Structural and functional studies on Gram-negative bacteria lipid transport and modification proteins

PhD Thesis for

ZHENGYU ZHANG

(100065867)

Faculty of Medicine and Health Sciences

Norwich Medical School

Submission year 2017

ABSTRACT

Gram-negative bacteria contain a double membrane system which is made up of the inner membrane (IM) and the outer membrane (OM). The OM, the asymmetric membrane, is an important contributor towards resistance against various toxic molecules which provides extensive protection of cell viability. The OM consists of lipopolysaccharides (LPS) at the outer leaflet and glycerophospholipid (GPL) at the inner leaflet. LPS is a large amphipathic molecule that is essential for most Gram-negative bacteria and is composed of three moieties: lipid A, core oligosaccharide and O-antigen. Precursors of LPS, the rough LPS and O-antigen units, are synthesized at the inner leaflet of the IM and are assembled into mature LPS at the outer leaflet of the IM. Seven proteins, lipopolysaccharide transporters (LptA-G), form a periplasmic bridge and transport mature LPS from the IM to the outer leaflet of the OM, where LPS functions as an impermeable barrier. The inner leaflet of the OM is also an obligate component for the stability and functionality of OM. GPL molecules consist of phosphatidylethanolamine (PE), phosphatidylglycerols (PG) and cardiolipin (CL) in an approximately 75: 20: 5 ratio in most bacterial membranes. The biogenesis and transport of LPS and GPL are often the highlights of novel drug development; however little is known about the mechanisms of the actions of the proteins involved in these processes.

Protein X-ray crystallography is a technique that determines the molecular structure of a protein or a protein complex from the diffraction data of a protein crystal generated under X-ray beam. Knowing a protein's molecular structure provides us with the information which may include possible molecular motion and conformational changes or interactions between residues, substrates and domains. These important details may provide hints into how to perform further studies on broad-ranged specific targets.

This thesis includes four Chapters related to the X-ray crystal determination of proteins involved in biosynthesis, modification and transport of LPS and the transport of CL. In Chapter 1 I have included the successful cloning, expression and structure

determination of the ABC transporter $LptB_2FG$, which is responsible for LPS extraction from the inner membrane. Also, the corresponding *in vivo* and *in vitro* assays are reported to support the LPS extraction and transportation functionality of $LptB_2FG$. Chapter 2 contains the successful expression of glycosyl transferase WaaB and the determination of its structure. Mutagenesis and enzyme activity assays have also been performed to determine the galactosyl transferase activity site of WaaB (work by Gareth James Asheworth). Chapter 3 describes the successful cloning and structural determination of LapB, which may control an intermediated step of lipid A biosynthesis via balancing the level of LpxC & FtsH, Finally, Chapter 4 reveals the protein expression and structural determination of the globular domain of PbgA. PbgA is regarded as the cardiolipin transportation protein and we have also performed *in vivo* mutagenesis functional assays to identify the important residues for its functionality.

Table of contents

Acknowledgment

First, I must thank my supervisors, Prof. Changjiang Dong and Prof. Tom Wileman, for their mentorship and support. Changjiang has helped me grow into a scientist. He has constructed a lab of free mind and free will, in which I could apply my ideas into researches; Tom helped me with my personal development and socialization with department members. In particular, I would like to thank Haohao Dong for helps on overcoming the difficulties I encountered in my Ph.D. He was also a close collaborator for the main parts of this thesis, and I thank him for his help and support in making this work possible. Next, I would like to thank I would like to thank Gareth Ashworth (G.J.A) for his editing assistance and contribution towards WaaB functional assays, who in addition to being one of my closest friends in lab. I would also like to thank Huanyu Li, with whom I enjoyed a happy life during the Ph.D studies. I would also like to thank all of the current and former members of the Changjiang Dong lab for their help and support throughout the years. In particular, Bo Peng, Xuejun Li, Saannya Sequeira, Rob, Yinghong Gu and Yi Zeng have all either provided me with materials or offered valuable scientific discussion during my time in the Changjiang laboratory. In addition to group members, I thank Yingxue Wang and Weijiao Zhang; without them, I would feel much lonelier. I would also like to thank Charles Brearly for the use of his plate reader. I sincerely feel grateful to all stuffs from Diamond Light Source, particularly our collaborator Neil Patterson who shared me with endless of advanced knowledge on crystallography. At last, I would thank my parents Yiguo Zheng and Hongkun Zhang for their love and care given to me since my childhood. And I would thank for the waiting for my wife Qianru Zhao who has kept for waiting for me in the Far East for four years.

List of Abbreviations

ABC	ATP-binding cassette
ADP	Adenosine diphosphate
AMP-PNP	Adenylyl imidodiphosphate
ATP	Adenosine triphosphate
β-OG	n-Octyl-β-D-Glucopyranoside
BAM	Outer Membrane Beta Barrel Assembly Machinery
C_8E_4	tetraethylene glycol monooctyl Ether
Cymal-5	5-Cyclohexyl-1-Pentyl-β-D-Maltoside
CL	Cardiolipin
DDAO	n-decyl-N,N-dimethylamine-N-Oxide
DDM	n-dodecyl-β-D-maltopyranoside
DM	n-decyl-α-D-maltopyranoside
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
Fos-12	n-dodecylphosphocholine
Gal	D-Galactose
Glc	D-Glucose
GPL	glycerophospholipid
Hep	L-glycero-D-manno-heptose
IM	Inner membrane
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kdo	3-deoxy-D-manno-oct-2-ulosonic acid
Kleb	*Klebsiella pneumoniae*
LAPAO	3-dodecylamido-N,N'-dimethylpropyl Amine Oxide•3-Laurylamido-N,N'-Dimethylpropyl Amine Oxide
LB	Luria broth
LDAO	n-dodecyl-N,N-dimethylamine-N-Oxide
LMNG	Lauryl Maltose Neopentyl Glycol
LPS	Lipopolysaccharide
Lpt	Lipopolysaccharide transport
MBP	Maltose-binding protein; or MalE
MSC	Multi-cloning site
NG	n-nonyl-β-D-glucopyranoside
NM	n-nonyl-β-D-maltopyranoside
OD600	optical density at a wavelength of 600 nm
OM	Outer membrane
PCR	Polymerase chain reaction
PEZ	PEZ candy dispenser model
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
SAD	Single wavelength anomalous dispersion/diffraction

SDS-PAGE	Sodium dodecyl sulfate Polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
Sec	Secretion route machinry
Tat	Twin-arginine translocation pathway
Shi	*Shigella flexneri* 2a
TEV	Tobacco Etch Virus nuclear-inclusion-a endopeptidase
UDM	n-undecyl-β-D-maltopyranoside
UDP	Uridine diphosphate
WT	Wild type

List of Figures

List of Tables

1 CHAPTER 1 Lipopolysaccharide inner membrane ABC transporter $LptB_2FG$

1.1 INTRODUCTION

1.1.1 The Gram-negative bacteria outer membrane

Gram-negative bacteria contains a double membrane system, the outer membrane (OM) and the inner membrane (IM). There is a water-filled periplasm between the two membranes and the cytoplasm is surrounded by the inner membrane. The periplasm contains peptidoglycan, a polymer network that helps determine the cell`s shape and protects it from osmotic stress (2,3). The asymmetric OM, a unique membrane of Gram-negative bacteria, faces the extracellular environment acting as a strong barrier which protects the cells from some antibiotics and hydrophobic small molecules, providing heightened resistance to antibiotics compared to Gram-positive bacteria (4,5). The asymmetry of OM is caused by the component difference between the outer leaflet and the inner leaflet of the OM. The periplasm facing inner leaflet is comprised of phospholipid while the extracellular facing outer leaflet is comprised of lipopolysaccharide (LPS). LPS, an essential polyglyco-lipid in most Gram-negative species, is generally described as the source of the OM`s relative impermeability (4,5). This double membrane system carefully protects Gram-negative bacterial from harsh environments and allows bacterial cells to survive under severe circumstances (6-9). The two membranes have an entirely different structure and composition. Additionally, these membranes differ with respect to the structure of the integral membrane proteins. The IM typically contains α-helical transmembrane proteins, while the integral OM proteins (OMPs) generally consist of anti-parallel amphipathic β-strands that adopts the cylindrical β-barrels with hydrophilic interior channel and hydrophobic surface. The hydrophobic surface around the barrel structure to help be anchored in the OM (10). The OM also contains lipoproteins anchored into the periplasmic leaflet of the OM via an N-terminal N-acyl-diacylglycerylcysteine, in most cases, facing into periplasm. Unlike the IM, the OM is not energized by a proton gradient and ATP is no available in

the periplasm (11). Nutrients usually pass through the OM passively by diffusion via porins formed by the OMPs (12). Porins form water-filled channels that allow the entry of small hydrophilic molecules with molecular weights up to ~600 Da. All the components of the OM are synthesized in the cytoplasm or at the cytoplasmic face of the IM, and they have to be transported across the IM and through the periplasm to reach their destination and to be assembled into the OM (11) (Figure 2). For the OMPs, they are synthesized in the cytoplasm as precursors with an N-terminal signal sequence. The N-terminal signal sequence is recognized by the Sec system, which is a Sec protein complex machinery that translocate the unfolded proteins from the cytoplasm to the periplasm (13). After reaching the periplasm, chaperons immediately interacts with the OMPs helping stabilize and fold of the OMPs (14). With the help of the chaperons, OMPs are folded and inserted into the OM by BamA-E protein complex machinery (15-18) (Figure 2). Similar to OMPs, lipoprotein precursors are synthesized at the cytoplasm and translocate to the periplasm via Sec machinery. Then, Lgt (19), LspA (20) and Lnt (21) worked together to maturate the lipoproteins (Figure 3). The mature lipoproteins are sorted to determine that whether they should stay in the IM or be translocated into the OM by LolCDE machinery (22). If a lipoprotein is sorted to be transported to the OM, the LolCDE machinery transport this lipoprotein to a periplasmic chaperon LolA (23). Then the lipoprotein is delivered to LolB, an outer membrane receptor, and finally being inserted into the OM by LolB (24) (Figure 3).

The majority of OM phospholipids of *E. coli* are phosphatidylethanolamine and phosphatidylglycerol. Phospholipids are synthesized at the cytoplasmic side of the IM (25,26). The phospholipids needs to be flipped to the periplasmic leaflet first before being transported to the OM. The flipping process may not require a specific transporter but require the help of typical α-helical fold transmembrane segments of some IM proteins (11). The transport of phospholipids may be achieved via spheroplasts in a different way of lipoproteins (described above) or LPS (discussed in the section 1.1.7).

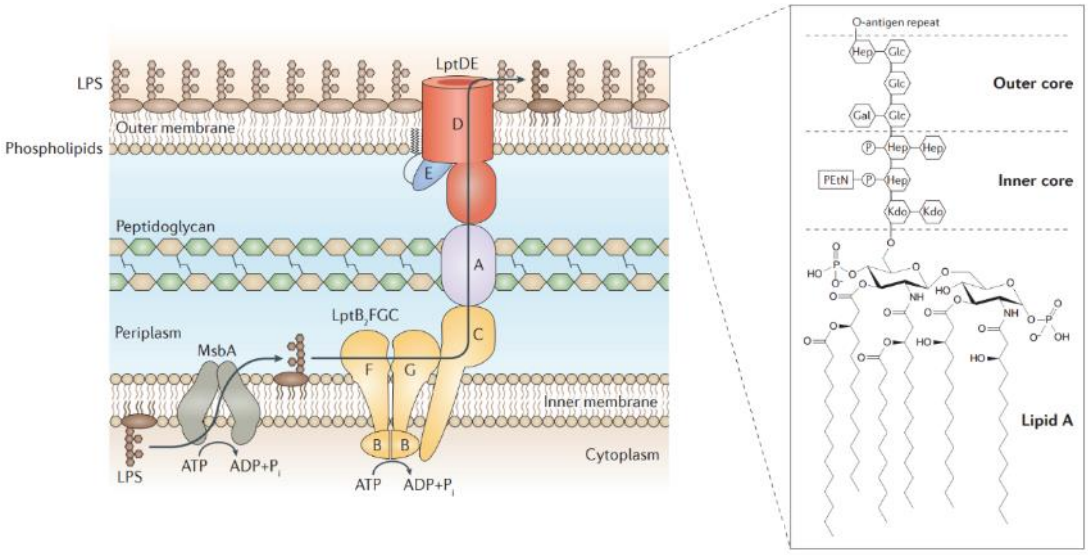


Figure 1 PEZ model describing LPS transport path (1).

A PEZ candy dispenser model of LPS transportation and assembly system. The model is proposed by Prof. D. Kahne (1). In this model, LPS transport from the IM to the OM in a similar way as the PEZ candy dispenser. After flipped by MsbA (27) and matured in the periplasmic leaflet of the IM, LPS is extracted from the IM and transported to LptC by IM ABC transporter $LptB_2FG$. Then LptC, LptA and the N-terminal of LptD form a trans-periplasm bridge and LPS is delivered along this path. The sharing of the similar β-jelly-roll folds of LptC, LptA and the N-terminal of LptD indicate this possibility and further evidence are from the crosslinking assays (28). Finally, LptD/E outer membrane complexes would receive LPS from LptA, then insert and assembly LPS into the outer leaflet of the OM (29).

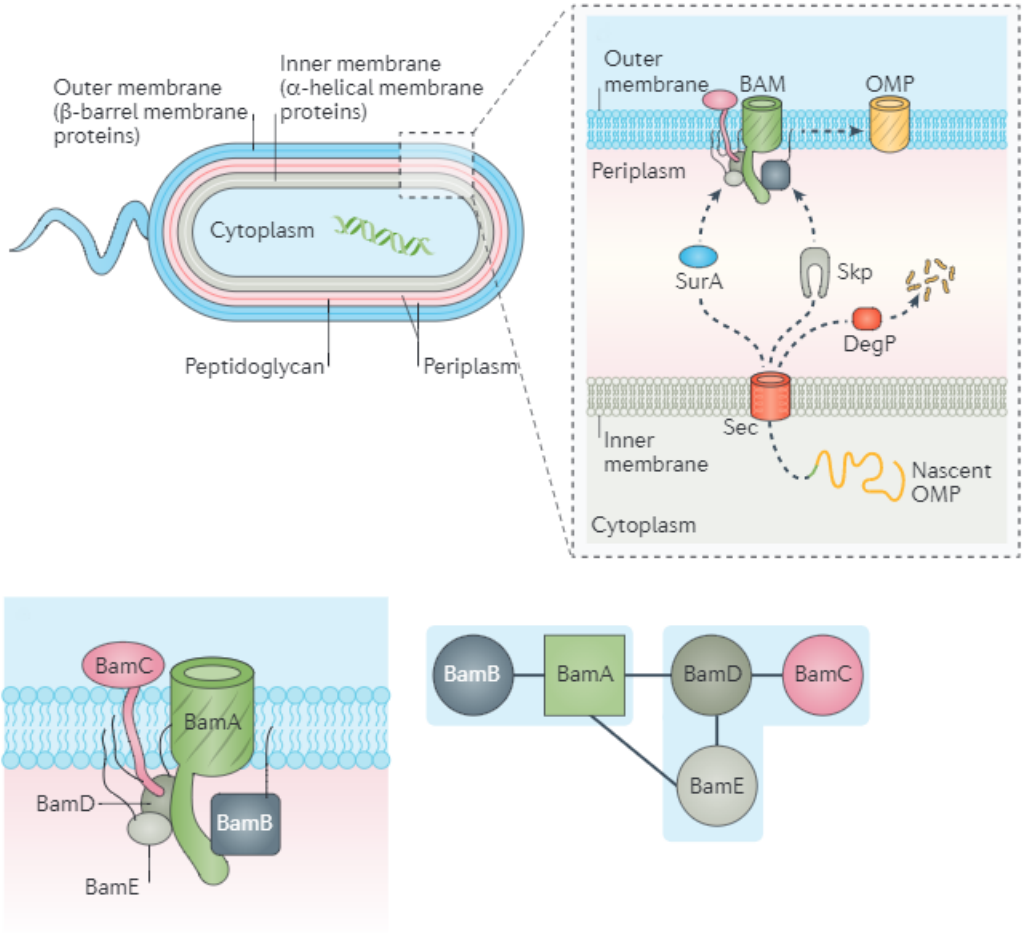


Figure 2 Outer membrane protein folding and insertion pathway (30).

Gram-negative bacteria contain both an inner membrane and an outer membrane, of which the inner membrane partitions the cytoplasm, periplasm and peptidoglycan. For the transmembrane proteins, the inner membrane contains almost α-helical membrane proteins, whereas the outer membrane contains almost β-barrel outer membrane proteins. The biogenesis of an OMP begins in the cytoplasm, where it is translated with a signal peptide (green). It is then translocated across the inner membrane into the periplasm by Sec machinery (31). The periplasmic chaperones SurA or Skp (24) then stabilize and further escort the OMP to the β-barrel assembly machinery (BAM) complex for final insertion into the outer membrane (17). Once OMPs become wrongly folded and cannot be rescued by Skp, they will be degraded by DegP to prevent damaging to the cell. In *E. coli*, the BAM complex consists of five components: the essential core component BamA, which is an OMP itself, and four

accessory lipoproteins named BamB, BamC, BamD and BamE, which are anchored into the OM.

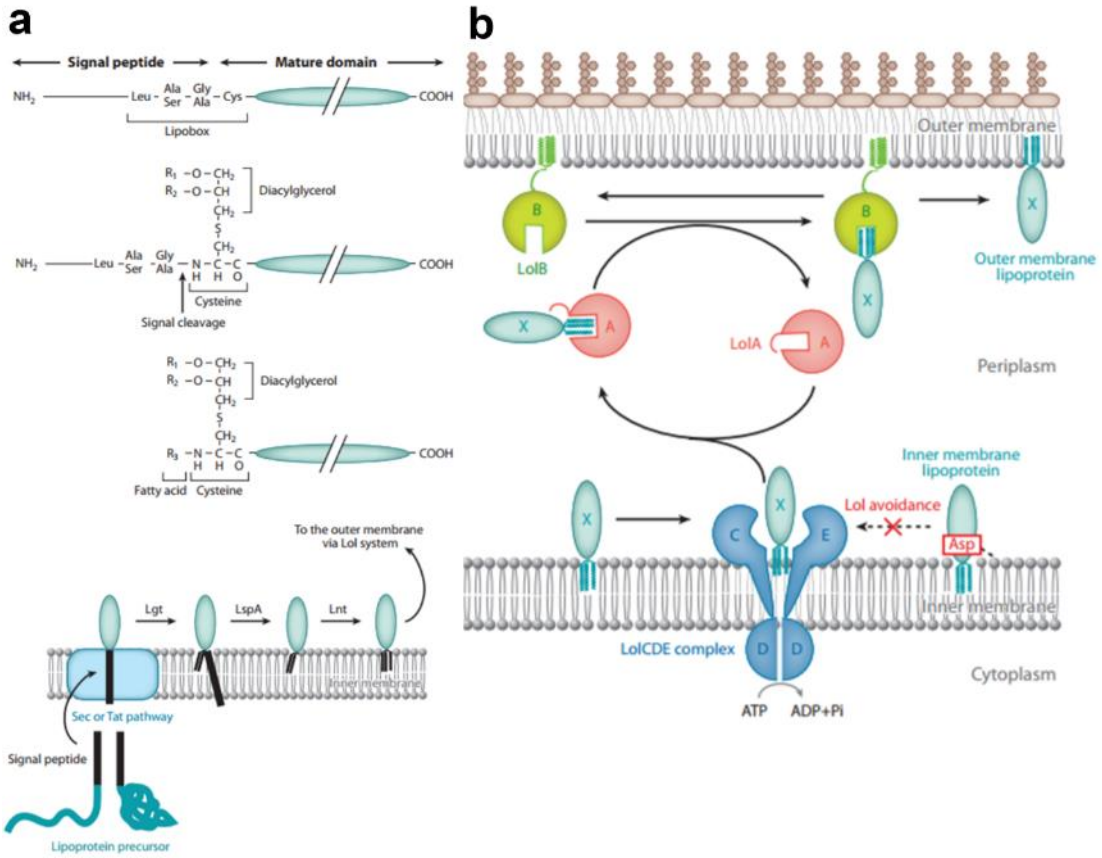


Figure 3 Lipoprotein maturation and transportation (32)

(a) Maturation of lipoproteins. Lgt turns the lipoprotein precursor into a diacylated prolipoprotein. LspA cleave the product of Lgt at the arrow indicated the cleavage position. Lnt then mature the lipoprotein by adding another pair of fatty acid chains on to the LspA product. Same as OMPs, lipoprotein precursors are translocated across the IM via the Sec (Secretion machinery) (31) or Tat (twin-arginine translocation) (33) pathway. (b) The Lol system consists of an ABC transporter LolCDE complex; a periplasmic molecular chaperone, LolA; and an OM receptor, LolB. The OM specific lipoproteins released from the IM by LolCDE form a hydrophilic complex with LolA and then cross the periplasm. Lipoproteins are transferred from LolA to LolB and are then incorporated into the OM.

1.1.2 LPS is made up of lipid A, core oligosaccharide and O-antigen

The biogenesis of LPS starts with the synthesis of precursors in the inner leaflet of the IM and followed by their flipping across the IM into outer leaflet of the IM, subsequently maturing into LPS in the periplasmic leaflet of the IM, and transported from the IM across the periplasm into the outer leaflet of the OM (Figure 1). LPS contains three moieties, lipid A, core oligosaccharide and O-antigen. lipid A, a β-1,6 glucosamine disaccharide that is phosphorylated at positions 1 and 4` and hexa-acylated via modifications at positions 2, 3, 2`, and 3`, forms the lipidic part of LPS that anchors it into the membranes. There are ~10^6 lipid A residues and ~10^7 glycerophospholipids in one cell of *E. coli* (34). Two monomers of Kdo (3-deoxy-D-manno-oct-2-ulosonic acid) are attached to lipid A forming the minimum essential viability unit of LPS (Figure 4) (35-37). The core oligosaccharide of LPS contains two regions, the inner core region and outer core region. The inner core region is composed of several glycosyls. An essential double Kdo, together with lipid A is called Re-LPS. Connecting to the inner core is an outer core which is also made up of glycosyls as those of the inner core. The rough LPS containing all core regions with a lipid A is called Ra-LPS. The most extracellular part of LPS is the highly varied O-antigen (made up of repeat polysaccharide units) which provides extra protection for Gram-negative bacterial cells, and accounts for immune evasion and antigenicity when the bacteria infects mammals. In wild-type strains, additional core and O-antigen glycosyl residues may be present, although they are not essential for growth in laboratory conditions.

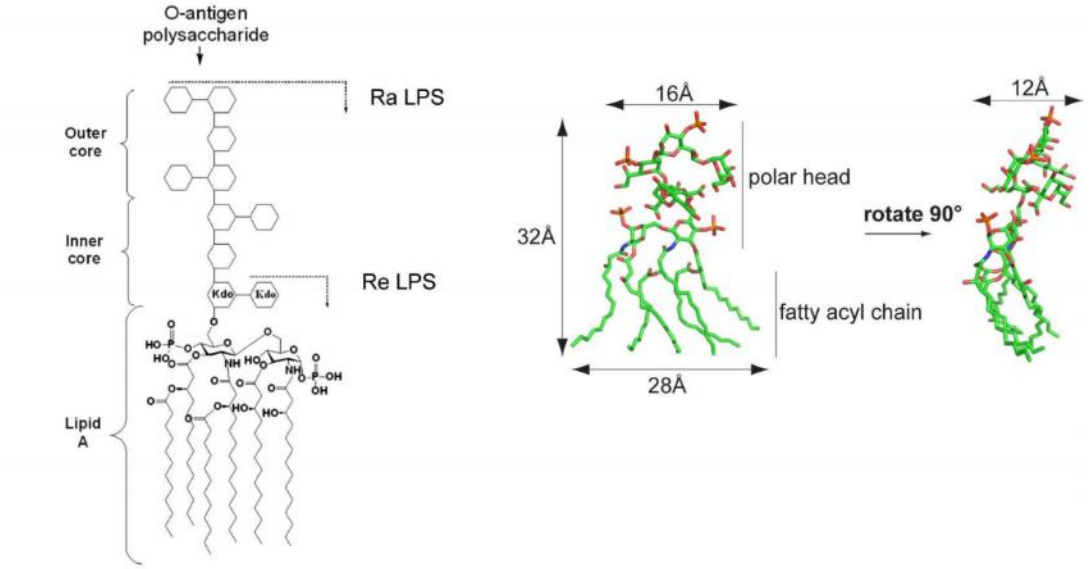


Figure 4 A rough LPS model.

A rough LPS can be divided into two parts the lipid A and the core oligosaccharides. The Kdo_2-lipid A is called Re-LPS, the whole core with lipid A is called Ra-LPS. The polar head in the image refers to the core region of the LPS, the fatty acyl chain refers to the lipid A. A Ra-LPS is about 28 Å and 12 Å in length and width, and about 32 Å in height overall. The core size is smaller than lipid A with 16 Å in length. LPS is in ball-and-stick model [O atom (red), N atom (blue)]

1.1.3 The biogenesis of lipid A of LPS

Many Gram-negative bacteria synthesize lipid A and the lipid A product resembles the one found in *E. coli* (35-37). The enzymes that participated in the steps of lipid A biosynthesis have been characterized well in *E. coli* (38,39) (Figure 5). All these enzymes worked individually to finish their own steps. The details of the biosynthesis of lipid A are further discussed in section 3.1 in Chapter 3. LpxA is the enzyme for the first step of lipid A synthesis. The next step is the deacetylation of the product by zinc metalloenzyme LpxC. Following deacetylation, a second β-hydroxymyristate moiety is

incorporated by LpxD to generate UDP-2,3-diacylglucosamine (40). UDP-2,3-diacylglucosamine is cleaved at its pyrophosphate bond by the highly selective pyrophosphatase LpxH to form 2,3-diacylglucosamine-1-phosphate (lipid X) (41). A specific kinase next phosphorylates the 4' position of the disaccharide to form lipid IV_A (42). *E. coli* LPS contains two Kdo glycosyl residues that are transferred to lipid IV_A by a bifunctional enzyme, WaaA (43). The last steps of *E. coli* lipid A biosynthesis involve the addition of lauroyl and myristoyl residues to the distal glucosamine unit, generating acyloxyacyl moieties (44). These steps are shown in Figure 5.

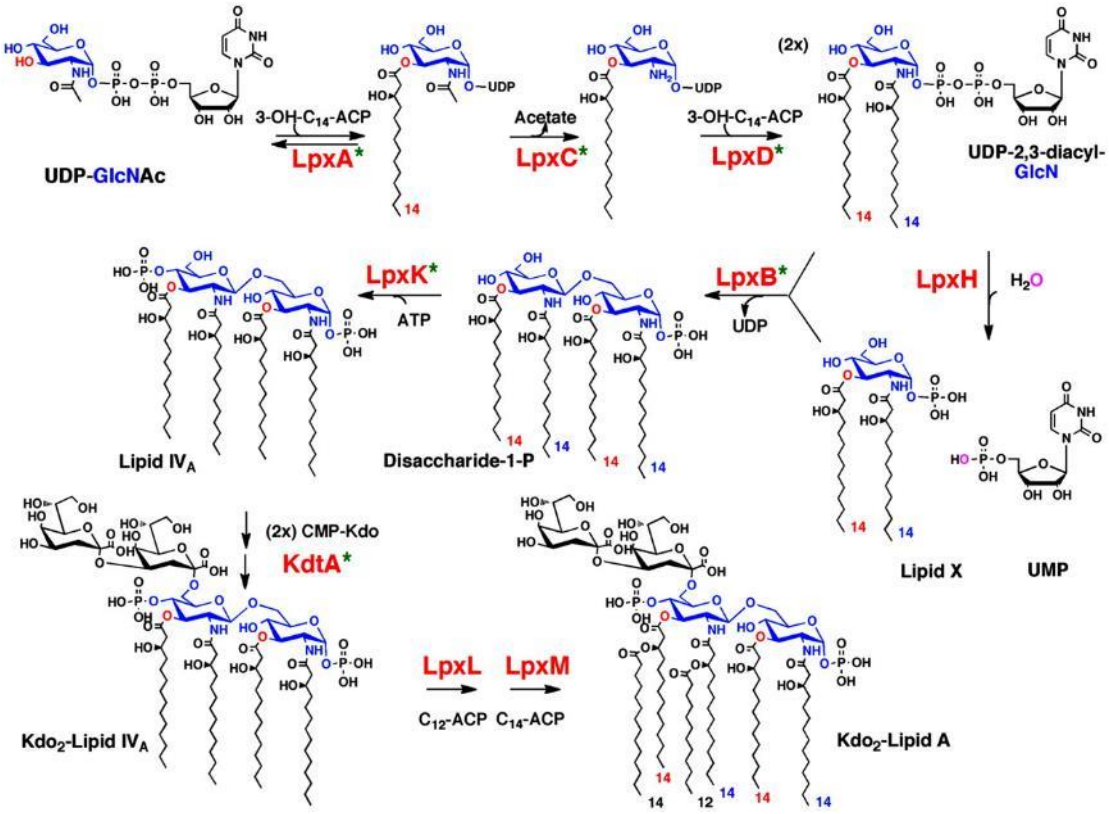


Figure 5 Structure and biosynthesis of Kdo_2-lipid A in K-12 *E. coli*. (5)

The enzymes committing each step of Kdo_2-lipid A synthesis. (39). A single enzyme catalyses each single reactions. In almost all cases, as illustrated by *E. coli* as an example, the genes encoding the enzymes of lipid A biosynthesis are present in single copy in the genome (45). At the protein level, LpxA and LpxC are the most highly conserved among bacteria. KdtA is also named as WaaA.

1.1.4 Lipid A and rough LPS flipping by MsbA

MsbA is an essential ABC transporter, resembling eukaryotic multi-drug resistance (MDR) proteins. The recent Cryo-EM structure of MsbA in complex with a Ra-LPS molecule elucidates the mechanism of lipid A and rough-LPS flipping (27,46). The structure demonstrates that LPS binds deep inside MsbA through extensive hydrophilic and hydrophobic interactions, and reaches the level of the target leaflet without flipping (27). As reported in the latest paper, LPS enters MsbA and is lifted to the periplasmic leaflet. The flipping then happens after LPS reaches the periplasmic leaflet and LPS is inserted into the periplasmic leaflet finally (27).

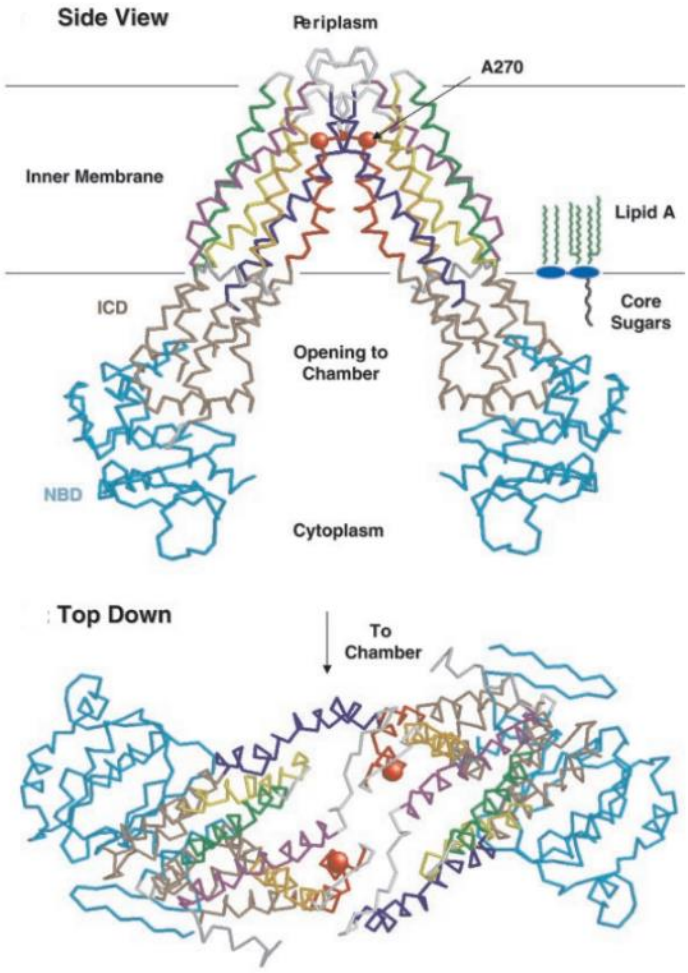


Figure 6 The ABC transporter MsbA dimer crystal structure (3B60) (46) .

Trans-membrane helices 1-6 of MsbA are coloured in purple, blue, yellow,

green, red, and orange, respectively. The transmembrane domain (TMD) is brown, and the nucleotide-binding domain (NBD) is cyan (5).

1.1.5 The importance and biosynthesis of the core oligosaccharide of LPS

The LPS contains a core oligosaccharides in the middle of its structure. The structure of core oligosaccharides of LPS from a broad range of species has been identified (Figure 7) (5), but studies mainly focused on those from *E. coli* and *S. typhimurium* (Figure 7). The core oligosaccharides are theoretically divided into two regions: inner core and outer core. The LPS that lacks O polysaccharides are called rough LPS. Within a bacterial family, the glycosyl sequence is usually conserved and the conservation can sometimes expand to the distantly related bacteria, suggesting the importance of the glycosyl residues of the inner core for the outer membrane functionality (5). In *E. coli* and several close species, the inner core typically contains three essential glycosyl residues, two Kdo and one Hep. The base structure of inner core is often decorated with sugars and with phosphate, pyrophosphorylethanolamine (2-aminoethanol; PPEtN), or phos-phorylcholine (PCho) residues (5). The variety of these modifications contributes to the heterogeneity of LPS. In many cases, we could only identify the carbohydrate backbone of a core oligosaccharide structure even in known species, and the details modifications of the core oligosaccharide are hard to identify. In contrast, the outer core structure is less conserved, because the outer core is sometimes expected to expose under selective environmental pressures. However, this variation is still limited, such as the *E. coli*. In *E. coli* there are five known core types (R1, R2, R3, R4 and K-12) (Figure 7). They have different infection specification and endotoxin level (47-50). The two known cores from *Salmonella* are quite similar to those of *E. coli* (Figure 7). In *K. pneumonia* only one major core structure has been discovered (Figure 7), and this is distributed among different serotypes. The limited structural variation in the core oligosaccharide within a genus is in striking contrast to the highly variable O polysaccharides and has stimulated interest in the possibility of targeting the core

oligosaccharides to generate immunotherapeutic antibodies (51-54). The deep-rough phenotype is displayed by *E. coli* and *S. typhimurium* that lack the Hep region of the inner core. The phenotype is actually a series of characteristics that collectively reflect changes in both structure and composition of the outer membrane leading to its instability (36). These deficiency mainly includes the surface hydrophilicity, resulting in the increasing in the sensitivity to hydrophobic dyes, detergents, hydrophobic antibiotics and polycyclic hydrocarbons (5). Also, the deep-rough mutants would release periplasmic enzymes into the medium unless sufficient amount of Mg^{2+} are added. The outer membrane of the deep-rough mutants show a decreased protein to lipid concentration and would influence biofilm formation (5,55). These previous studies has explained the importance of the cores.

The biosynthesis of core oligosaccharide starts from the Kdo_2-lipid A as the first receptor of glycosyl transferase. A sequential glycosyl transferases uses UDP-glycosyl as sugar residue donor and transfer the sugar on to Kdo_2-lipid A to synthases the core oligosaccharide of LPS. It is assumed that rapid and efficient core oligosaccharide synthesis reflects a coordinated complex of membrane-associated glycosyl transferases, although the existence of such a complex was only based on various predictions and computer analysis (56). The chromosomal *waa* region (formerly *rfa*) contains the major core oligosaccharide assembly operons, and *E. coli K-12* provided the first *waa* region sequenced (5).

In *E. coli*, the enzymes that participated in core oligosaccharide synthesis are shown in Figure 8. Besides the glycosyl transferases that add sugar residues onto the core oligosaccharide, the modifications of the core region are also important. In *E. coli* and *S. typhimurium*, the modification of the core region requires three enzymes WaaP, WaaY and WaaQ (Figure 8). The WaaP is a LPS kinase and is the most important one because the modification process must happen in WaaPQY order (57). The waaP mutant exhibit deep-rough phenotype in *E. coli* and *S. typhimurium* and this mutant is even lethal in *P. aeruginosa* (58). However WaaP protein sequence share extensively high similarity in these three species (5). All of the *E. coli* and *S. typhimurium* cores have a glucose residue as the first sugar in the outer core. Previous biochemistry data showed

that WaaG was identified as the UDP-glucose:(heptosyl) LPS α1,3-glucosyltransferase in *E. coli* (59) and *S. typhimurium* (60). The following enzymes worked in sequence to add residues onto the outer core of LPS (Figure 8).

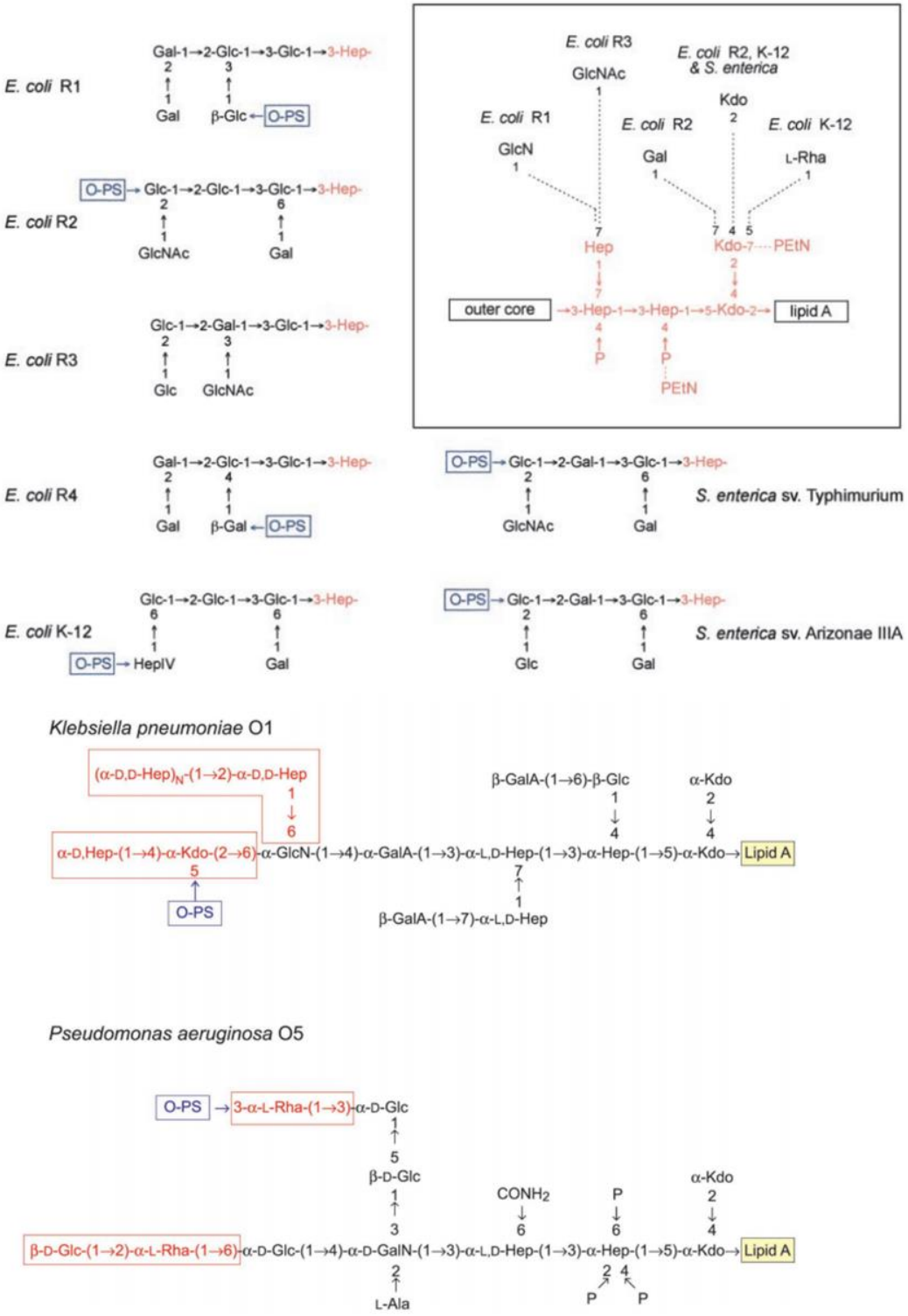


Figure 7 Structure of core oligosacchrides of several species (5).

In *E. coli* and *S. typhimurium* species, the outer cores are shown, together with one Hep residue of the inner core. O-PS refers to O polysaccharides (O-antigen) (60-62). The figure shows the conserved base structure of the inner core, and type-specific nonstoichiometric additions to the inner core are identified by dotted lines. Residues and linkages that are conserved in each of the core oligosaccharides are shown in red colour. The *K. pneumoniae O1* and *P. aeruginosa O5* core oligosaccharide are reported in previous literatures (63,64). In *K. pneumoniae O1* and *P. aeruginosa*, the boxed regions of the core structures are those that differ between S-LPS (smooth LPS) and R-LPS (rough LPS).

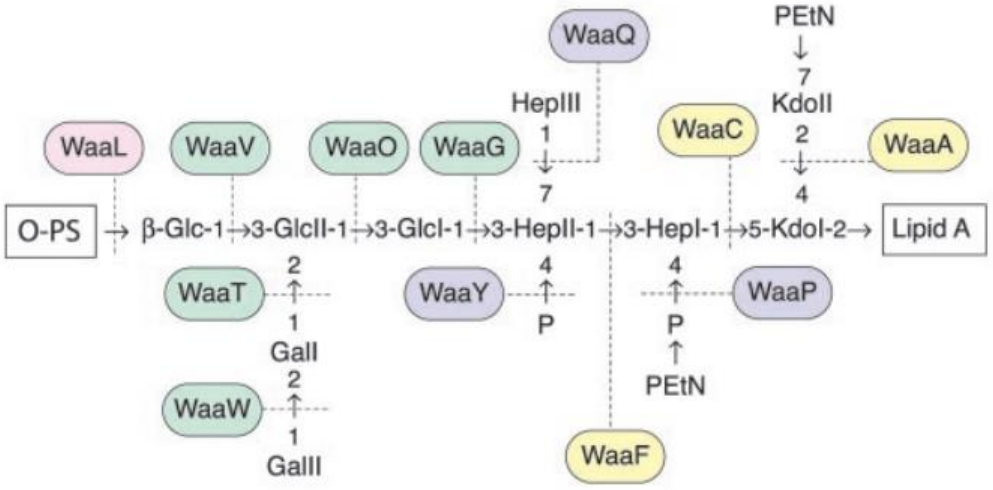


Figure 8 Structure and biosynthesis of the *E. coli* R1 core. (5)

Glycosyl transferases that form the inner core backbone are in yellow boxes, and enzymes that modify the structure are in blue boxes. Green boxes identify outer core glycosyl transferases, and the ligase enzyme is in pink.

1.1.6 The functionality and the biosynthesis of the O-antigen of LPS

Smooth LPS are produced by lots of Gram-negative bacteria species. The structure diversity of O antigens is huge; more than 60 monosaccharaides and 30 different noncarbohydrate components have been identified (65). The O polysaccharide repeat unit structures can differ in monomer sugars, the position and the stereochemistry of the linkage manner (5). LPS extracted from S-LPS strain shows a strain specific ladder pattern in the SDS-PAGE, suggesting that the O antigen units differs from single LPS molecules in one cell (5). The synthesis of O antigen units are also little conserved with some species encoding the O antigen from chromosome and some from mobile plasmid (66,67). The location of O polysaccharide at the cell surface where the bacteria contact the environment directly. The primary role of O-antigen appears to be protective. In animal pathogens. O-antigen may contribute to bacterial evasions and the evasion could be affected by the component and the length of O-antigen (68,69). With some exceptions, the enzymes involved in O polysaccharide assembly are encoded by genes at the locus historically known as *rfb*. These loci encode enzymes for the synthesis of sugar nucleotide precursors, glycosyl transferases that synthesize the O polysaccharide, and enzymes required for export processes (5). The diversity of the O-antigens also comes from the diversity of the *rfb* loci. Unlike biosynthesis of the core oligosaccharide, the repeating unit structures of O polysaccharides are assembled on the membrane-bound carrier, undecaprenyl phosphate (und-P) (5). O polysaccharides are synthesized from sugar nucleotides by glycosyltransferase enzymes that are often soluble or periplasmic membrane proteins, indicating that assembly of lipid-linked O-repeat units occurs at the inner face of the cytoplasmic membrane (5). However, O polysaccharides are transferred from the carrier lipid and ligated to lipid A-core at the periplasmic leaflet of the inner membrane (70). There are three known pathways for O polysaccharide biosynthesis: Wzy-dependent (71), ABC-transporter dependent, and synthase-dependent (5). In *S. typhimurium* and several other species, the biosynthesis of O oligosaccharide is Wzy-dependent pathway (Figure 9). The corresponding gene and gene products have been identified in the process (72). The synthesized O-antigen

repeat units are flipped by Wzx and polymerized by Wzy, with the length controlled by Wzz (Figure 9). The ABC-Transporter-Dependent pathway is confined to linear O polysaccharide structures and involves chain extension by processive addition of glycosyl residues to the nonreducing terminus of the und-PP-linked growing chain (5). The O polysaccharides biosynthesis process is initiated by WecA resulting in und-PP-GlcNAc product as a primer for chain extension (5). The involvement of WecA is well established for the polymannose O polymers of *E. coli* O8, O9, and O9a and their structural counterparts in *K. pneumonia* O5 and O3 (5). In the next step in biosynthesis, an O polysaccharide specific glycosyl transferase adds the adaptor between the und-PP-GlcNAc primer and the repeat unit domain and commits the lipid intermediate to the chain extension pathway (5). As with the WecA-mediated transfer, this reaction occurs only once per chain, and the residue is confirmed in the linkage regions for *K. pneumonia* O3 and O5 LPS (73). In the biosynthesis of D-galactan I, the corresponding process is mediated by the double-functional galactosyl transferase, WbbO. Both activities of WbbO are involved in formation of the adaptor, but one also participates in subsequent chain extension reactions (74,75). In the chain extension step, the repeat unit is assembled by transfer of residues to the nonreducing terminus of the und-PP-linked acceptor in sequence (76,77). Following polymerization, surface assembly requires an ABC transporter (Wzm and Wzt complex) encoded by the O polysaccharide biosynthesis cluster Figure 10. The ABC-transport-dependent pathway does not have a Wzz chain-length determinant, yet S-LPS in bacteria with this pathway do exhibit strain-specific model distributions, which requires a termination mechanism to control the chain length (78).

For the synthase dependent O polysaccharide biosynthesis, synthases are glycosyl transferases that have the capacity to synthesize polymers within a single polypeptide (5). The plasmid-encoded O:54 antigen of *S. typhimurium* Borreze is currently the only known example of a synthase-dependent O polysaccharide (67,79,80). Synthesis of the poly-N-acetylmannosamine homopolymer that comprises the O:54 antigen is initiated by WecA (Figure 11). The first ManNAc residue commits the und-PP-GlcNAc primer to O:54 biosynthesis and is transferred by the nonprocessive ManNAc transferase

WbbE (81). These first two steps resembles those of the ABC-transporter-dependent process. The second transferase, WbbF, is a member of the HasA family of enzymes, and it is proposed that this enzyme performs the chain-extension steps. Similar to the ABC-transporter– dependent pathway, the product of the synthase pathway is ligated to rough LPS by ligase (67).

The impact of nonstoichiometric substitutions on O antigenicity is widely recognized. The phenomenon is well studied in *S. typhimurium* and *S. flexneri*, where additions of acetyl groups or glucose residues create novel O factors. The *oafA* gene product encodes the O-acetyltransferase conferring O:5 specificity in *S. typhimurium*(82). Other well-characterized O-acetyltransferases include Oac, encoded by the *S. flexneri* seroconverting bacteriophage SF6 (83), and the O-acetyltransferase carried by *P. aeruginosa* bacteriophage D3 (84). The modifications on the O antigens usually involves in Wzy-dependent pathway, but there is evidence for similar and often nonstoichiometric glycosylation and O-acetylation modifications in ABC-transporter pathways too (85-89).

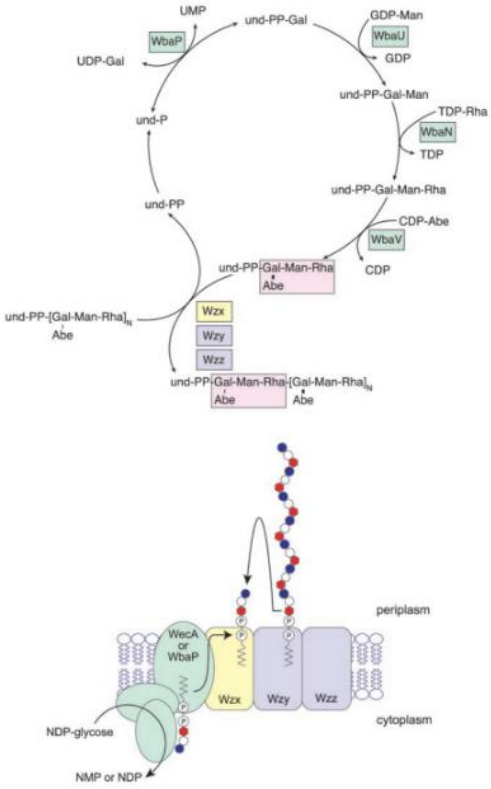


Figure 9 Biosynthesis and assembly of O-antigen in a Wzy-dependent pathway (5)

Upper panel shows the sequence of reactions involved in the formation of the O-antigen repeat units and their polymerization in *S. typhimurium*. The individual glycosyl transferase enzymes are identified in green boxes. Lower panel shows a model for the polymerization of O-antigen. Individual repeat units are flipped across the membrane by a process involving Wzx (yellow). These intermediates provide the substrates for the putative polymerase, Wzy (blue), acting in the periplasm. Chain extension occurs at the reducing terminus with the nascent chain being transferred from its undecaprenyl carrier to the nonreducing terminus of the “new” O-antigen repeat unit. The chain length is determined by Wzz (blue). The polymer is then ligated to lipid A-core and mature LPS is translocated to the outer membrane by $LptB_2FG$.

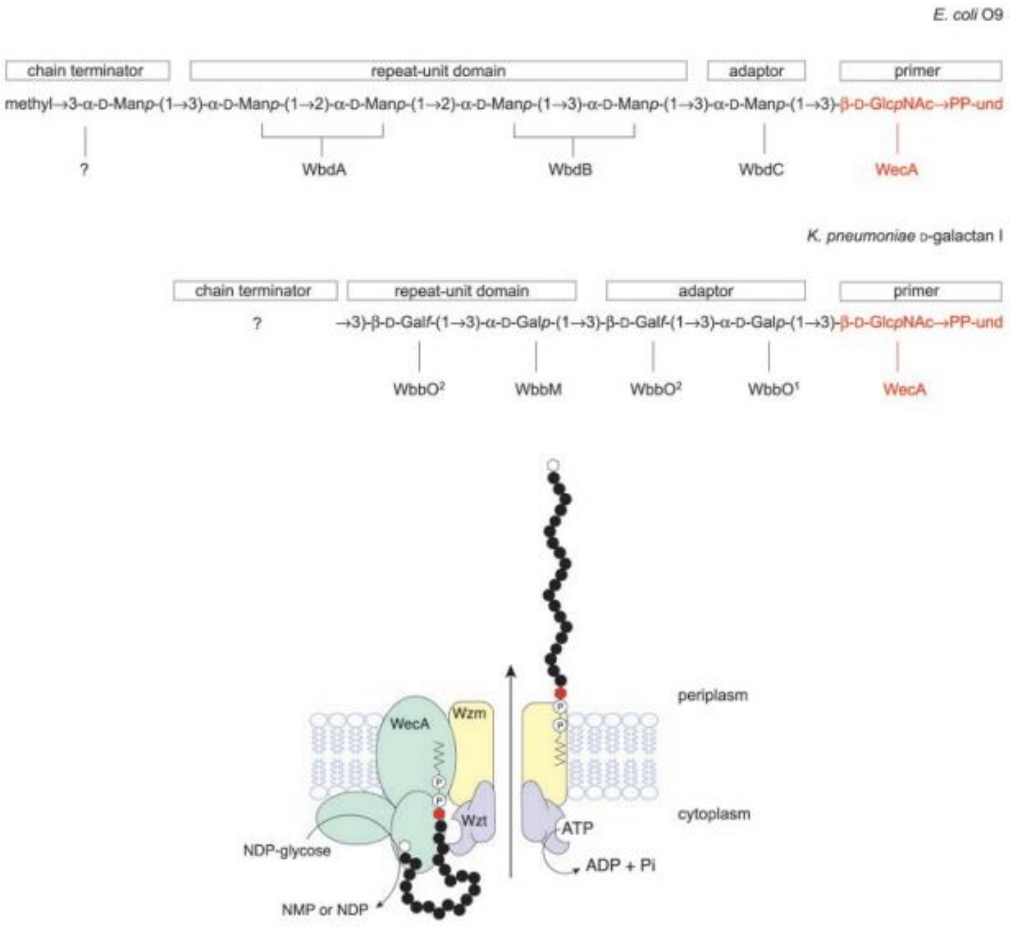


Figure 10 Biosynthesis and assembly of O-antigen in an ABC-transporter-dependent pathway (5).

