of LptD undergoes conformational opening, consisting of a concerted weakening of hydrogen bond networks. This conformational fluctuation is coupled with a simultaneous opening of

transmembrane β-strands adjacent to the putative LPS exit gate. We also showed that an antimicrobial peptide, thanatin, strongly binds to LptDE in a non-competitive manner towards LPS. In contrast, we observed that thanatin induced opposite conformational fluctuations in the periplasmic region compared to LPS, thus stabilising the hydrogen bond networks and making the complex less prone to LPS transport. Based on our results, we present a model whereby LPS insertion in the OM relies on combined opening movements of specific periplasmic and transmembrane domains of LptD. Importantly these results indicate a means of designing antimicrobial therapeutics to target this essential process in Gram-negative bacteria. This work showed the potential of HDX-MS as a powerful tool to examine local structural rearrangements and obtain semi-quantitative insights on the dynamics induced by substrate or inhibitor binding. The presence of the EX1 behaviour also allowed to quantify the rate of the conformational change of LptD periplasmic domain and the combination with MD simulations allowed us to obtain structural information at single-residue resolution.

In Chapter 5, we investigated a system that summarises the previously described ones. Indeed, the TA protein insertase complex is a transporter that inserts proteins in the membrane using the energy provided by ATP, similar to ABC transporters. We integrated nMS and lipidomics with cryo-EM and *in vivo* experiments to elucidate the structural features and lipid binding properties of *S. cerevisiae* Get1/Get2/Get3 complex and its human orthologue WRB/CAML/TRC40. Eukaryotic proteins add further challenges for characterisation using nMS. Indeed, they possess a different lipid environment and may also possess PTMs, resulting in highly heterogeneous samples. Nonetheless, the methods employed in this thesis were successfully utilised to dissect the oligomeric state and lipid binding proteins of such complexes. We provided first clear evidence that that the complexes have a 2:2:4 stoichiometry, whereby Get3 (TRC40) binding to the membrane proteins supports Get1/Get2 (WRB/CAML) heterotetramer formation. Analysis carried out using nMS indicated preferential binding of phosphatidylinositol, interacting at the heterotetramer interface as confirmed by structure-based mutagenesis of both yeast and human complexes. The structural similarities of the TA insertase complex with YidC and EMC suggest an evolutionarily conserved insertion mechanism for

divergent substrates involving a hydrophilic groove. We also identified an unanticipated, conserved Get2/CAML cytoplasmic helix that forms a 'gating' interaction with Get3/TRC40 and promotes TA insertion. This work provided a detailed structural and mechanistic framework for TA insertion in the ER membrane.

In summary, this dissertation exemplifies the power of nMS and, generally, MS-based techniques to gain insights in protein-protein interactions, protein complex stoichiometry and organisation and protein-ligand interactions, especially lipids. We combined nMS with a diverse range of other techniques, yielding unparalleled information on protein structure, dynamics and regulation. In Chapter 2-4 of this thesis, we elucidated the assembly and regulation of proteins that may represent ideal targets for the development of new antibiotics to tackle the problem of anti-microbial resistance. Advances in MS-based methods may have a critical role in the drug discovery field, particularly in the case of membrane proteins where other biophysical techniques present drawbacks such as the need for non-physiological tags. The data presented in this thesis open new routes for research in multiple areas of investigation, with the ultimate aim of gaining a deeper understanding of the finely-tuned modulation mechanisms that regulate protein and cellular homeostasis.