Upper panel shows the predicted structures of undecaprenyl-linked intermediates in the biosynthesis of the O9 O-antigen of *E. coli* and the D-galactan I polymer found in several serotypes of *K. pneumoniae*. The glycosyl transferase enzymes involved in each step are indicated below the structures. Biosynthesis is broken down into the formation of a common primer (shown in red), addition of an adaptor region, chain extension of the repeat unit, and addition of a chain terminator. In the O9 O-antigen structure, the chain is thought to be terminated by addition of 3-O-methylmannose, but the enzyme responsible has not been identified. Chain elongation occurs by glycosyl transfer to the nonreducing terminus. Several enzymes in these pathways are bifunctional. In the case of the galactosyl transferase WbbO, both of its activities are required for adaptor synthesis, but only one ($WbbO_2$) participates in chain extension. The lower panel provides a model for the transmembrane assembly system. The glycosyl transferases are shown in green. The ABC-transporter formed by Wzm and Wzt is required for transfer of the undecaprenyl-linked polymer to the periplasmic face of the membrane, where it is ligated to lipid A-core and translocated to the outer membrane. Within the nascent polymer, the primer/adaptor is identified by the red hexagon, the residues of the repeating-unit domain by black circles, and the chain terminator by the blank hexagon.

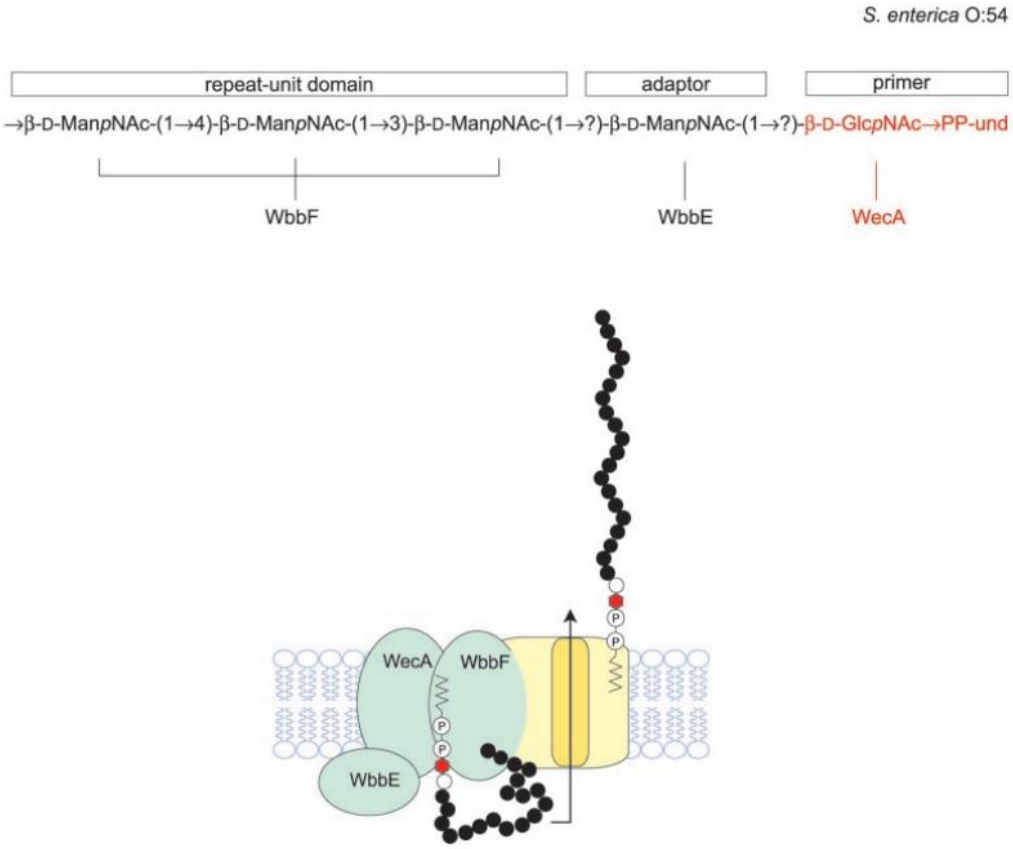


Figure 11 Biosynthesis and assembly of O polysaccharides in a synthase-dependent Pathway (5)

The only current example is the plasmid-encoded O:54 O-antigen of *S. typhimurium* Borreze. Upper panel shows the predicted structure of the undecaprenyl-linked intermediate. The primer is made by WecA. The WbbE enzyme then adds a single ManNAc residue as an adaptor. The synthase (WbbF) is then required for chain extension, generating a repeat unit domain with alternating β(1,3) and β(1,4) linkages. Growth is at the nonreducing terminus. Lower panel shows a model for the transmembrane assembly process, with glycosyl transferases identified in green and putative transport functions in yellow. The intermediate polymer is then ligated to lipid A-core oligosaccharide and translocated to the outer membrane.

1.1.7 The translocation of LPS from inner membrane to outer membrane

After the biosynthesis of the core oligosaccharide, the lipid A-core (called rough LPS in the following text) would be flipped from the cytoplasmic leaflet to the periplasmic leaflet by MsbA as described above (27,46). The completion of the S-LPS molecule involves the addition of O-antigen to the rough LPS, where it happened at the periplasmic face of the cytoplasmic membrane by the putative ligase WaaL (60).

Seven proteins, lipopolysaccharide transporter (Lpt) A-G, have been identified to form an LPS transport pathway to transport LPS from the periplasmic leaflet of the IM to the OM outer leaflet (29,90-94). And the translocation process does not discriminate between LPS with different O antigen composition or length (95). Firstly, the LPS is extracted by $LptB_2FGC$ complex from the periplasmic leaflet of the IM with the energy from ATP hydrolysis by LptB and is then transported to LptA, an oligomerised periplasmic chaperon (28,96). There is evidence that LptA interacts with LptC in a head to tail fashion (97), which suggests that LPS is passed by LptC to LptA then passed to the LptD/E complex and finally inserted into OM outer leaflet (98-103). Structural and functional studies of LptD/E have been completed previously (29,90,104-108). The LptA and LptC periplasm bridge has been thoroughly studied during 2008-2014 (97,109-112). Until today, the whole LPS transporting system still keeps one unsolved mystery that is $LptB_2FGC$ machinery (113). Works at Harvard University and University of Tokyo have already identified and purified by the inner membrane $LptB_2FGC$ complex in 2008 and 2009 (114,115). In this study, we have generated new constructs that produce better protein than them, which greatly helped with the crystallography studies.

However the mechanism of LPS extraction and transporting to LptA by the $LptB_2FGC$ complex remains unknown. More importantly, different from other ABC (ATP Binding Cassette) transporters, the IM protein ABC transporter complex $LptB_2FGC$ does not transport substrate across the membrane, instead, it extracts the substrate laterally from the periplasmic side of the IM and transport it to the periplasmic chaperon LptA (28). This unusual transporting path indicates that the $LptB_2FGC$ may adopt a special

machinery to laterally extract and vertically transport LPS. This project mainly focused on the structural and functional studies of the mechanism of how LPS is laterally extracted by $LptB_2FGC$ complex and further delivered to LptA. It has been reported that LptB is an ATPase from the ATP-binding cassette protein (ABC) $LptB_2FG$ (28,115). $LptB_2FGC$ consist of LptB: LptF: LptG: LptC at molar ratio 2:1:1:1 (115). Within $LptB_2FGC$, LptC is an inner membrane anchored protein, which binds to $LptB_2FG$ (a typical ABC transporter) forming $LptB_2FGC$ complex.

1.1.8 Bacterial ATP-binding cassette (ABC) transporter

The ATP-binding cassette (ABC) transporters form a superfamily in both prokaryotes and eukaryotes, which couples the energy from ATP hydrolysis to the translocation of substrates across a membrane (116). Their substrates include polypeptides, sugars, lipids, other hydrophobic compounds, and amino acids (116). All known eukaryotic ABC transporters are exporters while the majority of prokaryotic ABC transporters act as importers. In this Chapter, the $LptB_2FGC$ protein complex transports LPS from the lateral periplasmic leaflet of the IM to the periplasmic LptA protein, representing a different type of ABC exporter indicative of a novel subfamily.

The archetypal ABC transporter is made up of four core domains: two hydrophobic transmembrane domains (TMDs), and two nucleotide-binding domains (NBDs). These can be arranged in a variety of topologies (117). The TMDs consist of multiple transmembrane α-helices which span across the lipid bilayer and form the binding site/path for substrates. It is speculated that the TMDs may possess one or more pairs of high affinity and low affinity binding sites, with substrate transferring from one to the other after entering the transporter (118). There are 48 ABC transporters in humans and 80 in the *E. coli* (119). The ABC transporter family is currently further divided into 22 subfamilies of prokaryotic importers, 24 subfamily of prokaryotic exporters, and 10 subfamilies of eukaryotic proteins. Despite the relatively high diversity in the TMDs, 25%-30% sequence identity is shared through most of the NBDs in ABC transporter

superfamily, suggesting a similar mechanism for utilising of ATP power to drive the machinery.

In the bacterial uptake systems, the TMDs of bacterial ABC importers typically interact with a periplasmic binding domain/protein, which delivers the substrate afterwards and ensures specificity. These binding protein-dependent transporters take up a wide variety of substrates, range from small sugars, amino acids, anions, iron chelatros vitamin B_{12} and etc. In Gram-negative bacteria, small substrates enter the periplasm via diffusion through the outer membrane OMP porins (120). Large compounds, such as vitamin B_{12} and iron-siderophore complexes, are transported across the outer membrane through high-affinity transporters by consuming energy from electrochemical gradient across the cytoplasmic membrane (121). A typical example of this kind of domain/protein is in the $MalFGK_2$ transporter which transport maltose from the periplasm to cytoplasm (122). The periplasmic protein MBP usually had low affinity with $MalFGK_2$ transporter, however when the binding affinity of MBP with $MalFGK_2$ transporter increased significantly when ADP vanadate (a kind of ATP mimic) was trapped inside the NBDs of $MalFGK_2$ transporter (122). This kind of mechanism allow the periplasmic protein to search for substrate and promotes the efficiency of substrate transporting.

ABC transporters also functions in efflux mode in some cases, including surface components of the bacterial cell (such as capsular polysaccharides, lipopolysaccharides, and techoic acid), proteins involved in bacterial pathogenesis (such as hemolysin, heme-binding protein, and alkaline protease), peptide antibiotics, heme, drugs and siderophores (123). In this cases periplasmic protein/domain usually fuse with the TMDs. As an example, the inner membrane ABC tranporter HlyB secret HlyA with an inner membrane protein HlyD (124) and an outer membrane facilitator TolC (125). TolC forms a channel spanning both outer membrane and the periplasm of *E. coli* (126,127). Both TMDs and NBDs of HlyB and HylD are responsible for substrate recognition. Following the recognition, ABC transporter HlyB presumably transport the unfolded HlyA across the inner membrane (128-131) and into the tunnel formed by TolC.

The NBDs are intracellular, water soluble subunits which drive the whole machinery

transporting substrate by binding and hydrolysing of ATP. ATP hydrolysis is thought to drive a conformational change in the NBDs which propagate to the TMDs resulting in an energy consumable substrate translocation. Unlike TMDs, the NBDs are highly conserved, suggesting their universal role in powering the transport process. The structure of a NBD monomer can be divided into two subdomains: a larger RecA-like subdomain consisting of two β-sheets and six α-helices and a smaller helical subdomain formed by three to four α-helices (119). They are comprised of an ATP binding cassette domain, which contains a Walker A and Walker B motif, the Signature sequence (C motif) and several possible functional sites which together defines the ABC transporter. The Walker A and B motifs are supposed to bind phosphates of ATP and the Mg^{2+} ion when ATP binds in the NBDs (Figure 12, Figure 24 and Table 1). The Walker A motif, also known as the P loop, follows β-strand 3 and forms a loop that binds to the phosphates of ATP or ADP. The aspartic acid residue in Walker B coordinates the Mg^{2+} ion in the nucleotide binding through H_2O (132-134). A glutamic acid residue binds to the attacking water molecule and the Mg^{2+} ion (135,136). This glutamic acid residue is the catalytic base for hydrolysis functionality. In LptB E163 is the catalytic site. The Q loop, also known as the lid (137), contains a glutamine residue that binds to the Mg^{2+} ion and attacking water (134,135). The structure of the Q loop is part of the Rec-like domain and appears to be highly flexible (119). The H loop following the β-strand 8 is referred to as a switch and also contains conserved histidine residue that interact with γ-phosphate of ATP (135,138) (Figure 2). The signature motif is also named as LSGGQ motif, linker peptide or C motif. It has been used to identify ABC transporter and is the only major highly conserved motif that does not have direct interaction with nucleotides in the monomer structure (119).

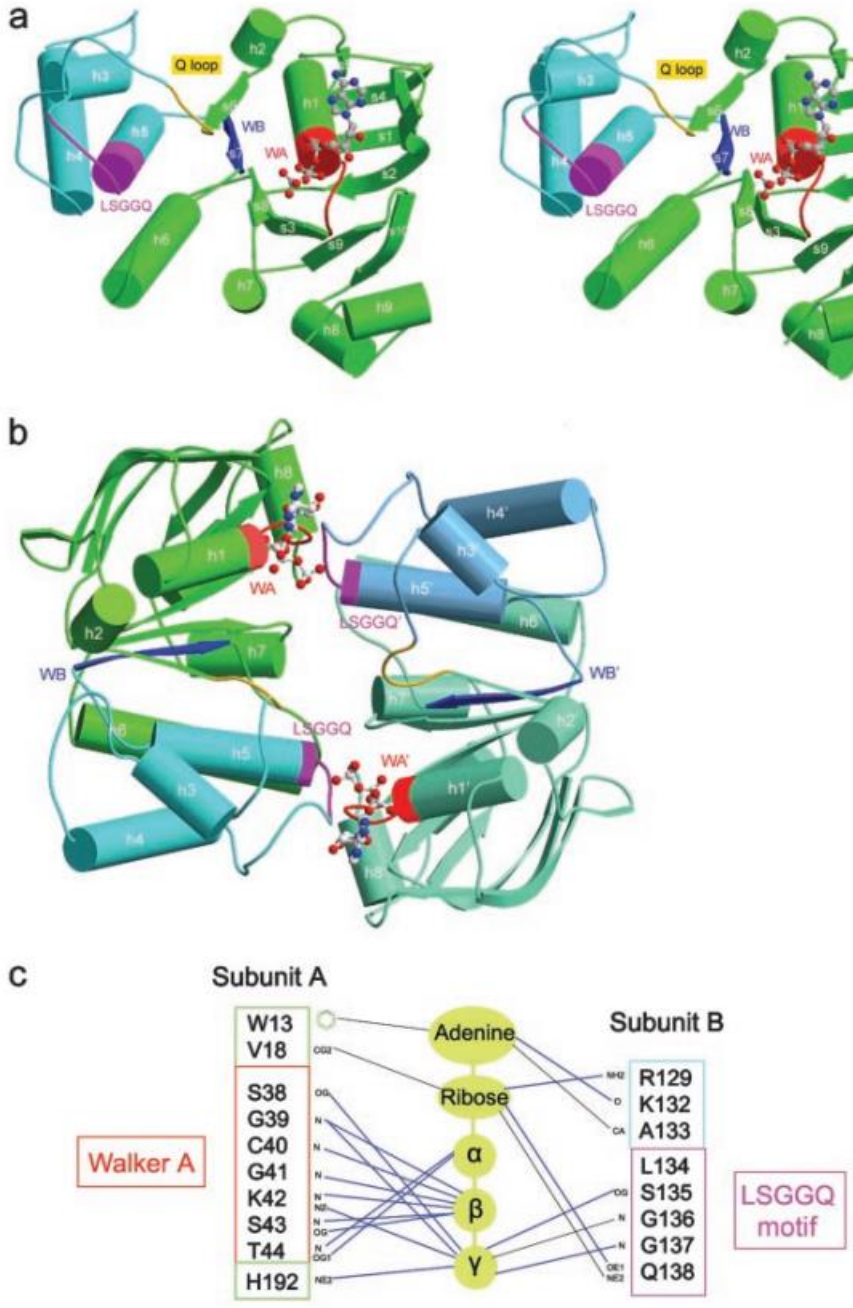


Figure 12 *E. coli* MalK structure in ATP bound form (119).

MalK is the NBD domain of the $MalFGK_2$ ABC transporter that is responsible for maltose transport (a) Stereo view of the monomer E. coli MalK in ATP bound form. (b) The homodimer of MalK, viewed down the local twofold axis. The RecA-like subdomain is green, and the helical subdomain is cyan. Different colors further distinguish the conserved segments: Walker A motif (red), LSGGQ motif (magenta), Walker B motif (blue), and the Q loop (yellow). The ATP is represented in ball-and-stick model [O atom (red), N atom (blue)]. (c)

Schematic diagram of the interaction between one of the two ATPs bound to the homodimer. Black lines represent van der Waals contacts, and blue lines correspond to hydrogen bonds and salt bridges. (138).

Motif	**Consensus sequence**	**Function**	**Referencing structures**
Walker A or P loop	GXXGXGKST	ATP binding	HisP, MJ0789, MJ1267, Rad50, TAP1, GlcV, MalK
Q loop or lid	Q	a. TM subunit interacion	a. BtuCD
		b. Q H-bond to Mg^{2+}	b. MJ0796 (E171Q), GlcV-ADP
		c. Binding to the attacking water	c. MJ0796 (E171Q)
LSGGQ or linker peptide or signature motif	LSGGQXQR	ATP binding	Rad50, MJ0796 (E171Q), MalK
Walker B	hhhhD	D makes a water-bridged contact with Mg^{2+}	GlcV (Mg^{2+}ADP, Mg^{2+}AMP-PNP), MJ1267 (Mg^{2+}ADP) MJ0796 (Mg^{2+}ADP)
	E following Walker B	a.Binds to attacking water	a. MJ0796 (E171Q)
		b.Binds to Mg^{2+} through a water	b. GlcV (Mg^{2+}ADP, Mg^{2+}AMP-PNP)
H motif or switch region	H	His H-bond to γ-phosphate	MJ0796 (E171Q), MalK

Table 1 Function of conserved motifs in the nucleotide-binding domain NBDs

This table described the conserved motifs in the NBDs of the ABC transporter, references for the structure mentioned above are: HisP (139) , MJ0796 (132), MJ1267 (133), Rad50 (138), TAP1 (140), GlcV (141), MalK (138), BtuCD (142), MJ0796 (E171Q) (135).

1.2 METHODS AND MATERIALS

1.2.1 $LptB_2FG/LpB_2FGC$ complex plasmids generation

In order to screen for the best expression level and stability of $LptB_2FG/LptB_2FGC$ from different strains, I have generated constructs for 9 species $LptB_2FGC$ in total, including *E. coli K12, Salmonella typhimurium LT2, Shigella flexneri 2a, Klebsiella pneumoniae, Enterobacter aerogenes, Vibrio cholera, Acinetobacter baumannii, Neisseria gonorrhoeae, Citrobacter freundii* during these 3 years using different methods and vectors. We finally focused on *Shigella flexneri 2a* and *Klebsiella pneumoniae* $LptB_2FG/LptB_2FGC$, of which one had the highest expression level and the other one had a unique SDS-PAGE pattern, respectively. To generate the most simplified, modified and expression-level qualified constructs, we amplified *E. coli, S. flexneri 2a* and *K. pneumoniae lptCAB* and *lptFG* fragments respectively using primers (see Table 11) with restriction enzyme sites EcoR I/ Kpn I for *lptCAB* fragments and Kpn I/ Xba I for *lptFG* fragments. PCR reaction mixtures were prepared as per Table 2 and PCR program was set according to Table 3

Component for 25 μl reaction system	
5 × Q5 Reaction Buffer	5 μl
10 mM dNTPs	0.5 μl
10 μM Forward Primer	1.25 μl
10 μM Reverse Primer	1.25 μl
Template DNA	variable
Q5 High-Fidelity DNA Polymerase	0.25 μl
Nuclease-Free Water	to 25 μl

Table 2 PCR reaction component

This is a general PCR reaction mixture preparation recipe. All the PCR reactions throughout this thesis resemble this one with slight changes on the volumes of some components. For multi-sample PCR, a Master Mix (a mixture of shared component of all reactions) would usually be prepared and divided into separated tubes to speed up the preparation procedure. dNTPs are bought from ThermoFisher Scientifc. Q5 DNA polymerases and related buffers are

bought from NEB Company.

Step	Temperature	Time
Initial denaturation	98 ℃	30 seconds
35 Cycles	98 ℃	10 seconds
	68(Tm-5) ℃	10 seconds
	72 ℃	30 seconds/kb
Final Extension	72 ℃	10 minutes
Hold	12 ℃	∞

Table 3 PCR program

This is the PCR program used here. All the other PCR programs resemble this one, and may have some changes on the cycles or annealing temperatures depending on different cases. The PCRs are all run on Bio-Rad PCR machine.

The PCR products were loaded into 1% agarose gels supplemented with 0.1% EtBr dye, which were run for 30 min at 100 V to check purity and yields. The PCR products agarose bands were excised and purified using GeneJET Gel Extraction Kit (Life Technologies). 25 μl of purified *lptCAB* PCR products and 0.5 μg pTRC99a plasmids were digested with EcoR I and Kpn I (FD digestion enzyme with Green Buffer) for 25 min at 37 ℃. The digested *lptCAB* fragment and plasmids were purified by 1% agarose gel with 0.1% EtBr. The final ligations were performed as Table 4 (Thermo Scientific T4 DNA ligase, 5U/μl).

Linear vector DNA	8 ng
Insert DNA	3 times of vector in molar
10×T4 DNA Ligase Buffer	1 μl
T4 DNA Ligase	1 μl
Nuclease-free Water	to 10 μl
Reaction temperature: room temperature	

Table 4 Ligation mixture component

This is the ligation mixture recipe for this experiment, other ligations resemble this one and may have some changes at the DNA amount or total volume. The ligation can happen both at room temperature or 16°C for 3 to 16h. T4 ligase is purchased from ThermoFisher Scientifc Company.

The ligation reactions were performed at room temperature for 3-4 h or at 4℃ overnight. 5 µl of the ligation mixture was added into 100 µl high efficiency competent cell (Home-made *Top10*) for transformation. The transformation mixture was gently tapped in the tube and incubated on ice for 30 min followed by heat shock for 90s at 42℃ using a water bath. 500 µl sterilized SOC (Super Optimal broth with Catabolite repression) medium was added into the transformation mixture and rocked at 200rpm and 37℃ for 1 hour. 200 µl transformed cells were plated on to LB medium plates that contained appropriate antibiotics. Single colonies were picked and inoculated into 10 ml of LB with 100 µg/ml ampicillin, at 37℃ 200 rpm overnight shaking. The plasmids pTRC99a-*lptCAB* were purified using GeneJET Plasmid mini prepare Kit (Life Technologies) and the clones were verified by double restriction enzymtic digestion and DNA sequencing. After the *lptCAB* fragment was inserted into the pTRC99a vector, *lptFG* was subsequently inserted in the same way and confirmed by sequencing, generating pTRC99a-*lptBCFG* plasmid (LptA was a soluble protein and would not affect the membrane protein $LptB_2FGC$ complex purification since we separate and extract proteins from the membrane fractions, which is insoluble before detergent extraction). To generate pTRC99a-*lptBFG* constructs, deletion mutation method(143) was used to delete *lptCA* fragment off the original pTRC99a-*lptBCFG* plasmid. Finally, *lptBCFG*/*lptBFG* genes were cloned into single plasmid, transcribed and expressed into a single mRNA polycistron, which was common in prokaryotic cells and could magnify the expression level of complexed protein expression.

1.2.2 Functional assay plasmids generation and site direct mutagenesis

The *E.coli* LptFG deletion NR1113 strain contains ampicillin resistance and our *lptBFG*-pTRC99C also contains ampicillin resistance gene. In order to select the transformtants, I performed resistance gene modification on pTRC99a plasmid and related plasmids. The pTRC99a-*E.coli-lptBFGC* plasmid ampicillin resistant gene was replaced by kanamycin resistance gene to generate pTRC99a-Kan^R-*E.coli-lptBFGC*

using GeneArt® Seamless PLUS Cloning and Assembly Kit with primers listed in Table 11. A Flag tag at position 230 for LptF (LptF-230-Flag), plus a c-Myc tag at 228 for LptG (LptG-228-Myc) or a Flag tag at position 138 for LptF (LptF-138-Flag), plus c-Myc tag at 144 for LptG (LptG-144-Myc) were inserted to generate two tagged plasmids named *lptBF230G228C* and *lptBF138G144C* respectively. This was so that the expressed protein LptF and LptG could be detected by appropriate antibodies during Western Blotting. These two plasmids and the plasmid pTRC99a-Kan^{R}-*E.coli-lptBFGC* were all confirmed to be able to rescue the lethal phenotype of NR1113 strain. The *lptBF230G228C* and *lptBF138G144C* were later used as the template to generate all the mutations. All single or double mutations were generated following Liu`s protocol(143). Primers for all the mutations are listed in Table 11 Table 11 Primers used in this thesis. All mutations have been confirmed by DNA sequencing.

1.2.3 Expression and purification of $lptB_2FG$/$lptB_2FGC$ complex

Recombinant plasmids pTRC99a-*lptBFG/lptBFGC* from different species were transformed into *C43 (DE3)* strain (Novagen) respectively for protein complex expression. Single colonies of all the transformations were inoculated in 10ml overnight culture which was then prepared into 10% glycerol stock for further experiments. The glycerol stock was first inoculated into 500 ml LB overnight culture with antibiotic (ampicillin 100μg/ml). Then the bacterial cell culture was amplified into 12L LB supplemented with antibiotic (ampicillin 100μg/ml) at 37 ℃ until the optical density of the culture measured at a wavelength of 600nm (OD_{600}) reached 0.6-0.8. The co-expression of LptB, LptF and LptG was induced with 0.1mM isopropyl β-d-thiogalactopyranoside (IPTG) for 16 hours at 20 ℃. Cells were harvested into pellets at 5,000rpm (JLA8.1000 Beckman) for 15min centrifugation the next day, and re-suspended in buffer (20 mM Tris-Cl, pH 7.8, and 150 mM NaCl) supplemented with cOmplete EDTA-free protease inhibitor tablet (Roche), 1 μg/ml DNase (Sigma-Aldrich) 100μg/ml lysosome. The cells were broken by passing twice through a cell

disrupter at 30,000psi (Constant Systems Ltd). Unbroken cells and cell debris were removed by centrifugation at 18,000g for 15 min at 4℃. The cell membrane was then pelleted with ultracentrifugation at 100,000g for 1h at 4℃. The membrane pellet was solubilized with Extraction Buffer (20 mM Tris-Cl, pH 7.8, 300 mM NaCl and 10 mM imidazole) supplemented with 1% (w/v) n-Dodecyl-β-D-Maltopyranoside (DDM) (Anatrace) supplemented with cOmplete protease inhibitor(Roche) at room temperature for 20 min. The suspension was then ultracentrifuged at 100,000g for 30 min before being loaded onto a 5 ml HP HisTrap column (GE HealthCare) and washed with Wash Buffer (20 mM Tris-Cl pH 7.8 300 mM NaCl and 60 mM imidazole) supplemented with 0.6% (w/v) 5-cyclohexyl-1-pentyl-β-d-maltoside (Cymal-5) (Anatrace), 0.03% (w/v) DDM and 0.4% n-nonyl-b-D-maltopyranoside (NM) mixed detergent. The $LptB_2FG$/$LptB_2FGC$ complex protein were eluted with Elution Buffer (20 mM Tris-Cl, pH 7.8, 300 mM NaCl and 300 mM imidazole) with 0.6% (w/v) Cymal-5, 0.03% DDM and 0.4% NM. The protein was further purified using size exclusion chromatography with a HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare) in 20 mM Tris-Cl, pH 7.8, 150 mM NaCl, 0.6% (w/v) Cymal-5, 0.03% DDM and 0.04% NM. Protein fractions of highest purity were collected and concentrated to 10 - 15 mg/ml using 100kDa concentrator.

1.2.4 Appropriate detergents screening and selecting of $LptB_2FG$

From the preview experience and literature reviewing, inner membrane protein crystallization relied heavily on a correct choice of detergents. To select the best detergents, 1 L of *S. flexneri* $LptB_2FG$ was cultured, whose membrane pellet was solubilized in 10 ml Extraction Buffer followed by ultra-centrifugation again to remove insoluble debris. Instead of loading all of the supernatant into a single Ni-NTA column, the supernatant was evenly distributed into 10 aliquots. One ml aliquots were incubated with 50 µl Ni-NTA beads for 15 min and loaded to a spin column. After centrifugation at 15,000rpm (12,000g) for 1min at bench top centrifuge to remove the non-binding

supernatant, the 10 columns that contained the Ni-NTA beads, were washed with Washing Buffer supplemented with various single detergents of certain concentration (see results for details), achieving a detergent-exchange screening. Finally the 10 samples were eluted with Elution buffer containing their own certain detergent. The eluted samples were checked by visible precipitation degree first, then analysed on SDS-PAGE. The detergent that could successfully elute $LptB_2FG$ and protect the complex from precipitation for at least 6 hours were identified as a suitable detergent.

1.2.5 $LptB_2FG$ crystallization, optimization and crystal soaking

$LptB_2FG$ complex was successfully purified and crystallized using 1-3 kinds of suitable detergent or detergents mixture from all the possible combination. The best crystals from the screening were found from (0.1 M Sodium chloride 0.1 M MES 6.5 36 % v/v PEG 300) and (0.2 M Sodium acetate trihydrate 0.1 M MES 6.5 28 % v/v PEG 400). Crystallization was performed by sitting drop vapour diffusion method using 1 µl protein with 1 µl reservoir solution. Less than 3 kinds of single detergent were suitable for crystal forming. After the crystals appeared, the crystals optimisation and reproducing were attempted by designing and preparing 2-dimension expansions of pH and PEG concentration of crystallization conditions. The prepared optimization solutions were arrayed in deep well block for storage. The crystallization trails were performed at room temperature, 22 ℃.

Nucleotides (ATP, ADP, AMP or AMP-PNP) were attempted for co-crystallization with the $LptB_2FG$ complex. The nucleotides and $MgCl_2$ were prepared into 200mM stock solution in ddH_2O respectively. Before setting onto crystallization, the $LptB_2FG$ complex were incubated with nucleotides and $MgCl_2$ at the final concentration of 2.5-10mM for both of the chemicals for 40-60min on ice.

For the crystals that had promising resolution for structure determination, the anomalous signal data, which was required to determine the substructure, could only be collected on crystals containing anomalous signal scattering element. To obtain

anomalous signal scattering element incorporated crystals, we soaked crystals in different kinds of solutions that contained heavy metal ion or bromide/iodide ion. The soaking was performed at a broad range of conditions, from durations of 10 s to 4 h; from concentrations of 1mM to 100mM.

1.2.6 LptB$_2$FG Data collection and structure determination

All datasets were collected at beamline I03, Diamond Light Source, UK. High redundancy platinum (potassium tetranitroplatinate (II) $K_2Pt(NO_2)_4$ soaked crystals) SAD data were collected at the platinum L3 edge (1.07226Å) using parameter oscillation 0.1 °, transmission 10% and image number 7,200 or 9999 for each anomalous signal scattering crystal and processed using XIA2 (144) with DIALS (145). XIA2 is an expert system for macromolecular crystallography data reduction, which recruit existing software to automatically complete data reduction process from images to merged structure factor amplitudes. This can automatically identify multi-wedge, multi-pass and multi-wavelength data sets and includes specific procedures to test for crystallographic special cases. With the push towards high efficient crystallography at synchrotron beamlines and automation of structure determination, the ability to reduce data with auto input fills and without manual operation is an important gap in the pipeline (144). DIALS (Diffraction Integration for Advanced Light Sources) is a new software project aiming at analysis of crystallographic diffraction images. DIALS mainly focuses on versatility and modularity, making best use of CPU and GPU to increase speed and accuracy analysis based on a comprehensive physical model (145). Several Hg-derivative data were combined together and scaled using AIMLESS (146) to amplify anomalous scattering signal (147). AIMLESS is a program scales together multiple observations of reflections, and merges multiple observations into an average intensities and it is a successor program to SCALA (147). In this case, we put several unmerged data processed by XIA2 (144) into AIMLESS (146), which were then analysed and merged into a single reflection file (.mtz file). The reason why for merging

data from same crystal and even several different crystals was to trying to increase anomalous redundancy to increase the anomalous signal to noise ratio. In total, about 60000 images were merged together to maximum the anomalous signal. SAD data of Hg-derivative crystals were collected at its peak wavelength 1.00068Å, using parameter oscillation 0.1 °, transmission 10% and image number 7,200 or 9999 for each anomalous signal scattering crystal and processed using XIA2 (144) with DIALS (145). SAD data of selenomethionine incorporated crystals were collected at its peak wavelength 0.9795Å, using parameter oscillation 0.1 °, transmission 10% and image number 7,200 or 9999 for each anomalous signal scattering crystal and processed using XIA2 (3) with DIALS (162). Native data were collected at the wavelength 0.976 Å.

We observed severe anisotropic problem during X-ray diffraction and data collection and was characterized as largely affecting the short axis of the cell with data extending to approximately 2.5 Å less along this direction. Anisotropy correction of the raw dataset was performed using the STARANISO webserver (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi) with a surface threshold of 1.2I/σI and approximate vector 0.1a* + 0.6b* + 0.8c*. The selenomethionine crystals from *S. flexneri* belonged to space group $I2_12_12_1$ with unit-cell dimensions: a = 105.5 Å, b = 210.8 Å, c = 258.9 Å and α = β = γ = 90° (the unit cell dimensions may vary a little bit between different crystals). All crystals from *K. pneumoniae* belonged to space group $P2_12_12_1$ with unit-cell dimensions: a = 110.15 Å, b = 124.53 Å, c = 398.09 Å and α = β = γ = 90°. The platinum positions in the Pt-derived *K. pneumoniae* crystal of the $I2_12_12_1$ space group were determined by the SAD method using the SHELX (148) suite that were successful in locating platinum sites but did not result in a readily interpretable map.

The phase determination of protein crystal X-ray diffraction contains several common ways: molecular replacement, experimental phase and density extension.

In recent experimental phase determination methods, MAD (Multi-wavelength Anomalous Diffraction/Dispersion) and SAD (Single-wavelength Anomalous Diffraction/Dispersion) are the most widely used and helpful ways to solve phases. The atomic scattering factor contains three components: a normal scattering term f_0 that is

dependent on the Bragg angle and two terms *f'* and *f"* that are not dependent on the scattering angle but on wavelength. The *f'* and *f"* of an element represent the anomalous scattering that happens at the absorption edge during X-ray crystal diffraction.

In the MAD method, data were usually collected from a single crystal or crystals at several wavelengths (2-4 kinds wavelength) (149), in order to maximize the absorption and dispersive effects. Usually, wavelengths are chosen at the absorption peak (γ_1), at the point of inflection (γ_2) on the absorption curve and two remote wavelength (γ_3 and γ_4) (remote data are optional) (Figure 13).

Single-wavelength anomalous diffraction/dispersion (SAD) is a technique used in X-ray crystallography that facilitates the determination of the structure of proteins by allowing the solution of the phase problem. In contrast to MAD, SAD uses a single dataset at a single appropriate wavelength. One advantage of the technique is the minimization of time spent in the beam by the crystal, thus reducing potential radiation damage to the molecule. Although SAD is called as single-wavelength anomalous dispersion, but no dispersive differences are used in this technique since the data are collected at a single wavelength. The SAD experiment only provides measurements of the anomalous, or Bijvoet, differences $\Delta F^{\pm} = |F_{PH}(+)| - |F_{PH}(-)|$. These are then used as estimates of the heavy-atom contribution to the scattering and enable direct or Patterson methods to be used to derive the positions of the heavy-atom substructure (149).

It is rare that experimental determined phases are sufficiently accurate to give a ready to use electron density map. Experimental phase are usually requires further phase improvement by density modification. The common techniques used in the density modifications contains solvent flattening, solvent flipping, histogram matching and noncrystallographic averaging. These methods are encoded into various programs, such as DM (150) RESOLVE (151) and CNS (152). The program ARP/wARP is particularly useful and performing the placing of atoms into electron density maps followed by refinements, model building and update (153). These method requires the data to be at least 2.7 Å resolution for extensive automatic interpretation, particularly in side chain assignment. Programs like BUCCANEER can work at lower resolution (154). SHELXE uses a novel approach on density modification (148). Density modification

is a cyclic procedure, involving the back-transformation of the modified electron density map to produce the improved phases (149). If native data have been collected at higher resolution, the density modification can work beyond the resolution limit of the experimental phase data and during cycles in the density modification. The phase information can be finally transfer to high resolution data (149).

In this $LptB_2FG$ cases, six data sets from three isomorphous crystals were combined to increase the anomalous multiplicity to around 60-fold with a diffraction limit 3.7 Å. These steps helped to produce a map where transmembrane helices were visible as well as broken density for the LptB domains. Phasing and density modification from this modified dataset with PHENIX (155) resulted in a significantly more interpretable map. The *S. flexneri* selenomethionine incorporated crystal $LptB_2FG$ data was collected and the structure was solved using molecular replacement using the *K. pneumoniae* $LptB_2FG$ model. And the anomalous difference Fourier map was calculated with this model for registering and validating the amino acids sequence.

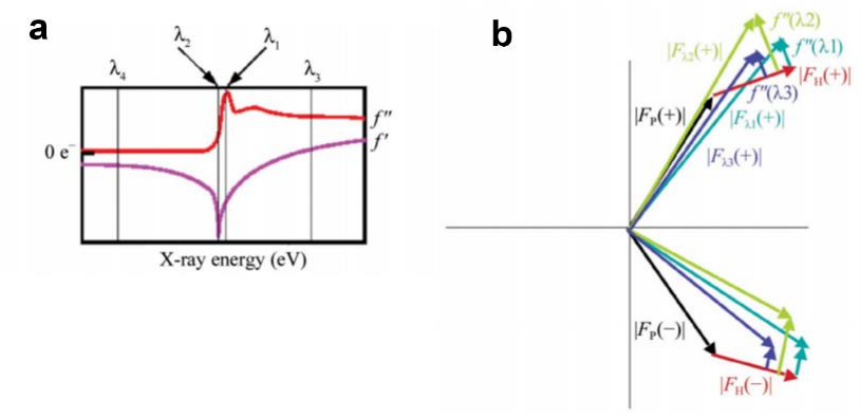


Figure 13 MAD phasing (149)

(a) A typical absorption curve for an anomalous scatter element. This kind of information can be gained by the Fluoresces Scan on a certain element at the Synchrotron before data collection (b) Phase diagram. |FP| is not measured, so one of the data sets is chosen as the 'native'.

Two copies of the high resolution structures of LptB (PDB 4QC2) (156) were placed into the density at the base of the transmembrane region and adjusted to fit by COOT (157). The transmembrane region of LptF and LptG were built using standard helices. In addition, the platinum sites are found adjacent to methionine, arginine and histidine residues which follows the property for this element. For the sequence registry, we have collected sulphur anomalous and mercury-derived crystal anomalous data trying to register the amino acids sequence, but they are too noisy to give a good registry. We then tried extensively to produce several selenomethionine incorporated crystals of $LptB_2FG$ from *S. flexneri* to register amino sequence of LpB_2FG structures.

Final refinement of the model was performed using REFMAC5 rigid body and restrained refinement(158). REFMAC5 utilizes different likelihood functions depending on the diffraction data, the presence of twinning and the availability of SAD experimental diffraction data. To ensure chemical and structural integrity in the refined model, REFMAC5 offers several classes of model parameterization. Low resolution data refinement can be achieved thanks to low resolution refinement tools such as secondary structure restraints, restraints to known homologous structures, automatic global and local NCS restraints, 'jelly-body' restraints and the use of novel long-range restraints on atomic displacement parameters (ADPs). When high-resolution data are available, REFMAC5 additionally offers TLS parameterization, fast refinement of anisotropic ADPs (Atomic Displacement Parameters). Refinement in the presence of twinning is performed in a fully automated fashion. REFMAC5 is a flexible and highly optimized refinement package that is ideally suited for refinement across the entire resolution spectrum encountered in macromolecular crystallography (159).

To improve the agreement of the protein structure model with the electron density map, refinements are needed which is an important step to adjusting the coordinate of atoms that refines the model to fitting better with the diffraction data. After refinement the electron density map will be improved and more details can be observed so that a more accurate model can be then build. After several rounds of this process, a model that fits with the original data best will be generated.

The R-factor is a measure of agreement within a structural refinement by measuring the difference between current model structure factor amplitude $|F_{cal}|$ and the observed structure factor amplitude $|F_{obs}|$ from the dataset.

$$R_{factor} = \frac{\Sigma \, ||Fobs| - |Fcal||}{\Sigma \, |Fobs|}$$

The R-free is used as an important model quality control and measures the quality of the current model according to the entire dataset. It is calculated with a randomly selected 5-10% data set, which is not used during refinement.

Coot (Crystallographic Object-Oriented Toolkit) is for macromolecular model building, model completion and validation, particularly suitable for protein modelling using X-ray data. Coot displays maps and models and allows model manipulations such as idealization, real space refinement, manual rotation/translation, rigid-body fitting, ligand search, solvation, mutations, rotamers, Ramachandran plots, skeletonization, non-crystallographic symmetry and more (157).

For $LptB_2FG$ structure determination, COOT was used at last to fix the errors and adjust bond angles to improve Ramachandran plots.

1.2.7 Functional assays and Western blotting analysis

All single or double mutants were transformed into the *E. coli lptFG* deletion NR1113 (114) strain by electroporation. The transformed *E. coli* cells were grown on LB agar plates supplemented with antibiotics (kanamycin 50 μg/ml) and 0.2% L-arabinose at 37°C for 12 h. Single colonies of each transformation were inoculated into 5 ml LB medium supplemented with above antibiotics and 0.2% (w/v) L-arabinose and incubated by shaking at 200 rpm at 37°C for 12 h. Subculture cells were used for functional assays. *E. coli* NR1113 with the empty plasmid pTRC99a-*Kan* was used as a negative control while plasmids *lptBF230G228C* and *lptBF138G144C* were used as positive controls. For functional assays, NR1113 cells harbouring mutation plasmids or control plasmids were harvested, washed 2 times and diluted with sterile LB medium to an absorbance OD_{600} = 0.08 and streaked onto LB agar plates supplemented with

kanamycin 50 μg/ml. Cell growth was recorded after overnight incubation at 37 °C. All the assays were repeated in triplicate.

Western blotting was performed to examine the $LptB_2FG$ leaky expression levels of mutants and controls during the functional assay experiments. 0.5 ml of overnight cultures of transformed NR1113 cells with respective control or mutation plasmid were inoculated in 50 ml LB supplemented with antibiotics (kanamycin 50 μg/ml) and 0.2% L-arabinose at 37°C for 6 h and pelleted. The cells were re-suspended in 1 ml buffer (20 mM Tris-Cl, pH 7.8, and 150 mM NaCl) supplemented with cOmplete (Roche), 1 μg/ml DNase (Sigma-Aldrich). The cells were lysed by sonication for 1 min on ice. The lysed samples containing whole cell membranes were further solubilized with Extraction Buffer supplemented with DDM to a final concentration of 2% using rotation for 20 minutes at room temperature. The undissolved debris was removed with desktop centrifugation at 13,000*g* for 10 minutes at 4 ℃. The supernatant was loaded into Ni-NTA agarose beads spin column, washed with 0.05% DDM Wash Buffer and eluted with 0.05% DDM Elution Buffer to further purify and concentrate samples. The eluted samples were supplemented with 4×SDS–PAGE loading buffer before been heating for 10 min at 98 ℃. After being centrifuged for 1 min at 13,000g, 10 μl of each sample were loaded into 4–12% Bis-Tris Plus Gel (Invitrogen) for SDS–PAGE. The proteins were transferred to PVDF membrane using Trans-Blot Turbo Transfer Starter System (Bio-Rad) at 20 V for 20 min. The PVDF membranes were blocked in 10 ml protein-free T20 (TBS) blocking buffer (Thermo Fisher Scientific) at 4 ℃ for 1 h. The membranes were incubated with 10 mL anti-Flag or anti-Myc monoclonal antibody (diluted, 1:300) (Sigma) at room temperature for 1 hour followed by washing with PBST (Thermo Fisher Scientific) four times and incubated with IRDye 800CW goat anti-mouse IgG (diluted, 1:20,000) (LI-COR) for 30 min. The membrane was washed with PBST four times and PBS twice. Images were acquired using LI-COR Odyssey (LI-COR).

1.2.8 ATPase activity assay

ATPase activity was performed using ATPase/GTPase Activity Assay Kit (Sigma). C43 (DE3) cells harbouring lptBFG (or lptBFG mutant) plasmid or LolCDE plasmid were cultured in 1L LB medium. The protein over-expression and the cell collection were performed using the same protocol as method section 1.2.3 described. Cells were twice disrupted by passing through a cell disruptor at 30,000 psi. The membrane fraction was harvested by ultra-centrifugation at 100,000g for 30 min and solubilized in 1% DDM followed by another ultra-centrifugation. The supernatants of each sample were loaded onto a gravity column containing pre-balanced 100 µl of Ni^{2+}-NTA beads. The $LptB_2FG$ or LolCDE complex was washed with 15 column volumes of wash buffer (40mM imidazole 20mM Tris-Cl pH8.0 300mM NaCl and 0.05% DDM), and eluted using the elution buffer (250mM imidazole, 20mM Tris-Cl pH8.0,300mM NaCl and 0.05% DDM).

The protein concentration of all samples were determined with detergent compatible Pierce BCA Protein Assay Kit (Thermo Scientific) according to the manufacture's instruction. Briefly, 2.5 µl of purified protein was diluted to 25 µl for the BCA assay. Albumin (BSA) was used as standard. 200 µl of Working Reagent (made by mixing Reagent A and Reagent B at 50:1 volume ratio) was added to each sample and incubation at 37 ℃ for 30min. The absorbance at the 562nm was measured and the protein concentration of each sample was determined.

The ATPase activity assay was performed in 96 well plates. 1 µl of each sample was mixed with 4 µl 0.5% DDM TBS (20mM Tris-Cl pH7.8, 150mM NaCl) and 5 µl Assay buffer (ATPase/GTPase Activity Assay Kit) to make 10 µl of ATPase activity assay sample. The phosphate standards and blank control for colorimetric detection was prepared according to the manufacturer's instructions of ATPase/GTPase Activity Assay Kit (Sigma). 30 µl reaction mix (made by 20 µl Assay buffer plus 10 µl 4mM ATP solution) was added into each ATPase activity assay sample. After incubation at room temperature for 15min, 200 µl reagent (ATPase/GTPase Activity Assay Kit) was added into each sample to terminate the reaction and all samples were incubated for additional

30min. The absorbance at 600nm was measured. All assays were repeated 6 times. ATPase activities of all samples were determined using the mean value of the samples according to the linear regression of standards.

1.3 RESULTS

1.3.1 The $LptB_2FG$ and $LptB_2FGC$ proteins form stable complexes

The $LptB_2FG$/$LptB_2FGC$ were co-expressed in their own single plasmid in polycistron as method described in 1.2.3. $LptB_2FGC$ of 9 species were included in the expression test during my PhD period. In all the expression plasmids, LptB contained 8 ×His-tag at its C terminus, which was used to purify the whole complex during Ni-NTA purification. The detergent DDM was chosen to protect complexes in the species expression screening. According to the result of $LptB_2FGC$ species screening, LptB from all the species could be purified after Ni-NTA purification but only samples of *K. pneumoniae* and *S. flexneri* had LptF/G bands on SDS-PAGE, compared to the *E.coli* sample control. Based on the screening result, both complexes were successfully purified from two species (*K. pneumoniae* and *S. flexneri*). The *K. pneumoniae* $LptB_2FG$ gel-filtration is shown in Figure 14. $LptB_2FG$/$LptB_2FGC$ protein complexes tended to aggregate and precipitate. The results shown here were under the best combination of detergents (see results part for details) that could stabilize the protein complex in solution form. With non-suitable detergents, $LptB_2FG$/$LptB_2FGC$ would precipitate immediately after being eluted from Ni-NTA column. *S. flexneri* $LptB_2FGC$ protein complex was analysed on SDS-PAGE (Figure 15). The SDS-PAGE analysis of *S. flexneri* $LptB_2FG$ and *K. pneumoniae* $LptB_2FG$ are shown in Figure 16 and Figure 17, respectively. Interestingly, the same protein complexes from two species that shared sequence identity of over 80% showed different SDS-PAGE patterns. The LptF of *K. pneumoniae* seemed to migrate slower than that of *S. flexneri* however *K. pneumoniae* LptF had only 1 residue less than the *S. flexneri* LptF. Of course, *S. flexneri* $LptB_2FG$ shared exact same pattern with *E.coli* $LptB_2FG$ on SDS-PAGE. However the highly

hydrophobic nature of transmembrane residues would attract extra SDS molecules, resulting in membrane proteins migrating at a faster rate than the ladder, which was supposed to estimate a protein mass under a similar mass-to-charge ratio. Since LptF and LpG each contained 6 transmembrane helices as inner membrane transmembrane proteins, the migration rate on SDS-PAGE would be much faster than for a normal soluble protein. Some residue differences within LptF or LptG between these two species might severely affect the group SDS binding of the transmembrane residues, influencing the migration rate of these two proteins. Regardless, LptF of *K. pneumoniae* displayed a strong distinguishable SDS-PAGE pattern from other species in the screening, which aroused our interests and *K. pneumonia* $LptB_2FG$ was set as the main target to be worked on.

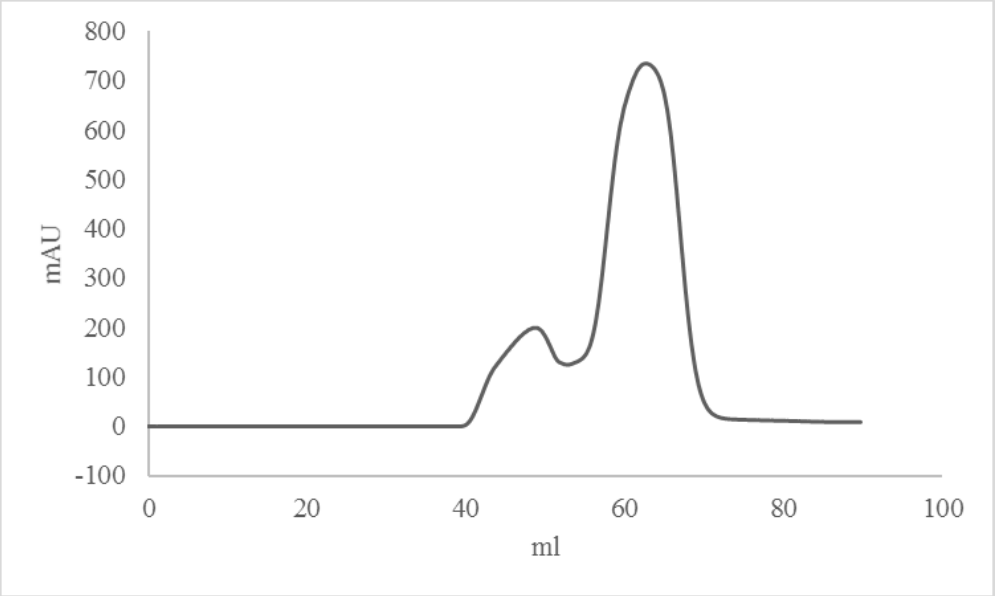


Figure 14 Gel-filtration result of *K. pneumoniae* $LptB_2FG$

10ml of the sample eluted from the Ni-NTA column were loaded into the HiLoad 16/600 SUPERDEX 200PG pre-equilibrated column. The $LptB_2FG$ protein complex from *K. pneumoniae* was purified with 0.05% DDM 20mM Tric-Cl pH 7.8 150mM NaCl during the gel-filtration, at 1ml/min and room temperature. The first peak at 48 ml is an aggregation peak and will precipitate within several hours after being collected in to the deep well block. And the $LptB_2FG$ main

soluble peak appeared at 62 ml. The gel-filtration result was further analysed on SDS-PAGE. When necessary, the main soluble peak may be harvested for further studies (crystallization or biochemistry studies). When using the detergents that can stabilize the $LptB_2FG$ complex, the gel-filtration results would be very similar to this one.

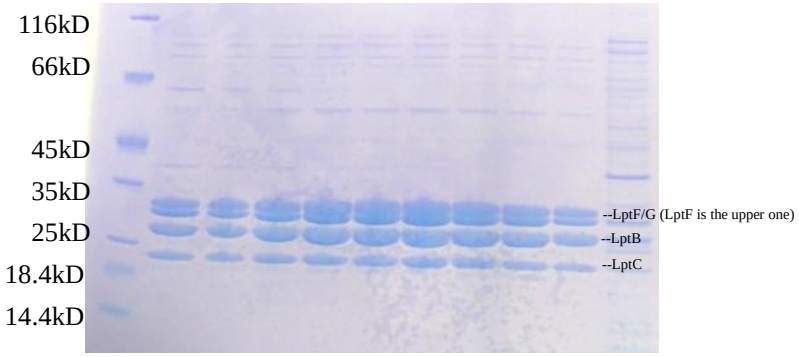


Figure 15 *S. flexneri* $LptB_2FGC$ SDS-PAGE analysis from gel-filtration

S. flexneri $LptB_2FGC$ was purified with 0.05% DDM 20mM Tric-Cl pH 7.8 150mM NaCl and samples were taken from gel-filtration fractions and analysed on SDS-PAGE. The strong bands from top to bottom were LptF, LptG, LptB and LptC, respectively. LptF and LptG bands are extremely close to each other and hard to separate clearly. 5µl protein samples were mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into The SDS-PAGE gel used is Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific) after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature. There are several non-specific bands above or below target bands and they are regarded as contamination proteins during purification.

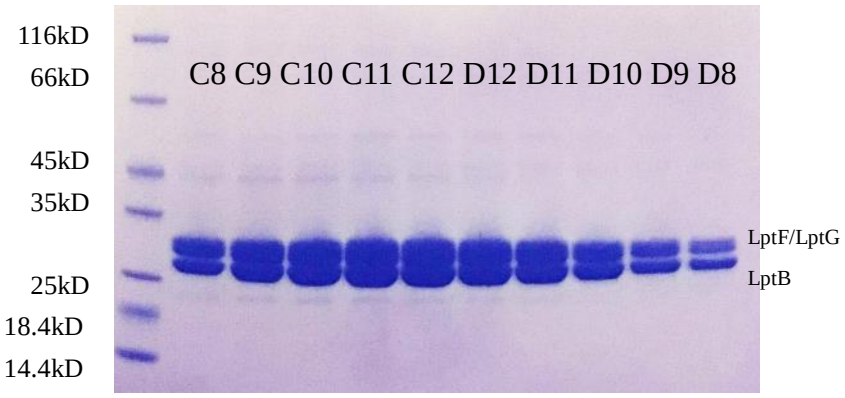


Figure 16 *S. flexneri* $LptB_2FG$ SDS-PAGE analysis from gel-filtration

S. flexneri $LptB_2FG$ was purified with 0.05% DDM during the gel-filtration and samples were taken from the main peaks of the gel-filtration fractions (C8-D8 cover volume range from 60-75ml) and analysed on SDS-PAGE. The protein bands distribution resembles the Figure 15 except for the absence of LptC. The LptF and LptG are still hard to separate and the protein ratio of each component of the substrate varies a little bit between the different wells (i.e. LptF/G relative amount in C8 is bigger than that in D8), suggesting that the purified proteins contains extra LptB besides $LptB_2FG$ (LptB-8His). 5µl protein samples mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loading into The SDS-PAGE gel used is Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific) after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature.

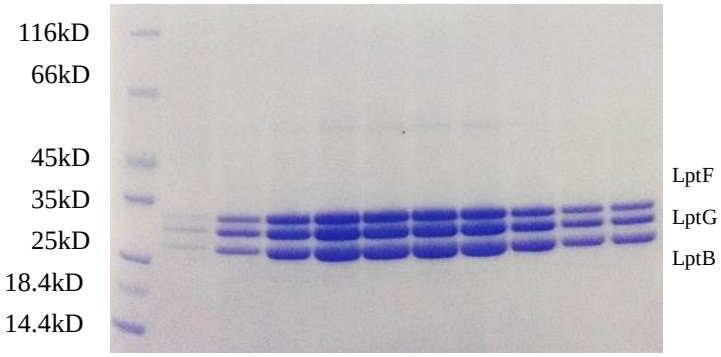


Figure 17 *K. pneumoniae* $LptB_2FG$ SDS-PAGE analysis

K. pneumoniae $LptB_2FG$ was purified with 0.05% DDM and samples were taken from gel-filtration fractions and analysed on SDS-PAGE. LptF from *K. pneumoniae* migrated a little bit slower compared to the *E. coli* one or *S. flexneri* one (Figure 16 and Figure 19) making the pattern of the bands different from them. 5µl protein samples mixed with 10µl ddH2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loading into The SDS-PAGE gel used is Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific) after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature.

1.3.2 Detergent stability screening of the inner membrane protein complexes $LptB_2FG$ and $LptB_2FGC$.

Crystallization of membrane proteins depended on the level of homogeneous distribution of the target protein in the final solution with little aggregation and high purity of target protein. And the choice of correct detergents made an essential difference. A good qualified detergent choice should stabilize the target protein at a reasonable level, preventing it from forming aggregations while not forming a strong micelle that affect crystallization. To find a good detergent, detergent screening was performed covering all the available detergents in my lab for $LptB_2FG$ and $LptB_2FGC$

from *E.coli* representing all other species $LptB_2FG$/$LptB_2FG$ protein complexes (Figure 18 and Figure 19). According to the SDS-PAGE result, around half of the detergents here could protect $LptB_2FG$ complex until protein complexes were eluted from Ni-NTA column during the detergent screening. However, most of the detergents failed to protect them after elution from the column as the eluted protein would precipitate within a few minutes in the tube. This precipitation would be more severe if large scale purifications were tried with those detergents. All the detergents were used at a concentration 2-3 times over normal usage amount. Those detergents that could keep the protein complexes from forming precipitates over 6 h were labeled as suitable detergents. After the detergents were chosen, *K. pneumoniae* $LptB_2FG$ was purified trying different combination of suitable detergents at different concentrations (1 CMC to 10 CMC) and set into crystallisation trails if applicable. After several rounds purification experiments, Cymal 5 and NM behaved better than others on $LptB_2FG$ complex protection. Desired aggregation levels were observed and the peak positions in gel-filtration of $LptB_2FG$ moved 3-4 ml later, compared to the DDM gel-filtration result. The SEC gel-filtration was the most efficient and practical way to roughly analyse a particle size, the gel-filtration peak position shift under the usage of these two detergent signifies a smaller average size of the micelles, implying an easier crystallisation. (GE Healthcare Data file 18-1100-52 AF)

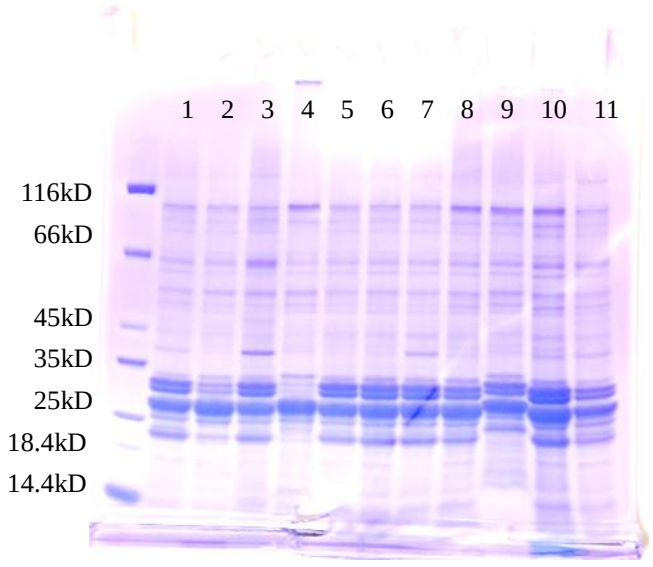


Figure 18 SDS-PAGE analysis of *E.coli* $LptB_2FGC$ detergent screening

In the detergent screening results of $LptB_2FG$, the amount of the inner membrane transmembrane domain (TMD) LptF & LptG bands quality decided whether a detergent was able to protect the complex. The supernatant were loaded into Ni-NTA columns separately and for each sample they were washed with different detergents. The eluted samples with the same detergents as the wash detergents were analysed here and also by precipitation observation. Lane 1: 0.05% DDM; lane2 0.5% C8E4; lane 3: 0.1% LDAO; lane 4: 1% β-OG; lane 5: 0.2% DM; lane 6: 0.06% UDM; lane 7: 0.01% LMNG; lane 8: 0.5% NM; lane 9: 0.05% Fos-14; lane 10: 0.3% Cymal 5; lane 11: 0.4% β-NG. The DDM detergent was used as a control to ensure that the complex was expressed and purified. The LptF/G bands were used to evaluate whether a detergent was suitable or not since the 8-His tag was attached to LptB. The lanes without or with little LptF/G refers to the unsuitable detergent choices.

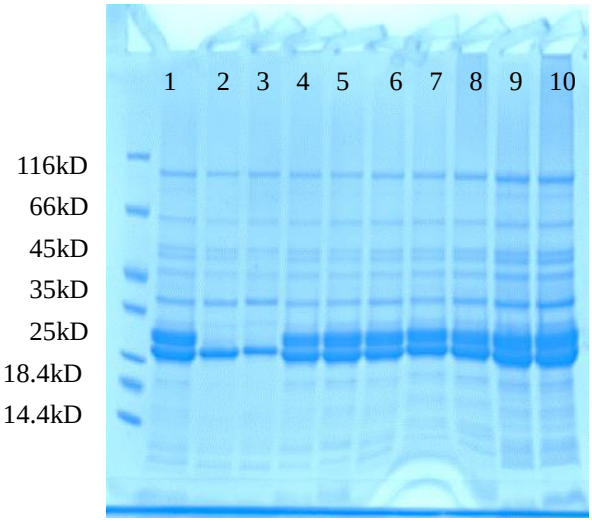


Figure 19 SDS-PAGE analysis of *E.coli* $LptB_2FG$ detergent screening

In the detergent screening results of $LptB_2FG$, the amount of the inner membrane transmembrane domain (TMD) LptF & LptG decided whether a detergent was able to protect the complex. The screening result resembled that of $LptB_2FG$, but the choices of detergent varied a little. Lane 1: 0.05% DDM; lane 2 0.5% C_8E_4; lane 3: 1% β-OG; lane 4: 0.2% DM; lane 5: 0.06% UDM; lane 6: 0.01% LMNG; 0.5% DDAO; lane 8: 0.1% LDAO: 0.5% NM; lane 9: 0.3%

Cymal 5; lane 10 0.2% LAPAO. The DDM detergent was used as a control to ensure that the complex was expressed and purified. The LptF/G bands were used to evaluate whether a detergent was suitable or not since the 8-His tag was attached to LptB. The lanes without or with little LptF/G refers to the unsuitable detergent choices.

1.3.3 Crystallization and optimization of $LptB_2FG$ protein complex

After trying DDM, NM and Cymal 5 individually or combined in pairs, crystal hits were obtained from our detergents candidates (Figure 20). With correct detergent choice, new hits from crystallization screening were obtained, however the resolution of the crystals stayed below 6 Å, which is not sufficient to determine the protein's structure. To break this deadlock and improve crystal quality, we optimised the whole $LptB_2FG$ project at three levels: protein purity; detergent; crystallization. For protein purity level, we increased the imidazole concentration during Ni-NTA washing step from 30 mM (universal protocol) to 60 mM leading to much higher purity, but also loss of half the product of each purification. For detergent, we broke the traditional limit of two detergent combination and tried mixing all the suitable detergents in a single purification experiment, which luckily did not produce any side effect from normal observable range (SDS-PAGE gel-filtration). For the crystallization level optimization, extensive optimisation solution were designed and prepared by ourselves or purchased from Molecular Dimensions to perform optimization on crystal forming conditions (0.1 M Sodium chloride 0.1 M MES 6.5 36 % v/v PEG 300) and (0.2 M Sodium acetate trihydrate 0.1 M MES 6.5 28 % v/v PEG 400) Certain substrates were tried to be added into final protein samples before setting onto crystallization (ATP, ADP or AMP-PNP & $MgCl_2$) as a kind of optimization method. After the advanced optimization of the whole procedure, crystals grown from the optimization well were much larger, and the best crystals were diffracted to 4.0 Å which finally made the structure determination possible.

To determine experimental phase of $LptB_2FG$, the crystals were soaked in 2 mg/ml potassium tetranitroplatinate (II) $K_2Pt(NO_2)_4$ for 4 hours at room temperature.

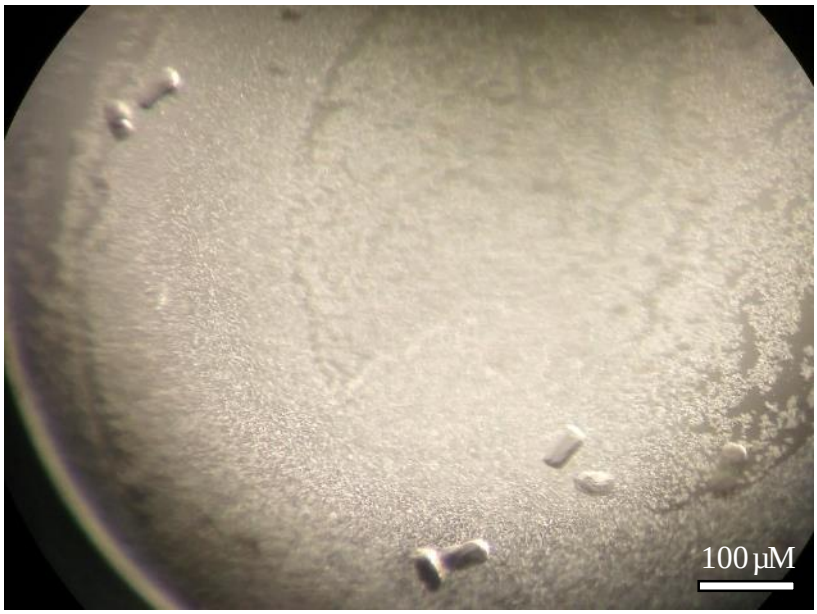


Figure 20 Crystals of the *K. pneumoniaem* $LptB_2FG$ grown from original screening

K. pneumoniae $LptB_2FG$ crystallized under detergents combination of 0.6% Cymal 5 and 0.05% DDM (20mM Tris-Cl pH8.0 150mM NaCl) at room temperature and appeared after 2 weeks, the original crystals could diffract up to 6 Å resolution. (0.1 M Sodium chloride 0.1 M MES 6.5 36 % v/v PEG 300)

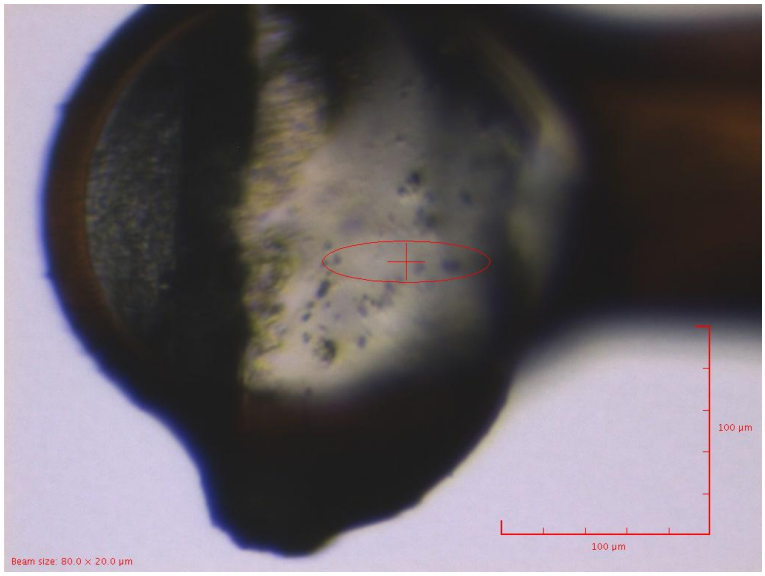


Figure 21 Crystals of the *K. pneumoniaem* $LptB_2FG$ after optimization

K. pneumoniae $LptB_2FG$ crystallized under detergents combination of 0.6% Cymal 5 and 0.05% DDM with buffer 20mM Tris pH8.0 150mM NaCl at room temperature. After optimization, the crystal appear after one week and become full size within one month. The size of the optimized crystals were about 100×100×40 μm^3 and the resolutions of these crystals were about 3.6-4.0 Å. This image was captured at I03 beamline Diamond Light Source before data collection. (0.1 M Sodium chloride 0.1 M MES 6.5 36 % v/v PEG 300)

1.3.4 Nucleotide-free form of the *K. pneumoniae* $LptB_2FG$ transporter

The crystal structure of the Gram-negative inner membrane LPS ABC transporter LpB_2FG complex was determined. The substructure was determined from the SAD data collected on Pt-derived crystals. After correction and modification the final structure was determined at 3.7 Å; residues registration was validated Se-Met (*S. flexneri*) data (Table 5). The anomalous difference Fourier map of Pt-derived data (*K. pneumoniae*) and Se-derived data (*S. flexneri*) are shown in Figure 22. There were about six platinum atoms found per transporter for the initial phasing and structure determination and the amino acid sequence registry was completed using the Se-derived data with about 21 selenium atoms found per transporter. Since the $LptB_2FG$ from *K. pneumoniae* reaches higher resolution, the following studies are based on *K. pneumoniae* $LptB_2FG$ structure. The $LptB_2FG$ complex of *K. pneumoniae* was co-expressed, purified and crystallised using methods described in 1.2.3. The crystals belonged to space group $I2_12_12_1$ with the cell dimensions a = 105.3, b = 210.5, c = 258.9 and $\alpha = \beta = \gamma = 90°$. There was one $LptB_2FG$ transporter complex molecule per asymmetric unit. The overall structure of $LptB_2FG$ covers approximately 86 Å in width and 128 Å in length. (Figure 23). As an ABC transporter, four typical domains (two TMDs and two NBDs) are clearly seen in the structure. LptF and LptG each contained six transmembrane α-helices, TM1~6F and TM1~6G respectively, a periplasmic β-jellyroll domain, three periplasmic loops and a pair of cytoplasmic turns. The transmembrane helices of LptF and LptG corresponds to

TMDs. LptF and LptG formed a hetero-dimer with a hydrophobic cavity between them (Figure 23 and Figure 25). The hydrophobic cavity covers a space of 25 Å in width and 8 Å in length extending from the IM into the periplasm (Figure 25). The centre hydrophobic cavity is predicted to be the LPS extraction site when the LPS transportation starts. Additionally, two LptB copies form a homo-dimer in a "V" shape at the cytoplasmic face of the TMDs region, sealing the bottom part of the complex. The two LptB corresponds to the two NBDs of a typical ABC transporter. The $LptB_2$, are in nucleotide free form in the *K. pneumoniae* $LptB_2FG$ complex structure. The typical Walker A, Walker B, D loop, Q loop, H loop and Signature motif can be clearly identified as in Figure 24. The super imposition of LptB protomer of *E. coli* (nucleotide binding form) with LptB protomor of $LptB_2FG$ of *K. pneumoniae* (nucleotide free form) with RMSD of 1.953 over 229 Cα atoms reveals some small conformational changes, possibly leading to further conformational changes of the whole $LptB_2FG$ complex. Apart from TMDs and NBDs, the periplasmic domains of $LptB_2FG$ makes the transporter special. The periplasmic domains adopt β-jellyroll fold, implying the function of the periplasmic domains as part of the LPS transporting bridge in the same manner of LptA and LptC (28). Despite the high similarity of the folding manner of the periplasmic domains, in our *K. pneumoniae* $LptB_2FG$ structure, the periplasmic domains of LptF and LptG lean to one side rather than in a mirror equal position (Figure 23). The periplasmic domains lean to TM1G-5F side, opening the TM5F-1G side (Figure 23). At this position, the hydrophobic cavity and the periplasmic domain of LptF forms an ideal path for LPS to pass through (Figure 23). In contrast the opposite side is closed and currently impossible for substrate to pass.

No structural homologues to LptF or LptG were found in the DALI server, suggesting that LptF/G represents a previously unreported membrane protein fold. The well-known ABC transporters $MalFGK_2$ (160), BtuCD-F (161), PglK (162) and MsbA (163) transported maltose, vitamin B, lipid-linked oligosaccharide and lipid A oligosaccharide across the IM, respectively. However, to our knowledge the $LptB_2FG$ complex is the first crystal structure of an ABC transporter to translocate a substrate laterally from the inner membrane's outer leaflet to periplasmic domain of another

protein (LptC), rather than directly across the IM, making it an example of its family. Another similar example is LolCDE (164) which recognizes and extracts lipoproteins from the periplasmic side of the IM and passes the substrates to LolA.

	K. pneumoniae $LptB_2FG$-Pt	*K. pneumoniae* $LptB_2FG$-Hg	*K. pneumoniae* LptBFG-sulfur	*S. flexneri* LptBFG-SeM
Data collection				
Resolution (Å)	29.96-3.70 (3.90-3.70)	22.99-5.14 (5.27-5.14)	29.15-4.23 (4.64-4.23)	29.96-6.00(6.61-6.00)
Wavelength (Å)	1.07226	1.00068	1.77120	0.9795
Space Group	$I2_12_12_1$	$I2_12_12_1$	$I2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, b, b (Å)	105.3, 210.5, 258.9	100.9, 215.9, 258.6	106.9, 212.1, 260.6	110.15, 124.53, 398.09
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Completeness (%)	99.8 (99.8)	99.0 (99.3)	97.3 (90.1)	99.2(100)
I/σ(I)	15.5 (1.7)	6.0 (1.1)	26.4 (7.9)	10.7(1.5)
Redundancy	120.8 (117.5)	26.2 (27.9)	173.6 (81.8)	19.4(20.2)
R_{merge} (%)	26.2(>100)	40.9(>100)	21.1(72.4)	22.0(>100)
$CC_{1/2}$ (%)	100 (94.6)	99.5 (76.0)	100 (99.6)	99.8(84.9)
Phasing				
Resolution (Å)	29.96-3.7			
Site (Pt)	4			
Figure of merit	0.332			
Refinement				
Resolution (Å)	29.96-3.7			
No. reflections	20311			
R_{work}/R_{free}	0.29/0.32			
No. atoms	8422			
Ligand/ion	2			
B-factors				
Protein	112.40			
R.m.s deviations				
Bond lengths (Å)	0.009			
Bond angles ( )	1.160			
Ramachandran statistics				
Allowed (%)	96.8			
Outliers (%)	3.2			
PDB code	5L75			

Table 5 Data collection and structure refinement statistics of *K. pneumoniae* $LptB_2FG$.

R_{factor} = Σ|| Fobs|−| Fcal||/Σ|Fobs|, where Fobs and Fcal are observed all reflection measured and calculated currently model as structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly selected not used in refinement.

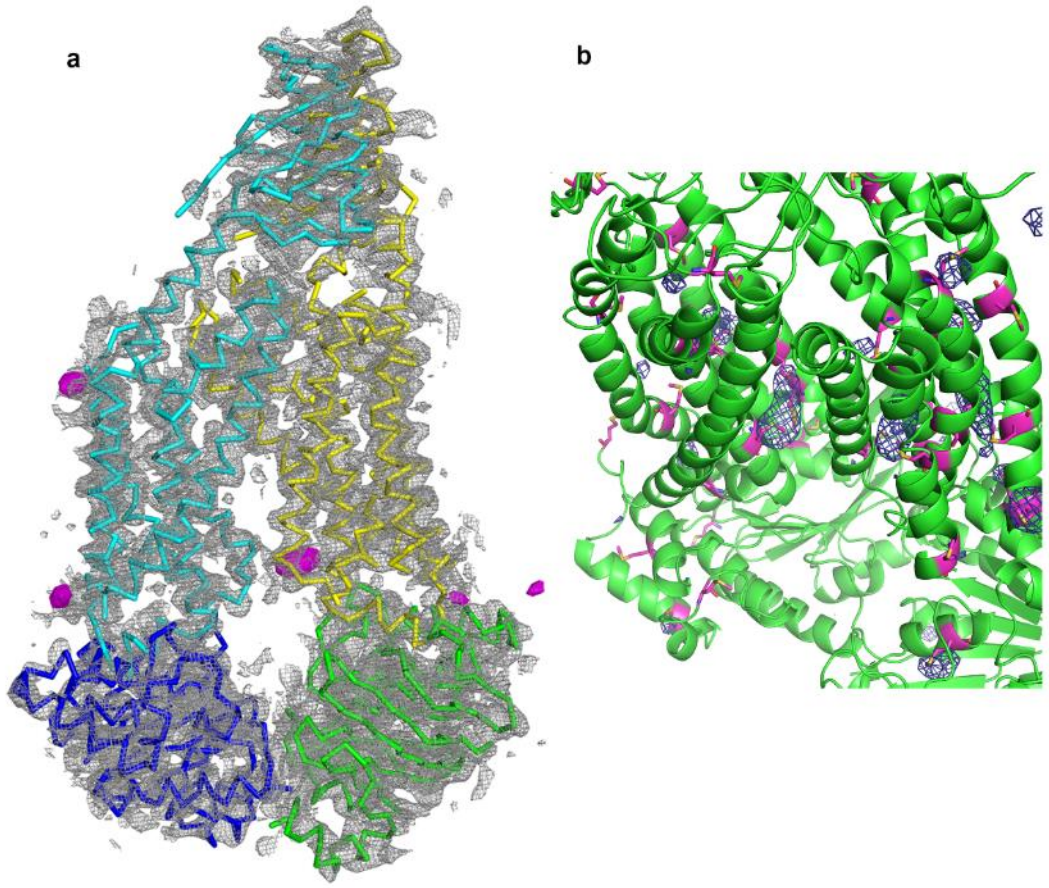


Figure 22 Anomalous difference map of *K. pneumoniae* $LptB_2FG$-Pt and anomalous difference map of *S. flexneri* LptBFG-SeM.

(a) The platinum derived $LptB_2FG$ SAD data from was *K. pneumoniae* collected at 1.07226 Å. The 2Fo-Fc and anomalous difference Fourier map were calculated with PHENIX (155). The 2Fo-Fc map was contoured at 1.2σ and displayed in grey colour and the anomalous difference Fourier map was contoured at 7.0σ in pink colour showing the platinum sites. The Cα trace structure of $LptB_2FG$ was shown here. LptF, LptG and two LptB are shown in cyan, yellow, green and blue respectively. (b) The selenomethionine-derived

$LptB_2FG$ SAD data from *S. flexneri* was collected at 0.9795 Å. The anomalous difference Fourier map was contoured at 3.0σ showing the registry of amino acids sequence registry of $LptB_2FG$ structure. The $LptB_2FG$ from *S. flexneri* was shown in green cartoon and methionine sites were highlighted in sticks with sulphur element coloured in deep yellow.

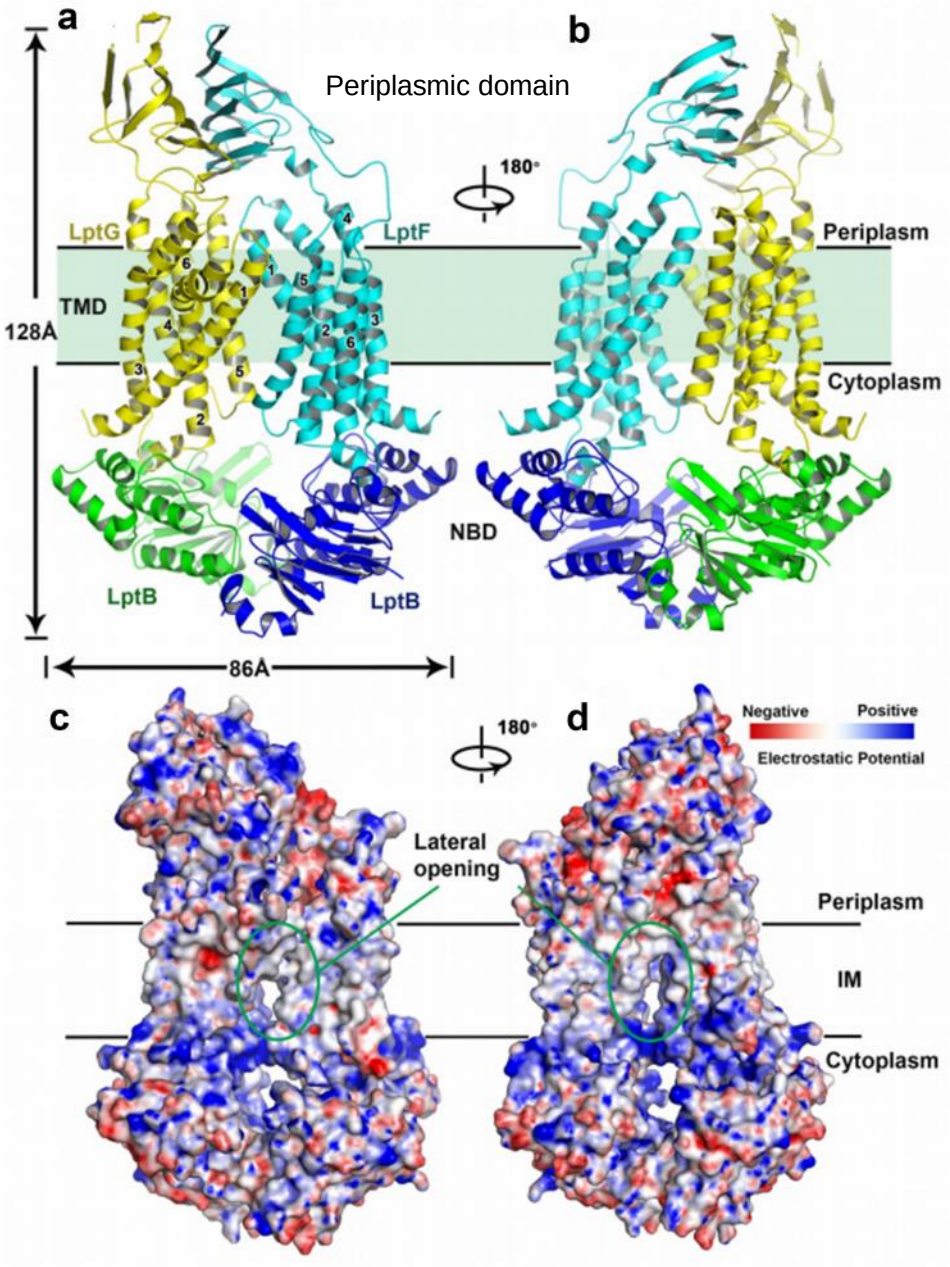


Figure 23 Crystal structure of LPS transporter $LptB_2FG$ from *K. pneumoniaem*

(A) Cartoon representation of $LptB_2FG$. LptF, LptG and two LptB are shown in

cyan, yellow, green and blue respectively. (B) 180° rotation view along Y-axis relative to left panel. (C) The electrostatic potential map of $LptB_2FG$. There is a lateral opening between TM1G and TM5F. Positive potential is labelled in blue and negative potential is labelled in red. (D) 180° rotation view along the Y-axis relative to the left panel. The lateral opening also opens between TM1F and TM5G.

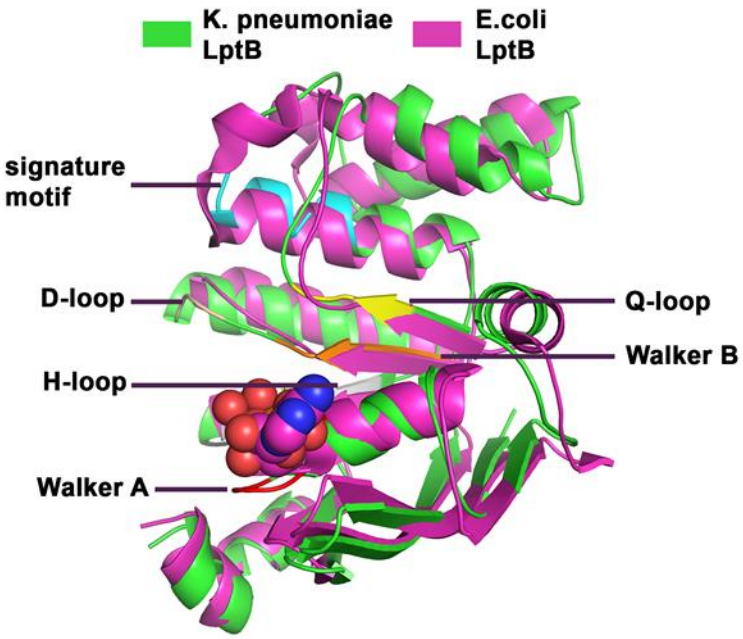


Figure 24 Superimposition of LptB structures of the *E. coli* and *K. pneumonae*

LptB of *E. coli* was only crystallized in presence of nucleotide (ATP, ADP or AMP). The ATP-free LptB structure of *K. pneumonae* was determined. Superimposition of structures of *E. coli* LptB (magenta) (PDB code 4QC2) and *K. pneumonae* LptB (green) with RMSD of 1.953 over 229 Cα atoms is presented here. The conformational changes were observed when ATP bound. The sequence identity of *E. coli* LptB and *K. pneumonae* LptB is 95.85%. The conserved and important motifs are labelled out.

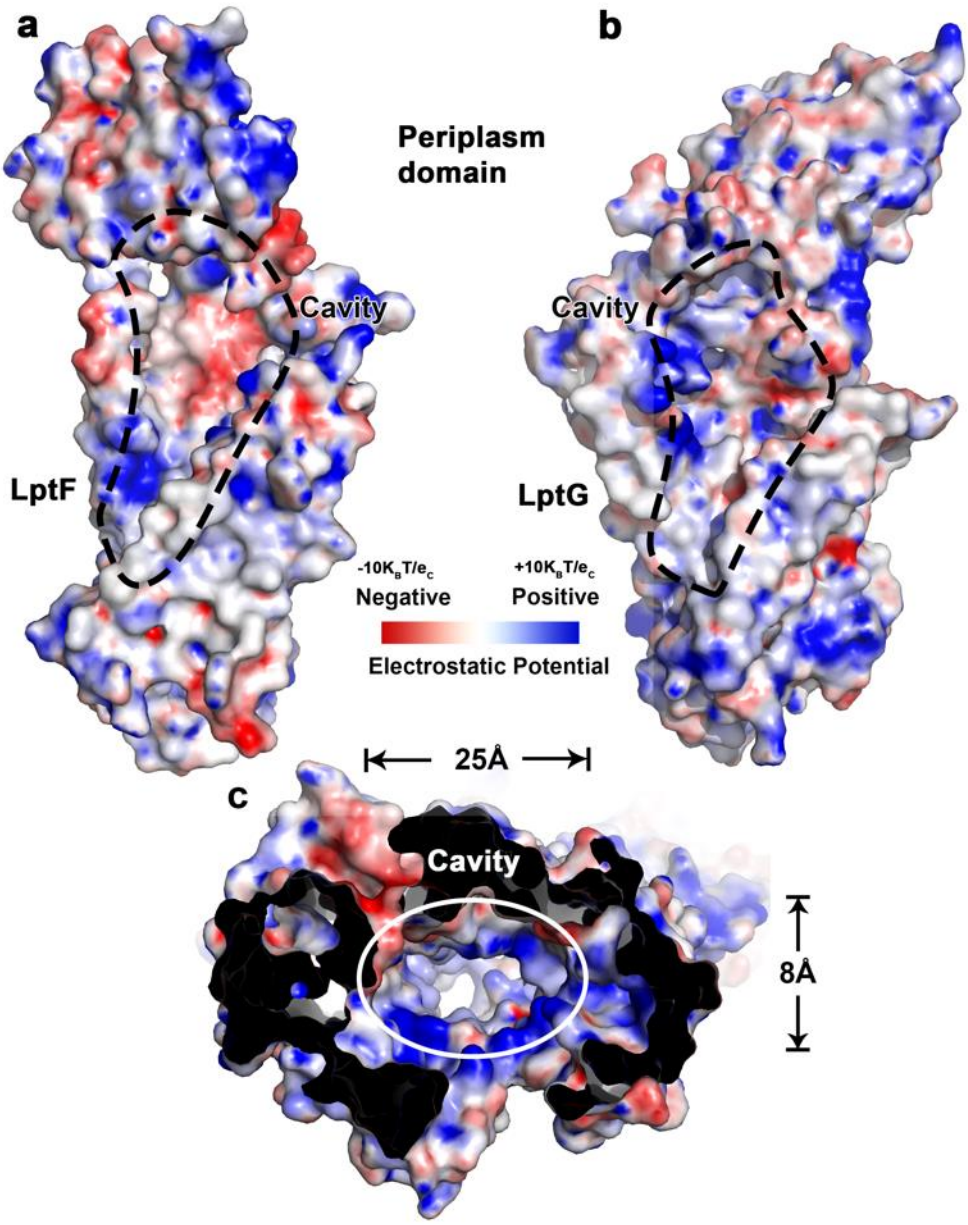


Figure 25 Electrostatic potential map of the cavity of $LptB_2FG$ of *K. pneumoniae*

The cavity of $LptB_2FG$, shown in the dotted circular lines (black). The cavity in the membrane section is very hydrophobic, while the cavity in the periplasm is highly positively charged. This feature may be important for LPS binding, as the lipid A of LPS is hydrophobic, while the core oligosaccharide is highly negatively charged. (a) The electrostatic potential map of LptF, showing the cavity side view. (b) The electrostatic potential map of LptG, showing the cavity side view. (c) The electrostatic potential map of the cavity of $LptB_2FG$, showing the cross section at the widest point of the cavity. The cavity is about 25 Å in length and 8 Å in width. The cavity is shown in the circular line (white).

1.3.5 Functional assays based on *K. pneumoniae* $LptB_2FG$ complex structure

Based on the crystal structure of *K. pneumoniae* $LptB_2FG$ complex and conserved residues analysis (Figure 26), we could identify some potential pathways of LPS extraction and transport by $LptB_2FG$. The TM1-F(G) crossed the IM at a 67 °(53 °) angle while TM1-G crossed the IM at a 67 °(53 °) angle relative to the membrane plane (Figure 27). These features resulted in a central cavity in between, which was predicted be an entry of LPS. Positively charged residues are required for lipids substrate specific binding in other membrane proteins, for example PglK (162). The highly conserved residues K34 (TM1G) and R136 (TM3G) are located in the upper cavity of the transporter. We hypothesised that the residues K34 and R136 may be involved in LPS extraction and transport. Functional assays indeed revealed that double glutamic acid substitution K34E/R136E caused cell death (Figure 27). In contrast, LptG residues K40 and K41 shared low sequence conservation among homologues, and the K40E/K41E of LptG substitution caused no impact on cell growth. We also expected that the hydrophobic cavity is important for $LptB_2FG$. Highly conserved hydrophobic residues F26 and L62 of LptF are located in the central cavity and mutagenesis of these two residues to alanine caused severe cell growth deficiency (Figure 27). These data suggest that the highly conserved residues in the cavity of $LptB_2FG$ are essential for the functionality of the complex, possibly involved in forming hydrophobic interaction or salt bridge with LPS lipid A or inner core phosphate group.

Besides the importance of the central hydrophobic cavity of $LptB_2FG$, the periplasmic domains of LptF and LptG should also play essential role for LPS transporting. The periplasmic domains of LptF and LptG adopt the β-jellyroll fold, which together comprised of 10 β-sheets, resembling the folding of the periplasmic domain of LptC, LptA and N-terminal domain of LptD, implying the similarity in potential functions as well. The periplasmic domains of LptF and LptG were in Y-axis-rotationally opposite position Figure 23, with the tunnels of both periplasmic domains connecting to the central hydrophobic cavity. These structural features suggests the possibility that the periplasmic bridge of LPS transport(165) might start with these two periplasmic

domains. To test it, we introduced the double proline mutations (F_D229P/Q231P, F_T225P/R223P, G_G228P/W230P and G_T225P/S223P) to disrupt the extracellular exposing β-sheets of LptF or LptG. Functional assays of all these mutants in Figure 29 showed severe growth deficiency. Western blotting detected leaky expression level of LptF and LptG Figure 29. To further study on the residual function of the periplasmic β-strands of LptF and LptG. We have also generated single mutants at the C-terminal β-strands of LptF periplasmic domain (F_D229P, F_Q231P, F_T225P and F_R223P) and at the C-terminal β-strands of LptG periplasmic domain (G_G228P, G_W230P, G_T225P and G_S223P). Functional studies showed that all the single proline mutants severely impaired cell growth, except F_T225P. Collectively, the periplasmic domain of last two β-strands of both LptF and LptG are important for LPS transport. The mutations are designed based on the amino acids sequence alignment results and the figure shows the structure of *K. pneumonia* and the mutations are based on the *E.coli* protein sequence.

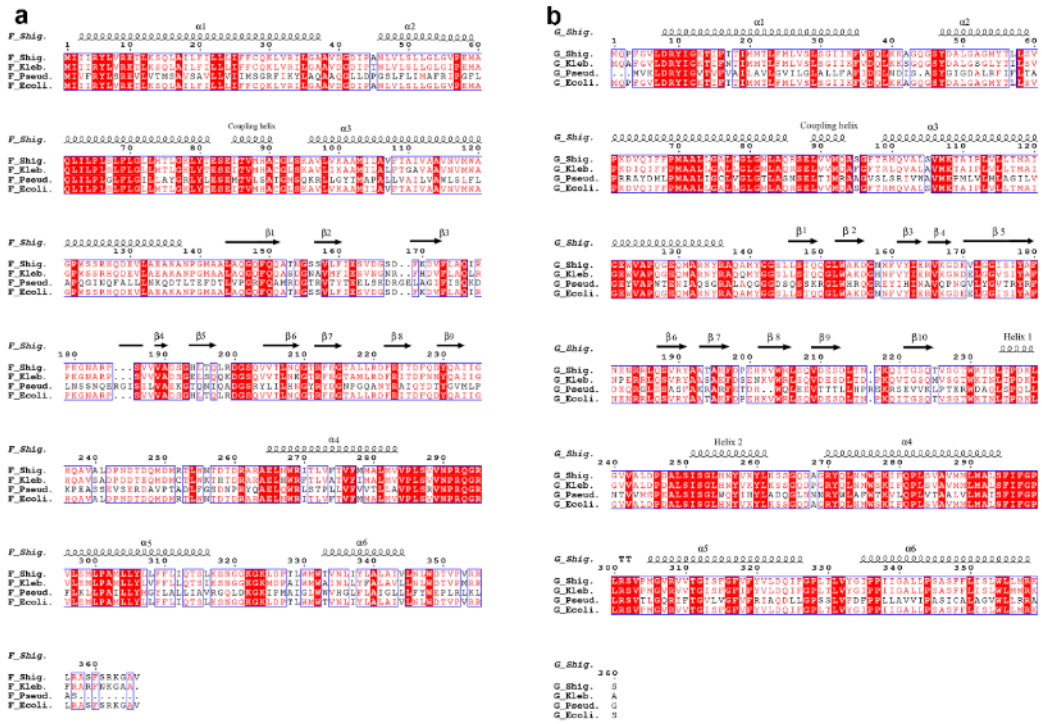


Figure 26 Amino acids sequence alignment of LptF and LptG

LptF and LptG amino acids alignment of *S. flexneri*, *K. pneumonia*, *P. aeruginosa* and *E. coli* respectively. The α-helices are labelled as α1-6. The β-strands are labelled as β1-10. Red boxes show conserved residues.

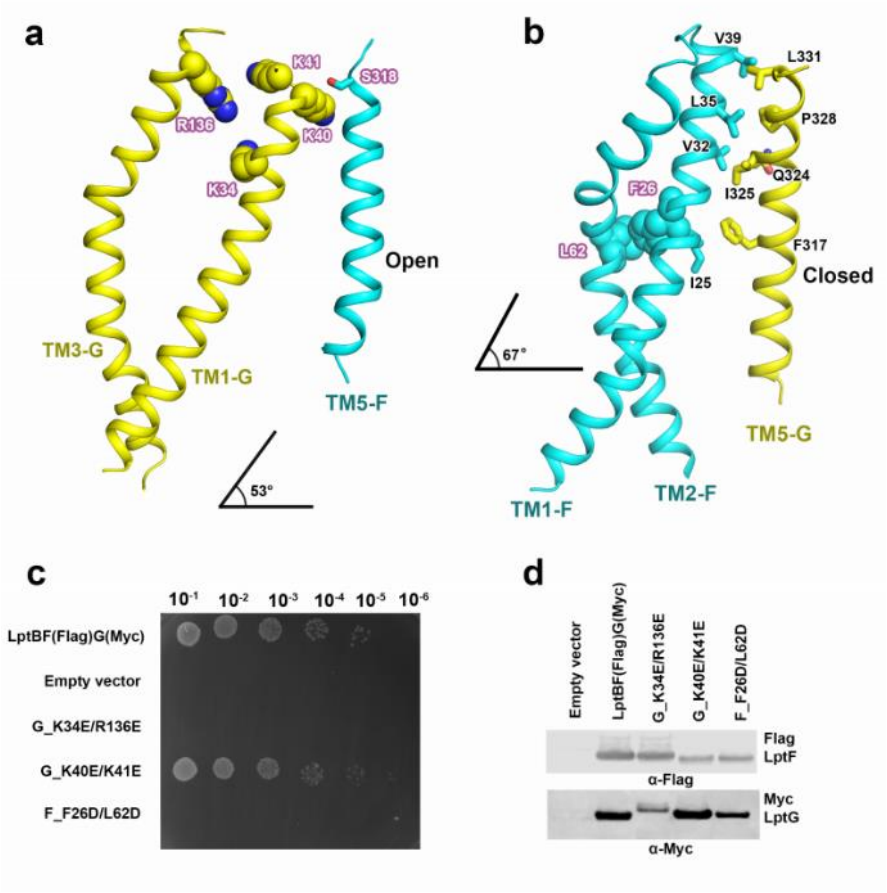


Figure 27 The transmembrane domains of $LptB_2FG$ and the lateral opening

The colour scheme is the same as in Figure 23. (A) The lateral gate of TM15F-1G. The residues K40 of TM1-G and S318 of TM5-F forming the contact are shown as sphere and stick, respectively. The positively charged and highly conserved residues K34 and R136 of LptG are also shown as sphere. (B) The lateral gate of TM1F-5G interaction. There are five residual interactions within TM1F-5G, these residues are shown in stick. The hydrophobic residues F26 and L62 of LptF are shown in sphere. The TM1-F(G) crosses the IM at an angle of 67°(53°). (C) Functional assays of the double mutants LptG_K34E/R136E, LptG_K40E/K41E and LptF_F26D/L62D in NR1113 *lptFG*

depletion strain. All bacterial cultures were adjusted to an OD600 of 0.5 and diluted 1:10 serially. (D) Detection of protein expression level of the samples in functional assays.

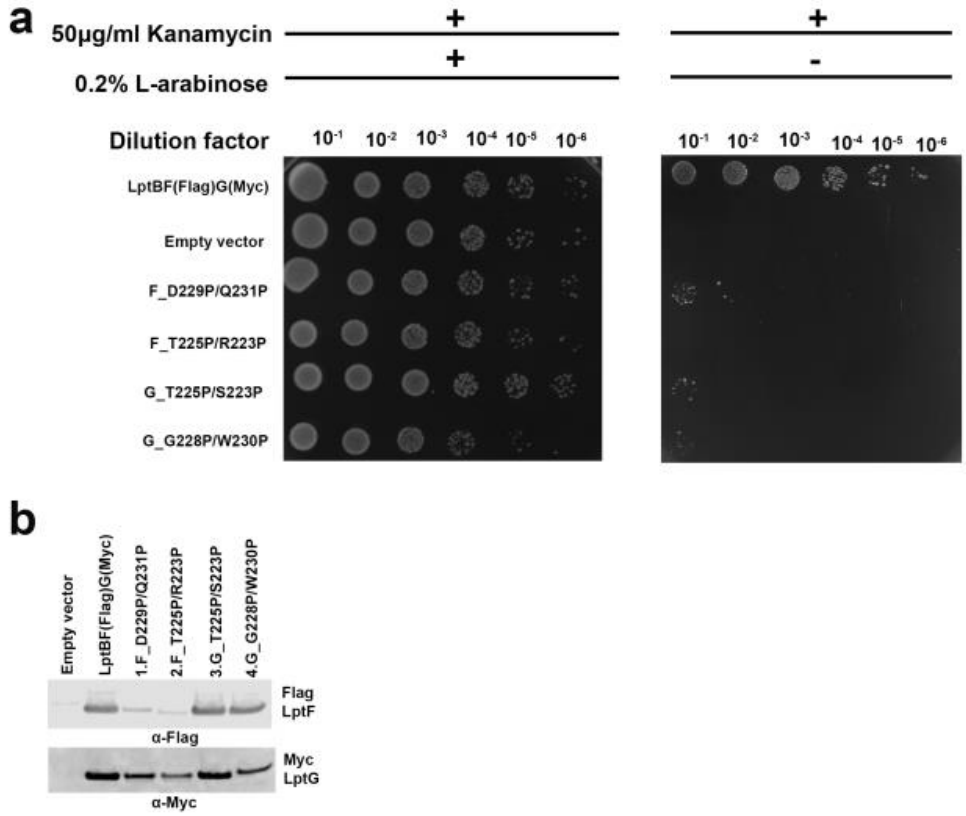


Figure 28 Complementation assays of NR1113 with wild type and mutant $LptB_2FG$ complexes

(a) Functional assays of the double proline mutants. All bacterial cultures were adjusted to an OD600nm of 0.5 with fresh LB medium, serially diluted 1:10 as indicated on the top of the figure and then replica plated in agar plates with (+) or without (-) 0.2% L-arabinose. (b) Detection of protein expression levels of LptF(Flag)G(Myc) of the positive control, negative control, and the double proline mutants by Western blotting. The bacterial cells for Western blotting were cultured in the presence of 0.2% L-arabinose to ensure that the dysfunctional mutant can grow.

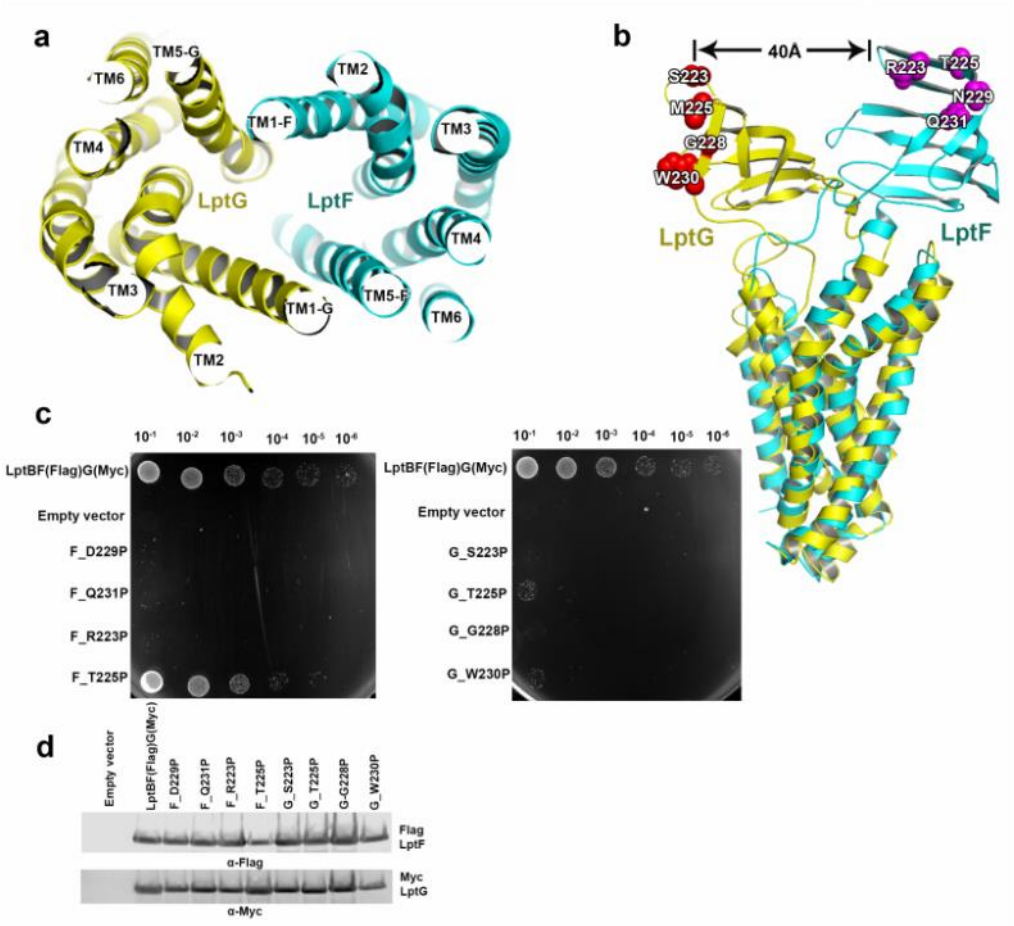


Figure 29 $LptB_2FG$ cavity and the periplasmic domains

The colour scheme is the same as in Figure 23. (A) The transmembrane domains of $LptB_2FG$, top view. Both LptF and LptG have six TM segments respectively, forming a cavity. (B) The superimposition of LptF and LptG from *K. pneumoniae*. The residues selected for proline substitution are shown in sphere. The mutations are designed based on the amino acids sequence alignment results (Figure 26) and the figure shows the structure of *K. pneumonia* and the mutations are based on the *E.coli* protein sequence. (C) Functional assays of the single proline mutants of periplasmic domains of LptF and LptG. The bacterial cells were diluted to an absorbance of 0.5 at OD600nm as the starting bacterial cells for the dilution assays. LptF mutants F_D229P, F_Q231P, F_T225P and F_R223P, and LptG mutants G_G228P, G_W230P,

G_T225P and G_S223P were used in the functional assays. Except F_T225P, all single proline mutants severely impaired the cell growth or killed the bacteria. (D) Detection of protein expression level of the samples in functional assays.

1.3.6 ATPase activity assay

To study whether *E. coli* $LptB_2F$(Flag)G(Myc) is functional *in vitro*, we purified the $LptB_2FG$ of *E. coli* and studied the ATPase activity of Wild type and mutants. The purification method is same as method 1.2.3. The LptB_E163A and LptB_K42A mutants are supposed to be dysfunctional mutants (166). As result $LptB_2F$(Flag)G(Myc) has ATPase activity and LptB mutants showed significant ATPase disability. LolCDE is another inner membrane ABC transporter used here as control.

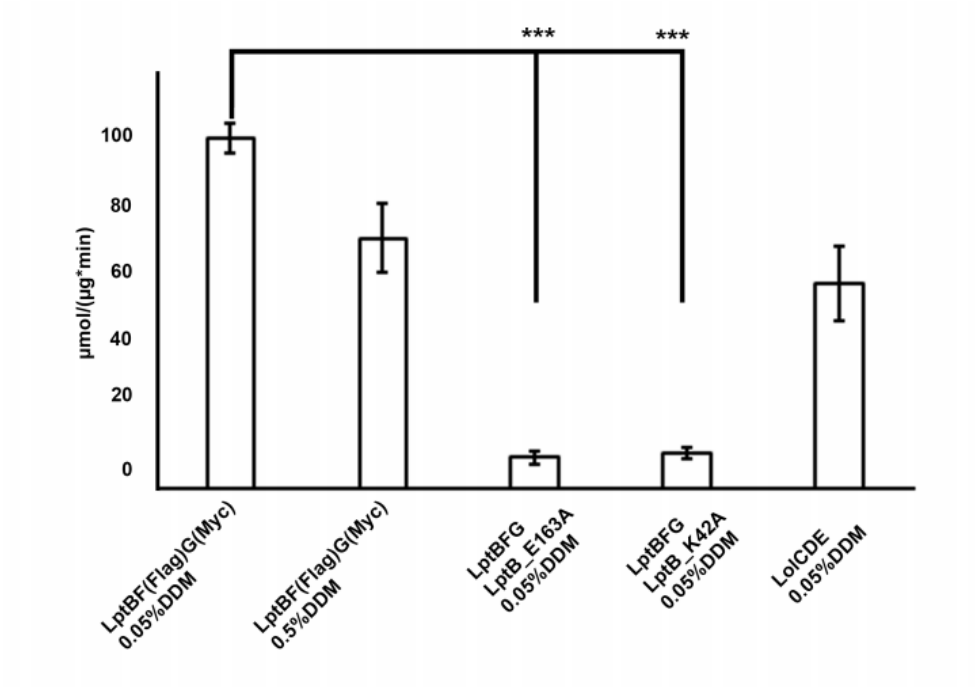


Figure 30 Purified $LptB_2FG$ has ATPase activity.

Purified *E. coli* LptBF(Flag)G(Myc) transporter has ATPase activity *in vitro*. The *E. coli* $LptB_2FG$ protein complex with LptB_E163A or LptB_K42A mutation

shows significant lower ATPase activity than that of LptBF(Flag)G(Myc). *E. coli* lipoprotein transporter LolCDE is included in the ATPase activity assay as a positive control. The 10-folds increasing of the concentration detergent DDM has a small amount affection on the ATPase activity of $LptB_2FG$ compared to $LptB_2FG$ in 0,05% DDM. The data was analysed using one-way ANOVA, n=3; P<0.01; graphs show mean±S.E.M.

1.4 CONCLUSION AND DISCUSSION

The unique composition of LPS of Gram-negative bacteria provides the cells extra resistance against various toxic compounds compared to Gram-positive bacteria that does not have the LPS and OM system. The biogenesis and transports of LPS has been studied as a hot spot throughout the years. This chapter revealed some unknown mechanism of the inner membrane ABC transporter $LptB_2FG$ by determine the crystal structure of $LptB_2FG$ from *K. pneumoniae*, which laterally extracts LPS from the periplasmic leaflet of the IM and delivers it to LptC. According to our structure, LptF and LptG each contain a coupling helix located within a highly conserved groove of LptB, and the coupling helices are playing important roles for the functionality of $LptB_2FG$ (167,168). The interactions between this residues will transfer the energy of ATP hydrolysis from LptB to LptF and LptG, leading to conformational changes that are required by LPS extraction and transport. According to our functional assays, the lethal phenotypes of the dysfunctional substitutions on the conserved positively charged and hydrophobic residues inside the cavity showed that they are important in the function of the complex. This cavity are suggested to capture the LPS after extraction. Further, the single and double proline mutants reveals the importance of both periplasmic domains of LptF and LptG, indicating that both periplasmic domains may be involved in the LPS transport pathway.

However, the most interestingly detailed mechanisms are still in the speculative. The ABC transporter motion models describe that an ABC transporter (like PglK (162) or

ButCD-F (169)) adopts three conformation states in cycle: taking in the substrate, ATP-hydrolysis and ADP-released/substrate-transported. If based on our current structure, the LPS could only enter the $LptB_2FG$ from the TM5F-TM1G side into the cavity then get delivered to the periplasmic domain of LptF. However, we speculate that LPS would be extracted passively into the hydrophobic centre cavity from both sides of the cavity. Then, ATP binding and hydrolysis by LptB brings the LptB dimer closer, resulting in a whole conformational change of the complex. The conformational changes is around the TM1F-5G and TM5F-1G of the cavity, pushing up the LPS molecule into the corresponding side of the periplasmic domain of LptF or LptG. Powered by the hydrolysis of ATP, the gate of TM1-5G or TM5F-1G opens and closes alternately, resulting both sides of the cavity extract and transport LPS alternately. The alternate conformational change is also supposed to affect the periplasm domains. In the structure both periplasmic domains lean to one side, opening the pathway of its side. We speculate that during conformational change, there should be a state when both periplasmic domains lean to the other side and open the other side as a pathway for LPS to pass through. In summary, we proposed an alternatively working model for LPS extraction and transportation pathway through a six-step cycle (Figure 31). The $LptB_2FG$ complex is first in a resting-state when the NBDs ($LptB_2$) exists in an open state as observed in our structure (nucleotide-free). When ATP binds into NBDs, LPS is extracted from the outer leaflet of the IM into the cavity via the lateral gate TM5F-1G, which induces the dimeric of LptB to move closer. Then, ATP hydrolysis of $LptB_2$ triggers conformational changes through the coupling helix, causes TM5F and TM1G to move closer. The lateral gate TM5F-1G then closes and pushes LPS through the cavity to the LptF's periplasmic domain. Another inner membrane protein LptC receives the LPS delivered by $LptB_2FG$ and the complex will return back to the resting-state. Next, the ATP binds to the NBDs again, LPS is extracted alternatively into the cavity through TM1F-5G and the ATP hydrolysis triggers the transmembrane helices TM1F and TM5G to move closer. Then the closed lateral gate TM1F-5G pushes LPS through the cavity to the LptG' periplasmic domain. Finally, the complex will return back again to the resting-state again and start a new cycle. The mechanism we propose

here represents one possibility of how $LptB_2FG$ uses ATP hydrolysis energy to transport LPS in an alternative double channel machinery (Figure 31).

Our proposed working model can partially be validated with the *P. aeruginosa* $LptB_2FG$ crystal structure (5X5Y) (168). We first tried to superimpose $LptB_2FG$ from *S. flexneri* and *K. pneumoniae* and found that they are in the same conformation (Figure 32) with RMSD of 0.660 over 1011 Cα atoms. Then we superimposed *S. flexneri* $LptB_2FG$ with the *P. aeruginosa* $LptB_2FG$ with RMSD of 2.678 over 884 Cα atoms. According to the comparison of $LptB_2FG$, the NBDs keep almost unchanged. The TMDs show some translational and rotational shifts of the α-helices by 3-10 Å, mainly around the TM1 and TM5 of LptF and LptG (Figure 32). The TM1G-5F open wider in our structure, while TM1F-5G open wider in *P. aeruginosa* $LptB_2FG$ (Figure 32). This is consistence with the difference of the position of the periplasmic domains, which is also the major difference. The periplasmic domains of our $LptB_2FG$ lean to TM1G-5F side and the periplasmic domain of LptF is at the position feasible for LPS transportation. In contrast, the periplasmic domains of *P. aeruginosa* $LptB_2FG$ lean to the other side, opening the LptG periplasmic domain and making it almost feasible for LPS transportation (Figure 32). This differences can be described as the rotational shift of the periplasmic domains by about 33.5 Å/22.5 Å for LptF/G at their widest points. It appears that our $LptB_2FG$ is in a state that LptF is responsible LPS transportation while *P. aeruginosa* $LptB_2FG$ is in the opposite state that LptG is responsible for LPS transportation. These conformational differences between our LpB_2FG and *P. aeruginosa* $LptB_2FG$ further supported the alternative LPS transportation working model, however the conformation of this working model requires further mutagenesis and crosslinking experiments.

The recognition of LPS is predicted to happen within the cavity. Our mutagenesis studies and the recently publish paper (168) suggest that the positively charged residues and the hydrophobic residues in the cavity may play important roles (Figure 27, Figure 28 and Figure 29). The hydrophobic interactions with the lipid A and the salt bridge between phosphate group of LPS and the positively charged residues both may be important for the recognition of LPS. Other lipids, which may be smaller than LPS,

should also enter the cavity. The complex should be able to distinguish between them. MsbA could flip both lipid A and lipid A core (46), however the components that make up the OM does not have lipid A but only LPS. So, the most possible explanation of the selectivity is from the core oligosaccharide of LPS, because the O-antigens share low similarity even within a single cell.

Another interesting issue is the interaction between $LptB_2FG$ and LptC, which I have tried to avoid mention throughout this chapter. It has been reported that $LptB_2FG$ could form complex with LptC at 1:1 ratio (1). Everyone who has heard about this project supports the thought that LptF and LptG would form a tunnel together and connecting to LptC in the similar manner of LptA oligomerization bridge, until we solved the structure. According to our structure, it seems that only LptF is responsible for LPS transportation, and LptG seems to be a scaffolding domain. Considering the high similarity of the folding manner of the periplasmic domain of LptF and LptG, we proposed the alternative LPS transportation working model (Figure 31). The only flaws of this mechanism is where we should put LptC at the 1:1 ratio with our $LptB_2FG$ complex. Since LptC and LptA form the LPS trans-periplasm bridge without forming a stable complex (28), one explanation of the $LptB_2FGC$ complex is that the affinity of LptC with $LptB_2FG$ is accidently about 50% and the native complex *in vivo* should be $LptB_2FGC_2$. The purification steps caused the complex to decompose partially. The other explanation is that LpB_2FGC is correct and the LptC interact with the periplasmic domain of LptF or LptG alternatively in the same pace as the alternation of the transportation path switch. The details of the relationship of LptC with $LptB_2FG$ worth further studies. By using Cryo-EM (Cryo-electron microscopy) techniques, we may identify the way LptC in complex with $LptB_2FG$.

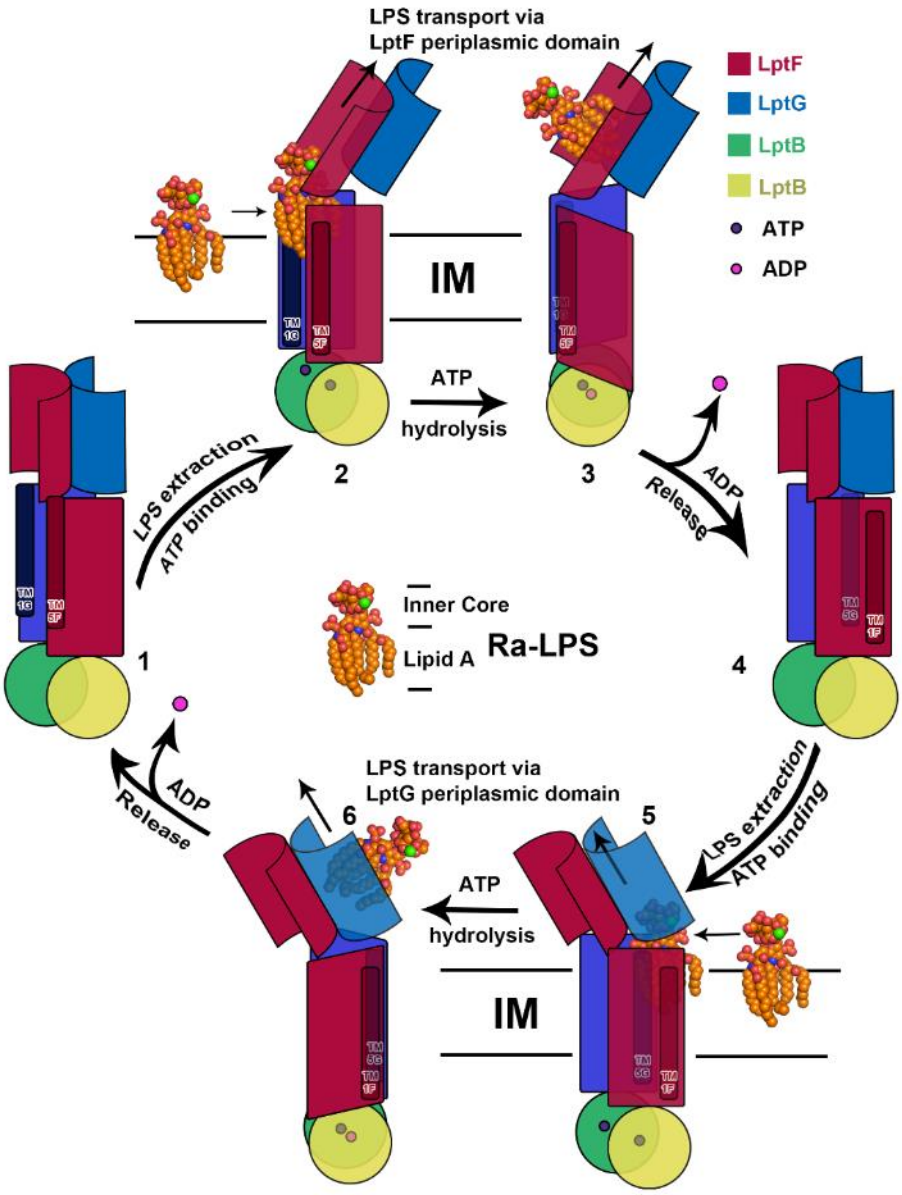


Figure 31 A proposed working model of $LptB_2FG$.

This model described the alternative mechanism of $LptB_2FG$ transport LPS from the lateral gates TM5F-1G and TM1F-5G to the internal cavity and then to the periplasmic domains of LptF and LptG. The LptB binds and hydrolyses ATP to provide the energy for LPS extraction and transport. The details of this working model are discussed in the discussions above. Ra-LPS is used as the substrate, smooth LPS is proposed to be transported in the similar manner with O-antigen passively dragged in the periplasm.

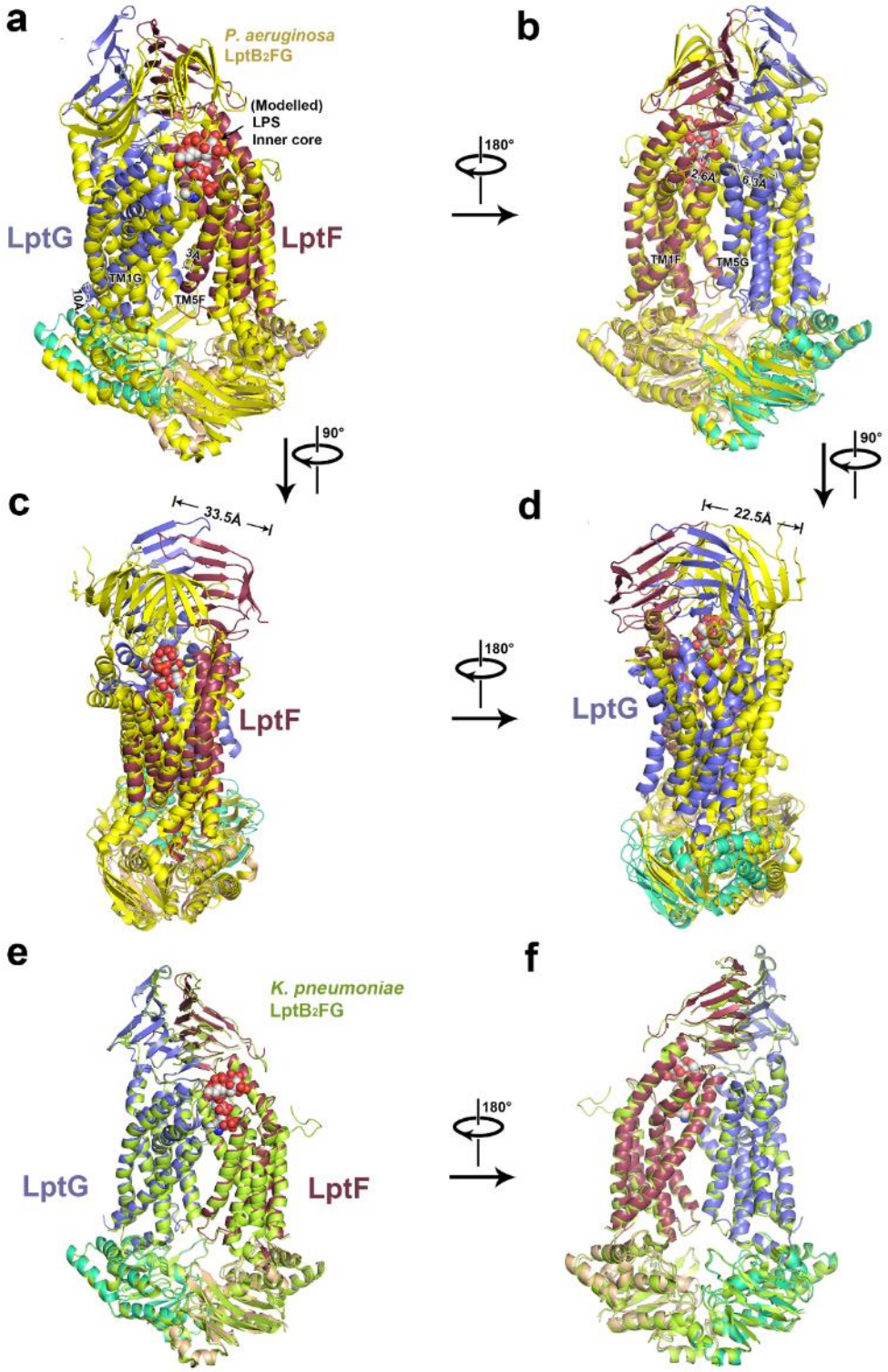


Figure 32 $LptB_2FG$ structure comparison between *K. pneumoniae*, *S. flexneri* and *P. aeruginosa* (5X5Y) (168).

$LptB_2FG$ structure of *S. flexneri* is similar to that of *P. aeruginosa* with a root mean squared deviation (RMSD) of 2.678 over 884 Cα atoms. A LPS inner core

motif is modelled into the internal cavity, which is sufficiently large to host it. (a) Front view of superimposition of $LptB_2FG$ structure of *S. flexneri* and *P. aeruginosa*. Cartoon representation of $LptB_2FG$. Two copy of LptB is shown in green and wheat. LptF is shown in dark red colour and LptG is shown in blue colour. The $LptB_2FG$ structure of *P. aeruginosa* is shown in yellow. (b) Rotation 180° along the y-axis relative to the left panel (c) Rotation 90° along the y-axis relative to the up panel. The periplasmic domain of LptF of *P. aeruginosa* is rotationally shifted of about 33.5 Å compared to the structure of *S. flexneri*, moving closer to TM5F-1G lateral gate. (d) Rotation 90° along the y-axis relative to the up panel. The periplasmic domain of LptG of *P. aeruginosa* is rotationally shifted of about 22.5 Å compared to that of the *S. flexneri*, moving far away from TM1F-5G lateral gate. (e) Front view Front view of superimposition of $LptB_2FG$ structure of *S. flexneri* and *K. pneumonia* with RMSD of 0.660 over 1011 Cα atoms. The $LptB_2FG$ of *K. pneumonia* is shown in lime colour. (f) Rotation 180° along the y-axis relative to the left panel. There is little conformational changes between the two $LptB_2FG$ structures from *S. flexneri* and *K. pneumonia.*

2 CHAPTER 2 Core oligosaccharide of lipopolysaccharide synthesis enzyme WaaB

2.1 INTRODUCTION

Lipopolysaccharide (LPS) is an essential component of the outer membrane outer leaflet for most of the Gram-negative bacteria viability. LPS contains three component, lipid A (the hydrophobic membrane anchor), core oligosaccharide (a nonrepeating oligosaccharide) and O-antigen (a polymer of repeating oligosaccharide unit). The core oligosaccharide can be further divided into two structural distinct regions: the inner core and outer core. Both cores play important roles for the survival and virulence of Gram-negative bacteria.

Among Gram-negative bacteria, *S. typhimurium* causes serious diseases in humans and a wide variety of farm animals (170-172). The effective impermeable outer membrane provide resistance against a large number of noxious agents (for example vancomycin), making *S. typhimurium* such a destructive pathogen. While the O-antigen portion of LPS has been recognized as an important virulence determinant in the pathogenesis of systemic salmonellosis (173-175), recent research suggests that the core oligosaccharide of LPS also contributes a lot to the maximum bacterial virulence (176-179).

The synthesis of the *S. typhimurium* LPS outer core region needs the sequential addition of glycosyls to a growing polysaccharide chain. The most studies example of LPS core oligosaccharide synthesis are *E. coli* and *S. typhimurium*. The inner core of the bacteria is highly conserved as shown in Figure 8, WaaA add the first two Kdo to lipid A, forming Re-LPS (Figure 8). Then WaaC adds one Hep to the product of WaaA (Figure 8). These three glycosyl residues are essential and conserved in most bacteria. In *E. coli* and *S. typhimurium* (5), WaaF are responsible for adding another Hep onto the product of WaaC (Figure 8), forming the Rd-LPS. At this time, the inner core of LPS has been synthesised. The outer core synthesis starts with the participation of WaaG, which add

a glucosyl residue onto the inner core of the LPS precursor. Then the product of WaaG is added by a glucosyl in *E. coli* by WaaO or a galactosyl in *S. typhimurium* by WaaI (180) (Figure 33). Then WaaR in *E. coli* or WaaJ in *S. typhimurium* transfer another glucosyl residues onto the previous products. Side chain glycosyl residues are transferred by WaaB onto the product of WaaG. In *E. coli* K-12, WaaU transfer another Hep and N-Acetylglucosamine to the WaaR product while in *S. typhimurium* WaaK transfer only N-Acetylglucosamine to the product of WaaJ at the similar outer core position as *E. coli* K-12. At this step the oligosaccharide of the outer core is fully synthesised. There are additional modifications on the core regions which, in some species, plays more important roles than the sugar groups. In *E. coli* and S. *typhimurium* the modifications, as described in 1.1.5, requires the strictly ordered enzymatic event by WaaP, WaaQ and WaaY. WaaP is a kinase that adds a phosphate group linked PEtN on to the product of WaaC (Figure 8). WaaQ transfers a Hep onto the product of WaaF (Figure 8). WaaY adds a phosphate group to the product of WaaQ (Figure 8). At this end, the synthesis of the outer core is completed and the combination of outer core, inner core and lipid A is calls a Ra-LPS (Figure 4). After being flipped from the IM cytoplasmic leaflet into the periplasmic leaflet, Ra-LPS would be ligated with the O-antigen by WaaL, forming a mature LPS (Figure 9, Figure 10 and Figure 11).

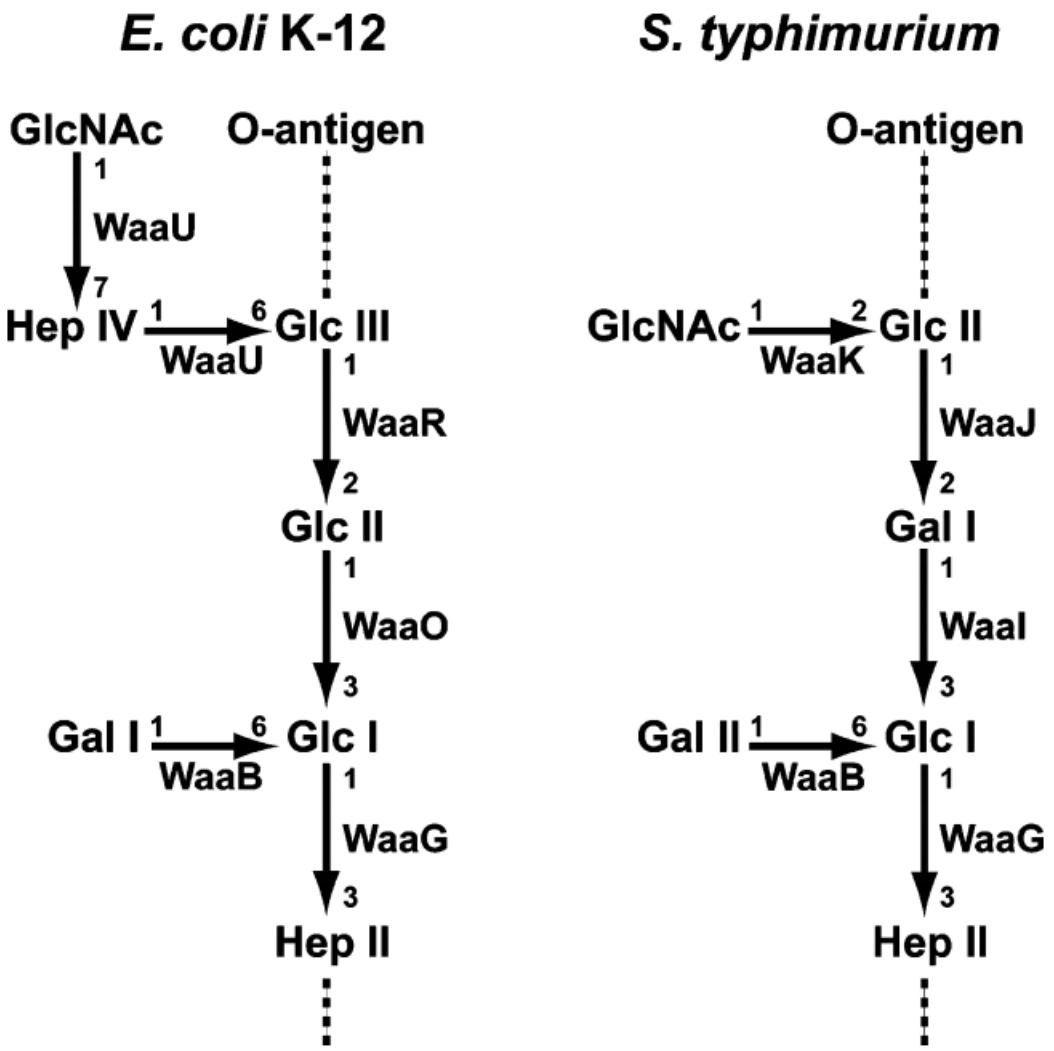


Figure 33 *E. coli* K-12 and *S. typhimurium* LPS outer core oligosaccharides (180). The enzymes proposed to catalyse each glycosylation are shown. The linkages among the sugars are indicated (5); all are α linkages except sugars added by WaaU from *E. coli* K12, which are β linkages. The dashed line attached to Hep II indicates where the rest of the inner core and lipid A is attached. The location at which the O-antigen is attached to the outer core oligosaccharide in dashed line.

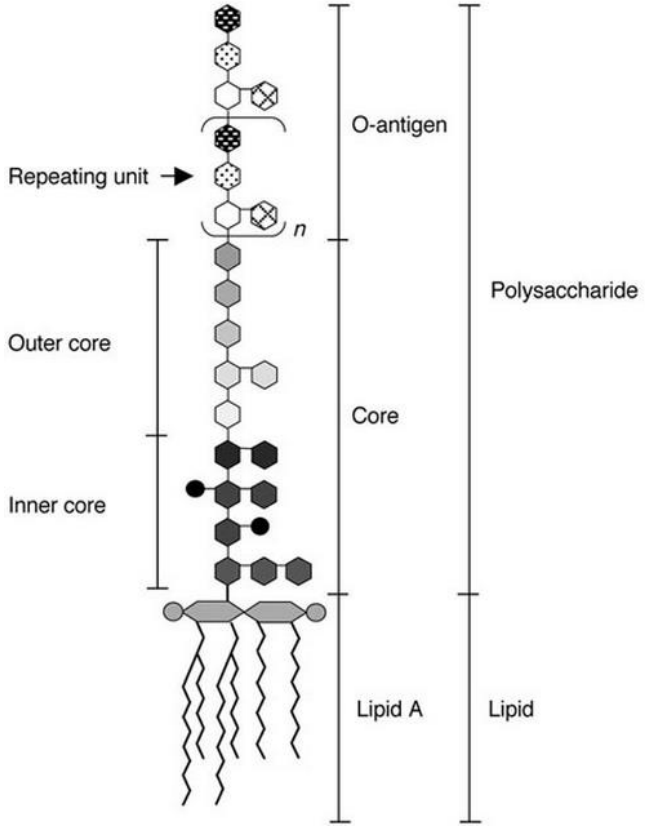


Figure 34 A LPS can be divided into three parts: lipidA, Core and O-antigen.

A LPS can be divided into three moieties: lipid A (the membrane anchor), core oligosaccharide and O-antigen. The core oligosaccharide can be further divided into outer core and inner core. The inner core is highly conserved throughout Gram-negative bacteria, while the diversity of outer core is higher. The O-antigen is highly variable among difference species and even within one single cell.

Except WaaY and WaaP, all other Waa proteins mentioned above belonged to glycosyl transferses. The genes responsible for the glycosyl transferases are located in the *rfa*GB1J gene cluster (36,181,182). In *S. typhimurium* the gene product of *rfaB*, also named *waaB*, is the α1, 6-linked D-galactosyl transferase WaaB. WaaB transfers D-galactosyl residue from UDP-galactose to the polysaccharide core. The *waaB* deletion mutants lacks the α1, 6-linked D-galactosyl residue of the core LPS. The deficiency of the LPS core region leads to attenuation in the virulence of *S. typhimurium*

(177,179,182). Further, *waaB* deletion strain is less tolerant to bile salts, which suggests that the glycosyl residue added by WaaB to the core lipopolysaccharide contributes to the detergent resistance of *S. typhimurium* (180,183,184).

WaaB is one of the glycosyl transferases (GTs) belonging to the GT4 family, that contains several enzymes of therapeutic significance. This family contains examples of enzymes that not only utilize the nucleotide-sugar donors, but also simple phospho-sugars and lipid-phospho-sugar donors, hinting at the ancient origin of this family (185). The mechanisms of the glycosyl transferase can be categorised into two kinds: inverting and retaining (186) (Figure 35). The inverting and retaining mechanism depends on whether the stereochemistry of the donor's anomeric bond is retained (α→α) or inverted (α→β), during the transfer. This allows the categorisation of glycosyl transferases as either retaining or inverting enzymes. The inverting mechanism is widely understood and acknowledged, while the retaining mechanism is still in debate with several kinds of proposed ways of achieving the retaining glycosyl transferase (Figure 35) (187). In the inverting mechanism, the conserved residues with carboxylic acid group usually serve as the catalyse site and facilitate the glycosyl transfer step. The hydrogen oxygen bond in the hydroxyl group of the acceptor is attacked by the negatively charged residue with carboxylic acid and the donor carbon and oxygen bond, activating the related reaction groups. After electrons moving within these bonds, the glycosyl residue is transferred from the donor to the acceptor in an inverting position (Figure 35). In the retaining mechanism, there are several kinds of possible mechanisms and we would describe the two most popular mechanisms. The reaction group activations are similar to the inverting mechanism. One is the double replacement, in this mechanism the carboxylic acid residue first accept the glycosyl residue from the donor, and then transfer it to the acceptor (Figure 35). The other is the S_NI like mechanism, in this mechanism the donor release the glycosyl for a short time and the acceptor receive the glycosyl residue (Figure 35).

Sharing the functional similarity with other LPS synthesis-related glycosyl transferases, WaaB possess a nucleotide-glycosyl (UDP-galactose for WaaB) binding site, an enzyme hydrolysis activity site and a potential sugar-acceptor substrate binding site

(intermediate LPS). Those glycosyl transferase matches lots of glycosyl transferase within LPS oligosaccharide biosynthesis such as D-glucose-transferase WaaG, the structure of which has been reported in 2006 (185). Interestingly, despite the similarity between LPS related glycosyl transferases, the highest sequence identity of WaaB, compared to all other functional similar glycosyl transferase, is only 17% suggesting that WaaB might have some special features. By evaluating the donor of WaaB and the product of WaaB, we could find that WaaB catalyse the galactosyl transferase in a retaining mechanism, which is in consistent with the GT4 family (Figure 36).

In this chapter, we present the first X-ray crystal structure of WaaB, providing insight into the addition of galactosyl on LPS biosynthesis. Our glycosyl-transferase activity assays has determined the UDP-galactose hydrolysis activity centre.

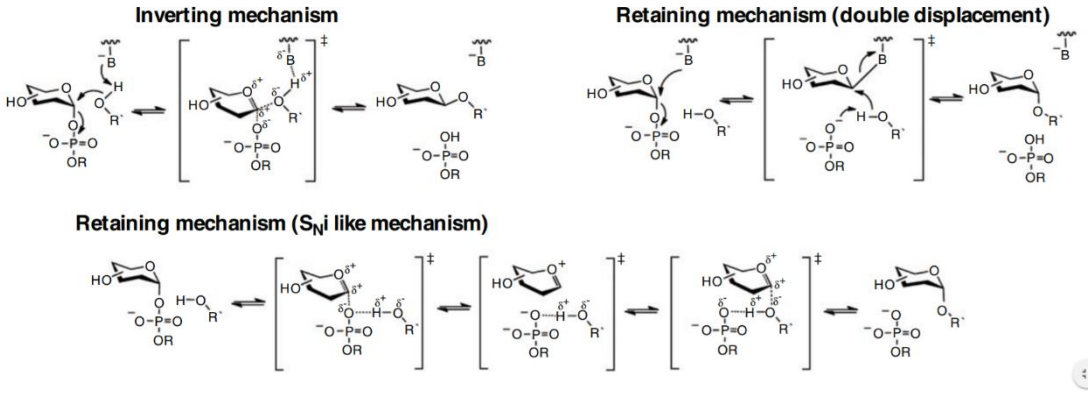


Figure 35 Two kinds of glycosyl transferase catalyze mechanism

Depending on whether the stereochemistry of the donor's anomeric bond is inverted or retained during the reacting, enzymes are classified into two kinds using different mechanisms. In the inverting mechanism, the negatively charged base group usually serve as the catalyse site and facilitate the glycosyl transferase step. The hydrogen oxygen bond in the hydroxyl group of the acceptor is attacked by the negatively charged residue and the donor carbon and oxygen bond, activating the reaction groups. Then, the glycosyl residue is transferred from the donor to the acceptor in an inverting position. In the retaining mechanism, there are several kinds of possible mechanisms. One is the double replacement, in this mechanism the base residue first accept the

glycosyl residue from then donor, and then transfer it to the acceptor. The other is the S_NI like mechanism, in this mechanism the donor release the glycosyl for a short time and the acceptor receive the glycosyl residue. The B in the image refers to the base, which is negatively charged and usually contain carboxylic acid group. The R refers to the rest part of the donor or acceptor except the shown reaction groups.

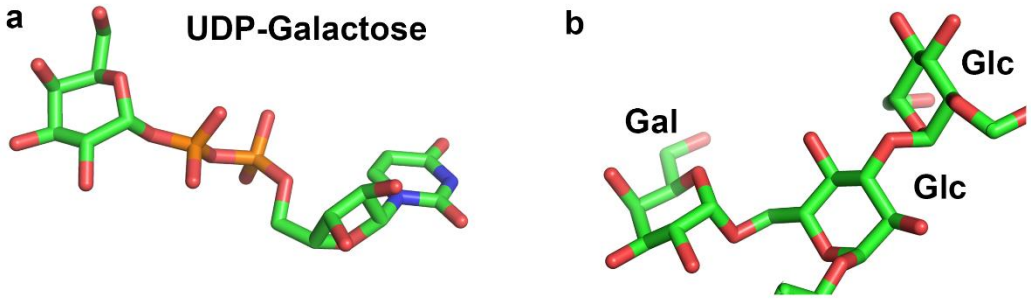


Figure 36 WaaB catalyze the galactosyl transferase via the retaining mechanism

Left panel is the UDP-galactose molecule. The right panel is part of the outer core oligosaccharide of *S. typhimurium*, galactose residue attached to the core as branch sugar chain as in Figure 33. Compare the stereochemistry of the galactose residue between the donor and acceptor, they are in the same orientation without bond inverting, suggesting that the WaaB enzyme catalyse this step in a retaining mechanism. Molecules are shown in ball-and-stick model [O atom (red), N atom (blue), P atom (orange)]

2.2 METHODS AND MATERIALS

2.2.1 *S. typhimurium* WaaB gene cloning

The *waaB* gene from *Salmonella typhimurium LT2*, was cloned into an expression plasmid pLOU3 (Amp^R) with double restriction enzyme digestion site NcoI and HindIII. The final construct was able to express the WaaB in a fusion protein form: 6×Histag-MBP-TEVcleavage site-WaaB. pLOU3 vector is a gift from Prof. Jim Naismith the map of this vector is shown in Table 12

2.2.2 *S. typhimurium* WaaB Expression strain

The pLOU3 plasmid harboring *waaB* was transformed into expression strain *E. coli BL21 (DE3)* (Novagene). A single colony was inoculated into 10 ml of LB for overnight culture. 20% glycerol (final concentration) was added into the overnight cultured cells to prepare a glycerol stock, which was kept in -80 ℃ freezer for further experiments.

2.2.3 Expression and purification of *S. typhimurium* WaaB

The glycerol stock (*BL21 DE3*) containing pLOU3-*waaB* plasmid) cells was inoculated into 500 ml LB overnight culture with 100 μg/ml ampicillin antibiotics. 12L LB medium were inoculated from 1:200 of overnight cell culture supplemented with 100 μg/ml ampicillin. The cells were cultured in a 37 ℃ shaker at 200rpm until the OD_{600} reached 0.8-1.0 and the protein expression was induced by adding 0.1mM IPTG final concentration of 0.1 mM into the culture for 3-4 h at 37 ℃ 200 rpm. Cell pellet was then harvested by centrifugation of 12 L culture at 5,000 rpm (4,300g) for 15 min. Followed by re-suspending in Balance buffer (20 mM Tris-Cl, pH 7.8, 300 mM NaCl and 10 mM imidazole 10% glycerol cOmplete EDTA-free Protease Inhibitor Cocktail Tablet (Roche) 50 μg/ml DNAse I (Sigma) and 100 μg/ml Lysozyme). After that the cells were lysed by passing through cell disruptor at 30,000 psi. The unbroken cells and

cell debris were removed by centrifugation at 20,000 rpm (40,000g) for 30 min. The supernatant was applied to pre-balanced (with Balance buffer) 5 ml HP HisTrap (GE Healthcare). The column was washed using Wash buffer which is prepared by adding 30mM imidazole to the Balance buffer to 30 mM final concentration, and then the fusion-protein was eluted by Elution buffer (20 mM Tris-Cl, pH 7.8, 300 mM NaCl, 300 mM imidazole and 10% glycerol). The fusion protein sample was desalted using a pre-equilibrated desalting column (GE Healthcare) with Balance buffer to remove the high imidazole concentration. TEV protease was added to the desalted protein at a TEV enzyme to protein mass ratio 1:100 and incubated overnight to cleave the MBP together with 6×His tag off from the fusion protein at the TEV cleavage site. After cleavage, the sample was loaded onto the pre-balanced 5 ml HP HisTrap again to remove the TEV protease and the cleaved MBP. During this step, the *S. typhimurium* WaaB would directly go through the column and be collected and concentrated to 5-10 ml. The protein was further purified by gel-filtration using pre-equilibrated SEC (HILOAD 16/600 SUPERDEX 200 PG, GE Healthcare) (with 150mM NaCl and 20mM Tris pH 7.8). The protein purity was checked on SDS-PAGE and the most pure fractions were pooled and concentrated to 5-10 mg/ml for crystallization.

2.2.4 *In situ* proteolysis of *S. typhimurium* WaaB

To increase the possibility of crystal hits during crystallization screening, we tried to use proteases to remove the disordered regions and obtain a stable soluble domain after purification. We screened seven proteases, Papain (1 mg/ml), Elastase (1 mg/ml), V8 (2 mg/ml), α-Chymotrypsin (2 mg/ml), Subtilisin (2 mg/ml), Thermolysin (1 mg/ml), and Trypsin (2 mg/ml). Firstly, 50 µg of the purified protein was prepared into separate tubes and in each tube individual protease at mass ratio 1:100 or 1:50 (protease: protein) was added and incubate at room temperature. Samples were taken every 30 min for SDS-PAGE analysis to select one or two of the best performing proteases, which would be chosen for *in situ* proteolysis during crystallization. The performance of the

proteases was analysed on SDS-PAGE and an ideal choice should exhibit almost single strong band on the SDS-PAGE, which should also display a small molecular weight shift compared to the non-digested control.

2.2.5 Crystal forming, optimization and chemical soaking of *S. typhimurium* WaaB

S. typhimurium WaaB crystallization was performed using the sitting drop diffusion method with drop size 0.3 µl (sample): 0.3 µl (mother liquid) using the crystallization robot Gryphon (Air). The screening solutions were purchased from Hampton research and Molecular dimensions (Table 10).

After crystal appeared, the optimizations were performed by designing and preparing 2D extension of pH and PEG concentration of crystallization condition (1% tryptone; 0.001M sodium azide; 0.05 M HEPES pH7.0; 20% w/v PEG 3350), developing several 5×5 home-made optimization solution in deep well blocks.

In order to determine the experimental phase of WaaB, which is crucial for structure determination and model building, we soaked crystals grown from optimization conditions with solutions containing different kinds of anomalous signal scattering elements (including Hg, I, Au and etc.). Crystals were soaked with substrates (UDP-galactose or UDP) for substrate binding structure determination. Before being flash frozen by liquid nitrogen, crystals soaking condition were screened at different concentrations, from 1 mM to 500 mM metal ion solution or sodium iodide solution, and at different time durations, from 10 s to 24 hour.

2.2.6 Data collection and structure determination of *S. typhimurium* WaaB

All crystals from different treatments were flash frozen in liquid nitrogen with certain cryoprotectant (0.05 M HEPES pH7.0; 20% w/v PEG 3350 20% glycerol) and kept in pucks stored in a dewar, which were both pre-cooled by liquid nitrogen. The crystals

were sent to Diamond Light Source for data collection at beamline I02.

The native crystal data and iodide-derivative data were collected at wavelength 1.82325 Å, Oscillation 0.1 ° and 360 °. The iodide-derived SAD data were collected at the peak wavelength of iodide elements according to the X-ray Anomalous Scattering Information (Ethan A Merritt ©1996-2011/ merritt@u.washington.edu / Biomolecular Structure Center at UW) and according to the fluorescent signal scan result at I02. The original data was processed by Diamond Light Source automatic pipeline (including XIA2 (144) with Dials (145)), to generate a MTZ file. The crystals soaked with 100 mM NaI for 10 min behaved best and the iodide-derived LapB structure were further solved by Crank (CCP4i) (188), which was a pipeline that gather lots of programs together. Crank is an automatic program that can solve the macromolecular structure of SAD, MAD and SIRAS data. Crank interfaces with several crystallographic programs. For substructure detections, Crank can interface with CRUNCH2(189) and SHELXD (190). For substructure phasing, BP3(191) can be used. For density modification, SOLOMON (192), DM(193), SHELXE(194), PARROT, PIRATE (195)and RESOLVE(196) can be used. For model building, RESOLVE (197), BUCANEER (154) and ARP/wARP (198) are available in CRANK. For refinement ARP/wARP (198) and REFMAC5 (199) can be used. To calculate FA values needed for substructure detection, SHELXC (200) or AFRO including SFTOOLS and TRUNCTATE can be interfaced by CRANK. Also, COOT (157) is available to visualize the map or the model if necessary.

In this case, the substructure of iodide-derived WaaB crystal was determined and refined by AFRO/CRUCH2 (148) and BP3 (191,201) respectively, hand determination and density modification was complete by SOLOMON (192), finished by model building by BUCCHANEER (154). After the Crank pipeline, REFMAC5 and were used to perform final refinement. Program COOT (157) was used to fix the rest errors. Iodide atom coordinates were removed from coordinates file of the structure to produce the primary model for molecular replacement of following structure determination.

The native data or substrate soaked crystal structure were determined using Molecular Replacement PHASER MR (202) (CCP4i) using the primary model and refined by

REFMAC5 (158). Bond angle adjusting for Ramachandran plots and final error corrections were achieved by COOT (157).

PHASER is a program for phasing macromolecular crystal structures with maximum likelihood methods. Molecular replacement programs always use a 3D atomic model obtained from a known structure to solve the crystal structure of an unknown structure. By computing the agreement of the measured diffraction data with the computed model, the program tells whether a structure is solved. Likelihood measures the agreement of the model with the data by using probabilities. For molecular replacement, the 'model' tested includes not only the structure itself, but also the orientation of the template in the unit cell, as well as the sizes of difference sources of errors. In refinement, the error can be calculated by comparing the F_{ob} and F_{calc}. In molecular replacement, the errors are set aside until the replacement problem is resolved. In molecular replacement, the components of the structure are treated as rigid bodies that are rotated and translated to place them correctly in the unit cell of target unknown structure. These components can be proteins, complexes and flexible domains. In PHASER, each type of component that is treated as a rigid body referred to an ensemble. For each ensemble, PHASER computes a set of statistically-weighted average structure factors, which is used in the molecular replacement calculations. In most cases, we use automatically PHASER to do molecular replacement. PHASER runs several steps in sequence: anisotropic correction, cell content analysis, rotation search, translation search, packing check and rigid-body refinement. The anisotropic corrections will remove overall anisotropy from the diffraction data. Rotation search is the likelihood-based fast rotation function followed by rescoring of the op peaks using rotation likelihood function. Translation search is a likelihood-based fast translation function followed by rescoring of the top peaks using the translation likelihood function. Packing search is testing for overlaps and selecting symmetry-related copies of molecules to create a tightly packed assembly. The automated search can search for multiple copies of molecules or multiple kinds of molecules and can test multiple space groups as well. After PHASER, the top solution would be written into the output coordinates file by default. The MTZ reflection file after automate PHASER search contains the data from the input reflection file and new

columns (FC/PHIC, FWT/PHWT, DELFWT/PHDELWT, FOM and HLA/HLB/HLC/LD).

2.3 RESULTS

2.3.1 *S. typhimurium* WaaB purification

S. typhimurium WaaB was first purified at a 6×Histag-MBP-TEVcleavagesite-WaaB fusion form, which was confirmed by gel-filtration (Figure 37) and SDS-PAGE (Figure 38). In the Figure 37 the peak B was the fusion protein while the peak C mainly consisted of MBP. In Figure 38 the band that was located around 85 kD was the fusion protein while the band that appeared at molecular weight of 44 kDa was the MBP which is refered to as the peak B in Figure 37. After overnight TEV cleavage, the mixture was passed through a Ni-NTA column. The non-tagged WaaB protein went through the Ni-NTA column, and was then concentrated. The WaaB was further purified by gel-filtration. The gel-filtration of WaaB showed almost a single peak (Figure 39), which was verified on SDS-PAGE as well (Figure 40).

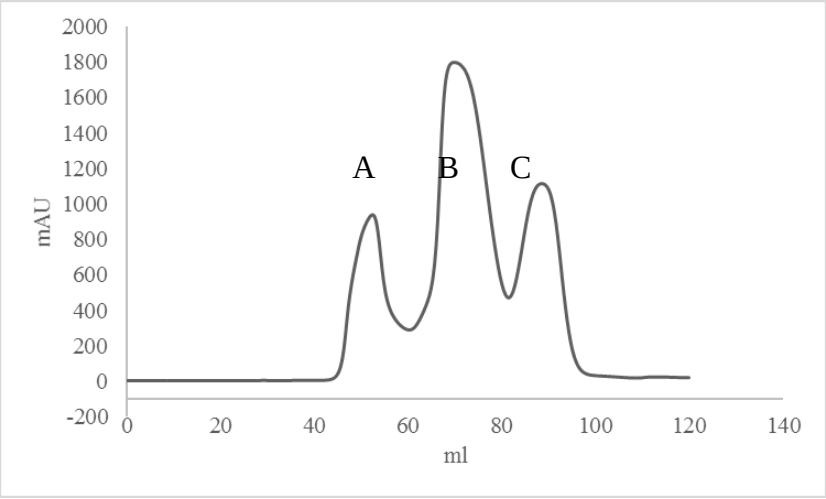


Figure 37 Gel-filtration of *S. typhimurium* WaaB-MBP fusion protein**.**

7.5ml fusion WaaB-MBP protein eluted from Ni-NTA column was loaded into gel-filtration during first purification to verify the fusion protein expression using

buffer (20mM Tris-Cl pH 8.0 300mM NaCl and 10% glycerol) at 1ml/min flow rate and room temperature. The SEC column used throughout the thesis was (HILOAD 16/600 SUPERDEX 200 PG, GE Healthcare). The peak A was an aggregation peak, the peak B was WaaB-MBP fusion protein peak and the peak C was MBP peak, which were verified in SDS-PAGE in Figure 38. This result was the first time we purify WaaB. In the following similar repeating purifications, the checking of the expression step and removing of imidazole for the following TEV protease digestion were achieved via desalting, which is a simpler method.

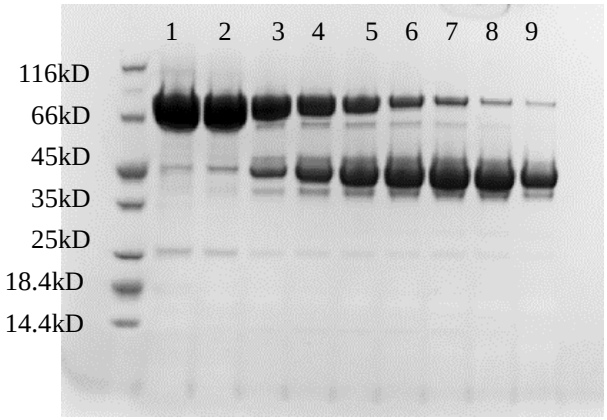


Figure 38 SDS-PAGE analysis of *S. typhimurium* WaaB-MBP fusion protein**.**

Lanes 1-4 are sample fractions of the peak B of Figure 37 and lane 5-9 were from fractions of the peak C. The peak B contained the WaaB-MBP fusion protein, while the peak C was mainly made up of MBP. The peak A aggregation peak was not analysed here. 5µl samples were mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific) after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature

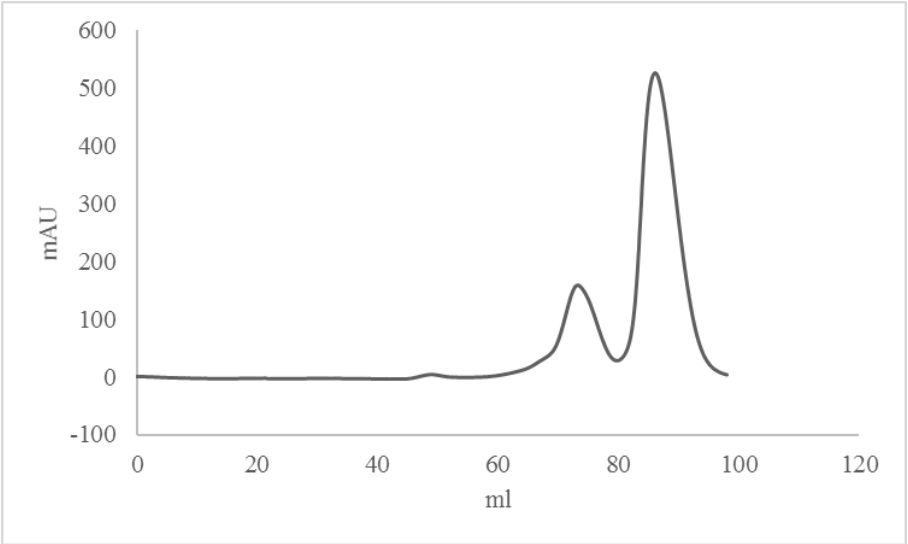


Figure 39 Gel-filtration of *S. typhimurium* WaaB after TEV protease digestion

After TEV protease digestion on the fusion protein, the mixture were purified through a reverse Ni-NTA column to remove the TEV protease. The flow through that contained non-tagged native WaaB was concentrated to 7.5-10ml and loaded on to SEC column (HiLoad 16/600 Superdex 200 pg) using buffer 20mM Tris-Cl pH8.0 and 150mM NaCl. The main peak at 85ml was the WaaB. There is a small peak at 75ml, which may correspond to the WaaB-MBP that survived during the TEV cleavage and reverse Ni-NTA chromatography purification.

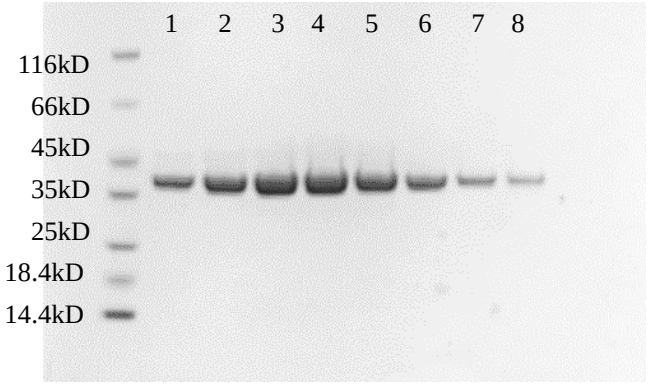


Figure 40 SDS-PAGE analysis of *S. typhimurium* WaaB.

SDS-PAGE samples were taken from the final gel-filtration fractions from Figure 39 and boiled with 4×reducing SDS-Loading buffer. After finishing all steps of purifications, the purity of WaaB was absolutely enough for cystallization. The lanes 1-8 were sample from D3, D2, D1, E1, E2, E3, E4 and E5 (volume range 82ml-95ml). 5µl samples were mixed with 10µl ddH2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into the SDS-PAGE gel after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature. Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific)

2.3.2 Limited proteolysis of *S. typhimurium* WaaB

Besides the normal native purification and crystallization, I also tried limited proteolysis to remove possible disordered regions of the protein that might interfere with crystal formation of WaaB. According to the results of the 45 min limited proteolysis (Figure 41), none of the proteases were treatment showed a desired results. They neither destroyed WaaB protein into peptide pieces nor failed to remove any residues off WaaB. However, we still believed in trying the non-cleavage protease, because these protease mightremovetwo or three flexible residues, which facilitated crystal forming. And no visible molecular weight changes could be observed on SDS-PAGE. (203).

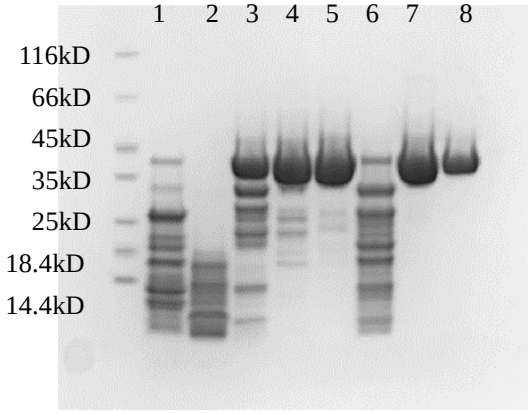


Figure 41 Limited proteolysis screening of *S. typhimurium* WaaB.

Proteolysis screen of *S. typhimurium* WaaB. Lane 1: α-Chymotrypsin; lane 2: Subtilisin; lane 3: Trypsin; lane 4: Thermolysin; lane 5: Elastase; lane 6: Papain; lane 7: V8; lane 8: control. All the proteases were used at 1:100 mass ratio of protease to *S. typhimurium* WaaB. The reactions were performed at room temperature. SDS-loading buffer was added into every sample after 45 min to terminate all the reactions. 5μl samples were mixed with 10μl ddH_2O and 5μl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into the SDS-PAGE gel after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature. Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific)

2.3.3 Crystallization of *S. typhimurium* WaaB

Native *S. typhimurium* WaaB crystallization trials were performed using the sitting drop diffusion method and Gryphon crystallization robot. The none of the proteases from the proteolysis performs satisfactorily enough. The native WaaB crystals were successfully grown in a few conditions at room temperature (Figure 42). Some crystals diffracted well, with high resolution up to 1.89 Å. Extensively to optimize the crystals by varying crystallization conditions had been tried. The crystals were reproducible and had been optimized with hand-made 5×5 optimization solutions to produce enough good quality

crystals for further soaking experiments (Table 6).

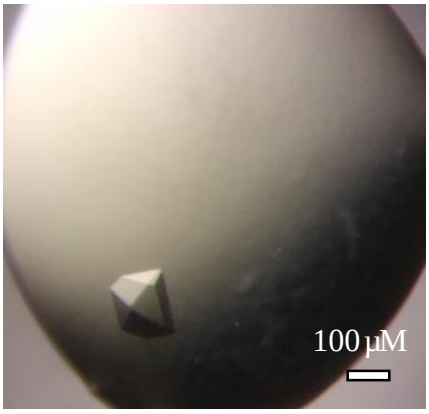


Figure 42 The crystal image of *S. typhimurium* WaaB

Crystal was grown in 1% tryptone; 0.001M sodium azide; 0.05 M HEPES pH7.0; 20% w/v PEG 3350 at room temperature. The crystals appeared after 1 month. Native WaaB protein concentration was 9.7 mg/ml; drop size 0.3ul: 0.3ul vapour diffusion sitting drop.

PEG 3350 w/v / pH (HEPES) 0.05 M pH	16%	18%	20%	22%	24%
7.4					
7.2					
7.0					
6.8					
6.6					

Table 6 optimization condition array of *S. typhimurium* WaaB

The concentration of all other components stayed the same as the original one. 1% tryptone; 0.001 M sodium azide. ddH_2O was used to fill the rest volume to 1ml of each conditions of the optimization. And the 5×5 solutions were prepared into the deep well block. All other home-made optimization conditions are similar to this one, we vary the pH and the PEG concentration of the crystal growing in most cases. And the optimization sizes vary from 5×5 to 8×12 depending on the optimization depth.

2.3.4 Crystal structure of *S. typhimurium* WaaB and UDP-WaaB

The substructure and an initial model of WaaB was determined from iodide anomalous scattering SAD data as described in method 2.2.4 and Table 7. The iodide anomalous difference Fourier map contoured at 10σ is shown in Figure 43, ten iodide atoms are located in the structure. Buccaneer (154) was used to build the initial model. The final models were generated by molecular replacement using this initial model as template for both apo and UDP bound WaaB high resolution datasets. Refinement and final model improvement were performed by REFMAC5 (158) and PHENIX (155).

The overall crystal structure of WaaB, a galactosyl transferase, adopted a two domain "Rossmann-like" (β/α/β) domains, the N-terminal domain (the acceptor substrate binding domain) and the C-terminal domain (the donor substrate binding domain), which is a typical GT-B fold. (Figure 44). The N-terminal domain comprises residues Met1-Val161 and displays seven parallel β-strands sandwiched by six α-helices with two helices at one side and four helices at the other side. The C-terminal domain contains residues Ala181-Gln358, which consists of six parallel β-strands forming a β-sheet, six α-helices with three of them at one side and the remaining three at the other side of the β-sheet, and an additional two C-terminal α-helices. The first C-terminal α-helix associates with the C-terminal domain, while the last C-terminal helix (residues Asp342- Gln358) joins to the N-terminal domain (Figure 44). Therefore, the N-terminal domain and the C-terminal domain are connected by the C-terminal helices and a long loop consisting of residues Tyr 162 to Pro180 (Figure 44).

The structure is consistent with its glycosyl transferase function, which captured the potential substrates between these two domains while hydrolyzing UDP-galactose into galactosyl residue and then transferred it onto the sugar core part of LPS. Looking into the hydrophobic cavity between the C-terminal domain and N-terminal domain, several conserved residues form a potential active site. The UDP-bound structure of WaaB also confirmed that the cavity should also be responsible for UDP-galactose bound (Figure 47). Unfortunately, neither of the UDP-galactose soaked crystals nor co-crystallization provided us any data that contained galactose or galactosyl molecules inside WaaB. In

fact, we did collected crystal diffraction data from UDP-galactose soaking treatment, however all those data turned out to be UDP binding conformation rather than the UDP-galactose bound conformation. This finding suggests that the UDP-galactose hydrolysis by WaaB, occurs even inside the crystal.

	S. typhimurium apo WaaB	*S. typhimurium* UDP-WaaB	*S. typhimurium* WaaB NaI (SAD)
Data collection			
Space group	$P4_32_12$	$P4_32_12$	$P4_32_12$
Cell dimensions			
a, b, c (Å)	104.27, 104.27, 88.44	104.44, 104.44, 89.64	104.09, 104.09, 89.00
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	1.823	0.979	1.823
Resolution (Å)	67.64-1.81 (1.85-1.81)	73.85-1.92 (1.97-1.92)	67.64-1.90 (1.95-1.90)
R_{merge} (%)	12.5 (86.3)	16.0(133)	14.5 (192.2)
$CC_{1/2}$ (%)	100 (60)	100 (50)	100 (70)
$I/\sigma(I)$	29.0 (3.1)	16.1 (1.9)	18.5 (1.4)
Completeness (%)	98.5 (86.2)	99.9 (99.5)	99.8 (97.5)
Redundancy	60.1(18.2)	23.4 (11.8)	34.7 (18.6)
Anomalous Completeness			99.8 (97.6)
Anomalous Redundancy			18.1 (9.4)
Phasing			
Resolution			67.64-1.90
Site (I)			10
Figure of merit			0.6828
Refinement			
Resolution (Å)	67.45-1.81 (1.85-1.81)	73.85-1.92 (1.97-1.92)	
No. reflections	44277 (3879)	38400 (3752)	39073 (2766)
Rfactor / Rfree	0.2294/0.2591	0.2392/0.2713	
No. atoms			
Protein	2837	2849	
Ligand/ion	0	25	
Water	92	81	
B-factors	29.30	31.00	
Protein	29.30	30.90	
Ligand		49.20	
Solvent	30.00	30.00	
R.m.s.deviations			
Bond lengths	0.008	0.009	

(Å)		
Bond angles (°)	1.34	1.28
Ramachandran statistics		
Allowed (%)	94	96
Outliers (%)	1.7	0.56
PDB code	5N7Z	5N80

Table 7 Data collection and structure refinement statistics of *S. typhimurium* WaaB. R_{factor} = Σ|| Fobs|−| Fcal||/Σ|Fobs|, where Fobs and Fcal are observed all reflection measured and calculated currently model as structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly selected not used in refinement.

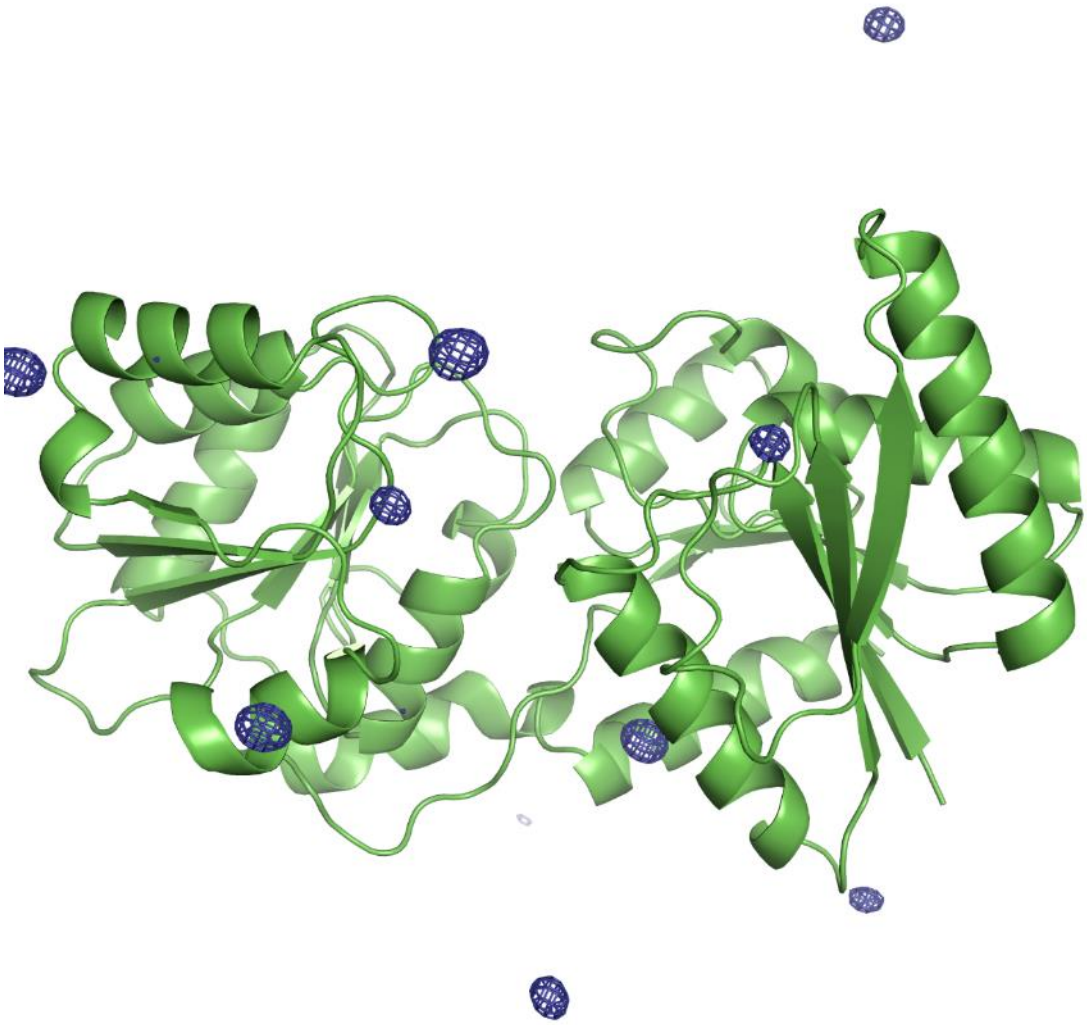


Figure 43 Anomalous difference Fourier map of iodide derived WaaB

The anomalous difference Fourier map of iodide-derived WaaB data was calculated using PHENIX (155) and contoured at 10.0σ. The peaks in the images represented the iodide atom that were soaked into the crystal of WaaB. WaaB structure was shown in cartoon and in green colour.

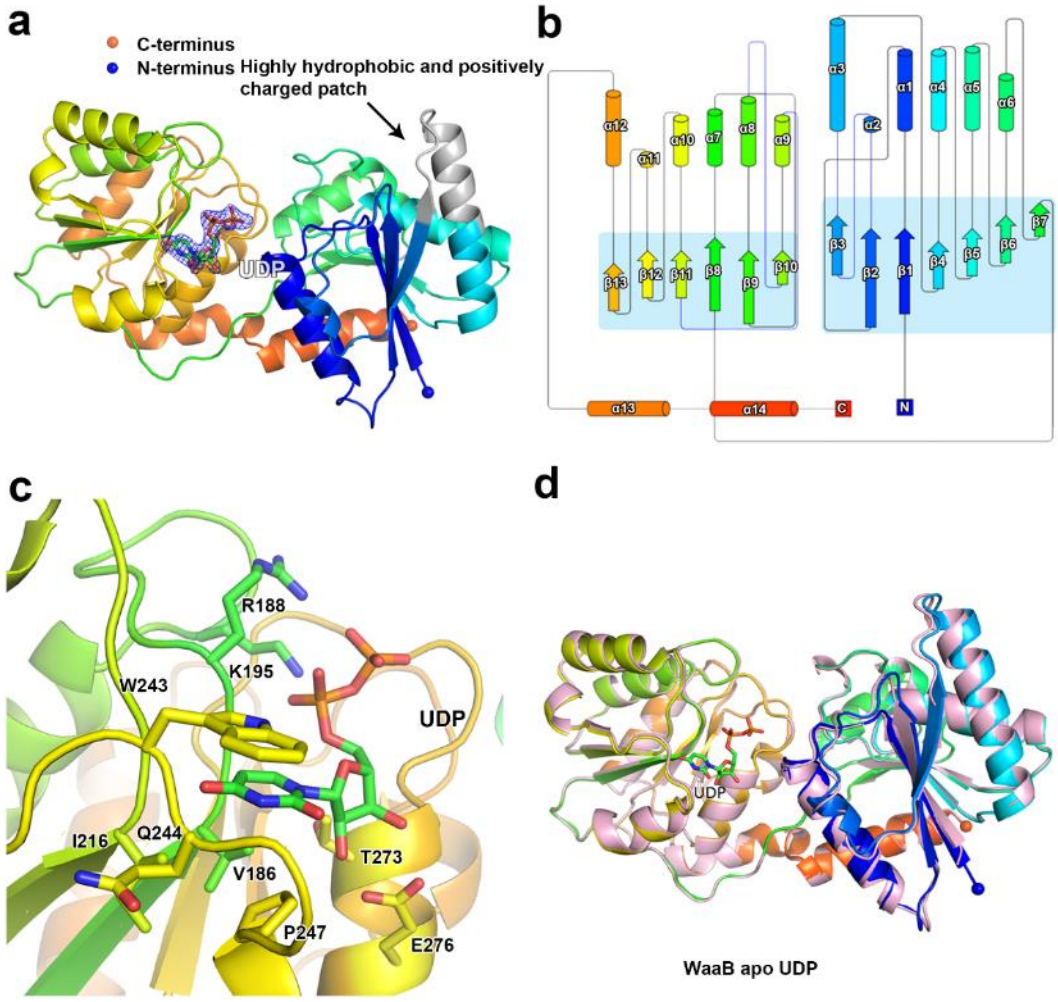


Figure 44 Structure of WaaB

(a) Cartoon representation of WaaB. WaaB can be divided into N-terminal domain and C-terminal domain. UDP is bound at the C-terminal domain (donor substrate binding domain). The highly hydrophobic and positively charged patch in grey may be the membrane anchor of WaaB. The Fo-Fc map contoured at 3.0σ matches the UDP molecule perfectly. LPS related biosynthesis glycosyl transferases are predicted to anchor into the IM. The highly hydrophobic and positively charged patch is the region we predict to help

WaaB get anchored into the IM (b) Topology of WaaB. (c) WaaB residues involved in UDP binding. UDP and the shown side chains are in ball-and-stick model [O atom (red), N atom (blue), P atom (orange)]. (d) Superimposition of structures of the apo WaaB and WaaB in complex with UDP. The overall structure of WaaB in complex with UDP is almost identical to the apo structure of WaaB with a RMSD of 0.5624 Å over 357 C atoms

2.3.5 Donor substrate and potential receptor substrate binding site

The UDP molecule has been successfully located in the UDP-WaaB. The UDP molecule binds to the C-terminal domain and is located in the cleft formed by the C-terminal and the N-terminal domains. The uracil of UDP is located in a hydrophobic pocket consisting of residues W243, I216, P247 and V186, where the uracil is anchored by two hydrogen bonds formed between the uracil and the amide and carbonyl groups of Q244. The ribose of UDP is stabilized by two hydrogen bonds with E276 and by hydrophobic interactions with T273, while the two phosphate groups form salt bonds with the side chains of K195 and R188 (Figure 45 and Figure 47).

The overall structure of WaaB in complex with UDP is almost identical to the apo structure of WaaB with a RMSD of 0.5624 Å over 357 Cα atoms, suggesting that binding the donor product UDP does not induce significant conformational changes of the apo WaaB structure (Figure 45). The N-terminal domain structure showed a minor shift while the C-terminal domain structure remained mostly unchanged to that of the apo state. The helix α5 was slightly stretched into a loop and α2 turned from a loop into a helix after UDP entered the WaaB structure (Figure 45). Amino acid sequence analysis showed that residues W243, P247, V186, Q244, E276, I216, T273, K195 and R188 are highly conserved in WaaB (Figure 48), suggesting that these residues play important roles in UDP-galactose binding and hydrolysis.

We tried to compare structures of WaaB in complex with UDP with WaaG in complex with UDP-2-deoxy-2-fluoro glucose. WaaG is a UDP-D-glucosyltransferase

participating in LPS core oligosaccharide biosynthesis directly upstream of WaaB. Although WaaB shares low identity in amino acids sequence with WaaG, the WaaB structure shares a high degree similarity to WaaG (Figure 45), with a RMSD of 3.2589 Å over 358 aligned Ca atoms. By comparing the structures of WaaG in complex with UDP-2-fluoro-Glc and WaaB in complex with UDP, we may identify the possible UDP-glycosyl hydrolysis centre, as well as the different residues needed for the different substrate bindings. Superimposition of WaaG and WaaB by overlapping their own ligands (Lignalign Pymol) (204), allowed us to identify conserved spatial arrangements local to the UDP-sugars. Accordingly, WaaB and WaaG structures share a high degree similarity in the C-terminal domain, which is mainly responsible for UDP-glycosyl binding (Figure 45). The diphosphate forms salt bridges with K209 and R208. However, F13 of WaaG is located in its N-terminal domain and UDP-glucose binding requires both the N-terminal and the C-terminal residues of WaaG. This is different from that of the WaaB where superimposition of the two structures showed that the WaaG's N-terminal domain is rotated toward to a closing conformation of WaaG (Fig. 4A). WaaB also forms a pocket as WaaG, but formed by E268, G269, F120, F270, G15, and Q194, which may be responsible for galactose binding (Figure 45). These differences between WaaB and WaaG for the donor substrate binding sites may be important for the selectivity of the donor substrates. To test whether this pocket is important, the potential galactose binding site residues E268 and Q194 were selected for mutagenesis, where E268 of WaaB is conserved and Q194 of WaaB is highly conserved. Galactosyltransferase assays showed that the mutant E268A activity is significant reduced (Figure 49), while the mutant Q194A has the same activity as the wild type WaaB, indicating that the E268 is critical for the WaaB's activity.

The di-phosphates of the UDP binding at the donor substrate binding domain point to a groove of the N-terminal domain, the acceptor substrate binding domain. This groove is in close proximity to the UDP. We speculate that this groove is the putative acceptor substrate binding site for the premature lipopolysaccharide (LPS), Glc-Hep_2-1-dephospho-Kdo_2-lipid A (Figure 46). In particular, E17 and G15 of WaaB are absolutely conserved in WaaB homologues and occupy the same positions as E31 and G30 of

WaaA (Figure 46). The E31 residue is proposed to act as a general base for WaaA, suggesting that E17 of WaaB may be important for the WaaB functionality. This groove is highly charged, which is ideal for binding the acceptor substrate Glc-Hep_2-1-dephospho-Kdo_2-lipid A (Figure 46). The potential acceptor substrate binding groove of WaaA is formed by residues E31, S54, L74, P75, D77, E98, E100, W102, R56, R99 and K121 (Figure 46). Although the two grooves share similarity, they are different in term of electrostatic potential (Figure 46).

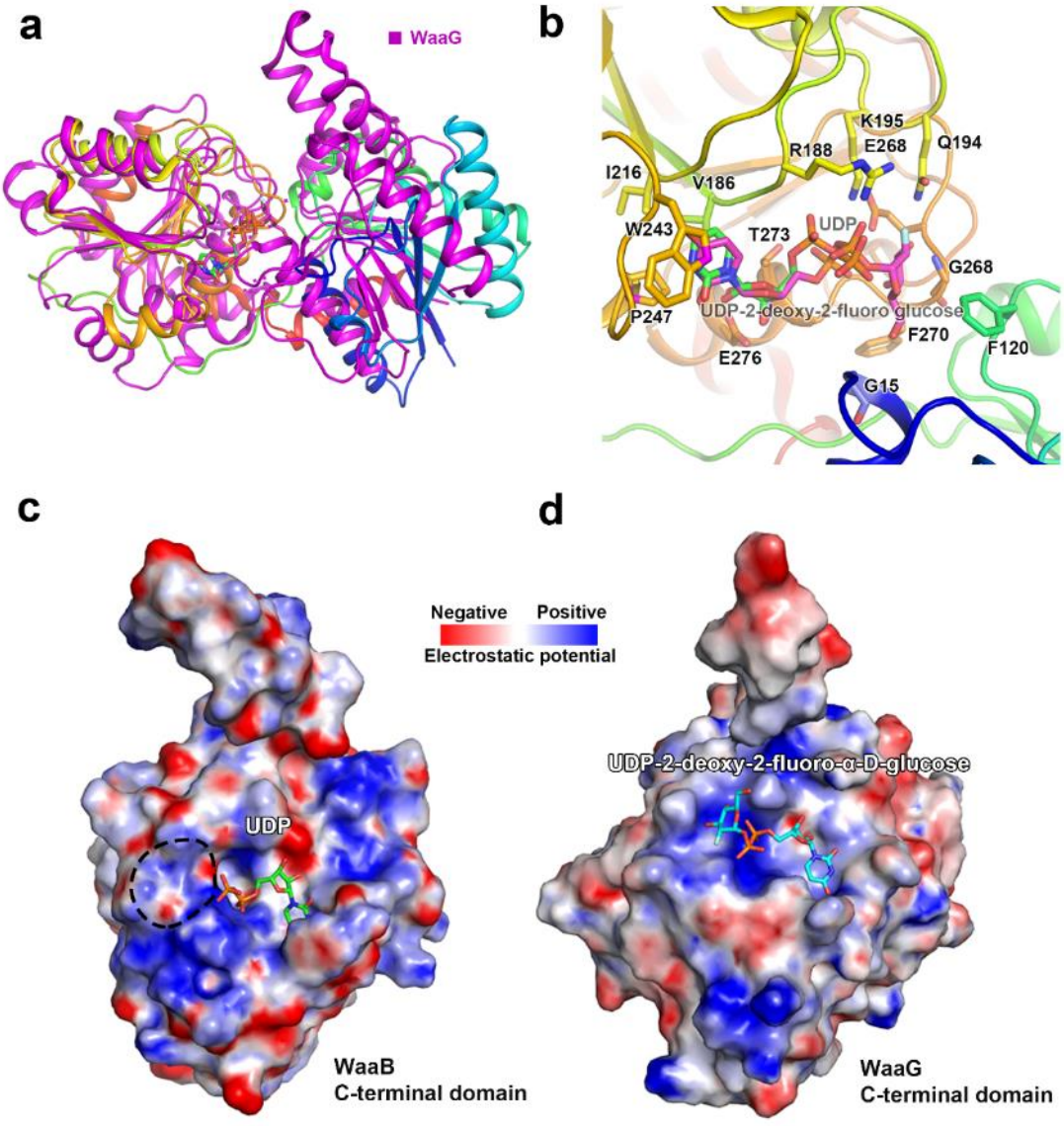


Figure 45 The Donor substrate binding groove of WaaB

(a) WaaB and WaaG (2IV7) structures are superimposed over their ligands with a RMSD of 3.2589 Å over 358 aligned Ca atoms. (b) The potential galactose

binding site of WaaB is identified by the structure superimposition. (c) The WaaB electrostatic potential map of the potential donor substrate UDP-galactose binding site. Substrates and the shown side chains are in ball-and-stick model [O atom (red), N atom (blue), P atom (orange)] (d) The WaaG electrostatic potential map of the donor substrate binding site. Red represents negatively charged which blue represents positively charged.

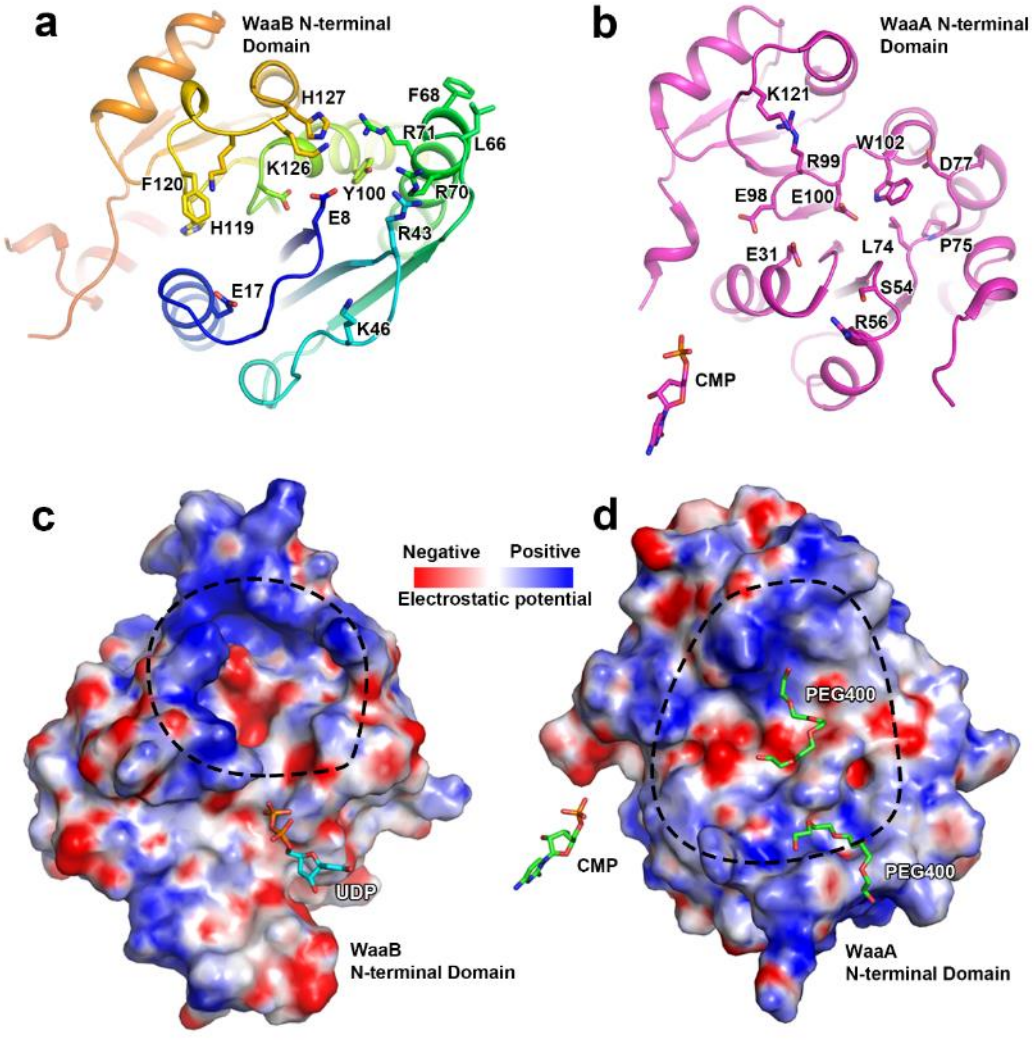


Figure 46 The potential acceptor substrate binding groove of WaaB.

(a) The N-terminal domain of WaaB is the acceptor binding domain. WaaB potential acceptor substrate binding groove is highly positively charged. The shown side chains are in ball-and-stick model [O atom (red), N atom (blue)] (b) WaaA (2XCU) acceptor substrate binding groove. (c) The electrostatic potential

map of the potential WaaB acceptor substrate binding grove. (d) The electrostatic potential map of WaaA acceptor substrate binding groove. Red represents negatively charged which blue represents positively charged.

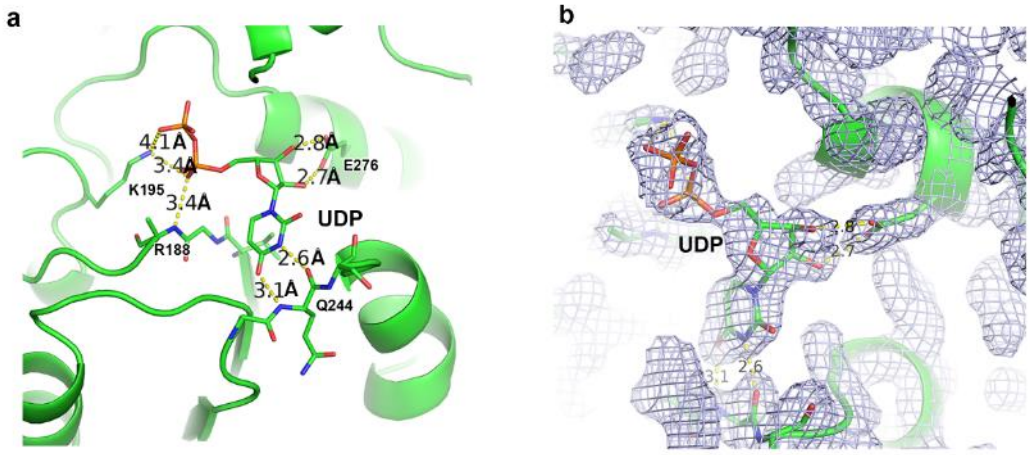


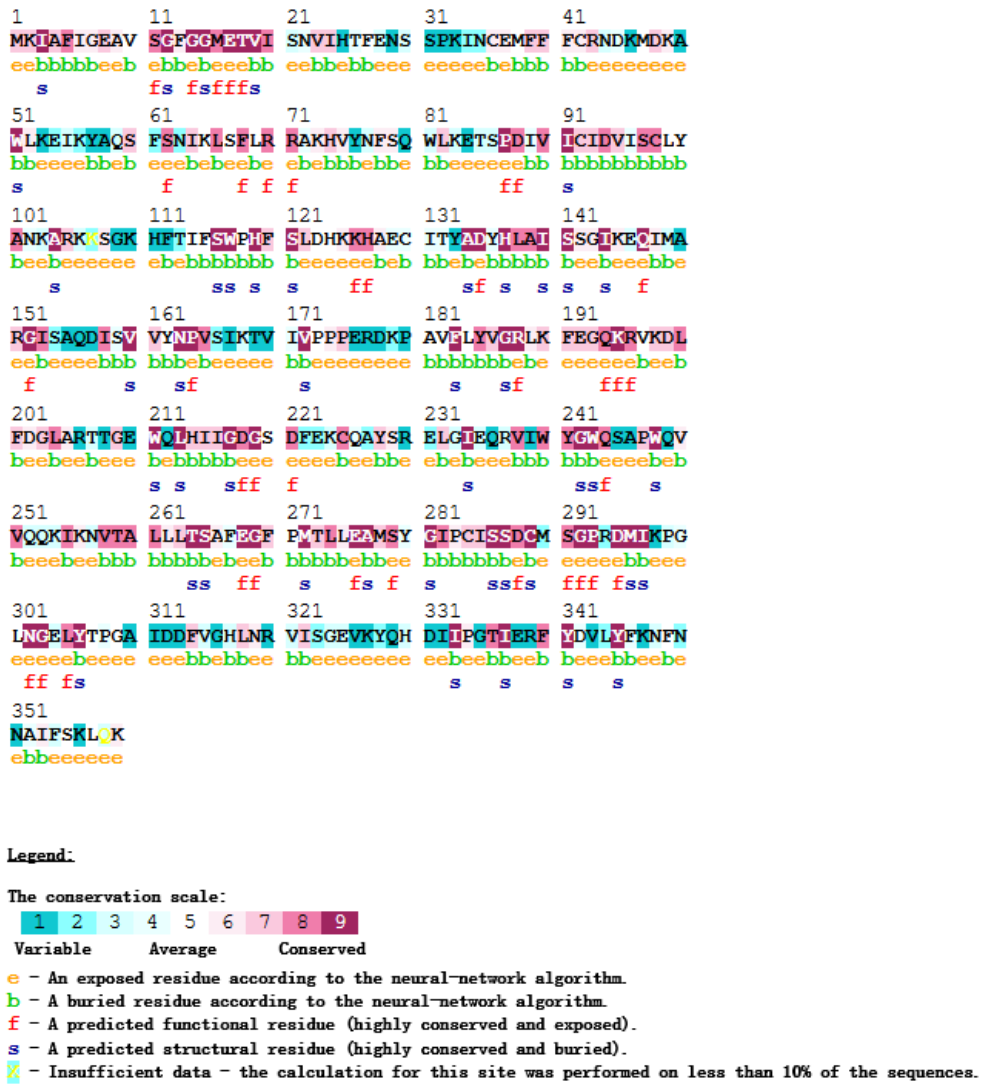
Figure 47 UDP molecule in *S. typhimurium* WaaB structure at 1.92Å resolution

S. typhimurium WaaB crystal was soaked with 100mM UDP for 60 min at room temperature before flash frozen into liquid nitrogen. The extra electron density in this map perfectly matched the UDP molecule, and no additional electron density was present part the diphosphate. The electron density is the 2Fo-Fc map contoured at 1.0σ after refinement with UDP molecule. The residues interacting with UDP were shown in sticks and the distances were measured. Substrates and the shown side chains are in ball-and-stick model [O atom (red), N atom (blue), P atom (orange)].

2.3.6 *S. typhimurium* WaaB enzyme activity sites determination

According to the conserved residue analysis by Consurf (205) and superimposition of WaaG and WaaB structure, we focused on several conserved residues that interact with UDP or might interact with the UDP-galactose and generated mutants trying to disable the enzyme activity of the protein (Figure 48). Those mutants were Q194A, K195A,

I216A, W243A, E268A, T273A and E276A. After performing an enzyme activity determination assay, the result was almost as expected (Figure 49). Wild type WaaB represented a positive control activity. The residues Q194, K195, I216, E268, and E276 are located at the hydrophobic pocket next to the UDP. Mutants Q194A, K195A, I216A, E268A, and E276A showed relatively inhibited activity in Figure 49, suggesting that the hydrophobic pocked next to the UDP binding site was important and possibly was the activity site of WaaB. By locating the dysfunctional mutant residues in WaaB structures, E268 and K195 were at the position for performing a galactose recognizing and hydrolysis. In contrast, the galactosyl transferase activity of mutant Q194A was little impaired, indicating that Q194 might not participated much in the transferase activity of WaaB. Interestingly, mutant T273 displayed an extremely higher activity even than wild type WaaB (Figure 49).



Legend:

The conservation scale:

1 2 3 4 5 6 7 8 9

Variable **Average** **Conserved**

e – An exposed residue according to the neural-network algorithm.
b – A buried residue according to the neural-network algorithm.
f – A predicted functional residue (highly conserved and exposed).
s – A predicted structural residue (highly conserved and buried).
X – Insufficient data – the calculation for this site was performed on less than 10% of the sequences.

By uploading *S. typhimurium* WaaB protein sequence into the online conserved

Figure 48 Conserved residue analysis of sequence alignment of *S. typhimurium* WaaB

residue analysis program, Consurf (206), each residues were listed and labelled according to its conserved level from ten homologous protein sequences. The residues above level 7 were considered as conserved residues and could selected for mutagenesis.

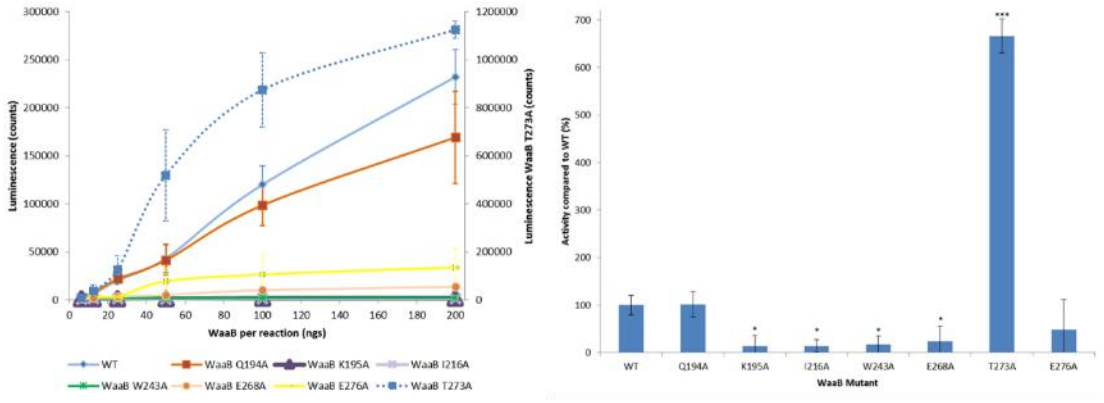


Figure 49 WaaB mutants UDP-galactosyl transferase relative activity

Left panel: The UDP-galactose transferase ability of wild type WaaB and its mutants were assayed. The concentration of the enzymes formed a gradient and the recorded luminescence reading of all the samples represented their activity. n = 4. (Work by G.J.A) Right panel: Relative activity of all the mutants were calculated using the wild type WaaB as positive control, n = 4. T273A mutant showed high activity and was not included in the relative activity graph. (Work by G.J.A)

2.4 CONCLUSION AND DISCUSSION

The Gram-negative bacterial double membrane system provides them with extra protection from toxic compounds. The asymmetric outer membrane is especially important, which is regarded as the source of the unique barrier of Gram-negative bacteria (4,5). The difference between the component of the inner leaflet and the outer leaflet causes the asymmetry of the OM. While the inner leaflet is composed of GPL similar to other membranes, the outer leaflet is made up of a large amphipathic lipopolysaccharide (LPS) (207). LPS, distinct from other membrane component, consists of three moieties: lipid A, core oligosaccharide and O-antigen (207). The biogenesis of LPS is a complicated process. In brief, lipid A is synthesized at the inner leaflet of the IM, then glycosyls are added to lipid A to form rough LPS (180). The rough LPS is flipped to the periplasmic leaflet of the IM by MsbA (208). In the meantime, the O-antigen unit is synthesized at the inner leaflet of the IM as well and is flipped to the periplasmic side of the IM by Wzx (209). Finally O-antigen units are polymerized by Wzz-Wzy (210) into O-antigen, which is then ligated to rough LPS by WaaL (211-215) to form mature LPS. Some of glycosyls on the rough LPS are essential for the survival of Gram-negative bacterial cells, such as the Hep and two early Kdo in the inner core region. The other glycosyls in the core oligosaccharide region also have various special functions contributing to the maintenance of cellular morphology or extra resistance against different kinds of toxic molecules. The biosynthesis of the core oligosaccharide is a process requiring the co-operation of a number of UDP-glycosyl transferases, each of them transfer a glycosyl onto the product of the previous transferase producing a precursor for the next glycosyl transferase. Amongst these, WaaB transfers a galactose from UDP-galactose onto the intermediate rough LPS (180,184,216).

This project focused on digging into the mechanism of this important step. Through crystallizing, X-ray diffraction and data processing of WaaB, we successfully determined the first structure of WaaB and WaaB-UDP. WaaB, sharing the features of all other glycosyl transferase, is structurally a clamp-shaped molecule, which may be

responsible for the WaaB molecule to capture the substrates and facilitate the galactosyl transferase. WaaB belongs to the GT-B family and folds into two "Rossmann-like" (β/α/β) domains, namely C-terminal domain and N-terminal domain. Similar to other members in the GT-B family, two domains binds donor and acceptor separately. In WaaB, C-terminal domain can bind UDP, which is confirmed by the UDP-bound WaaB structure (Figure 47), suggesting that the N-terminal domain corresponds to the acceptor binding domain.

Comparing the UDP-galactose (donor) structure and the outer core oligosaccharide (acceptor) structure of *S. typhimurium*, WaaB should catalyse the galactosyl transfer by retaining mechanism according to structure superimposition (Figure 45) and conserved sequence alignment analysis (Figure 48), we focused on the residues that may form the galactosyl transferase activity site. Single residue site direct mutants Q194A, K195A, I216A, W243A, E268A, T273A and E276A were generated aiming at dysfunction the potential activity site. K195, I216, W243, T273 and E276 formed interactions with UDP (Figure 47). According to the superimposition, Q194, K195 and E268 are located next to the potential galactosyl-UDP linkage position (Figure 45). However, Q and K are positively charged and could not catalyse the UDP-galactose hydrolysis, which usually require negatively charged residues. We still generated mutations on it as a negative control. Through the galactosyl transferase assay results, WaaB mutants K195A, I216A, W243A, E268A, and E276A showed deficiency in the UDP-galactose hydrolysis assay (Figure 49). K195 is the key residue of forming interactions with the phosphate group of UDP and the mutation K195A would disable the UDP binding thus prevent the hydrolysis by WaaB. And residue Q194 is not essential for this activity. Further K195 and E268 are at the position to be a galactosyl recognizing or hydrolysis residues while the others mainly interact with UDP (Figure 45). In Figure 47, residue K195 forms interactions with the β phosphate group of UDP and appears not feasible to participate in the hydrolysis when keeping the interactions at same time. The E268 residue is located perfectly for hydrolysis, its high conservation property enhanced the possibility of its catalysing ability (Figure 48). The catalysing site of glycosyl transferases are made up of the conserved negatively charged residue, usually glutamic

acid, which has a negatively charged carboxylic acid group in the side chain. Thus, we propose that the E263 residue is the key residue for the hydrolysis of UDP-galactose by WaaB. The negatively charged carboxylic acid group may attack and activate the donor and acceptor to initiate the galactose transferase.

Notably, T273A mutant displays extra activity compared to the wild type WaaB, which may be explained in a speculation that the alanine substitution of threonine on 273 residue position will facilitate the release of UDP, thus speeding up the UDP-galactose hydrolysis cycle. Possibly WaaB has a substrate retention mechanism to keep either substrates inside the WaaB for a while to ensure the galactosyl transfer rather than release the substrates immediately after hydrolysis as the *in vitro* assay shows. T273 and E276 both locates around the oxygen atoms of pentose sugar ribose, E276 formed interaction with ribose by hydrogen bond using the carboxylic acid end with the hydroxyl group of the ribose. The hydroxyl group on the side chain of T273 may form hydrogen bond with E276 using hydroxyl group of T273 with carboxylic acid group with E273. This interaction may keep E276 in position for capturing the ribose of UDP. The recognition and initial capture of UDP is multi-residual selection on the entering molecules, particularly V186, R188, Q244, E276 and K195. The T273A mutant loses the ability of keeping E276 in position, thus the WaaB is less able to lock UDP within it and the releasing of UDP is faster than wild type. However the key residues of hydrolysis and UDP recognizing/binding keep unchanged, thus the hydrolysis reaction of WaaB stays as usually with the entering and releasing speed of UDP-Gal/UDP increased, making WaaB able to hydrolyse more substrate in same time.

By determining the WaaB crystal structure and studying the WaaB mutation catalytic efficiency, we have found the residue E268 that is the catalytic site of WaaB. And we have also understood further on how proteins recognize UDP and perform biological functions. We also tried to propose the acceptor binding site of WaaB by comparing WaaB with other LPS synthesis related glycosyl transferases, and a speculative acceptor binding site is found in WaaB.

3 CHAPTER 3 Lipopolysaccharide synthesis regulation heat shock protein LapB

3.1 INTRODUCTION

Gram-negative bacteria cell envelop is made up of three distinct layers: the outer membrane (OM), the periplasm peptidoglycan wall, and the inner membrane (IM). The OM is an effective relative impermeability barrier to keep out most of the hydrophobic molecules and is responsible for the resistance to antibiotics, detergents and dyes (4). As has been described in the introduction in 1.1.1, the OM is an asymmetric lipid bilayer, of which the outer leaflet is made up by LPS, an essential component for cell survival (217). Any perturbations in synthesis, assembly or function of LPS affects normal barrier properties of the OM, leading to cell death or cell growth deficiency. The two precursors of the LPS are synthesized on the cytoplasmic leaflet of the IM which include the O-antigen units and rough LPS.

The detailed synthesis of lipid A is characterized in *E .coli*, and here we take it as an example to illustrate the biosynthesis pathway of lipid A. The final product of the fatty acid membrane anchor chains of LPS is lipid A. The first step of the biosynthesis of lipid A is the acylation of the sugar nucleotide UDP-GlcNAc (Figure 50). The enzyme for this step is LpxA, which uses the acyl carried protein (ACP) thioester as the donor and transfer the acyl chain (β-hydroxymyristate) on to UDP-GlcNAc (218). The active centre of LpxA functions as an accurate hydrocarbon ruler. The LpxA incorporation affinity of 14-carbon acyl chains is two times higher than 12- or 16- carbon acyl chains (219,220). The balance for the acylation of UDP-GlcNAc is important (221). LpxC, a zinc metalloenzyme, committed the step of deacetylation of the product of LpxA (222-226). Following deacetylation, LpxD incorporated another β-hydroxymyristate moiety onto the product of LpxC using the same donor as LpxA (40). The sequences of LpxA and LpxD are related due to the similar function and similar secondary structures (40). UDP-2,3-diacylglucosamine is cleaved at its pyrophosphate bond by LpxH to form

lipid X (Figure 50) and a UMP is released (41). LpxB then condensate another UDP-2,3-diacylglucosamine with lipid X to generate a β,1'-6 linked disaccharide (227,228). The kinase, LpxK, then phosphorylates the disaccharide to form lipid IV_A (Figure 50) (42). WaaA then add two Kdo glycosyl residues on to lipid IV_A (43). The last steps of lipid A synthesis are adding of lauroyl and myristoyl residues to the distal glucosamine unit, resulting in acyloxyacyl moieties (44) (Figure 50), using similar donor as LpxA and LpxD. In addition, there is another gene product termed LpxP which function at the same site as LpxL in *E. coli* under cold shock (12 ℃). Instead of insertion of laurate (C-12) by LpxL, LpxP inserts palmitoleate (C-16) at the same position (229). Perhaps this mechanism can adjust the outer membrane fluidity at low temperature (229).

There are additional modifications on lipid A with various kinds of moieties, including phosphoethanolamine, 4-amino-4-deoxy-L-arabinose (L-Ara4N) and/or palmitate groups. The corresponding enzymes are usually latent in *E .coli* K-12 but can be expressed in presence of metavanadate (230). Addition of phosphoethanolamine and L-Ara4N moieties are induced by exposure to mildly acidic conditions or by mutation on activation PmrA transcription factor (231-235). The palmitate modification of lipid A is controlled by PhoP/PhoQ system (236,237). *S. typhimurium* mutants without palmitate to their lipid A are more sensitive than wild type to some cationic antimicrobial peptides (236). The palmitate is a C-16 chain fatty acid.

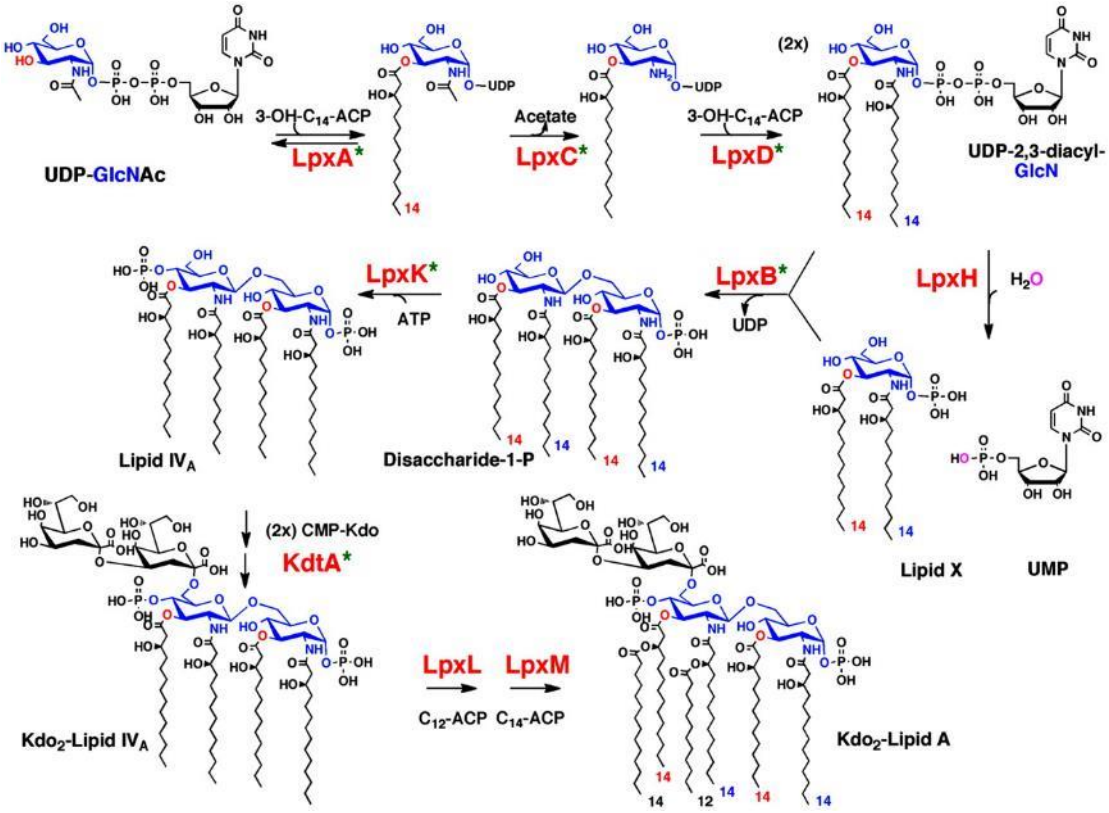


Figure 50 Biosynthesis of Lipid A in *E.coli* (5,238).

All the enzymes participated in synthesis of lipid A and Kdo_2-lipid A. KdtA is now named as WaaA. Every enzyme completes its step by themselves. ACP: acyl carried protein. GlcNAC: N-Acetylglucosamine. GlcN: Glucosamine.

The synthesis of the lipid A is regulated at the deacetylation step catalyzed by LpxC as this is the first committed step of the lipid A biosynthetic pathway (223,239). LpxC is modulated at the level of proteolysis by an essential metalloprotease FtsH in *E. coli* (240). Absence of FtsH stabilizes LpxC, leading to enhanced levels of LPS that causes cell death. Furthermore, LpxC will accumulate in the cell if the early step of lipid A biosynthesis are blocked. All these previous reports together point to the existence of a possible feedback regulation of LpxC in the LPS synthesis, which sets a balance in between phospholipid and LPS biosynthesis (217,241,242). However, there are additional proteins participate that in the biosynthesis. In 2013 and 2014 LapB (YciM) and LapA (YciS) were identified, which were both found to be able to participate in lipid A biosynthesis with unknown function (217,242). Deletion of LapB, the conserved

tetratricopeptide repeat (TPR) transmembrane inner membrane protein, is found to increase the LpxC enzyme levels resulting in an abnormal increase of LPS, ultimately leading to cell death. More interestingly, the *lapB* gene is found to be essential under laboratory growth conditions which means that the deletion strain of *lapB* gene could only be constructed under minimal medium at 30℃ (242). Conversely, a modest increase of LapB significantly reduces the amount of LpxC leading to a lower LPS levels and an increased OM permeability. Consisted with a role in the LPS assembly, during pull-down experiments, LapA and LapB were co-purified with LPS, FtsH, and Lpt proteins. Those result together indicate that LapB participates in LPS synthesis in negative feedback (243-245). It has also been reported that *lapA* and *lapB* gene belong to the RpoH regulon and their transcription will increase at 42℃ high temperature, confirming that LapA and LapB are heat shock proteins (242). LapB is anchored into the inner membrane via an N-terminal 20 residues long transmembrane helix and is facing into the cytoplasm (244). Topological and sequence analysis shows the soluble domain of LapB contains the conserved tetratricopeptie repeat (TPR) and a C-terminus domain (Figure 51). Interestingly, the C-terminus domain contains two cysteine-X-X-cysteine motifs that are supposed to be coordinate a redox-sensitive iron centre of the rubredoxin type. However the incorporated metal ion type has not been yet identified accurately, Zn^{2+} or Fe^{3+}. It has been reported that the over expressed and purified LapB cytoplasmic domain is pink under aerobic condition (colour of Fe^{3+}) and, if purified under anaerobic conditions it is colourless (non-metal ion, Zn^{2+} or Fe^{2+}) (244). In addition mass spectral data shows that, when over expressed, the concentrations of the medium ions determine the final incorporation ion ratio (LapB might contain multiple metal ion) (242). The non-specific metal ion incorporate ion makes LapB “passive” on metal ion binding relying on surrounding ion concentrations. In fact, Zn^{2+} concentration in native bacteria is much higher than Fe^{3+}, which indicates that LapB is possibly a Zn^{2+} binding protein (246).

This project is focused on generating *S. typhimurium* soluble domain LapB expression plasmids, solving the crystal structure of LapB soluble domain and performing functional studies within reach. In this Chapter, we report a successfully LapB soluble

domain structure determination at 2.2Å resolution.

Unfortunately, just one week after we solved the crystal structure, another group published that of the *E. coli* Zn^{2+} binding LapB soluble domain (156). Differently, our structure is in a totally different conformation, an ambiguous lipid substrate binding state. The conformational changes and the lipid substrate may illustrate the mechanism of LapB on LPS synthesis.

3.2 METHODS AND MATERIALS

3.2.1 *S. typhimurium* and *E. coli lapB* gene cloning

LapB cytoplasm domain of *S. typhimurium* and *E. coli* were amplified from *E. coli K12 MG1655* genome with primers listed in Table 11, respectively, with 5`-end NcoI and 3`-end HindIII restriction enzyme sites. The gene fragments were digested with restriction enzymes NcoI and HindIII and the same treatment was applied to vectors pLOU3 and pHISTEV as well. The gene fragments were then inserted into these vectors by T4 DNA ligase (Table 4). All plasmids were sequenced and confirmed the correct.

3.2.2 Expression and purification of LapB from *S. typhimurium* and *E. coli*.

The plasmids were transform into *BL21 (DE3)* cells with 100 μg/ml ampicillin. A single colony was picked and cultured overnight in 5ml LB to prepare a glycerol stock. The glycerol stock *BL21 (DE3)* with pLOU3 or pHISTEV harboring gene fragments encoding LapB cytoplasmic domain) cells were cultured in 500 ml LB overnight. Each 1L LB medium was inoculated 40 ml of overnight culture with 50 μg/ml ampicillin (pLOU3) or 50 μg/ml kanamycin (pHISTEV). The cultures were grown at 37 ℃ until the OD_{600} reached 0.8-1.0 and the proteins were induced by adding 0.1 mM IPTG for 3-4 hours at 37 ℃ with shaking at 200 rpm. 12L of cell culture were then harvested by centrifugation at 5,000 rpm (4,500 g) for 15 min. Followed by re-suspending the cell

pellet in Balance buffer (20 mM Tris-Cl, pH 7.8, and 300 mM NaCl 10mM imidazole 10% glycerol, cOmplete EDTA-free Protease Inhibitor Cocktail Tablet (Roche) 50 μg/ml DNAse I (Sigma) 100 μg/ml lysozyme). The cell pellets were lysed by passing them through a cell disruptor at 30,000 psi. The unbroken cells and cell debris were removed by centrifugation at 20,000 rpm (40,000g) for 30 min. The supernatant was applied to pre-balanced (with Balance buffer) 5 ml HP HisTrap column (Ge Healthcare). The column was washed using Wash buffer which was made by increasing imidazole concentration to 30 mM in Balance buffer; the fusion protein was eluted by Elution buffer (20 mM Tris-Cl, pH 7.8, and 300 mM NaCl 500mM imidazole and 10%glycerol). The fusion protein sample was desalted using a pre-equilibrated desalting column (GE Healthcare) with Balance buffer to remove imidazole. TEV protease was added into the desalted protein at a TEV to protein mass ratio 1:100 and incubated overnight to cleave the MBP together with 6×Histag off from the fusion protein at the TEV cleavage site. The post-cleavage mixture was loaded onto the pre-balanced 5 ml HP HisTrap column again to remove the TEV protease and the cleaved MBP during this step the WaaB protein would directly go through the column and been collected and concentrated to 5-10ml. The protein was further purified by gel-filtration using a pre-equilibrated SEC (HILOAD 16/600 SUPERDEX 200 PG, GE Healthcare) (with 150 mM NaCl and 20mM Tris pH 7.8). The protein purity was checked using SDS-PAGE and the most pure fractions were pulled and concentrated to 5-10 mg/ml for crystallization.

3.2.3 *In situ* limited proteolysis crystallization of LapB

To increase the possibility of crystal hits during crystallization screening, we tried to use proteases to remove the disordered regions and obtain a stable soluble domain after purification. We screened 7 proteases, Papain (1 mg/ml), Elastase (1 mg/ml), V8 (2 mg/ml), α-Chymotrypsin (2 mg/ml), Subtilisin (2 mg/ml), Thermolysin (1 mg/ml), and Trypsin (2 mg/ml). Firstly, 50 μg of the purified protein was prepared into separate

tubes and in each tube individual protease at mass ratio 1:100 or 1:50 (protease: protein) was added and incubate at room temperature. Samples were taken every 30 min for SDS-PAGE analysis to select one or two of the best performing proteases, which would be chosen for *in situ* proteolysis during crystallization. The performance of the proteases was analysed on SDS-PAGE and an ideal choice should exhibit almost single strong band on the SDS-PAGE, which should also display a small molecular weight shift compared to the non-digested control.

The target protein was digested using the best protease at low concentration 1:500-1:1000 rather than 1:50 which was used during the screening. Then a longer time course trail that spanned from 30 min to 4 hour at room temperature was conducted. Through this analysis, we could find the best desired condition that would digest the target protein into a stable protein without flexible fragments. It gives us a general idea on when the over digestion started because the protease could still have minor activity in crystallization condition. The promising proteases were incubated with *S. typhimurium* LapB at 1:500 for 2 hour at room temperature before we set the mixture onto crystallization experiments directly.

3.2.4 LapB crystallization and crystal sodium iodide soaking

Native or limited proteolysis treated LapB crystallization experiments were performed using the sitting drop diffusion method with drop size 0.5 µl (sample): 0.5 µl (mother liquid) using the crystallization robot Gryphon (Air). The screening kits were purchased from Hampton research and Molecular dimensions and Hampton research Table 10. After crystals appear the optimization was performed by designing and preparing 2D expansion of pH and PEG concentration from crystallization condition resulting in several 5×5 home-made optimization solution arrays in 96 deep well plates.

In order to determine the experimental phases, which was crucial for final structure determination and model building, the crystals grown from optimization conditions were soaked with sodium iodide which was quite successful in the WaaB project. To

find the best match between sodium iodide solutions with target crystals, we tried sodium iodide at 100 mM from 10 second to 15 min. The sodium iodide solution was prepared in cryoprotectant in advance the day before crystal harvesting and rapid freezing in liquid nitrogen.

3.2.5 Data collection and structure determination of LapB

All crystals, including native and sodium iodide soaked, were flash frozen in liquid nitrogen with cryoprotectant (21.25% Ethylene glycol 15% glycerol) and kept in pucks stored in a dewar, which were both pre-cooled by liquid nitrogen. The crystals were sent to Diamond Light Source I02 for data collection.

The native crystal data were collected at 1.82 Å, Oscillation 0.1 °, and 3600 images. Iodide-derived SAD data were collected at its anomalous signal scattering wavelength (1.8000Å) according to X-ray Anomalous Scattering Information (Ethan A Merritt ©1996-2011/ merritt@u.washington.edu / Biomolecular Structure Centre at UW). The original data was processed by Diamond Light Source automatic pipeline (including XIA2 (144) with Dials (145)), to generate a merged reflection file. The crystal soaked with 100 mM sodium iodide for 10 min behaved best and the experimental phase and initial structure was determined by Crank (CCP4i) ((188)), which was a pipeline that gather lots of programs together. Both native crystals and iodide-derived crystal belonged to space group $P2_12_12_1$ with cell dimensions: a=124.06Å, b=70.07Å, c=78.96Å, α=90°, β=90°, γ=90° for native crystals and a=124.34Å, b=71.86Å, c=79.09Å α=90°, β=90°, γ=90° for iodide-derived crystals. The highest resolution native crystals reach 2.2 Å and the highest resolution iodide-derived crystals reach 3.2 Å. For my case the substructure of iodide-derived LapB crystals was determined and refined by AFRO/CRUCH2 (188) and BP3 (247) respectively, hand determination and density modification was completed by SOLOMON, and model building by BUCCHANEER (154) model building. After the Crank pipeline, REFMAC5 (159) and COOT (157) were used to make final refinement removal of iodide coordinates from

the output coordinates file to produce the primary model for following Molecular Replacement. The high resolution native data were determined using PHASER MR (CCP4i) (202) using the primary model, refined by REFMAC5 (159) and further improved by COOT (157).

3.2.6 Functional assay and the related plasmids generation

In order to study the functions of LapB. We have bought the *Δlapb* strain (#9144) from *E. coli* Genetic Resources at Yale University CGSC, The Coli Genetic Stock Centre. The strain was tested to be sensitive to 100 μg/ml vancomycin, while the wild type BW25113 strain could survive. To rescue the phenotype of *lapB* deletion strain, two plasmids were generated, pBAD24-LapB/pBAD24-LapB-LpxC. By inserting full length *E.coli lapB* gene into pBAD24 vector, we have generated pBAD24-LapB. And we have inserted another *lpxC* gene fragment into pBAD24-LapB to generate pBAD24-LapB-LpxC. We transformed these two plasmid respectively into *Δlapb* strain under 100 μg/ml ampicillin and 50 μg/ml kanamycin. The culture was diluted to OD600nm=0.1 before streaking onto the agar plates with 0.0002%-0.02% L-arabinose and 100 μg/ml vancomycin, trying to rescue the phenotype of the deletion strain.

3.3 RESULTS

3.3.1 Cloning of the cytoplasmic domain of *lapB* from *S. typhimurium* and *E. coli*

*G*ene fragments encoding for the *lapB* cytoplasmic domain from *S. typhimurium* and *E. coli* were successfully amplified from the respect genomic DNA by PCR using Q5 polymerase and primers listed in Table 11. PCR fragments were then digested by appropriate restriction enzymes (Figure 52). Followed by ligation with same restriction enzyme treated vectors. Finally the colonies from the ligation mixture transformation agar plate (with appropriate antibiotics) were inoculated into 10 ml of LB for plasmids mini-prep and validation (Figure 53). One plasmid from each cloning, confirmed the correct inserts in double restriction enzyme digestion, was sent for sequencing.

```
YASPIN PREDICTION RESULTS

Name: input
Length: 389

*Key
*   AA: Target sequence
* Pred: Predicted secondary structure (H=helix, E=strand, -=coil)
* Conf: Confidence (0=low, 9=high)
*Hconf: Confidence of helix  predictions
*Econf: Confidence of strand predictions
*Cconf: Confidence of coil   predictions

   AA: MLELLFLLLPVAAAYGWYMGRRSAQQTKQDEANRLSRDYVAGVNFLLSNQQDKAVDLFLD
 Pred: -HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH---HHHHHHHHHH
 Conf: 945659748589986866572476898978745211287888899832376699999999

Hconf: 099999999999999999999999999999999999999999994009999999999
Econf: 000000000000000000000000000000000000000000000000000000000000
Cconf: 900000000000000000000000000000000000000000005990000000000000

   AA: MLKEDTGTVEAHLTLGNLFRSRGEVDRAIRIHQTLMESASLTYEQRLLAVQQLGRDYMAA
 Pred: HHHH----HHHHHHHHHHHHH---HHHHHHHHHHHHHH----HHHHHHHHHHHHHHHHH-
 Conf: 985767448599996999993153779999999997254455102359999978899931

Hconf: 999900019999999999995019999999999999900007899999999999999995
Econf: 000000000000000000000000000000000000000000000000000000000000
Cconf: 000099980000000000004980000000000000999921000000000000004

   AA: GLYDRAEDMFNQLTDETEFRVGALQQLLQIYQLTSDWQKAIEVAERLVKLGKDKQRIEIA
 Pred: --HHHHHHHHHHHHH----HHHHHHHHHHHHH--HHHHHHHHHHHHH------HHHH
 Conf: 34779999999981462204899989999912278799999999946573310156 6

Hconf: 009999999999999902159999999999997209999999999999900000479999
Econf: 000000000000000000000000000000000000000000000000000000000000
Cconf: 990000000000000978400000000000027900000000000009999520000

   AA: HFYCELALQQMGNDDMDRAMALLKKGAAADKNSARVSIMMGRVYMARGDYAKAVESLQRV
 Pred: HHHHHHHHHHHH---HHHHHHHHHHHHH----HHHHHHHHHHHH---HHHHHHHHHH
 Conf: 888888989993157879999999998774857879999699999223888999999999

Hconf: 999999999996009999999999999900009999999999940099999999999
Econf: 000000000000000000000000000000000000000000000000000000000000
Cconf: 000000000003990000000000000999900000000000059900000000000

   AA: IVQDKELVSETLEMLQTCYQQLGKNAEWAEFLRRAVEENTGAGAELMLADILEAREGSDA
 Pred: HHH----HHHHHHHHHHHHHH--HHHHHHHHHHHHH---HHHHHHHHHHHH---HHH
 Conf: 967541352899996999991326768999999983657562699989999921388 9

Hconf: 999013199999999999999810999999999999990009999999999999600999
Econf: 000000000000000000000000000000000000000000000000000000000000
Cconf: 000986800000000000001890000000000000999000000000000399000

   AA: AQVYITRQLQRHPTMRVFHKLMDYHLNEAEEGRAKESLMVLRDMVGEQVRSKPRYRCQKC
 Pred: HHHHHHHHHHH---HHHHHHHHHHHHHHH-HHHHHHHHHHHHHHHHHHH---------
 Conf: 999999999673645199998899997103313148899999999997415821000002

Hconf: 999999999900099999999999998808899999999999999999031672680
Econf: 000000000000000000000000000000000000000000000000000000012100
Cconf: 000000000009990000000000000119110000000000000000000968214209

   AA: GFTAYTLYWHCPSCRAWSTIKPIRGLDGQ
 Pred: -----EEEEE--------EEE--------
 Conf: 651111353227666442122142622 9

Hconf: 000100000000000000000000000
Econf: 004009999300000279840000000
Cconf: 995780000069999997201599999 99
```

Figure 51 *S. typhimurium* LapB secondary structure analysis by YASPIN server (248)

S. typhimurium LapB amino acids sequence was uploaded into YASPIN server (248) and the data shows that the LapB contains several α-helices forming a typical tetratricopeptide repeat (TPR) domain. The N-terminal is an anchor of 20 residues into the cell membrane and The C-terminal contains four conserved cysteine residues which correspond to a metal binding centre. (244). 1-9 refers to the possibility from low to high.

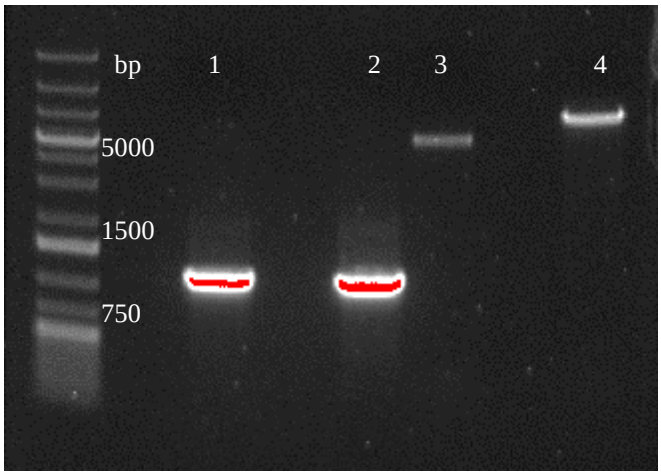


Figure 52 Agarose gel analysis of double restriction enzyme treated LapB PCR products and vectors

Gene fragments encoding *E. coli* and *S. typhimurium* LapB were amplified using PCR, digested with restriction enzyme NcoI and HindIII. Vectors pLOU3 and pHISTEV were treated with same restrictions as well. Digested PCR products and vectors were checked and purified on 0.8% agarose gel. Lane 1: *E. coli lapB* gene fragment 1101bp; lane 2: *S. typhimurium lapB* gene fragment 1101bp; lane 3: pHISTEV vector 5432bp; lane 4: pLOU3 vector 6702bp. All PCR products or vectors had been digested with restriction enzymes NcoI and HindIII.

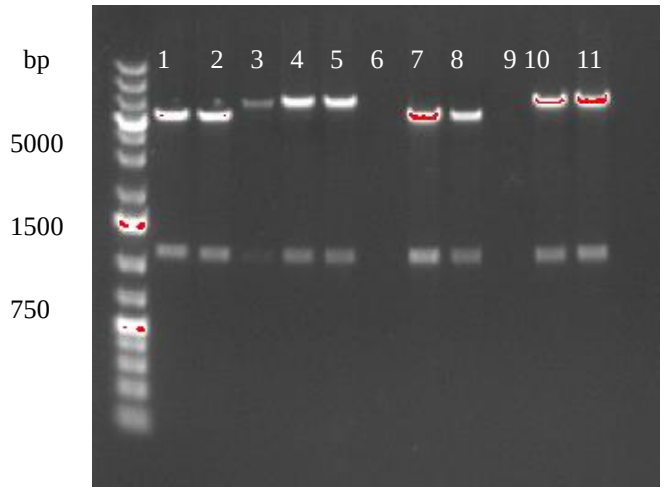


Figure 53 Double restriction enzyme digestion validation of *lapB* gene cloning.

lapB gene amplified from *E.coli* and *S. typhimurium* were inserted into pHISTEV or pLOU3 respectively by T4 ligation. The single colonies from the transformation of the ligation products were cultured. The plasmids extracted from theses colonies were double digested to check whether the insertions were correct. The restriction sites used here were NcoI and HindIII. Lane 1, 2 were *E. coli lapB* in pHISTEV; lane 4, 5 were *E. coli lapB* in pLOU3; lane 7, 8 were *S. typhimurium lapB* in pHISTEV; lane 10, 11 were *S. typhimurium lapB* in pLOU3.

3.3.2 *E. coli* LapB in pLOU3 and pHISTEV expression and purification

E. coli LapB was purified from expression vector pLOU3, in which the LapB cytoplasmic domain was expressed at a fusion protein form, 6×Histag-MBP-TEVcleavagesite-LapB, and then the LapB protein was separated from the fusion state using TEV protease and Ni-NTA purification *in vitro*. Interestingly once the protein was eluted from the column, it showed crimson color. The special crimson color indicated that the protein contains Fe^{3+} ion inside. We could not exclude the presence of other metal ions, like Zn^{2+} or Fe^{2+}. The metal binding ability of LapB was also consistent with previous reports (156). From the gel-filtration (Figure 54 and Figure 55) and SDS

PAGE (Figure 56), the LapB was purified successfully. According to the gel-filtration, the peak started earlier than expected which is estimated to be about 39kD. However, in Figure 56, LapB still presented a normal molecular weight band of 39kD, LapB probably formed dimer in solution which were confirmed later from the structure determination of LapB. Although the protein was successfully expressed and purified, some degradation bands were apparently observed after TEV protease digestion and final gel-filtration (Figure 56).

After detecting the degradation, I optimized some details in purification method, and also tried to express LapB in pHISTEV plasmid on the other hand. Unlike pLOU3, pHISTEV vector did not carry a 42kD MBP fused with your protein (Figure 57). To minimize the degradation of LapB during, the TEV protease digestion was performed on ice in the following experiments. As a result, the degradation was well controlled.

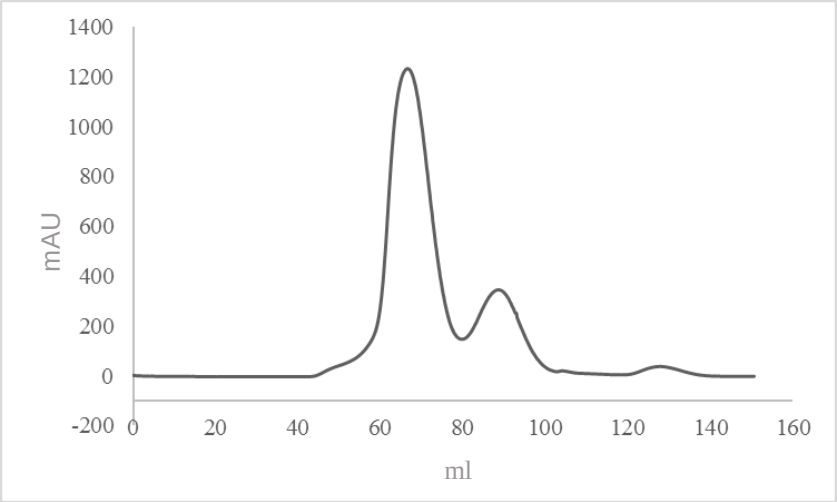


Figure 54 Gel-filtration of *E. coli* LapB-pLOU3 fusion protein

E. coli LapB-pLOU3 fusion proteins were eluted from the Ni-NTA columns. 10ml of the eluted fusion proteins were loaded into the gel-filtration with buffer 20mM Tris pH8.0 300mM NaCl and 10% glycerol. The columns used was 16/600 SUEPERDEX 200PG (GE health care). The first peak is *E. coli* LapB-MBP fusion protein peaked at 70ml (molecular weight 100 kD-200 kD). The second peak appeared at 90ml (30-60 kD) is MBP.

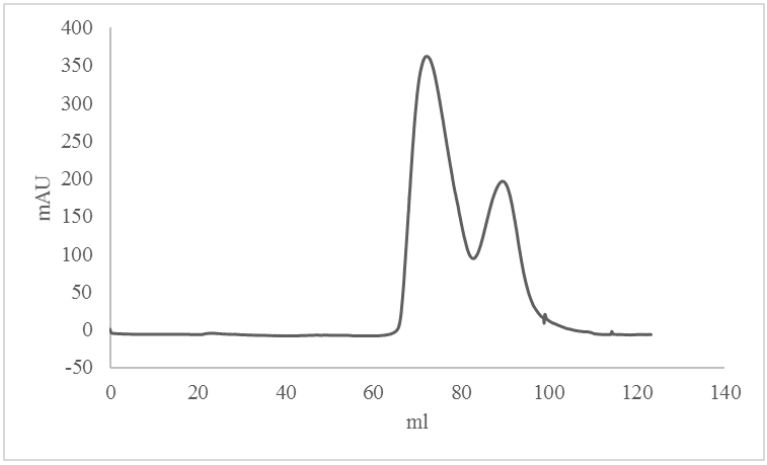


Figure 55 Gel-filtration of *E. coli* LapB-pLOU3 fusion protein after TEV digestion

After TEV cleavage and the reverse Ni-NTA purification to remove TEV protease from the LapB, MBP and TEV mixture. The flow through containing LapB was concentrated to 7.5ml before being loaded into the gel-filtration with buffer 20mM Tris-Cl pH8.0 150mM NaCl. The first peak was the purified *E. coli* LapB protein confirmed by SDS-PAGE analysis in Figure 56. The second peak was MBP that was not fully captured by reverse Ni-NTA purification after TEV protease digestion.

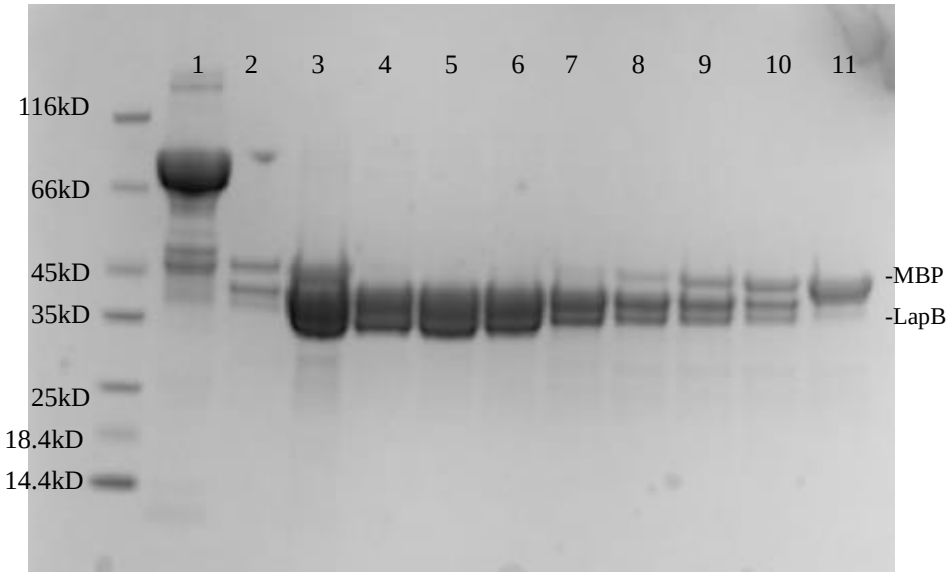


Figure 56 SDS-PAGE analysis of *E. coli* LapB-pLOU3 fusion protein purification

Several different samples during *E. coli* LapB purification was prepared and analysed with SDS-PAGE. Lane1 1: *E. coli* MBP-LapB fusion protein; lane 2: After TEV protease digestion; lane 3: After TEV protease digestion Ni-NTA flow through; lane 4-10: Gel-filtration first peak from fractions in Figure 55; lane 11: MBP peak sample. MBP bands is a little above LapB bands, and gel-filtration can almost separate them because LapB formed a dimer. LapB showed obvious degradation bands and form a ladder-like pattern from about 40kD to 35kD. 5µl samples were mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into SDS-PAGE gel after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature. Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific)

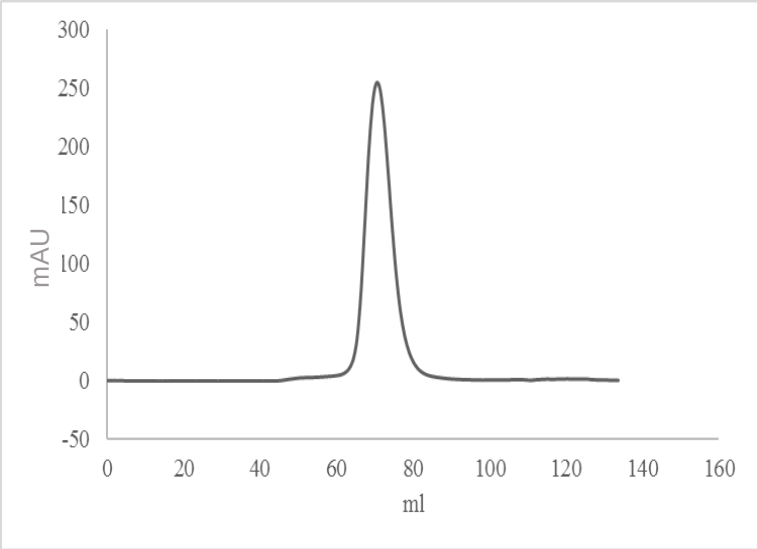


Figure 57 Gel-filtration of *E. coli* LapB-pHISTEV after TEV protease digestion

E. coli LapB purified from pHISTEV and the purification step was similar with that from pLOU3. The LapB after TEV cleavage was purified again through a reverse Ni-NTA column and loaded into gel-filtration subsequently using buffer 20mM Tris-Cl 150mM NaCl. The gel-filtration was performed on 16/600 SUPERDEX 200PG GE Health care at 1ml/min room temperature.

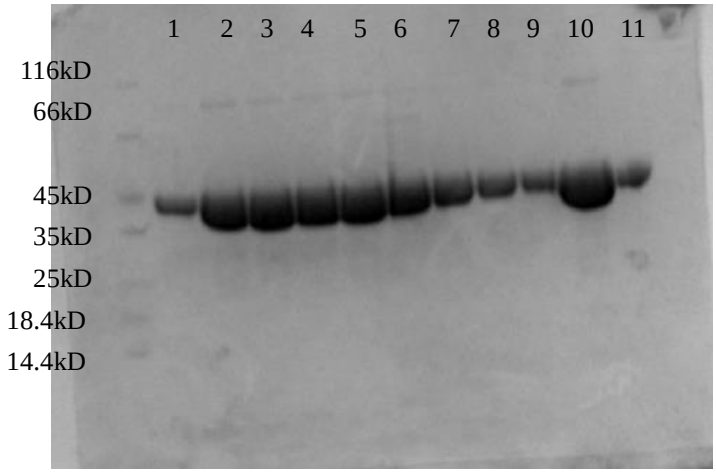


Figure 58 SDS-PAGE of *E. coli* LapB-pHISTEV after TEV digestion at 4℃

Lane 1-10: Gel-filtration fraction samples from Figure 57; lane 10: flow through after TEV protease digestion; lane 11: *E. coli* LapB before crystallization. This purification differs from the batch shown in Figure 56, the TEV digestion was performed on ice overnight rather than at room temperature. Some nonspecific degradations of LapB were prevented. LapB appeared to be stable at this a this stage. 5µl samples were mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into SDS-PAGE after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature. Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific)

3.3.3 *S. typhimurium* LapB expression and purification in pHISTEV vector.

While continuing trying to grow crystals of *E. coli* LapB, I have also tried to crystallize LapB from *S. typhimurium* which showed some slight differences. During the protein purification, the pink color of *S. typhimurium* LapB showed lighter compared to the *E. coli* LapB and minor precipitation appeared after cell lysis which did not affect purification. As a result, we could still produce high yield protein (Figure 59). During gel-filtration, the target peak appeared at around 67 ml with an aggregation peak before

it and the target fractions were collected for TEV protease digestion. After TEV digestion and reverse chromatography to remove TEV protease and cleaved 6×Histag, the protein was further purified by another gel-filtration. The protein purity was finally checked by SDS-PAGE (Figure 60 and Figure 61). As we can see from the SDS-PAGE, the *S. typhimurium* LapB displayed a small shift before and after TEV cleavage which confirmed the efficiency of TEV protease digestion. Although the TEV protease digestion was performed on ice, small degradation bands were still visible just below LapB, which meant that *S. typhimurium* LapB, similar to the *E. coli* one, was not very stable. Considering that *S. typhimurium* LapB showed lighter pink colour, we could conclude that this protein contains less Fe^{3+} compared to *E. coli* LapB. Since bacterial cells contained much higher concentration of Zn^{2+} than Fe^{3+}, the native LapB is likely to be a Zn^{2+} binding protein rather than a Fe^{3+} binding protein. The Fe^{3+} binding could be a non-specific binding. In that case, *S. typhimurium* LapB was chosen as the main target to focus on.

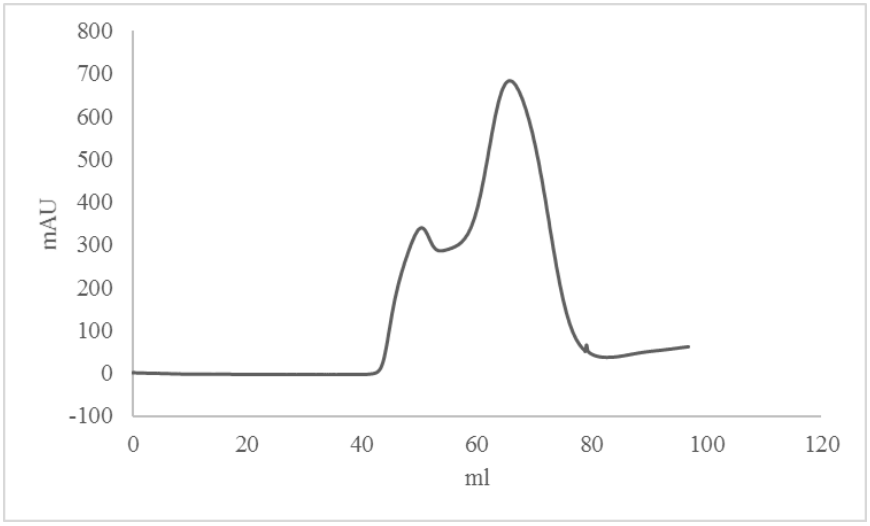


Figure 59 Gel-filtration of pHISTEV *S. typhimurium* LapB after Ni-NTA purification and before TEV protease digestion

S. typhimurium LapB was expressed in pHISTEV plasmid and purified using the same method as the *E.coli* one. After eluted from the Ni-NTA, 10ml LapB-Histag sample was loaded into the gel-filtration using buffer 20mM Tris pH8.0 300mM NaCl 10% glycerol at 1ml/min and room temperature. The first peak is

the aggregation peak caused by slight precipitation in the supernatant of LapB before purification. The second peak is 6×Histag-TEVcleavagesite-*Sal*-LapB. The fractions from volume 65ml to 80ml were collected for TEV digestion, and the first aggregation peak was discarded. The reason for slight precipitant in the supernatant remained unknown, but it had never happened in the repeating purifications. In the following purifications, desalting was used in place of gel-filtration to remove the imidazole from the fusion protein for TEV cleavage.

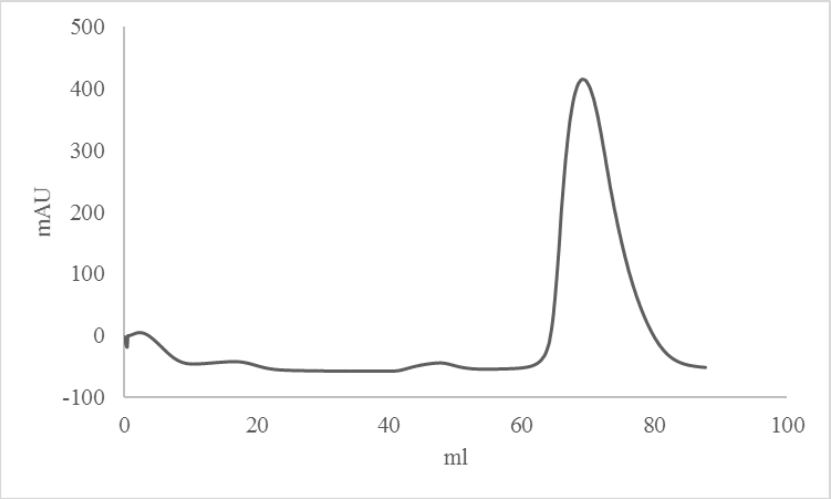


Figure 60 Gel-filtration of *S. typhimurium* LapB

After TEV protease cleavage and a reverse Ni-NTA purification to remove TEV protease, *S. typhimurium* LapB in the flow through was concentrated into 10ml before further purified by gel-filtration in 20mM Tris-Cl pH8.0 150mM NaCl. The protein samples were confirmed by SDS-PAGE (Figure 61). The slightly asymmetry of the peak could be caused by the slight degradation of *S. typhimurium* LapB. The protein harvested after this gel-filtration were concentrated to 5-10mg/ml for crystallization.

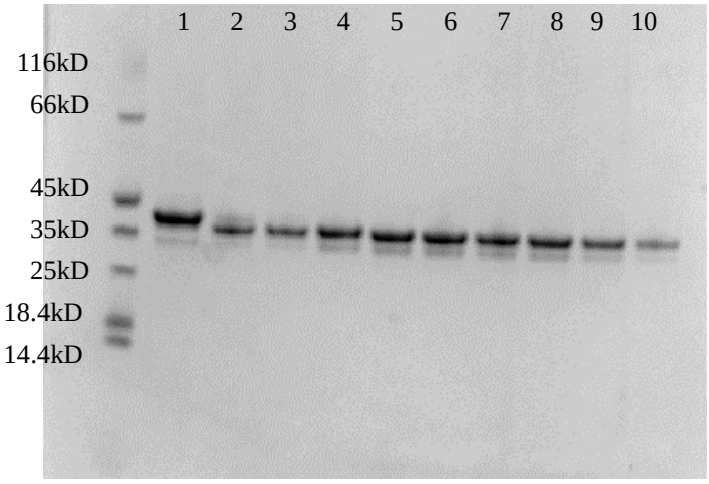


Figure 61 SDS-PAGE analysis of *S. typhimurium* LapB

Lane 1: *S. typhimurium* LapB purified in pHISTEV plasmid before TEV protease digestion; lane 2-10: gel-filtration fraction samples from Figure 60. The band of *S. typhimurium* LapB was similar to *E. coli* one. Although the protein is purified with high purity, the degradation problem still existed. 5µl samples were mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into SDS-PAGE after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific)

3.3.4 Limited proteolysis of LapB from *E.coli* and *S. typhimurium*.

Observing the degradation of LapB from both species, I tried to remove the flexible regions by limited proteolysis treatment to increase the possibility of crystallization. Seven different proteases was used on both LapB at 1:100 mass ration for 30 min room at temperature (Figure 62 and Figure 63). From the limited proteolysis result, proteins contain an unstable domain, which could be easily removed by several proteases. However, some proteases behaved better on LapB of two species, such as Elastase and Thermolysin. These two proteases can remove a certain part of peptide from the LapB protein and were used in situ proteolysis digestion crystallization for both *S.*

typhimurium and *E. coli* LapB. As a result, there were actually more crystal hits appearing in wells compared to untreated crystallization and the quality of some crystals was increased to about 2.4 Å. However, the high resolution crystals were not very reproducible. By using the same method during crystal optimization, only small crystals could be grown. The resolution of the optimized crystals were worse compared to the original one and they were also quite fragile under X-ray diffraction. The small crystals disappeared after two week with lots of cracks, finally became debris. The proteases were possibly still active at and could still digest the protein during the crystal formation, resulting in such low crystal quality. I decided to screen different proteolysis condition. A time course experiment with 1:500 mass ratio of Thermolysin to LapB was performed and samples at 1 h, 4 h and 24 h were taken from the proteolysis reaction and visualized by SDS-PAGE (Figure 64). According to the SDS-PAGE result of time course, it was apparent that the Thermolysin could digest about half of the *S. typhimurium* LapB at 1 hour using 1:500 ratio, and would fully digest the sample by 4 hour and over digestion was observed after 1 day. Based on the time course result and considering the possible protease activity during crystal growth, 2 hour digestion at 1:500 ratio was chosen for further *in situ* proteolysis crystallization.

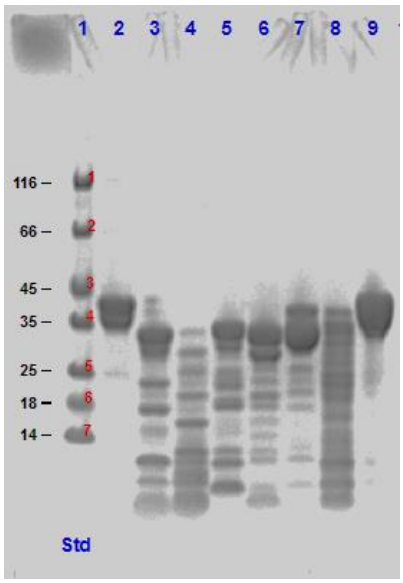


Figure 62 SDS-PAGE analysis of *E. coli* LAPB limited protease digest screen

Proteolysis screen of *E. coli* LapB. Lane 2: Control; lane 3: α-Chymotrypsin;

lane 4: Subtilisin; lane 5: Trypsin; lane 6: Thermolysin; lane 7: Elastase; lane 8: Papain; lane 9: V8. All the proteases were used at 1:100 mass ratio of protease to *E. coli* LapB at room temperature for 30 min. 4×SDS-PAGE Loading buffer was added into each samples to terminate the reactions. The SDS-PAGE samples were loaded into SDS-PAGE after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific). The control LapB showed some instability and there are two strong bands next to each other. The proteases that could produce a band shift with least protein loss were selected for proteolysis treatment, namely Thermolysin and Elastase.

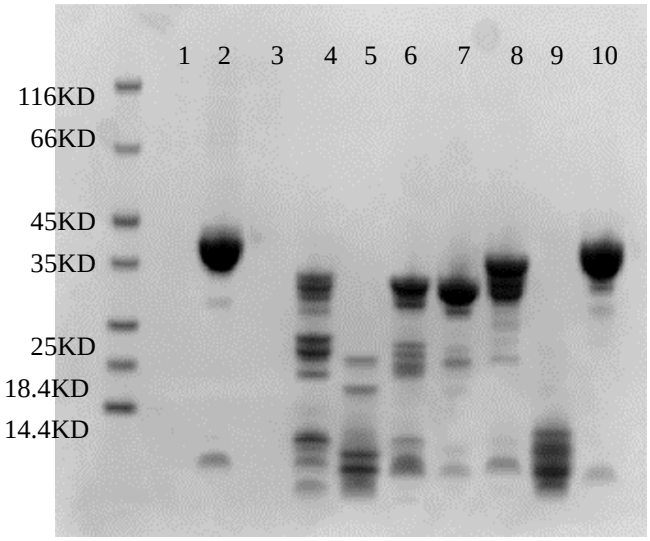


Figure 63 SDS-PAGE analysis of *S. typhimurium* LapB limited protease digest screen

Proteolysis screen of *S. typhimurium* LapB. Lane 2: Control; lane 3: α-Chymotrypsin; lane 4: Subtilisin; lane 5: Trypsin; lane 6: Thermolysin; lane 7: Elastase; lane 8: Papain; lane 9: V8. All the proteases were used at 1:100 mass ratio of protease to *S. typhimurium* LapB at room temperature for 30 min. 5µl samples were mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into SDS-PAGE. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for

15min at room temperature. Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific) after heating on 90°C for 10min. The LapB control seemed to be stable with little degradation, but after protease treatments the flexible part of it would be easily removed, which could be found in several kinds of proteases. The proteases that could produce a band shift with least protein loss were selected for proteolysis treatment, namely Thermolysin and Elastase in this case. After tested both for crystallization, Thermolysin can help produce crystals while Elastase cannot.

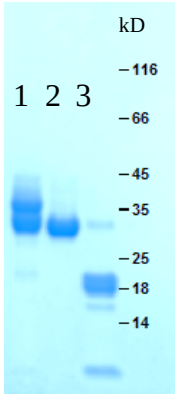


Figure 64 *S. typhimurium* LapB Thermolysin time course proteolysis digestion

S. typhimurium LapB was digested with Thermolysin at 1:500 mass ratio and room temperature. Samples was taken at 1 hour, 4 hour, 24 hour, and loaded into lane 1, 2, 3, respectively. 5µl samples were mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into SDS-PAGE after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature. Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific). This is a time course limited proteolysis digestion trying to find the best combination of concentration and duration to achieve best effect. As shown in the figure, 4h and 1:500 appeared to be a promising choice. 1h seemed to be not enough, we could still identify two bands which may refer to before/after digestion. 24h is too long, about half of LapB disappeared.

3.3.5 Crystallization of LapB of *S. typhimurium*/*E. coli* and seeding optimization

Both *S. typhimurium* and *E. coli* LapB could crystallize (Figure 65 and Figure 66), but only *S. typhimurium* LapB could produce large and high quality crystals. In contrast, the resolution of the crystals from *E. coli* LapB were very poor. However, optimization and limited proteolysis of *S. typhimurium* LapB did not work sufficiently well, with problems mainly about the crystal size and reproducibility (Figure 66). Microcrystal seeding could be an alternative choice, the *S. typhimurium* LapB crystals, together with the crystal growing solution, were pipetted out from one well and the mixture was dilute in 20 µl reservoir solution. Then after pipetting the diluted mixture for hundreds of times to fully destroy the crystals into microcrystals, the solution was diluted again to a volume of 200µl using buffer (20mM Tris-Cl pH8.0 and 150mM NaCl). This 200 µl stock was stored at -80℃ and named as Seeding Stock for *S. typhimurium* LapB. 1 µl of the Seeding Stock was added into 100 µl purified *S. typhimurium* LapB solution that was treated with limited proteolysis before setting onto crystallization. Combining limited proteolysis and crystal seeding methods, we have improved the quality and the reproducibility of the LapB crystals (Figure 67). The original crystals were obtained from 21.25% Ethylene glycol 15% glycerol at room temperature; the crystals appeared after 1-2 weeks.

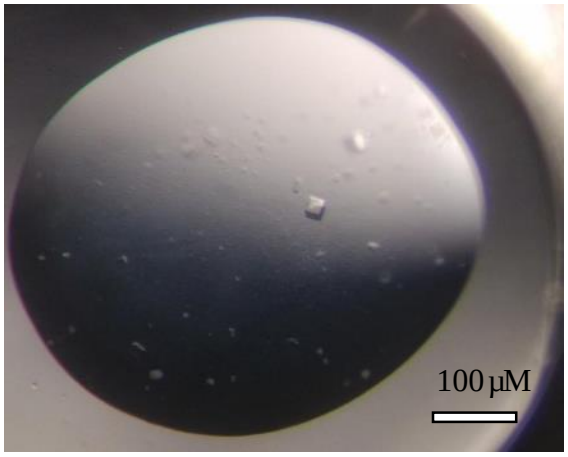


Figure 65 *E. coli* LapB crystals after Elastase proteolysis repetition

E. coli LapB formed small crystals after Elastase proteolysis repetition. These

crystals were bad and could only diffract up to 15 Å. There were about 3-4 hits like this through all the screens. Compared to the *S. typhimurium* LapB crystals (Figure 67), *E. coli* ones seemed to be difficult to optimize to reach high resolution. So we focused on the *S. typhimurium* LapB in the following experiments.

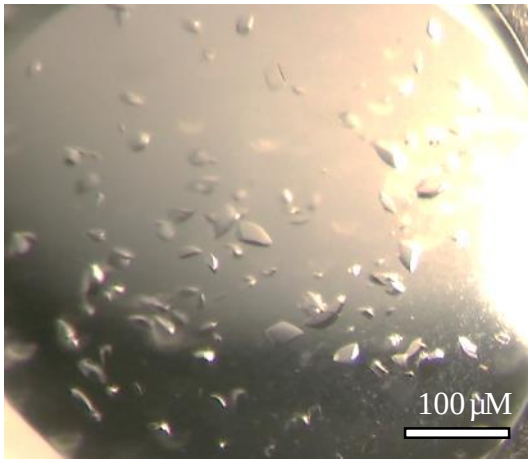


Figure 66 Small crystals of *S. typhimurium* LapB after advanced limited proteolysis

Based on the time course of Thermolysin proteolysis digestion, *S. typhimurium* LapB crystals became reproducible using the 1:500 mass ratio and 4h room temperature digestion condition, however the occurrence was 1 out of 10 and the crystals were not large enough for high resolution data collection. (21.25% Ethylene glycol 15% glycerol)

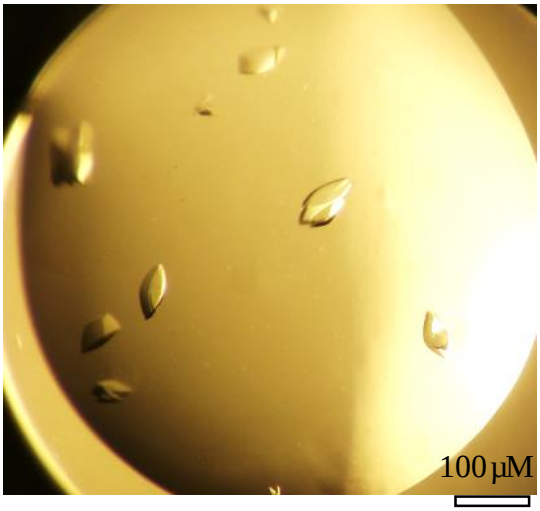


Figure 67 High resolution crystals of *S. typhimurium* LapB

By using optimized proteolysis method based on time course proteolysis result and microcrystal seeding, large crystals were grown with occurrence hit of 7 out of 10. The best of these crystal have good resolutions of 2.2 Å. 1:500 mass ratio protease, 4h at room temperature, 1:100 Seeding. It took 1 week for these crystal to form and 3 weeks to grow into full size. (21.25% Ethylene glycol 15% glycerol)

3.3.6 *S. typhimurium* LapB structure contains a lipid molecule

To determine the experimental phases of *S. typhimurium* LapB, crystals were soaked with cryoprotectant solution containing 100 mM sodium iodide (21.25% Ethylene glycol 15% glycerol 100 mM sodium iodide). The crystals were harvested and flash frozen in liquid nitrogen after 10 s, 1 min, 3 min, 5 min, 10 min and 15 min soaking. The soaking solution was added at 1:1 (drop size to soaking solution). All crystals were sent to Diamond Light Source beamline I04 for data collection. Native data were collected at wavelength 1.00 Å while sodium iodide soaked SAD data were collected at 1.80 Å with high multiplicity (Table 8). Initial model was determined from iodide-derived SAD data using CRANK (188). The iodide anomalous Fourier map was shown in Figure 68. Then Molecular Replacement was used to solve the high resolution native dataset using the initial model. COOT (157) was used to build and fix the final model. According to the overall structure in Figure 69, the *S. typhimurium* LapB structure was determined from dataset of *in situ* proteolysis crystals with the N-terminal 140 residues missing from the structure. Our construct truncates 20 transmembrane residues, so the rest about 120 residues were removed by proteolysis. The LapB protein was divided into two domains: the N-terminal domain made up with TPRs and the C-terminal metal binding centre (Figure 69). The N-terminal TPRs consists of several α-helix pairs (six pairs in our LapB structure) and may contain more in full length LapB. TPR domain is usually supposed to form scaffolds to mediate protein-protein interactions and may assembly multi-protein complex. As a rather new identified LPS biosynthesis related

protein, LapB may utilize the TPR domain to interact with enzymes from lipid A synthesis (Figure 50). The C-terminal domain of LapB consists of a four-cysteine motif located at the corners of a tetrahedron with metal binding site in the centre of this tetrahedron (Figure 71). *S. typhimurium* LapB is supposed to be a metal binding protein. However, our crystals from the microscope images (Figure 66 and Figure 67) did not contain any colour, which meant that our crystals did not have Fe^{3+} ions bound. These details indicated that the purified LapB protein was probably in a Zn^{2+} binding form. We tried to build a Zn^{2+} into our structure and refine with it against the high resolution data. 2Fo-Fc map contoured at 8.0σ still contains signal in the Zn^{2+} site, suggesting that our structure is in Zn^{2+} form (Figure 71). Comparing to the 4ZLH (249), the *E. coli* LapB crystal structure whose C-terminus metal binding centre captured a Zn^{2+}, our *S. typhimurium* LapB displayed a similar architecture (Figure 72). According to the superimposition result, *S. typhimurium* LapB showed a 31.9 Å rotational shift between the two monomer in the asymmetric unit, which opened the hydrophobic cavity between the two LapB monomers and might allow the entry of the lipid substrate (Figure 72). In our LapB structure we could observe an extra chain-shaped electron density between the two LapB protomer on the Fo-Fc map contoured at 3.0σ (Figure 69). Considering the relationship of LapB and LPS, we anticipate it to be a lipid or fatty acid molecule that may be involved in LPS biosynthesis. The tunnel space of the LapB dimer structure and the electron density both suggested that only single carbon chain could be put there (Figure 69). We modelled a palmitoleic acid molecule into it, and it fit well. It is no exclusive that this additional density can be assigned to other similar C-16 chain, including palmitic acid, or even C-14 chains. The synthesis of lipid A of LPS require the lipid donors in C12-ACP or C14-ACP form. And C16-ACP (palmitic acid) may also be required when modification on lipid A happens. This unassigned lipids may be this lipid or some precursor.

	LapB and lipid complex	LapB-iodide
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$

Cell dimensions		
a, *b*, *c* (Å)	124.06, 70.07, 78.96	124.34, 71.86, 79.09
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.91731	1.8000
Resolution (Å)	48.92–2.2(2.279–2.201)	73.85-3.12 (3.20-3.12)
R_{merge} (%)	14.3 (26.4)	22(81.9)
CC1/2 (%)	100 (82)	100 (100)
$I / \sigma(I)$	8.24(?)	19.6 (5.6)
Completeness (%)	100 (100)	100 (100)
Redundancy	13.3 (11.1)	47.2 (24.6)
Anomalous Completeness		100(100)
Anomalous Redundancy		25.2(12.7)
Phasing		
Resolution (Å)		73.85-3.12
Sites (I)		11
Figure of merit		0.612
Refinement		
Resolution (Å)[&]	48.72 – 2.50(2.589-2.5)	
No. reflections	23448(1675)	
R_{factor} / R_{free}	0.2346/0.2785	
No. atoms		
Protein	4067	
Ligand/ion	20	
Water	94	
B-factors		
Protein	58.10	
Ligand	54.4	
Solvent	49.5	
R.m.s.deviations		
Bond lengths (Å)	0.005	
Bond angles (°)	1.260	
Ramachandran statistics		
Allowed (%)	97.8	
Outliers (%)	2.2	

Table 8 Data collection and structure refinement statistics of *S. typhimurium* LapB lipid binding form.

R_{factor} = Σ|| Fobs|−| Fcal||/Σ|Fobs|, where Fobs and Fcal are observed all reflection measured and calculated currently model as structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly

selected not used in refinement.

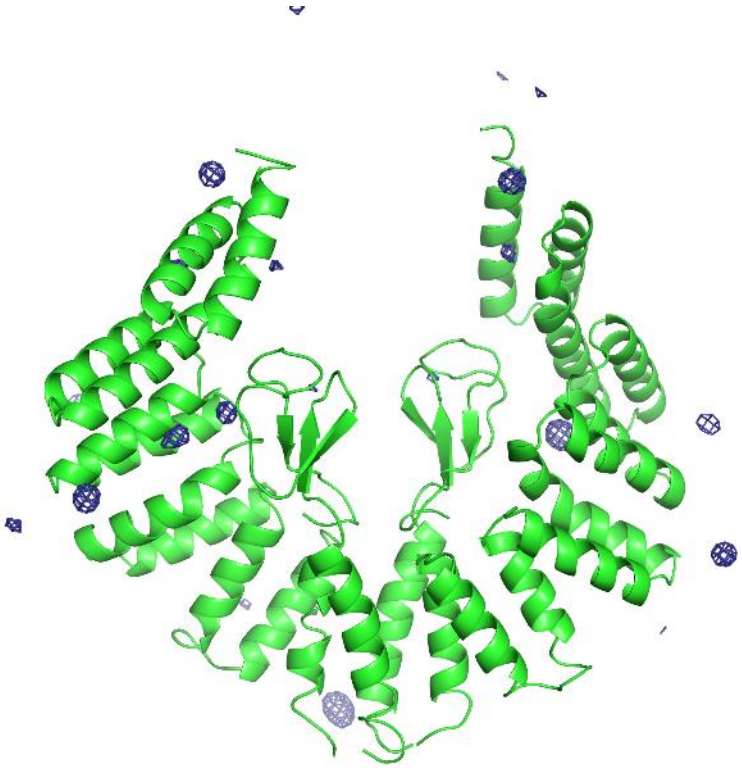


Figure 68 Anomalous difference Fourier map of iodide derived *S. typhimurium* LapB

The iodide derived *S. typhimurium* LapB SAD data was collected at 1.8000 Å. The anomalous difference Fourier map of iodide derived *S. typhimurium* LapB was calculated with PHENIX (155) and contoured at 7.0σ, by using the anomalous dataset and final model (without lipid substrate and Zn^{2+}). The peaks in the images shows the iodide sites that entered the crystals during the sodium iodide soaking.

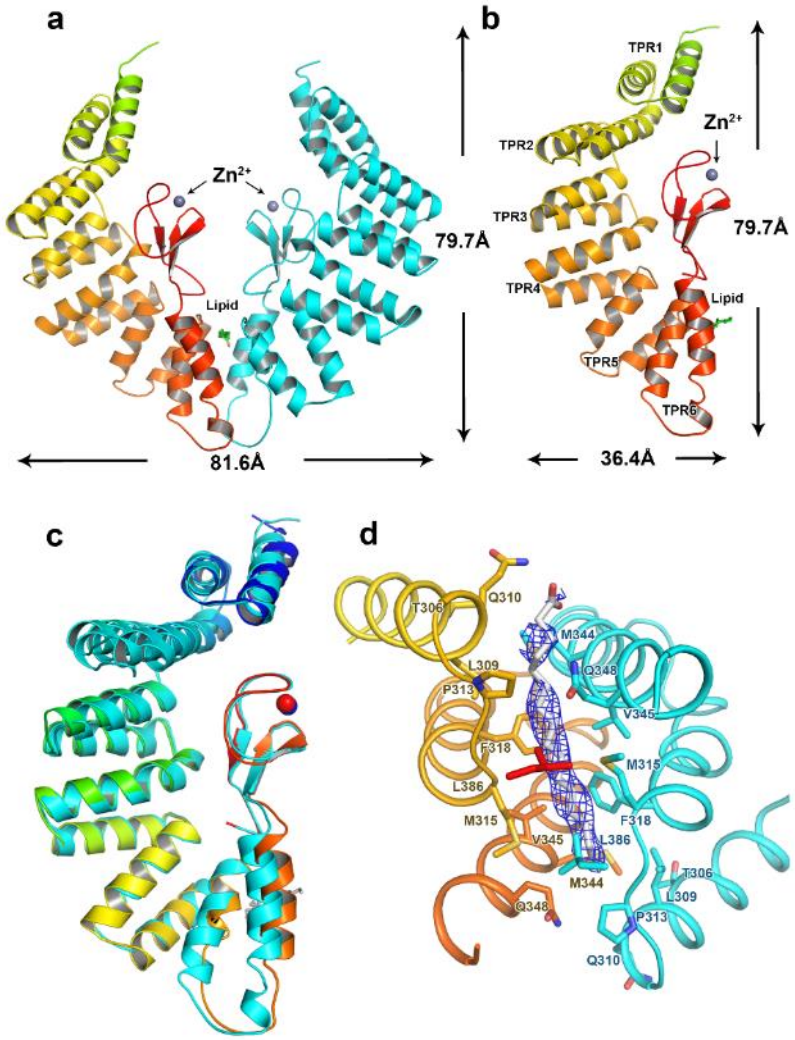


Figure 69 *S. typhimurium* LapB with lipid bound between two monomer

(a) There are two LapB molecules per asymmetric unit with dimensions of about 81.6 Å in length and 79.7 Å in height. There is one zinc ion per protomer located in the C-terminus domain. The lipid is located in the hydrophobic tunnel between the two protomers. (b) The LapB soluble domain can be divided into two domains; the N-terminal domain and the C-terminal domain (244) which form a bird wing shape with a dimensions of approximately 36.4 Å in length and 79.7 Å in height. There are six tetratricopetide repeats motifs (TPR1-6) while the C-terminal domain is the rubredoxin-like domain, with four cysteine

residues. (c) Superimposition of the two protomers of *S. typhimurium* LapB from one asymmetric unit. Two protomers in dimer LapB are in almost same conformation. (d) The unassigned electron density map and palmitoleic acid docking in the hydrophobic tunnel. The Fo-Fc map of the unassigned density is contoured at 2.5σ.

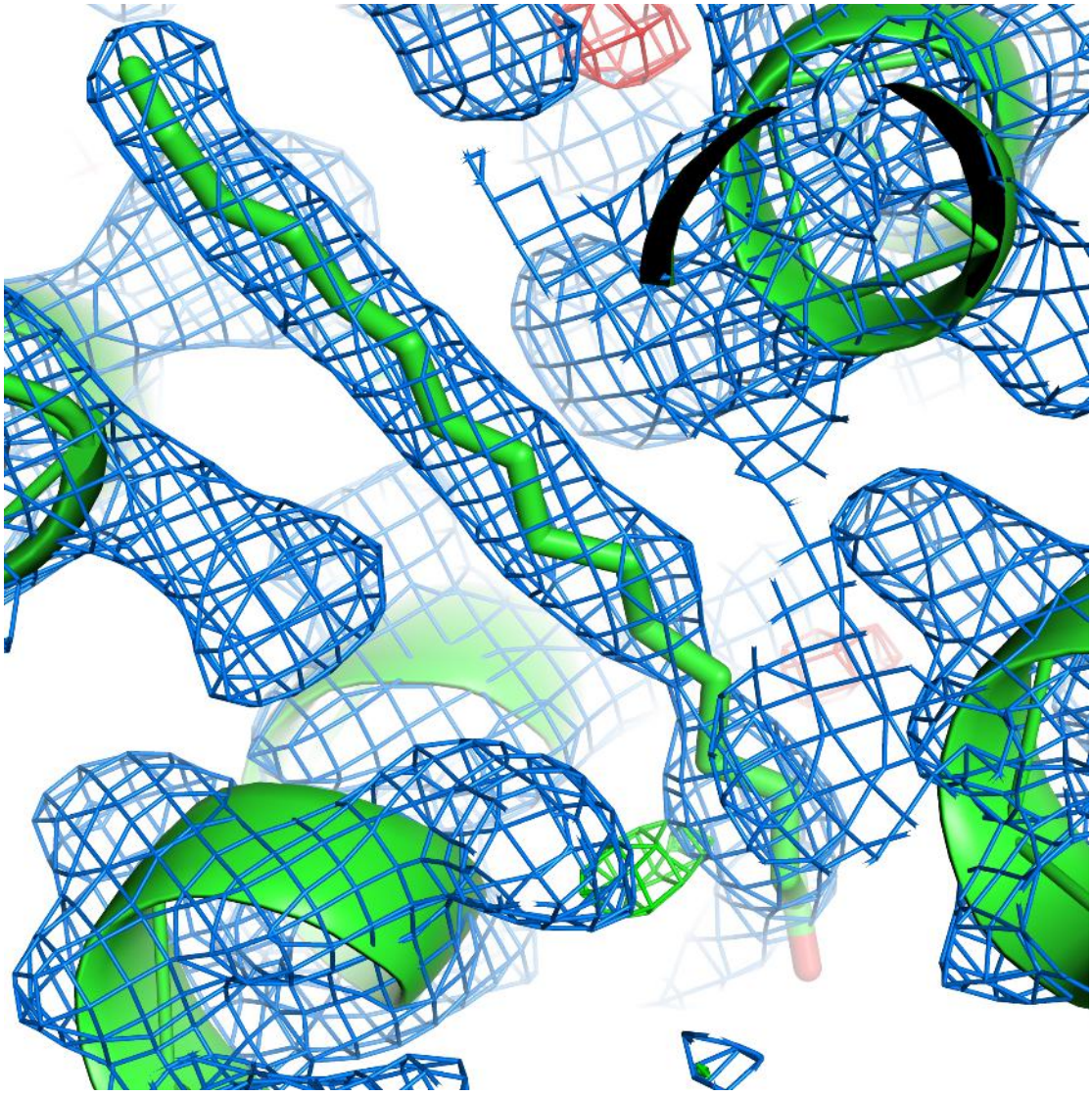


Figure 70 Electron density map of *S. typhimurium* LapB refined with lipid substrate

After refinement of the *S. typhimurium* LapB native data using the *S. typhimurium* LapB structure with palmitoleic acid substrate, the palmitoleic acid chain match the electron density well. And the extra electron density in Fo-Fc map (panel d in Figure 69) before refinement disappeared, suggesting that C-16 lipid may be the substrate. Due to the 2-fold symmetry of the LapB dimer

structure, it is hard for LapB to discriminate between the carboxylic acid end and the carbon chain end. This is also in consistence with our structure, there is no significant electron density difference between two ends of the lipid substrate region. And the direction of the lipid substrate in our structure is representing one possibility and there may be an opposite position of the lipid within the crystals. The image shows both the 2Fo-Fc and the Fo-Fc map. 2Fo-Fc map is coloured in blue and contoured at 1.0σ. Fo-Fc map is coloured in green/red and contoured at 3.0σ. The Fo-Fc map before refinement with the substrate is shown in (d) in Figure 69.

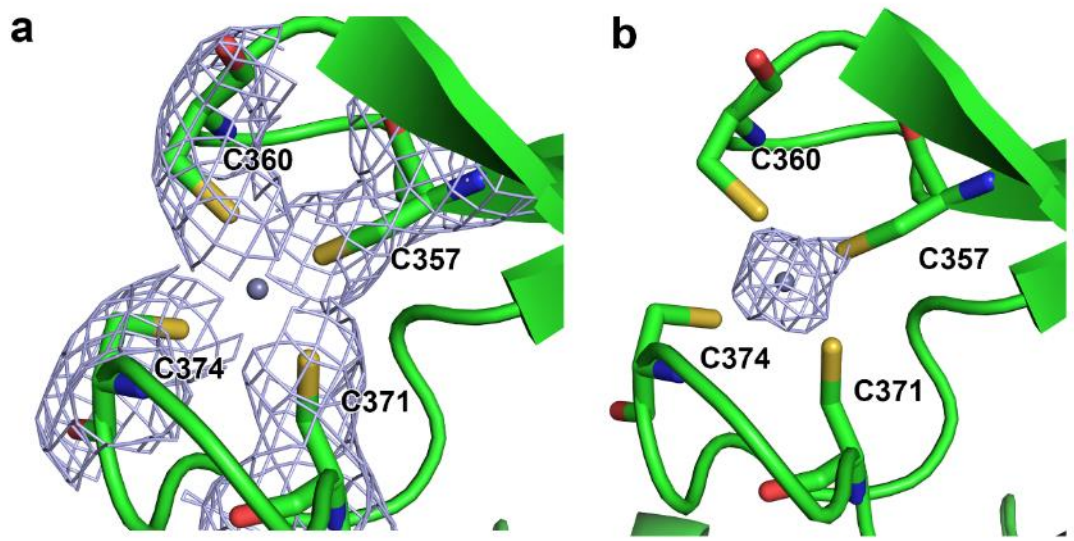


Figure 71 *S. typhimurium* LapB contains Zn^{2+} in the C-terminus

S. typhimurium LapB contains a zinc ion at the C-terminal domain and the zinc ion is captured by four cysteine residues. (a) The zinc binding site is shown. The *S. typhimurium* LapB is shown is cartoon labelled in green colour. The four cysteine residues are highlighted in sticks. 2Fo-Fc map contoured at 1.5σ shows the electron density at this zinc binding site. (b) 2Fo-Fc map contoured at 8.0σ still contains signal at the zinc site, suggesting that there is a zinc ion at this position. Zinc atom and the coordinating cysteine residues are shown in ball-and-stick model [O atom (red), N atom (blue), S atom (yellow), Zn atom (slate)].

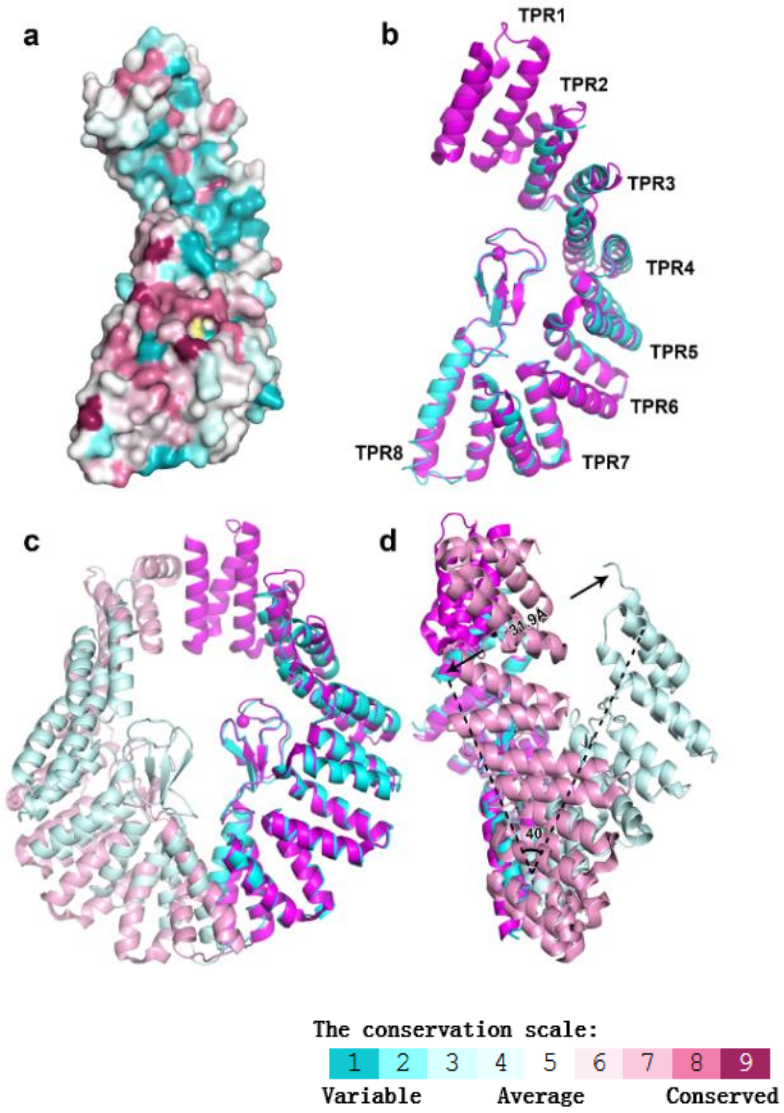


Figure 72 Superimpostion result of *S. typhimurium* LapB and *E.coli* LapB (4ZLH)

(a) Conserved residues surface map of one protomer of *S. typhimurium* LapB cytoplasmic domain. The online program Consurf (206) was used to calculate the amino acid conservation and PyMol (204) is used to display the conservation map following the Consuf online instructions. (b) Superimposition of one protomer of LapB from *E. coli* (4ZLH) and *S. typhimurium* (this study) with the RMSD of 1.51Å over 243 aligned residues. *E. coli* LapB contains eight TPRs, while *S. typhimurium* LapB has six, with two TPRs cleaved during proteolysis. There are few conformational changes when comparing one protomer between *E. coli* and *S. typhimurium* (c) Front view of the superimposition result of dimer LapB from *E. coli* and *S. typhimurium. S.*

typhimurium LapB is superimposed with one protomer of *E. coli LapB* with RMSD of 1.51Å over 243 aligned residues. The comparision is based on the dimer LapB (d) Side view of superimposition of dimer LapB from *E. coli* and *S. typhimurium*. It shows that the second protomer of the dimeric LapB from *E.coli* is rotationally shifted for about 40° and 31.9Å at the widest point compared to that from *S. typhimurium*.

3.4 CONCLUSION AND DISCUSSION

Gram-negative bacterial cell envelope is made up of a double membrane system. Within the two membranes, the OM, an asymmetric membrane, is essential for cell viability and provide drug resistance. The asymmetry of the OM is caused by the distinct difference of inner leaflet and outer leaflet. While the GPL composes the inner leaflet, the outer leaflet comprises of lipopolysaccharide (LPS) which is regarded as the source of the relative impermeability (4,5). The biogenesis of LPS is always a hot topic throughout these years and a promising target for novel drug development. LPS contains three components, lipid A, core oligosaccharide and O-antigen. Precursors of LPS are synthesized at the inner leaflet of the IM and matured at the outer leaflet of the IM into LPS. The synthesis of lipid A is a complicated multi-enzyme process (Figure 50). In this biosynthesis process, LpxC acts as a balancer between the biogenesis of GPL and LPS, both of which consume lipids, partially shared ingredients. Through the degradation of LpxC by FtsH, a degradation product from LpxC (250) is added onto the intermediate lipid A completing an essential step towards mature lipid A (Figure 50) (5).

This chapter reveals the crystal structure of the cytoplasmic domain of IM protein LapB that may participates in the LpxC feedback balancing. *S. typhimurium* LapB is a dimer in both the solution and crystal forms. LapB contains an N-terminal domain and a C-terminal domain. N-terminal is a TPR domain that may help LapB interact with multi proteins. C-terminal is a zinc binding domain with four cysteine residues locking a zinc ion in site. Our *S. typhimurium* LapB crystal structure revealed a new lipid bound form

compared to the *E. coli* LapB (4ZLH) (249). We have superimpose *S. typhimurium* LapB to one protomer of *E. coli* LapB. Structurally, the TPR region of the other protomer of the dimer *S. typhimurium* LapB is rotationally shifted by around 31.9 Å and 40 °compared to 4ZLH, opening the hydrophobic cavity between the dimer to hold a lipid-like substrate (Figure 72). The structural conformational changes indicate the motion cycle of LapB when anchored in the IM facing cytoplasmic side (244). Based on the two conformation of LapB structure, we anticipate that LapB could intake and release a lipid substrate when the LapB dimer is anchored in the IM inner leaflet. Through the TPR domain, LapB may interact with LPS synthesis or transport related proteins (i.e. LpxC or $LptB_2FG$) and control the donation of lipid substrate of LapB. We speculate that the conformation changes would be caused by the passive substrate intake of LapB. Notably, the lipid within the LapB has not been identified yet. Since that LapB soluble domain is expressed in the cytoplasm but not expressed as IM anchored protein, where the LapB may interact with related proteins to ensure its functionality. LapB soluble domain in cytoplasm may not be associated with the interaction proteins, so that the lipid it captures inside might be something similar rather than its biofunctional substrate. Considering the participation of LapB on LPS biosynthesis, the unidentified lipid could possibly be some precursor or products of LpxC related steps. C-12-ACP and C-14-ACP are donor of lipids of the six fatty acid chains of the lipid A. It is quite possible for the unassigned lipid to be a C-12, C-14 lipids or their precursors. C-16-ACP is also a lipid donor in the modification step of lipid A, so it is also possible for the unassigned density to be a C-16 lipid or the precursor. Another potential evidence is that C-14-ACP is also the donor of the lipids in GPL synthesis, the nature that LPS and GPL synthesis share C-14-ACP as a precursor indicates that LapB could act as a balancer between LPS and GPL synthesis (243).

Due to the 2-fold rotational symmetry in LapB dimer structure, the LapB protein itself could not discriminate between the carboxylic acid terminal and the carbon chain terminal. It is difficult to determine the exact length and the component of the lipid substrate from the crystal structure (the uncertainty of the direction of the lipids may blur the additional electron density of the lipid substrate). Based on the electron density

map, we modelled a palmitoleic acid molecule into the density, but it does not reflect the corresponding or functional substrate. Mass spectrum analysis on the purified full length inner membrane anchored LapB may answer this question better.

Our attempts on setting up the functional assay in LapB deletion system, unfortunately, failed. The LapB knockout strain JW1272-3 (#9144) (251), gained from CGSG *E. coli* strain collection, was confirmed to be conditionally lethal under 100 μg/ml vancomycin, while wild type BW25113 stain could still survive under the lethal conditions. This phenotype is consistent with that it participates in the outer membrane biosynthesis, and that outer membrane is an important barrier for hydrophobic antibiotics. To rescue the phenotype of *lapB* deletion strain, I generated two plasmids, pBAD24-LapB/pBAD24-LapB-LpxC. However the LapB deletion strain transformed with neither of these two plasmids failed to survive under any of the 0.0002%-0.02% arabinose concentration. This result could be explained by the LapB & LpxC feedback 'regulation'. The level of the native LapB is strictly controlled *in vivo* and the expression level of pBAD24-LapB could not be replaced *in vivo* but relied on the arabinose concentration. As a result, the LpxC-relied lipid A synthesis balance is broken, thus lead to the cell death. The plasmid pBAD24-LapB-LpxC, as described above, failed to rescue the phenotype again, even attempting to stabilize the LpxC level by co-expression. Additional experiments need to be performed to study into the mechanism of LPS synthesis controlling by LapB. An improved plasmid targeting at rescuing the conditional death phenotype has been designed as well: inserting *LapB* gene fragment together with its own upstream promoter regulation sequence into a low copy vector or even together with all the genes inside its operon. The LapB expression level in this modified plasmids should probably reach native level and can rescue the phenotype. In that case, site direct mutations and Western blotting can be then introduced to explore the detailed mechanism of LapB regulation in LPS synthesis. Additionally, the *lapB* deletion strain showed cell growth deficiency which can be observed under microscope. The intermediate LPS would also accumulate inside the cells when *lapB* is deleted. Experimental targeting in those directions would also illuminate some of the LapB participation pathway on LPS synthesis.

4 CHAPTER 4 Cardiolipin inner membrane transporter PbgA

4.1 INTRODUCTION

As discussed in the previous Chapters, Gram-negative bacteria contains two membranes, the inner membrane (IM) and the outer membrane (OM). Between these two membranes, there is a solvent filled periplasmic space and peptidoglycan. Comparing Gram-negative and Gram-positive bacteria, the peptidoglycan in the Gram-positive case is much thicker, however in Gram-negative bacteria the peptidoglycan binds to the OM by lipoproteins and contributes to the protection of the bacteria (207,252). The OM is an asymmetric membrane, in which the inner leaflet is composed of glycerophospholipids (GPL) while the outer leaflet is composed of lipopolysaccharide (LPS). The biosynthesis regulation, and transport of LPS has be described in the previous Chapters. In this Chapter, the mechanism of the transportation of GPL from IM to OM is studied.

The glycerophospholipids (GPL) molecules consist of phosphatidylethanolamine (PE), phosphatidylglycerols (PG) and cardiolipin (CL) in an approximately 75: 20: 5 ratio in most bacterial membranes. Phospholipids are synthesised at the inner leaflet of the IM (25,26). It is not clear whether a flippase is required, while LPS and lipid A have been confirmed to be flipped by MsbA. In species like *N. meningitides*, MsbA deletion strain appears to be viable and still produce a double membrane system, showing phosphoplipid transport does not require MsbA at least in this strain (253). Moreover, various α-helical IM proteins may induce phospholipids translocation, which means that phospholipids flipping may not require a specific transporter but a range of α-helical IM proteins (254).

CL is also known as diphosphatidyglycerol because structurally it appears to contain two overlapping PG units. CL is also located in the OM of most Gram-negative bacteria under the regulation of the PhoPQ two-component system (255,256). PhoPQ is a protein complex that regulates the OM remodelling in *S. typhimurium*. *E. coli* and S. *typhimurium* rely on the PhoPQ regulators to active genes and downstream pathways

that function to increase OM hydrophobicity and decrease OM negative charge, to protect the bacteria (255) .It is also essential for bacterial pathogenesis and survival in various environments. However, CL is synthesized at the IM, which means that there should be a pathway for transporting CL from the IM to the OM (257,258). CL contains a tetra-acylated dimeric structure which allow cardiolipin to self-associate and form non-bilayer hexagonal phases (259). Some bacterial proteins bind CL for membrane fission during sporulation and cytoskeletal arrangement during cell division (260).

Interestingly, CL is also present in the IM of mitochondria, where it occupies 20% of the total lipids and is required for the normal function of mitochondria (261,262). In yeast, the UPS1-Mdm35 complex has been identified to be responsible for transporting phosphatidicacid (PA) from the mitochondria outer membrane to inner membrane (263,264). In Gram-negative bacteria the Mla pathway (MlaA-F) has been speculated to be able to transport phospholipids form the OM to the IM (265). According to a report in 2015, in *S. typhimurium*, an inner membrane protein named PbgA (*yejM* gene product) was identified, which binds CL to promote PhoPQ-regulated trafficking of CL from the IM to the OM (256). It has also been reported that purified PbgA oligomers contain five transmembrane helices and a globular domain which binds to the OM in a PhoPQ-dependent manner (256).

However, it remains unknown how PbgA interacts with cardiolipin and whether PbgA can bridge the envelope for regulated cardiolipin delivery. In this chapter, we will describe the successful cloning of PbgA and the PbgA globular domain. Furthermore, the first crystal structure determination of PbgA and related functional assays of PbgA are also included.

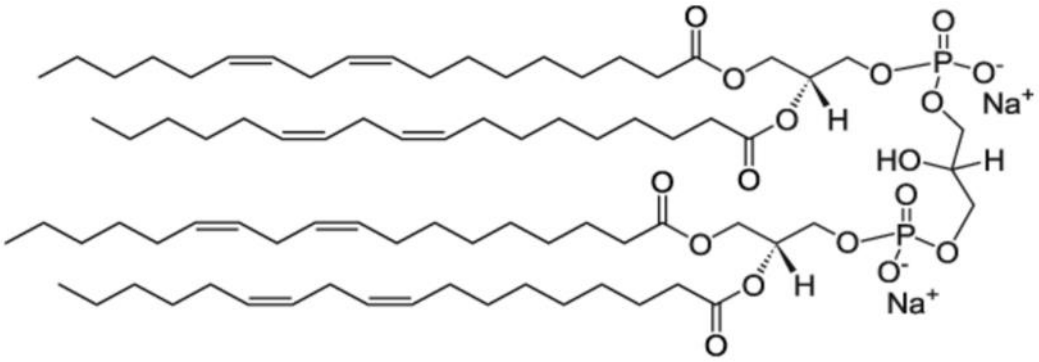


Figure 73 Cardiolipin sodium salt from bovine heart.

Cardiolipin sodium salt from bovine heart is purchased from Sigma Aldrich (#C0563) and is used as the experimental substrate of PbgA in this chapter.

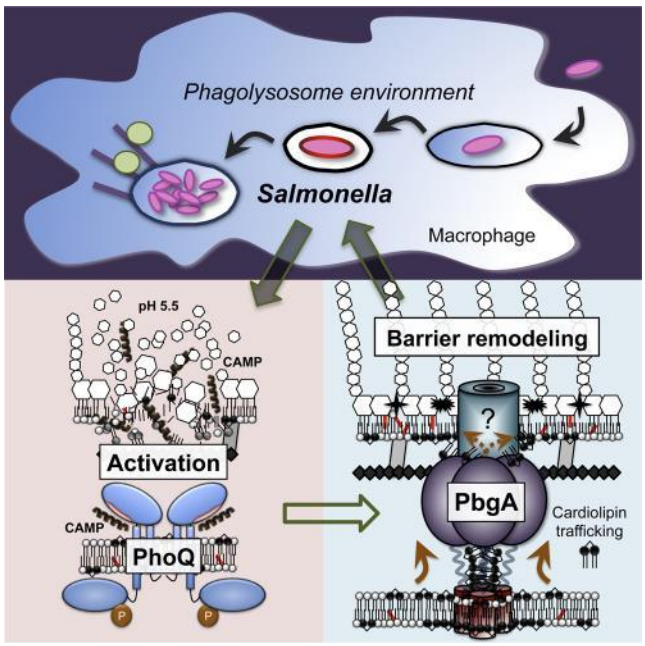


Figure 74 PbgA binds to the OM in a PhoPQ-dependent manner. (256)

A working model for how *S. typhimurium* regulate cardiolipin trafficking to the OM. Host environments that activate PhoPQ (left panel) induce PbgA to bind the OM and deliver cardiolipin that promote the barrier necessary for bacteria survival and replication within host tissues. When PhoPQ are not activated (right panel), PbgA is mostly unbound to the OM and OM-cardiolipin levels remain at baseline. (255)

4.2 METHODS AND MATERIALS

4.2.1 Generation of PbgA expression plasmids from *S. typhimurium* and *E. coli*

S. typhimurium and *E. coli ycjM* gene fragments from these two species encoding truncated PbgA protein from 191-586 and 245-586, respectively were amplified by PCR with restriction sites 5`-end NcoI and 3`-end EcoRI . The fragments were inserted into the same restriction enzymes treated plasmid pHISTEV which contains a 6×Histag and a TEV protease cleavage site between 6×Histag and the N-terminus of the cloned genes. Four plasmids encoding for the PbgA globular domain were generated named *Sal*-PbgA$_{245\text{-}586}$ *Sal*-PbgA$_{191\text{-}586}$ *E.coli*-Pbga$_{245\text{-}586}$ *E.coli*-PbgA$_{191\text{-}586}$ respectively. The full length PbgA gene from *S. typhimurium* and *E.coli* were inserted into pBAD24 in the same way, generating *Sal*-PbgA and *E.coli*-PbgA for full-length PbgA expression. All plasmids were sequenced and found to be correct.

4.2.2 Expression and purification of PbgA globular domain

All the different plasmids of the PbgA globular domain were transformed into *SoluBL21 (DE3)* (Invitrogen) and a single colony was picked, and inoculated, stored as glycerol stocks. The glycerol stock (*SoluBL21* harbouring *Sal*-PbgA$_{245\text{-}586}$, *Sal*-Pbga$_{191\text{-}586}$, *E.coli*-Pbga$_{245\text{-}586}$ or *E.coli*-PbgA$_{191\text{-}586}$) was inoculated into 500 ml LB overnight culture with 50 μg/ml kanamycin antibiotics. 12L LB medium were inoculated by 1:200 of overnight cell culture with 50 μg/ml kanamycin. The cells were cultured in a 37℃ shaker at 200 rpm until the OD_{600} reached 0.8-1.0 and the protein expression was induced by adding IPTG at 0.1 mM final concentration for 20 hours at 20℃ and 200 rpm shaking overnight. 12L of cell culture were then harvested into pellet by centrifugation at 5,000rpm (4,300g) for 15 min. Followed by re-suspending the cell pellet in Balance buffer (20 mM Tris-Cl, pH 7.8, and 300 mM NaCl 10mM imidazole 10% glycerol, cOmplete EDTA-free Protease Inhibitor Cocktail Tablet (Roche) 50 μg/ml DNAse I (Sigma) and 100 μg/ml lysozyme). The cell re-suspension were lysed by passing a cell disruptor at 30,000 psi. The unbroken cells and cell debris were

removed by centrifugation at 20,000 rpm (40,000g) for 30 min. The supernatant was applied to pre-balanced (with Balance buffer) 5ml HP HisTrap (Ge Healthcare). The column was washed using Wash buffer which is prepared by increasing imidazole concentration of the Balance buffer to 30mM, and then the PbgA protein was eluted by Elution buffer (20 mM Tris-Cl, pH 7.8, and 300 mM NaCl, 300 mM imidazole and 10% glycerol). The protein sample was desalted using a Balance buffer pre-balanced desalting column (GE healthcare) to remove imidazole. Subsequently, TEV protease was added into the desalted protein at a TEV protease to protein mass ratio 1:100 and incubated overnight to cleave the 6×Histag off the fusion protein at the TEV cleavage site. After cleavage, PbgA protein sample was loaded into the pre-balanced 5ml HP HisTrap to remove the TEV protease (TEV protease contains a 6×Histag) and the cleaved 6×Histag. During this step PbgA globular domain would directly go through the column into a collection tube and be concentrated below 10 ml. The PbgA protein was further purified by gel-filtration using a pre-equilibrated SEC (HILOAD 16/600 SUPERDEX 200 PG, GE Healthcare) with (150 mM NaCl 20 mM Tris pH 7.8). The gel-filtration fractions were analysed using SDS-PAGE and the most pure fractions were pulled and concentrated to 5-10 mg/ml for crystallization trails.

4.2.3 Selenomethionine labelling of *Sal*-PbgA$_{245-586}$

The selenomethionine mediums and related solutions are listed in Appendix 1 (247). The glycerol stock (*SoluBL21* harbouring *Sal*-PbgA$_{245-586}$) was inoculated into 500ml LB overnight culture with 50 μg/ml kanamycin. 500 ml cell culture was harvested under sterile conditions. The pellet was washed with sterile PBS 2 times and re-suspended into 200 ml PBS for subsequent cultures. In each 1L M9 medium 50 μg/ml kanamycin, 50 ml Selenomethionine Medium Nutrient Mix solution and Additional solution were added before inoculation of 20 ml PBS re-suspended overnight culture. The cells were cultured in 37℃ shaker at 200 rpm until the OD_{600} reached 0.5 just in the same way as for the native protein and then 10ml SeMet inhibitor amino acid solution was added to

each 1 L culture. After 15 min incubation, 1 ml Selenomethionine solution was added to each litre of culture and waited for another 15 min before *Sal*-PbgA$_{245-586}$ expression was induced by adding IPTG at final concentration at 0.1 mM for 20 hour at 20 ℃ and 200 rpm shaking overnight. The rest steps followed the native PbgA$_{245-586}$ purification method.

4.2.4 Full-length PbgA expression and purification

The plasmid pBAD24-*PbgA* encoding full length PbgA was transformed into *C43 (DE3)* strain (Novagen) for protein expression. The glycerol stock was first inoculated into 500 ml LB overnight culture supplemented with antibiotic (ampicillin 100 μg/ml). Then the bacterial cell culture were amplified into 12 L LB supplemented with ampicillin 100 μg/ml at 37 ℃ until the OD600 reached 0.6. Then PbgA expression was induced with 0.02% L-arabinose for 16 h at 20 ℃. Cells were harvested into a pellet by centrifugation at 5,000rpm (4,300g) the next day and the pellet was re-suspended in 20 mM Tris-Cl, pH 7.8, and 150 mM NaCl supplemented with cOmplete EDTA-free protease inhibitor tablet (Roche), 1 mM DNaseI (Sigma-Aldrich) 100 μg/ml lysosome. The cells were broken by passing twice through a cell disrupter at 30,000 psi (Constant Systems Ltd). The cell debris was removed by centrifugation at 18,000 g for 15 min at 4 ℃. The cell membrane was collected by ultracentrifugation at 100,000 g for 1 h at 4 ℃. The membrane fraction was solubilized with Extraction Buffer (20 mM Tris-Cl, pH 7.8, and 300 mM NaCl 10 mM imidazole) supplemented with 1% (w/v) n-Dodecyl-β-D-Maltopyranoside (DDM) (Anatrace) supplemented with cOmplete (Roche) at room temperature for 20 min. The suspension was then ultracentrifuged at 100,000 g for 30 min before loaded onto a 5ml HisTrap HP column (GE HealthCare) and washed with Wash Buffer (20 mM Tris-Cl pH 7.8 300 mM NaCl 60 mM imidazole) supplemented with 0.05% (w/v) DDM. The PbgA full length protein was eluted with Elution Buffer (20 mM Tris-Cl, pH 7.8, 300 mM NaCl and 300 mM imidazole) supplemented with 0.05% (w/v) DDM. The protein was further purified using size

exclusion chromatography with a HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare) in 20 mM Tris-Cl, pH 7.8, 150 mM NaCl and 0.05% (w/v) DDM. Protein fraction of the highest purity were collected and concentrated to 10-30 mg/ml.

4.2.5 Crystallization and data collection of PbgA globular domain

Protein crystallization trails were performed using 1 µl protein plus 1 µl reservoir solution by the sitting-drop vapour diffusion at room temperature using a Gryphon robot. The screening solutions were purchased as commercial membrane protein screens (Table 10). Optimization were prepared for several top quality crystals. The crystals were harvested after 21 days and cryoprotected before flash freezing by liquid nitrogen. The cryoprotectants were prepared by replacing 20% of water with glycerol in final crystallization condition. *Sal*-PbgA$_{245-586}$ native and Se-Met incorporated crystals were obtained from 0.1M Bis-tris pH 5.5 and 16% PEG3350. *E.coli*-PbgA$_{245-586}$ crystals were obtained from 0.2M lithium sulphate, 0.1M sodium acetate pH 4.3 and 48% PEG400. Sal-PbgA$_{191-586}$ crystals were obtained in 0.1M Bis-tris pH 6.5 and 25% PEG3350.

Native data were collected at Diamond Light Source, UK at beam station I04. All data were processed by the XDS pipeline at beam centre automatically. Selenomethionine derivation SAD data was collected at I02 using the peak wavelength resulting from a fluorescence scan of the selenomethionine-labelled sample.

4.2.6 Crystallization of full length PbgA and unexpected AcrB crystal

Unexpected crystal structure of *acrB* gene product was determined during our trails of crystallization the full length PbgA (266). During our purification and crystallization of PbgA full length experiments, we have obtained a small triangle shaped crystal and diffracted up to 4.2 Å in the initial screening. After optimizing, the crystals become bigger and diffracted up to 3.5 Å with an impressive unit cell dimensions, of which one

axis is over 550 Å. The Matthew cell content analysis of CCP4i shows that there should be at least three PbgA full length protomers in the asymmetric unit. However, we could not just ignore the unit cell dimensions, so we have put the crystal parameter into the nearest-cell (267) online program. The result contained several structure files with nearly exact unit cell parameters. Unfortunately, this crystal was AcrB, the 114kD multidrug resistance bacterial IM protein. We did not continued on PbgA full length inner membrane protein because there were no more crystal hits besides this one. It is a common difficulty to perfectly purify a bacterial inner membrane protein ($LptB_2FG$ complex is also not perfect pure as well). But we still didn`t expect such a faint bands on SDS-PAGE at top of lane 9 in Figure 75 would result in such a high resolution crystal.

4.2.7 Structure determination of PbgA globular domain

Selenomethionine anomalous SAD data of *S. typhimurium* was processed and scaled with XIA2 (144) automate pipeline. According to the Cell content analysis in CCP4i, the asymmetric unit may contain 3-4 PbgA molecules. The substructure and experimental phases of *Sal*-$PbgA_{245-586}$ were solved using the SHELXC/D/E suite (148). SHELX C generates column labels of the SAD data making preparation for SHELX D, which SHELX D determine the substructure of the SAD data. Finally density modification was performed by SHELX E to generate an initial electron density map. 35 selenomethionine sites were found by SHELX suit. According to the SHELX graphic figures, the inverted hand is correct. So, we selected the inverted reflection output file (.mtz file) for model building. The initial model of the *Sal*-$PbgA_{245-586}$ was automatically built using the program BUCCANEER (154) from CCP4i on the initial map. After several rounds of refinement by REFMAC5 (159), the initial PbgA structure was solved automatically, with R_{free}=0.35. In the asymmetric unit there are four PbgA molecules. Then we changed the MSE residue to MET residue in the coordinates file and save single PbgA as a separate coordinate file to generate a model for molecular

replacement to solve the native *S typhimurium* data and the *E. coli* PbgA. All the structures were finally refined by REFMAC5, improved by COOT and validated by Molprobilty (268),

The crystal structure of $PbgA_{191-586}$ belonged to the space group $P2_1$ with cell dimensions a = 51.48, b =195.65, c = 70.92, α = γ = 90°, β = 96.39°. The structure of *Sal*-$PbgA_{191-586}$ was determined by molecular replacement with a final R_{free} = 0.27 (Table 9). However according to the structure, there was no visible electron density for loop regions from residue 191 to residues 586. Similarly, there were 4 protomers within the asymmetric unit. The protomer A is superimposed well with protomers B, C and D with RMSD of 0.2642Å, 0.4338Å and 0.5627Å over 340 Cα atoms, respectively.

The crystals of *E. coli* $PbgA_{245-586}$ belonged to the space group $P3_12_1$ with cell dimensions a = b = 57.42 Å, c = 191.31 Å, α = β = 90°, γ = 120 °Table 9. This structure was also determined by molecular replacement. However there was only one monomer within the asymmetric unit rather than four. These three structures above were strikingly similar to each other with only limited conformational changes was observed after superimposition of each other. The *Sal*-$PbgA_{191-586}$ structure is very similar to *Sal*-$PbgA_{245-586}$ structure with RMSD of 0.5945 Å over 340 Cα atoms. The *E.coli*-$PbgA_{245-586}$ structure resembles Sal-$PbgA_{254-586}$ structure with a RMSD of 0.9264Å over 340 Cα atoms. In particular, the β4 of *E.coli*-$PbgA_{245-586}$ moves away approximately 8Å

4.2.8 Site direct mutagenesis functional assay in PbgA deletion strain

The NRD183 strain, used as PbgA deletion strain in this Chapter, contained kanamycin resistance gene during *ycjM* knockout generation and harboured a pBAD origin rescue plasmid pNR217 (Amp^R) at the same time. The NRD183 strain yet contained the part of the *ycjM* gene that encode the transmembrane domain of PbgA while only the part of gene, that encoded the periplasmic domain of PbgA, was deleted resulting in an essential transmembrane domain remained genotype. In fact, the NRD183 strain was an *ycjM* partial deletion strain and conditional lethal under 50 μg/ml vancomycin.

As for the functional assays, a leaky expression vector encoding full length *E. coli* PbgA was cloned into the pACYC-Duet plasmid (Novagen) using the same method as other gene cloning 4.2.1. This plasmid was used as the template to generate all the rest single or double mutations and used as positive control in the following functional assay experiment. All site direct mutations were generated following Liu`s protocol (143) and confirmed by sequencing.

All mutations, positive control and negative control (empty pACYC-Duet plasmid) were transformed into NRD183 strain respectively. The transformed cells were grown on LB agar plate supplemented with 0.2% L-arabinose and antibiotics (kanamycin 50 μg/ml, chloramphenicol 34 μg/ml, ampicillin 50 μg/ml). Single colony was picked into 5 ml overnight culture supplemented with same component as above for functional assay streaking. In functional assay analysis, cells were harvested, washed and diluted in sterile LB medium to OD_{600} = 0.5 and streak onto LB agar plates supplemented with 50 μg/ml vancomycin, 50 μg/ml, chloramphenicol 34 μg/ml and 5 mM IPTG. Cell growth was recorded after overnight incubation at 37 ℃. All the assays were repeated in triplicate.

4.2.9 PbgA-cardiolipin co-sedimentation assays

CL from sigma (Cat#: C0563) was first dissolved in chloroform using a glass tube at 500 mM concentration and dried up again under nitrogen gas flow, which protected it from being oxidized by O_2 from air. Then the fine powder was re-suspended in 20 mM Tris (pH 7.8) and 150 mM NaCl by sonication until solution became almost clear. After the cardiolipin aqueous solution been prepared, *Sal*-$PbgA_{245-586}$ was mixed with cardiolipin solution at 1:10 molar ratio (2 mM: 20 mM final concentration respectively) at 37 ℃ for 1 hour. After that the mixture was centrifuged at 13,200rpm (12,000g) for 15min until the supernatant looks clear without turbidity, the supernatants were pipetted out into an tube and the precipitants were re-suspended in 20 μl 20 mM Tris (pH 7.8) 150 mM NaCl. To each sample 4×SDS-Loading buffer was added until fully dissolved

and proteins were separated on SDS-PAGE after 10 min boiling at 95 ℃.

4.3 Results

4.3.1 Expression and purification of PbgA globular domain and full length PbgA

Both the *E. coli* and *S. typhimurium* PbgA globular domains ($PbgA_{245-586}$ & $PbgA_{191-586}$) have been successfully purified (Figure 75). The $PbgA_{191-586}$ protein degraded easier compared to $PbgA_{245-586}$ protein, which suggests that the loop region between residues 191 and 245 could be quite flexible and might affect crystallization.

PbgA is located in Gram-negative bacteria with the globular domain extending into the periplasm. The full length PbgA was purified from inner membrane manner, using detergent DDM to solubilize and stabilize PbgA (Figure 75). Both the full length *E. coli* and *S. typhimurium* PbgA were successfully expressed and purified (only *S. typhimurium* PbgA SDS-PAGE was presented here).

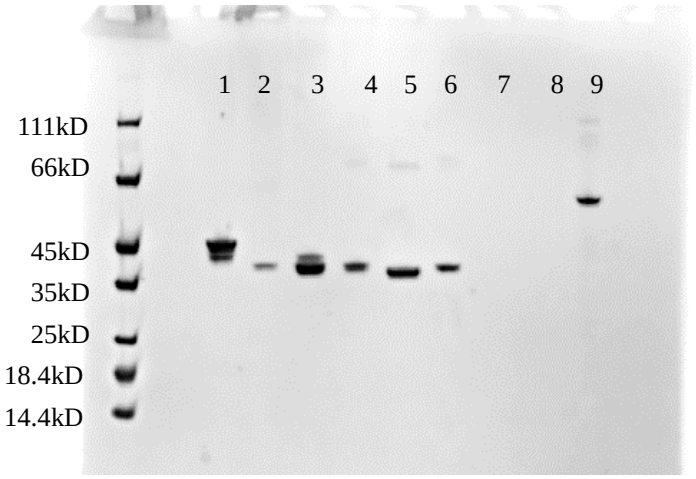


Figure 75 SDS-PAGE analysis of various PbgA samples

Various PbgA samples were analyzed on SDS-PAGE to verify the protein expressions. Lane 1: Purified *Sal*-PbgA191-586 protein; lane 2: Crystals of *Sal*-PbgA191-586; lane 3: Elastase digested purified *Sal*-PbgA191-586 protein; lane 4: Crystals of Elastase digested *Sal*-PbgA191-586; lane 5: Purified *Sal*-

PbgA245-586 protein; crystals of *Sal*-PbgA245-586; lane 9: Purified full length *Sal*-PbgA. All crystals from this SDS-PAGE analysis had been washed with reservoir buffer three times by immersing crystals inside for 1 min. 5µl samples were mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into SDS-PAGE gel after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature. Bolt™ 4-12% Bis-Tris Plus Gels, 12/15-well (Thermo Fisher Scientific)

4.3.2 Crystallization and optimization of the PbgA globular domain

Both the *E. coli* and *S. typhimurium* PbgA globular domains were crystalized using the sitting drop diffusion method by crystallization robot Gryphon. Lots of crystals were grown from various conditions of different PbgA proteins. Most of the crystals diffracted well, with resolution up to 1.64 Å. We had tried extensively to optimize the crystals by varying crystallization conditions. Those crystals were reproducible and had been optimized with hand-made 5 ×5 optimization solutions to produce enough crystals. Selenomethionine labeled *Sal*-$PbgA_{245-586}$ protein crystallized in both native crystal conditions and other screen conditions. Sal-$PbgA_{245-586}$ native and Se-Met incorporated crystals were obtained from 0.1M Bis-tris pH 5.5 and 16% PEG3350. E.coli-PbgA245-586 crystals were obtained from 0.2M lithium sulphate, 0.1M sodium acetate pH 4.3 and 48% PEG400. Sal-$PbgA_{191-586}$ crystals were obtained in 0.1M Bis-tris pH 6.5 and 25% PEG3350.

4.3.3 Crystal structure determination of the PbgA globular domain

The structure of the *Sal*-PbgA245-586 was determined using SAD data from selenomethionine-incorporated crystals. The crystal belonged to space group $P2_1$ and diffracted to high resolution, 1.64 Å, (Table 9). There were four protomers of $PbgA_{245-}$

$_{586}$ in the asymmetric unit. The structure of each protomer in the asymmetric unit shared high similarity with the other three with only minor differences. The final model of the PbgA$_{245-586}$ contained residues Q245-I563 and P568-585, while no electron densities for the loop between I563-P568 was visible, indicating that this loop might be disordered. Overall, the structure of PbgA$_{245-586}$ shaped like a burger with an α/β fold.

The crystal structure of PbgA$_{191-586}$ belonged to the space group $P2_1$ with cell dimensions a = 51.48, b =195.65, c = 70.92, α = γ = 90°, β = 96.39°. The structure of *Sal*-PbgA$_{191-586}$ was determined by molecular replacement with a final R_{free} = 0.27 (Table 9. However according to the structure, there was no visible electron density for loop regions from residue 191 to residues 586. Similarly, there were 4 protomers within the asymmetric unit. The protomer A is superimposed well with protomers B, C and D with RMSD of 0.2642Å, 0.4338Å and 0.5627Å over 340 Cα atoms, respectively.

The crystals of *E. coli* PbgA$_{245-586}$ belonged to the space group $P3_12_1$ with cell dimensions a = b = 57.42, c = 191.31, α = β = 90°, γ = 120° Table 9. This structure was also determined by molecular replacement. However there was only one monomer within the asymmetric unit rather than four. These three structures above were strikingly similar to each other with only limited conformational changes was observed after superimposition of each other. The *Sal*-PbgA$_{191-586}$ structure is very similar to *Sal*-PbgA$_{245-586}$ structure with RMSD of 0.5945 Å over 340 Cα atoms. The *E. coli*-PbgA$_{245-586}$ structure resembles *Sal*-PbgA$_{254-586}$ structure with a RMSD of 0.9264Å over 340 Cα atoms. In particular, the β4 of E.coli-PbgA$_{245-586}$ moves away approximately 8Å.

Due to the insufficiency on PbgA functional studies (potential CL transporter), we tried to identify whether there are similar structures of PbgA for functional information. The Dali server (269) was used with *Sal*-PbgA$_{245-586}$ as initial model. According to the result, the PbgA globular domain have similarities to sulfatase family enzymes, where arlsulfatase (PDB code 1HDH (270)) has Dali-sore of 29.9. The sulfatases hydrolyze different sulfate esters, utilizing an active site consisting a metal binding site combined with several conserved and oxidized cysteine or serine residues. Although the overall structure of PbgA was quite similar to the sulfatase, superimposition showed that *Sal*-PbgA$_{245-586}$ does not contain a cysteine or serine at the enzymes' active site (Figure 78).

The Dali-server also suggested another structure with Dali score of 28.4 a lipoteichoic acid synthase from Straphylococcus aureus (LatS, PDB code 2W5T). Despite only 17% sequence identity to *Sal*-PbgA, LatS is also located in the inner membrane with five transmembrane domains and a globular soluble domain which was strikingly similar to *Sal*-PbgA$_{245-586}$ (Figure 79). The LatS structure contained an active site pocket for binding a Mn^{2+} ion and β-glycerol phosphate. However, according to the superimposition, the residues that make up the pocket of *Sal*-PbgA$_{245-586}$ was totally different from the LtaS, suggesting that it is impossible for *Sal*-PbgA$_{245-586}$ to bind metal ion or bind PG.

	E. coli-PbgA$_{245-586}$ Native	*Sal*-PbgA$_{191-586}$ Native	Se-Met *Sal*-PbgA$_{245-586}$ (SAD)
Data collection			
Space group	*$P3_12_1$*	*$P2_1$*	*$P2_1$*
Cell dimensions			
a, b, c (Å)	57.21, 57.42, 191.31	51.48, 195.65, 70.92	51.60, 196.09, 70.27
α, β, γ (°)	90.0, 90.0, 120	90.0, 90.0, 96.39	90.0, 90.0, 95.79
Wavelength (Å)	0.9795	0.9795	0.9795
Resolution (Å)	47.83-1.67 (1.71-1.67)	65.22-2.19 (2.25-2.19)	65.85-1.64 (1.68-1.64)
R_{merge} (%)	14.3 (26.40)	16.5 (27.80)	15.2 (25.03)
$CC_{1/2}$ (%)	100 (59.5)	98.7 (54.5)	99.9 (49.5)
I /σ(I)	17.8 (1.1)	5.8 (1.8)	15.1 (1.1)
Completeness (%)	99.6 (99.2)	99.4 (99.4)	95.1 (89.2)
Redundancy	9.6 (7.6-9.9)	6.1 (6.4-4.8)	11.8 (13.1-6.1)
Anomalous Completeness			
Anomalous Redundancy			
Phasing			
Resolution (Å)	47.83-1.67	65.22-2.19	65.85-1.64
Site (Se)			35
Figure of merit			0.672
Refinement			
Resolution (Å)	47.83-1.67	65.22-2.19	65.85-1.64
No. reflections	44003 (4034)	51747 (5044)	66831 (6540)
R_{factor} / R_{free}	0.18/0.21	0.20/0.24	0.19/0.22
No. atoms			
Protein	2604	10724	10745

Water	498	961	1243
B-factors			
Protein			
Ligand			
Solvent			
R.m.s.deviations			
Bond lengths (Å)	0.014	0.02	0.017
Bond angles (°)	1.90	1.93	1.80
Ramachandran statistics			
Favoured (%)	97	93	95
Allowed (%)	3	7	5
PDB code	5I5H	5I5F	5I5D

Table 9 Data collection and structure refinement statistics of PbgA globular domian.

R_{factor} = Σ|| Fobs|−| Fcal||/Σ|Fobs|, where Fobs and Fcal are observed all reflection measured and calculated currently model as structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly selected not used in refinement.

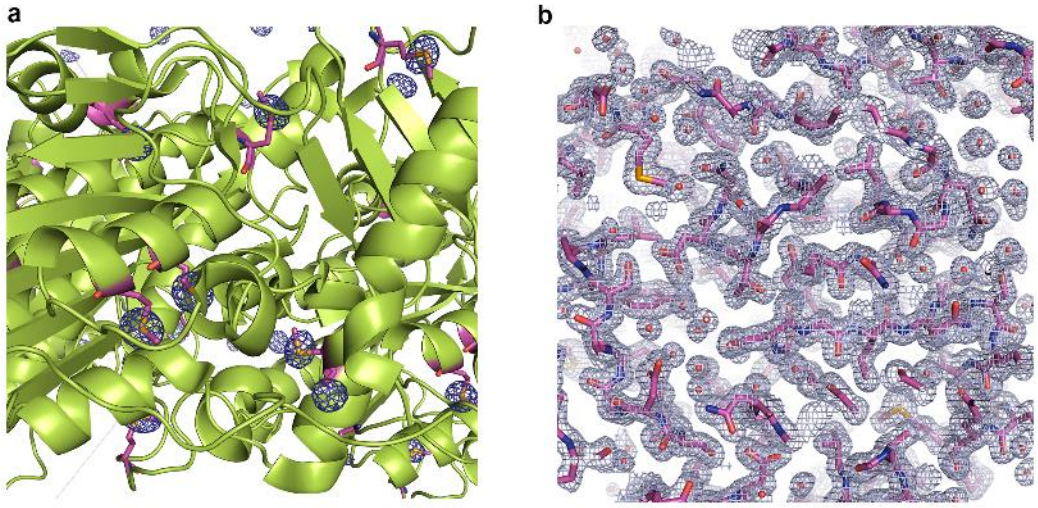


Figure 76 Amonalous difference Fourier map of selenomethionine deirved *Sal*-$PbgA_{245-586}$ and the 2Fo-Fc map of the 1.64 Å dataset.

(a) The SAD data of selenomethionine incorporated crystals were collected at 0.9795 Å. The anomalous difference Fourier map contoured at 8.0σ was calculated using PHENIX (155) with the anomalous dataset and the final Se-Met model. The peak position matches the Se atoms confirming the phasing is

accurate (b) 2Fo-Fc map of the 1.64 Å PbgA dataset is contoured at 1.0σ. PbgA structure is shown in ball-and-stick model [O atom (red), N atom (blue), S atom (yellow)].

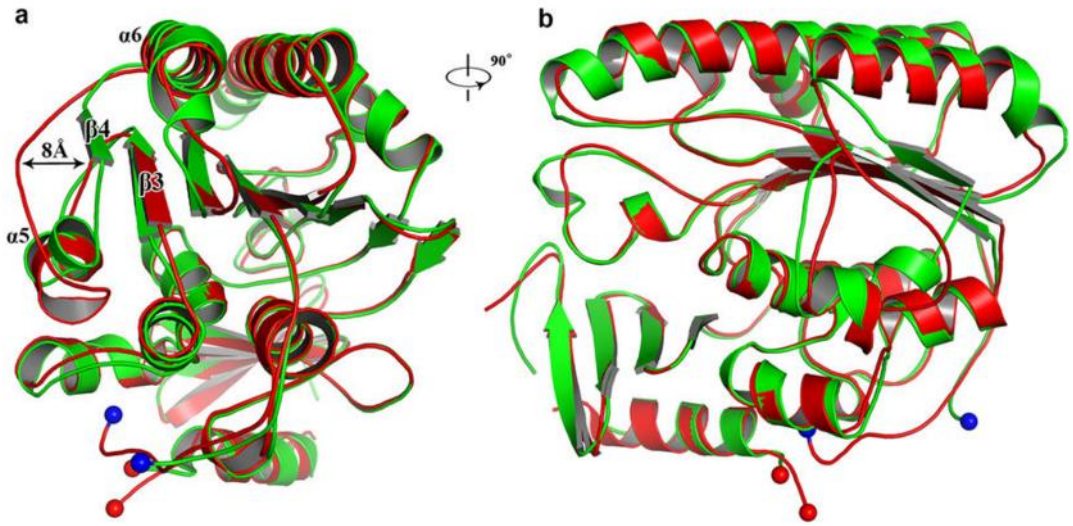


Figure 77 Superimposition of of structures of *Sal*-PbgA and *E. coli*-PbgA

(a) *Sal*-PbgA and *E. coli*-PbgA are superimposed well with each other. *Sal*-PbgA is coloured in red, and *E. coli*-PbgA is coloured in green. (b) Rotates 90° along y-axis relative to the left panel. Cα atoms of Loop S371-S360 moves away by a maximum of approximate 8Å from of *Sal*-PbgA to *E. coli*-PbgA. The *E. coli*-PbgA$_{245\text{-}586}$ structure resembles *Sal*-PbgA$_{254\text{-}586}$ structure with a RMSD of 0.9264Å over 340 Cα atoms.

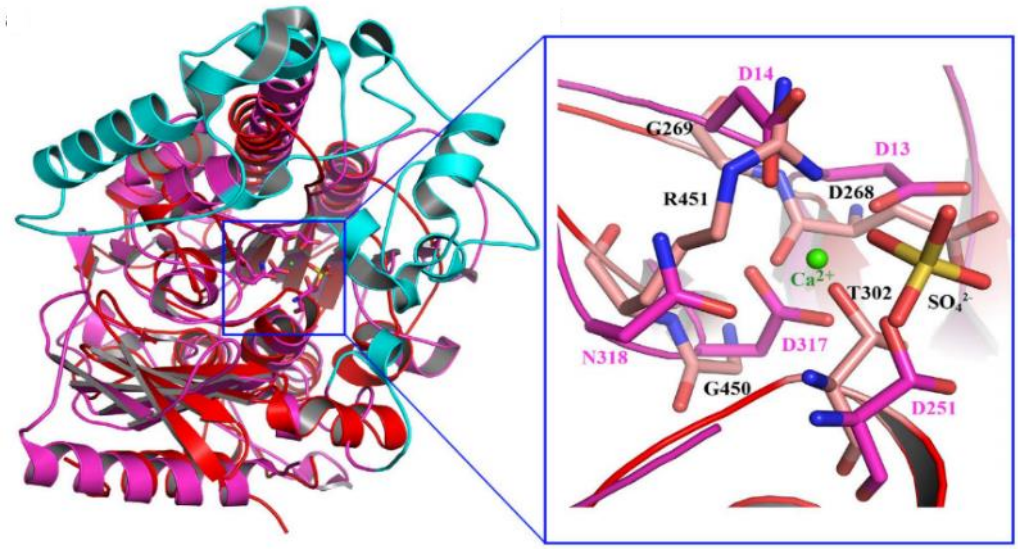


Figure 78 Superimposition of *Sal*-PbgA$_{245\text{-}586}$ with arylsulfatase

The Dali server showes that PbgA resembles the sulfatases that hydrolyse different sulphate esters, utilizing an active site consisting a metal binding site combined with several conserved and oxidized cysteine or serine residues. *Sal*-PbgA$_{245\text{-}586}$ (red) is superimposed to arylsulfatase (matanga) with RMSD of 2.9 Å over 296 Cα atoms. Although the overall structure of PbgA is quite similar to the sulfatase, superimposition shows that Sal-PbgA245-586 does not contain a cysteine or serine at the enzymes' active site. The arylsulfatase has additional α-helices and loops (blue) which might function in substrate binding. *Sal*-PbgA$_{245\text{-}586}$ is similar to it but did not contain the same catalytic residues. Side chains are shown in ball-and-stick model [O atom (red), N atom (blue), S atom (yellow)].

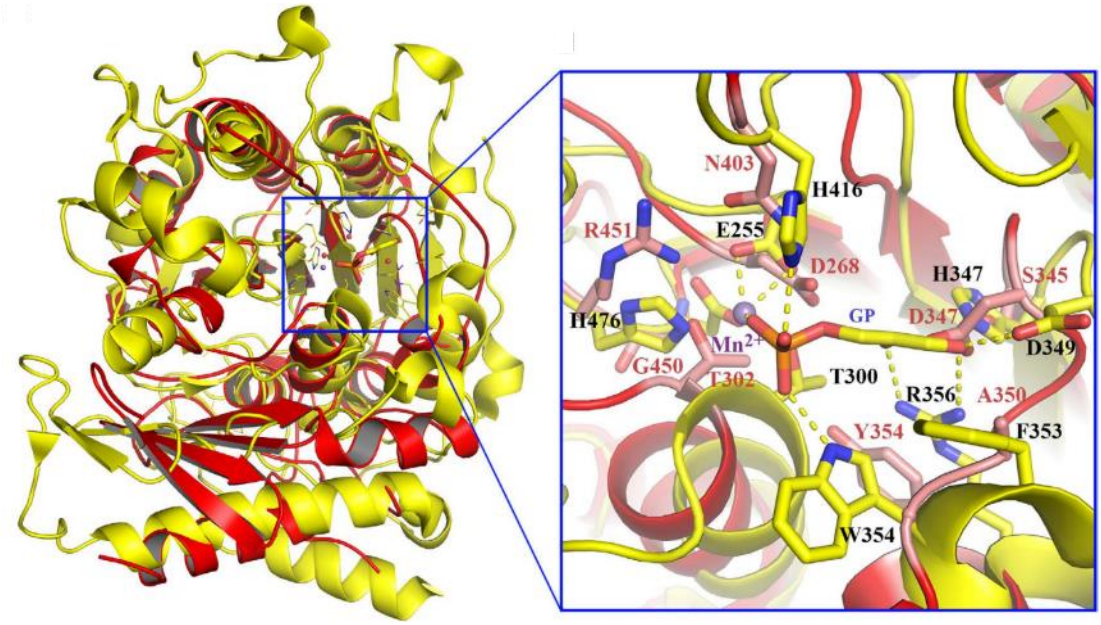


Figure 79 Superimposition of *Sal*-PbgA$_{245\text{-}586}$ with lipoteichoic acid synthase

The Dali server showed PbgA resembles the lipoteichoic acid synthase. *Sal*-PbgA$_{245\text{-}586}$ (red) is superimposed to lipoteichoic acid synthase (yellow) at the active site with RMSD of 3.0 Å over 284 Cα atoms. GP refers to the β-Glycerol phosphate (sn-glycerol-2-phosphate) which is the substrate for lipoteichoic acid synthase (271). However, the corresponding GP binding residues are not found in PbgA indication that PbgA does not have GP binding ability. Side chains and GP are shown in ball-and-stick model [O atom (red), N atom (blue), S atom (yellow) Mn atom (purple)].

4.3.4 Functional assay of PbgA in *E. coli* PbgA deletion strain

After determining the structure of PbgA globular domain, we designed a series experiments to explore the possible function and mechanism related to PbgA. Mutations: F349A, F292A, F519A, R215A/R216A, L343A, F394A, V471A, W477A, F285A, F362A, S351C/D303C, W396A, F310A, Y354A, F344A, F275A, L359A, W393D, W379C/S395C, D372C/F399C were generated by site direct mutagenesis, on pACYC-*E.coli-pbgA*. And pACYC-*E.coli-pbgA* was used as the positive control in the functional assays as well. These mutations cover the region of the three layers of PbgA globular domain. In particular, the second layer is made up of β1-7 and the third layer is made up of α-helices 2, 4, 5 and 10. This region fold into a hydrophobic core by residues L343, W396, F394, V262, W477, L473, V471, F292 from the β-sheet of the second layer, and residues F349, Y354, L334, F310, L309, F519 from the third layer (Figure 80 and Figure 81). We speculate that this hydrophobic core between the second and third layer may be responsible for CL transportation, so mutations L359A, F292A, F349A, F519A, L334A, L343A, F394A, W396A, V471A and W477A, as well as a double alanine substitution R215A/R216A were generated. It is also possible that the hydrophobic core between the first layer and the second layer is the CL transportation pathway. To explore this possibility, mutations F344A, W393D, and double mutants F399C/D372C and W397C/S395C were generated. The NRD183 strain was conditional lethal under 50 μg/ml vancomycin, which stimulates a stressful bacterial living conditions. Under this condition, the lethal phenotype of NRD183 could be rescued by positive control, which demonstrate that the leaky expression of *E. coli* PbgA could reach native level and was functional. From the similarity between *Sal*-$PbgA_{245\text{-}586}$, sulfatase and LtaS (Figure 78 and Figure 79), we speculated that the hydrophobic core between second and third layers might be responsible for cardiolipin binding. To explore this possibility, single alanine substitution of hydrophobic residues were generated as described. The F292A mutant was lethal, the F349A, R215/216A, W396A, F362A and F275A severely impaired bacterial growth under 50 μg/ml vancomycin, while mutants L343L, F394A, V471A and W477A had minor effect on

cell growth (Figure 81). Those results suggest that the residues F292, R215/216, W396, F362 and F275 in the hydrophobic core between second and third layer were important for cardiolipin transport. While, mutants F344A, W393D F399C/372D and W397C/S395C between the first and second layer are not involved in cardiolipin binding or transfer. This results suggested that it is the hydrophobic core between the second and third layer that may participate in the CL transport.

As a potential cardiolipin transporter, PbgA should possess a pathway, in which cardiolipin could pass. To explore that, we designed double cysteine disulfide bond mutants S351C/D303C to lock the flexible loop (D347-Q370). This loop had high B-factors between helices α4 and α6, indicating that this loop was flexible and might act as a lid for the cardiolipin binding and transport. S351C/D303C cause severe cell growth deficiency, indicating that this might be a pathway for Cl transport (Figure 81 and Figure 82).

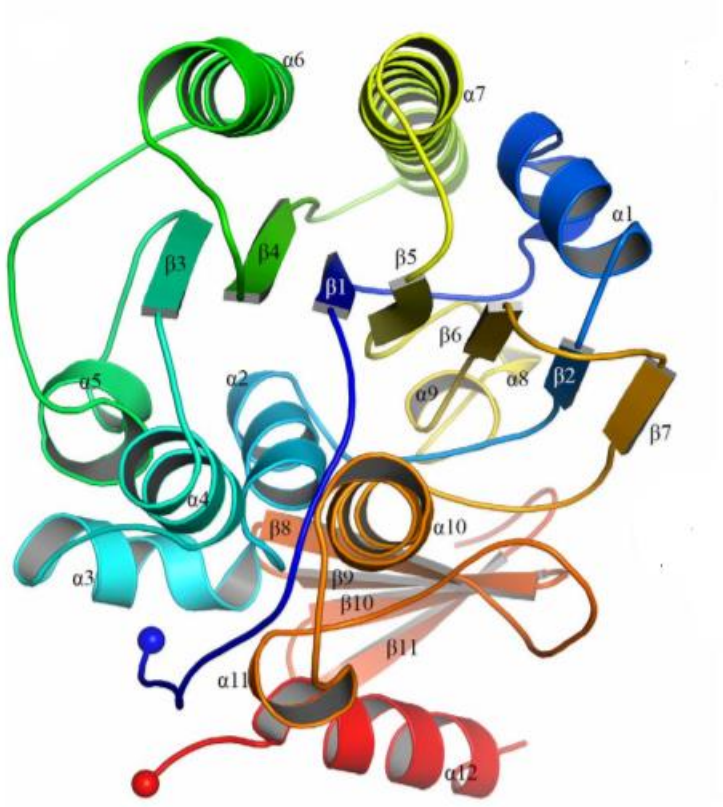


Figure 80 The *Sal*-PbgA$_{245\text{-}586}$ contains three layers.

Cartoon representation of crystal structure of *Sal*-PbgA$_{245\text{-}586}$ in rainbow. The

PbgA structure is coloured in rainbow. The structure of *Sal*-PbgA$_{245\text{-}586}$ has the α/β fold, which looks like a burger, where helices α1, α6 and α7 are at the top layer, β-strands 1, 2, 3, 4, 5, 6, 7, 8, α8 and α9 form the second layer, α-helices 2, 4, 5 and 10 form the third layer, and β-strands 8, 9, 10, 11, α11 and α12 form the bottom layer

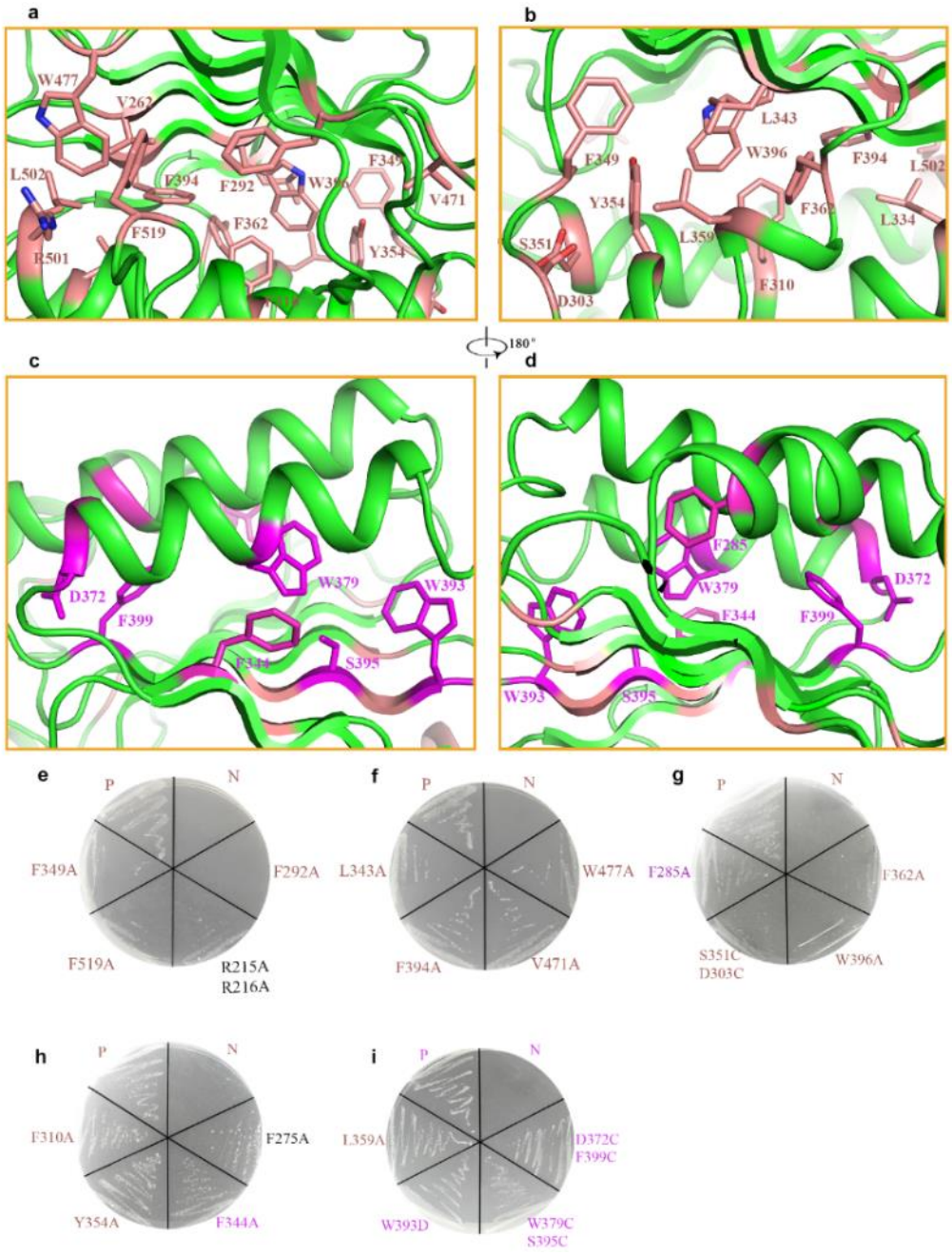


Figure 81 Mutations and *in vivo* functional assay of *E.coli* PbgA in NR183 strain

(a) The potential cardiolipin binding site between layer 2 and 3 (layer division shown in Figure 80). The hydrophobic residues were coloured in deep salmon. (b) Rotation of 180° along y axis relative to the left panel. (c) Hydrophobic residues between layer 1 and 2 are coloured in purple. (d) Rotation of 180° along y axis relative to left panel. (e–i) The functional assay plates of PbgA mutants against 50 µg/ml vancomycin environmental pressure. The functional assays are performed using NRD183 conditional lethal strain. WT or mutants were transformed into NRD183 respectively. P: positive control. (NRD183 with pACYC-PbgA) N: negative control (NRD183 with pACYC empty vector).

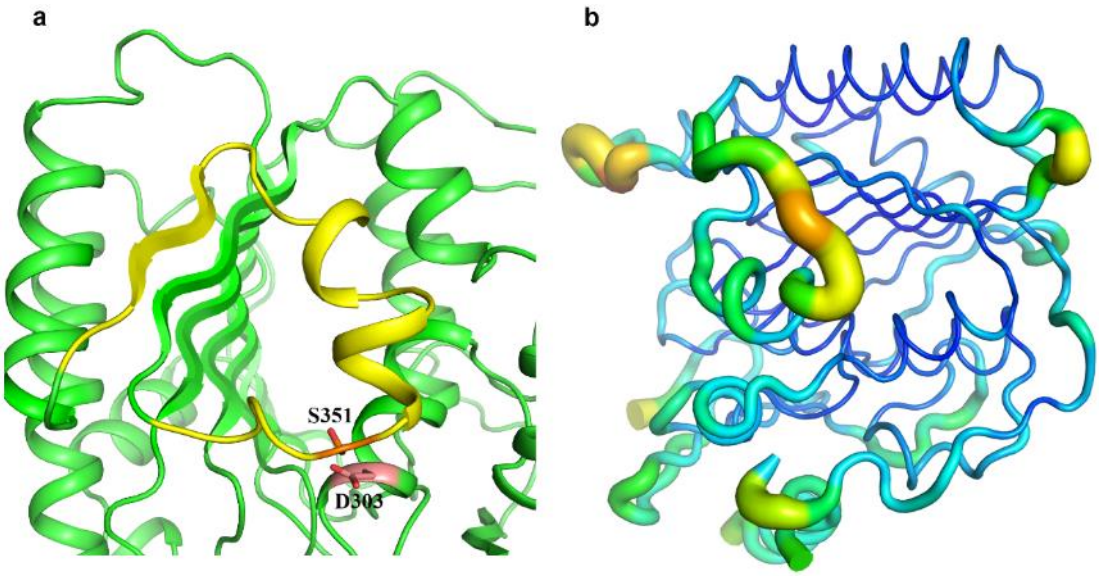


Figure 82 *Sal*-PbgA potential cardiolipin binding lid

(a) Cartoon demonstration of the potential cardiolipin binding lid coloured in yellow. The double cysteine mutants S351C/D303C are labelled and might form a disulphide bond closing the lid for cardiolipin binding. (b) The cartoon demonstration of the B-factor of *Sal*-PbgA structure. The radius of the cartoon structure represent the relative B-factor. The larger the radius of the region, the higher B-factor, suggesting the higher flexibility of the respect region. The PbgA loop, consisting of residues D347-Q370, had higher B-factor than other local secondary structure representing the flexibility of this loop. This structure feature is same in the *E. coli* PbgA globular domain, suggesting that the flexible

loops may have conformational changes when CL binds in.

4.3.5 PbgA-CL co-sedimentation assays

Cardiolipin binding was confirmed *in vitro* by PbgA-cardiolipin co-sedimentation assays. The wild type of *E.coli*-PbgA$_{191-586}$ can form co-sedimentation product at the presence of CL while R215A/216A mutant nearly cannot (256). Mutant proteins (F349A, R215A/216A, F292A and F362A) of *E. coli*-PbgA$_{191-586}$ were overexpressed and purified as described in method 4.2.2. Mutants F349A, R215A/216A, F292A and F362A showed low cardiolipin co-sedimentation levels compared to the wild type Figure 83. This result was consistent with our functional assays, supporting that the cardiolipin binding of these mutants was impaired. F519A mutant formed inclusion bodies during purification and failed to be purified like other mutants.

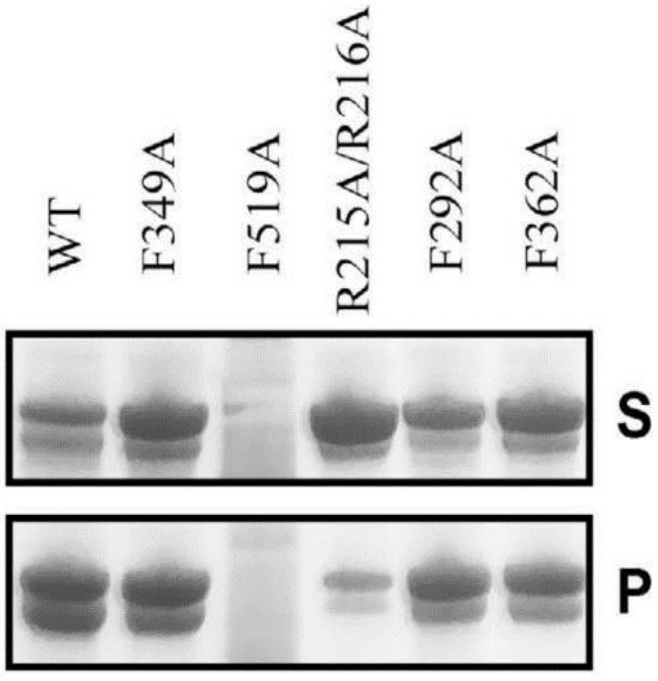


Figure 83 PbgA$_{191-586}$-cardiolipin co-sedimentation result

Wild type *E. coli*-PbgA$_{191-586}$ and its mutants were mixed with cardiolipin solutions and incubated at room temperature before centrifugation was applied. Supernatant and resuspended pellets were subjected to SDS-PAGE. Wild type

E. coli-PbgA$_{191-586}$ showed the highest level of co-sedimentation after incubation with CL at 1:10 (protein: CL) molar ratio. S: supernatant, P: pellet. Different mutant proteins showed different co-sedimentation level, R215A/R216A was the weakest one, while the other mutant proteins fell in between. The mutations that affected the CL-PbgA co-sedimentation levels were suggested to play important roles in CL binding. F519A PbgA mutant is not detected. The other mutants were found to decrease PbgA-CL co-sedimentation level, suggesting that they may participate in CL transportation. 5µl samples were mixed with 10µl ddH_2O and 5µl 4×SDS-PAGE Loading buffer. The SDS-PAGE samples were loaded into SDS-PAGE after heating on 90°C for 10min. The visualization of the bands was achieved by dying the gel in Quick Coomassie Stain (Generon) for 15min at room temperature.

4.4 CONCLUSION AND DISCUSSION

The OM of Gram-negative bacteria is an asymmetric membrane composed of phospholipids and lipopolysaccharide. Both lipids play important roles in the pathogenesis and drug resistance. Despite extensive studies on LPS biosynthesis and transport pathways, the mechanism for phospholipids transport and assembly in Gram-negative bacteria remains to be further studied at a structural level. According to the previous studies, it seems that GPL flipping from the inner leaflet of IM to the outer leaflet of the IM does not require a specific flippase but can be achieved via a broad range of α-helix fold IM transmembrane proteins (110).

However, the delivery of the GPL from the IM to the inner leaflet of the OM remains unknown. One possible explanation is that the inner leaflet of the OM and the periplasmic leaflet of the IM may form communication in bilayer lipids manner and lipids flow from IM to the inner leaflet of the OM and balancing the lipid amount. The other explanation is that there are some proteins or protein complexes responsible for transporting GPL from IM to the OM across the hyphophilic periplasm. By whichever

methods, the remodelling of the OM does exists and there are proteins that can adjust the GPL component ratio in the OM inner leaflet (272). Cardiolipin plays an essential role in the OM remodelling and is critical for mitochondrial function (272). PbgA transports CL from the IM to the OM in Gram-negative bacteria via the PhoPQ dependent mechanism (256). Mitochondria evolved from bacteria and the UPS1-Mdm35 (273) complex has been identified for phospholipid transport from the OM to the IM in yeast. The structure of UPS1 does not resemble PbgA. The UPS1 contains a concave 7-antiparellel stranded β-sheets, which hold PA molecules inside. In contrast, PbgA does not contain such cavity, implying that in PbgA it might be the third layer rather than the first or second layer that is responsible for CL transportation. This analysis is also consistent with our functional assays (Figure 81). Superimposition of $PbgA_{245\text{-}586}$ with arylsulfatases and lipoteichoic acid synthase showed striking overall similarity, but low structural identity in the active sites or metal binding sites. The result of superimposition suggest that PbgA is different from arylsulfatases or lipoteichoic acid synthase. Our functional assays and co-sedimentation results both point to the importance of residues F292, F349 and F362 in the functionality of PbgA and the mutants of them would impair cell growth or affect CL binding affinity. These three residues are located in the hydrophobic core between layer 2 and 3 suggesting that this it is a potential cardiolipin binding or transporting site. Additionally, the mutations that aimed at disrupting the first or second layer, cause relatively weak impairment of cell growth, which further confirm the importance of hydrophobic space between of layer 2 and 3. The structures of the *E.coli*-$PbgA_{245\text{-}586}$ and *Sal*-$PbgA_{245\text{-}586}$ are very similar without little conformational difference (Figure 77). The loop made up of residues D347-Q370 between these two structures is in different conformation between *E.coli*-$PbgA_{245\text{-}586}$ and *Sal*-$PbgA_{245\text{-}586}$, and the B-factor along this loop are higher than neighbouring residues (Figure 82), suggesting the flexibility of this loop. We speculate that this loop region may be the lid for CL binding, which can open or close the space between layer 2 and 3 for the entry of CL.

Our structural and functional studies have revealed that the hydrophobic core between the layer 2 and 3 may be responsible for CL binding, and that the residues D347-Q370

may form a lid for CL transport. This is the first structure study of PbgA, a bacterial protein involved in phospholipid transport from the inner membrane to the OM. However, the missing N-terminal transmembrane domain may also play an important role in the CL transportation function. There are five transmembrane α-helices in the transmembrane domain of PbgA (256). And the transmembrane domain of PbgA is essential rather than the globular domain (256). Our functional studies were based on the conditional lethal PbgA globular domain deletion strain NRD183, and the mutations were generated on the globular domain. If we could try to solve the full length structure of PbgA, we may study further into the mechanism of CL transportation. It is still interesting that how the CL is recognized from the IM and delivered from the IM to the OM. According to the previous studies, PbgA may for a tetramer to facilitate CL transport (255), which is consistent with our *S. typhimurium* PbgA crystal structure. There are 4 protormers in asymmetric unit. One hypothesis is that PbgA forms a tetramer by the interactions of four transmembrane domains, and forming a pathway for CL to pass through, the globular domain forms a CL pass channel that may facilitate the CL to be protected in the water filled periplasm. After PbgA transportation of CL, there should be an OM receptor that may receive CL. We still need to perform more biochemistry studies to identify PbgA related proteins. PbgA seems not to be able to work independently to transport CL from IM to OM. Identifying the full CL transportation path may facilitate the PbgA transportation mechanism.

Appendices

Selenomethionine (SeMet) labelling related medium/solution recipe:	
10 L M9 medium	Dissolve 64 g Na_2HPO_4-$7H_2O$, 15 g KH_2PO_4, 2.5 g NaCl and 5.0 g NH_4Cl in 2L ddH_2O and dilute into 10L then autoclave the medium.
20×Selenomethionine Medium Nutrient Mix	Dissolve 102 g powder in 1L sterile water and filter sterilized at last, (Molecular Dimensions Cat# MD12-502)
50×Additional solution	Dissolve 200g D-Glucose 12.5g $MgSO_4$ (1 M), 0.02g thiamine in 1 L ddH_2O and sterilized with filter at last.
100×Selenomethionine inhibitor amino acid solution	Dissolve 5g Lysine, 5g Phenylalanine, 5g Threonine, 2.5g Isoleucine, 2.5g Leucine and 2.5g Valine in 500ml sterile water and filter sterilized at last.
1000×Selenomethionine solution	Dissolve 1g Selenomethionine in 10ml sterile water.

Appendix 1 Selenomethionine (SeMet) labelling related medium/solution recipe (274)

This is the selenomethinone incorporated protein production solution recipie. Some of the solutions in the recipie can be omitted in some cases. The most essential component is the 10 L M9 medium, 20×Selenomethionine Medium Nutrient Mix, 100×Selenomethionine inhibitor amino acid solution and 1000×Selenomethionine solution

Appendix tables

Crystallization screen kits	Supplier company
MemGold	Molecular Dimensions
MemGold2	Molecular Dimensions
MemStart	Molecular Dimensions
Memsys	Molecular Dimensions
MemMeso	Molecular Dimensions
MemPlus	Molecular Dimensions
MIDAS	Molecular Dimensions
Morpheus	Molecular Dimensions
Morpheus II	Molecular Dimensions
MultiXtal	Molecular Dimensions
ProPlex	Molecular Dimensions
SG1 Screen	Molecular Dimensions
3D structure Screen	Molecular Dimensions

JCSG-plus	Molecular Dimensions
MemAdvantage	Molecular Dimensions
PEG/ION	Hampton Research
Index	Hampton Research
Crystal screen Cryo	Hampton Research
Crystal Screen 2 Cryo	Hampton Research
MembFcac	Hampton Research

Table 10 Commercial crytallizaion screening conditon kits used in all chapters

These crystallization screening kits were bought in the 96×10ml form and divided into deep well blocks (1ml 96well) by hands. The deep well blocks containing each screens can be applied on Gryphon Robot directly.

Ecoli_LapB_24_F	ATATCCATGGCGCAACAAAACAAGCAAGATGAAG
Ecoli_lapB_24_R	ATATAAGCTTTTACAGGCCATCAAGACCGCG
Sal_LapB_24_F	ATATCCATGGCGCAACAAACAAAACAGGATGAAG
Sal_lapB_24_R	ATATAAGCTTCTACTGCCCATCAAGTCCGC
Shi_lptCAB_F_EcoRI	ATATGAATTCGAAGCCAAAGGGCAATCGATATG
Shi_lptCAB_R_KpnI_8HIS	ATATGGTACCTCAGTGATGGTGATGGTGATGGTGATGGAGTCTGAAGTCTTCCCCAAGG
Shi_lptFG_F_KpnI	ATATGGTACCtttttacgggcgtatttaaaGTGATAATC
Shi_lptFG_R_XbaI	ATATTCTAGATTACGATTTTCTCATTAACAGCCACAG
Shi_delet_1-76C_F	TGGTTTACGCAGCCGGTACTTAC
Shi_delet_1-76C_R	CGGCTGCGTAAACCACATatcgattgccctttggcttc
Neisseria_lpt82CAB_F_EcoRI	ATATGAATTCgctttgaacgggtacatcaaATGGATTCGCCGCATCTCGTGTTCTTC
Neisseira_lpt82CAB_R_KpnI_8HIS	ATATGGTACCTTAGTGATGGTGATGGTGATGGTGATGATATTTGAAGTTCTCGCCCAAATAGACGG
Neisseria_lptFG_F_KpnI	ATATGGTACCGCCCCCCTGACCGCGATTTTATG
Neisseria_lptFG_R_XbaI	ATATTCTAGATTAACGTTTTTCCTGTTTGCGTATCAGCC
LT2_lpt76CAB_F_EcoRI	ATATGAATTCgaggccaaagggcaatcgatATGTGGTTTACGCAACCGGTATTAAC
LT2_lpt76CAB_R_KpnI_8HIS	ATATGGTACCTTAGTGATGGTGATGGTGATGGTGATGGAGTCTGAAGTCTTCCCCAAG
LT2_lptFG_F_KpnI	ATATGGTACCTTTTTACGGGCGTATTTAAAGTGATAATCATAAG
LT2_lptFG_R_XbaI	ATATTCTAGATTACGACTTACGCAGCATTAGCCAG
Klebsiella_lpt76CAB_F_EcoRI	ATATGAATTCgaggccaaagggcaatcaatATGTGGTTTACCAAGCCGGTGATG
Klebsiella_lpt76CAB_R_KpnI_8HIS	ATATGGTACCTTAGTGATGGTGATGGTGATGGTGATGGAGTCTGAAGTCTTCCCCAAG
Klebsiella_lptFG_F_KpnI	ATATGGTACCTTTTTACGGGCGTATTTATAGTGATAATCATAAGATATCTGGTTCG
Klebsiella_lptFG_R_XbaI	ATATTCTAGATTAGGCCTTGCGCATCATCAGCC

Legionella_lpt78CAB_F_EcoRI	ATATGAATTCatcacaggctatttaaaacaATGCTTAAAAACCCCCATATTATTATTAAGCAAG
Legionella_lpt78CAB_R_KpnI_8HIS	ATATGGTACCTTAGTGATGGTGATGGTGATGGTGATGCAAAGCAAATTCTTCACCCAAATAAAC
Legionella_lptFG_F_KpnI	ATATGGTACCGTCCTCTGTAGGGTATTGTAGTGATTATTTTTC
Legionella_lptFG_R_XbaI	ATATTCTAGATTATCTGACTCTTCGCATTAAATAAATTCCAATTAAAG
Klebsiella_lptCAB_F_EcoRI	ATATGAATTCGAGGCCAAAGGGCAATCAATATG
Klebsiella_lptCAB_R_KpnI_8HIS	ATATGGTACCTTAGTGATGGTGATGGTGATGGTGATGGAGTCTGAAGTCTTCCCCAAG
K12_CAB_sm_F	GAAGCCAAAGGGCAATCGATATGAGTAAAGCC
K12_CAB_HIS_sm_R	aattaGTGATGGTGATGGTGATGGTGATGGAGTCTGAAGTCTTCCCCAAGGTATACACG
K12_FG_sm_F	CACCATCACCATCACtaatttttacgggcgtatttaaaGTGATAATCATAAGATATCTG
K12_FG_sm_R	TTACGATTTTCTCATTAACAGCCACAGGC
pTRC99a_sm_F	TGTTAATGAGAAAATCGTAACGGGGATCCTCTAGAGTCGACC
pTRC99a_sm_R	ATCGATTGCCCTTTGGCTTCCATGGTCTGTTTCCTGTGTGAAATTGTTATCCG
Kan_SalI_F	tatagtcgacATACAAGGGGTGTTATGAGCCATATTCAAC
Kan_HindIII_R	tataaagcttttagaaaaactcatcgagcatcaaatgaaactgc
Sm_AmpKan_ptrc99a_F	TGCTCGATGAGTTTTTCTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTG
Sm_AmpKan_ptrc99a_R	CCCGTTGAATATGGCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGG
AcinetobacterLptB-FNcoI	CATGCCATGGAGCAAGCTCTTCAACAACCGC
AcinetobacterLptB-REcoRI	ATCGGAATTCTTAATGATGATGATGATGATGAACTGTGAAATCATCACCTAAATAGACT
AcinetobacterLptC-FNcoI	CATGCCATGGTTAATGAGACAATGAAGCATCAGTTT
AcinetobacterLptC-RHindⅢ	TCGCAAGCTTCACTTTGTAGAGAGTAAGCCGCCTTG
AcinetobacterLptF-FNdeI	GGAATTCCATATGATTATTCGGCGTTATCTCGTCAAGCAAGT
AcinetobacterLptG-NXhoⅠ	TATACTCGAGTTAACGGGCCCGATAAAGCAAATAACTTCC
citrobacterLptB-FNcoI	CATGCCATGGCAACATTAACTGCAAAGAACCTTGC
citrobacterLptB-RBamHI	ATCGGGATCCTTAATGATGATGATGATGATGCTCGAGTCTGAAGTCTTCCCCAAGG
citrobacterLptC-FNcoI	CATGCCATGGGCAAAACCAGACGTTGGGTTATC
citrobacterLptC-RHindⅢ	TCGCAAGCTTAAGGCTGAGTTTGTTTGTTTTGGATTTC
citrobacterLptF-FNdeI	GGAATTCCATATGATAATCATAAGATATCTGGTTCGGGAAACG
citrobacterLptG-NXhoⅠ	TATACTCGAGTTACGATTTACGCAGCATCAGCCACAGG
VibriocholeraLptB-FNcoI	CATGCCATGGCCATCTTAAAAGCACAGCATTTAGC
LptB-REcoRI	ATCGGAATTCTTAATGATGATGATGATGATGTCTGAGACGGAATTGTTCGCCGAGATAC
LptC-FNcoI	CATGCCATGGGTTTGTCGCGCATTGTTTACGTTCTGC
LptC-RHindⅢ	TCGCAAGCTTAAGGTGTGAGCGTTTCATATCTACCTTG
LptF-FNdeI	GGAATTCCATATGATTATTGTTAGATATTTGATCCGAGAAACAAT

LptG-NXhoⅠ	TATACTCGAGTTATAGCTTTCTACCCAATAGCCCTAAC
K12_G144_myc_F	GAACAAAAACTCATCTCAGAAGAGGATCTGTTGCTCTCTACCCAGCAAGGCTTATGG
K12_G144_myc_R	CCTCTTCTGAGATGAGTTTTTGTTCCGAGCCGCCGTACATCGCC
K12_G228_myc_F	GAACAAAAACTCATCTCAGAAGAGGATCTGACCTGGAAAACCAACCTCACGCC
K12_G228_myc_R	CCTCTTCTGAGATGAGTTTTTGTTCGCCGCTCACCGTCTGCGAAC
K12_F230_flag_F	GATTACAAAGATGACGACGATAAACAGGCGATCATTGGTCACCAGGC
K12_F230_flag_R	TTTATCGTCGTCATCTTTGTAATCATAATCCTGGAAGTCCGTAATGCGGAAATCACG
K12_F138_flag_F	GATTACAAAGATGACGACGATAAACCTGGCATGGCGGCGCTG
K12_F138_flag_R	TTTATCGTCGTCATCTTTGTAATCGTTCGCTTTCGCTTCTGCTAACACTTCATCCTG
Mutagenesis	
01_F_S318C_F	CCTGAAATGTAACGGCGGTAAAGGTAAGCTGGACCCGACG
01_F_S318C_R	CGCCGTTACATTTCAGGGAGGTCTGGATCAGGAAGAAAAGTAGATACAGC
01_G_K40C_F	CAGCTGTGTAAAGCCGGGCAGGGGAGTTACGACGCGTTAG
01_G_K40C_R	CCGGCTTTACACAGCTGATCGACAAACTTGATAATGCCCGACAGCG
02_F_L311C_F	CTTTTCTTCTGTATCCAGACCTCCCTGAAATCGAACGGCGGTAAAGG
02_F_L311C_R	GAGGTCTGGATACAGAAGAAAAGTAGATACAGCAGCATGGCTGGC
02_G_V36C_F	attatcaagtttTGTgatcagctgaaAAAAGCCGGGCAGGGGAGTTAC
02_G_V36C_R	ttcagctgatcACAaaacttgataatGCCCGACAGCGACACCAGC
03_F_S315C_F	ATCCAGACCTGTCTGAAATCGAACGGCGGTAAAGGTAAGCTGGAC
03_F_S315C_R	CGATTTCAGACAGGTCTGGATCAGGAAGAAAAGTAGATACAGCAGCATGGC
03_G_L39C_F	gatcagTGTaaaaaagccggGCAGGGGAGTTACGACGCGTTAG
03_G_L39C_R	ccggctttttACActgatcGACAAACTTGATAATGCCCGACAGCG
04_F_G36C_F	CTCTGTGCAGCGGTTGACGGCGATATTCCGGCGAATCTG
04_F_G36C_R	AACCGCTGCACAGAGGATCCTCACTAACTTTTGACAGAAGAAGATCAAAAGC
04_G_Q324C_F	ACTGGACTGTATCTTCGGCCCGCTGACGTTGGTTTATGGCATC
04_G_Q324C_R	CCGAAGATACAGTCCAGTACGTAGAAGACAAAACCGAAACTGATACCGG
05_F_L35C_F	TGAGGATCTGTGGCGCAGCGGTTGACGGCGATATTCCG
05_F_L35C_R	GCGCCACAGATCCTCACTAACTTTTGACAGAAGAAGATCAAAAGCAAGATGAAGAG
05_G_I325C_F	TGGACCAGTGTTTCGGCCCGCTGACGTTGGTTTATGGCATCC
05_G_I325C_R	GCCGAAACACTGGTCCAGTACGTAGAAGACAAAACCGAAACTGATACCGG
06_F_D42C_F	TTGACGGCTGTATTCCGGCGAATCTGGTGCTCTCCCTTCTCG
06_F_D42C_R	CCGGAATACAGCCGTCAACCGCTGCGCCGAGGATC
06_G_Q266C_F	TTGTGATGCCGGACGTTatcagctcaacatgtggagcaaaatcttccag
06_G_Q266C_R	AACGTCCGGCATCACAACCGCTCGACTTCAGATACTTCACATAGTTGTG

07_F_T153C_F	CAAGCGTGTAATGGCAGCTCGGTGCTGTTCATCGAAAGCGTTG
07_F_T153C_R	GCTGCCATTACACGCTTGCTGGAATTGCCCTTGCGCCAG
07_F_T246C_F	CGAACGATTGTGACCAGATGGACATGCGCACATTGTGGAACACTGAC
07_F_T246C_R	CATCTGGTCACAATCGTTCGGGTCGAGCGCCACCGCC
08_F_D229P_Q231P_F	AGCCGTATCCGGCGATCATTGGTCACCAGGCGGTGG
08_F_D229P_Q231P_R	CGCCGGATACGGCTGGAAGTCCGTAATGCGGAAATCACGTAAC
09_F_R223PT225P_F	TTTCCCGATTCCGGACTTCCAGGATTATCAGGCGATCATTGGTCACCAG
09_F_R223PT225P_R	GAAGTCCGGAATCGGGAAATCACGTAACAATGCAGTGCCTTCGAAG
10_G_K40E_K41E_F	ATCAGCTGGAAGAAGCCGGGCAGGGGAGTTACGACGCG
10_G_K40E_K41E_R	CGGCTTCTTCCAGCTGATCGACAAACTTGATAATGCCCGACAGCG
11_G_S223P_T225P_F	GTCCGCAGCCGGTGAGCGGCACCTGGAAAACCAACCTCAC
11_G_S223P_T225P_R	CTCACCGGCTGCGGACCGGTAATCTGTTTCGGATTGGTCAGATCAGATTC
12_G_G228P_W230P_F	AGCCCGACCCCGAAAACCAACCTCACGCCGGACAAAC
12_G_G228P_W230P_R	TTTTCGGGGTCGGGCTCACCGTCTGCGAACCGGTAATCTG
13_G_L234C_F	AAACCAACTGTACGCCGGACAAACTGGGCGTGGTGGC
13_G_L234C_R	CGGCGTACAGTTGGTTTTCCAGGTGCCGCTCACCG
13_G_D156C_F	GAAATGTGGCAACAACTTCGTCTACATTGAGCGGGTTAAAGGTGACGAAG
13_G_D156C_R	GAAGTTGTTGCCACATTTCGCCCATAAGCCTTGCTGGGTAGAGAG
14_F_F26D_F	CTTTTGATCGATTTCTGTCAAAAGTTAGTGAGGATCCTCGGCGCAGC
14_F_F26D_R	ACTTTTGACAGAAATCGATCAAAAGCAAGATGAAGAGTATCGCCAGCTGGC
14_F_L62D_F	AAATGGCGCAGGATATCCTGCCATTAAGCCTGTTCCTCGGG
14_F_L62D_R	AGGATATCCTGCGCCATTTCCGGCACGCCCAACCC
15_F_R295E_F	AGGGAGAAGTACTGTCGATGCTGCCAGCCATGCTG
15_F_R295E_R	TCGACAGTACTTCTCCCTGACGTGGGTTAACCACGCTCAG
15_F_K13E_F	ACGCTCGAAAGCCAGCTGGCGATACTCTTCATCTTGCTTTTGATCTTCTTCTG
15_F_K13E_R	CAGCTGGCTTTCGAGCGTCTCCCGCACCAGATATCTTATGATTATCAC
16_F_R295E_F	AGGGAGAAGTACTGTCGATGCTGCCAGCCATGCTG
16_F_R295E_R	TCGACAGTACTTCTCCCTGACGTGGGTTAACCACGCTCAG
16_F_K78E_F	CTGGGCGAACTGTATACCGAAAGTGAAATTACGGTAATGCATGCCTG
16_F_K78E_R	GTATACAGTTCGCCCAGCGTCATCAGCAGCCCGAGG
17_F_R269E_F	TGAACTGGGAAATCACGTTGGTATTCACCGTGTTTATGATGGCAC
17_F_R269E_R	ACGTGATTTCCCAGTTCAGTTCTGCGCGAGCACGATC
17_F_K317E_F:	CTCCCTGGAATCGAACGGCGGTAAAGGTAAGCTGGACC
17_F_K317E_R	GTTCGATTCCAGGGAGGTCTGGATCAGGAAGAAAAGTAGATACAGC
18_F_R269E_F	TGAACTGGGAAATCACGTTGGTATTCACCGTGTTTATGATGGCAC
18_F_R269E_R	ACGTGATTTCCCAGTTCAGTTCTGCGCGAGCACGATC
18_F_R126E_F	TCATCGGAACATCAGGATGAAGTGTTAGCAGAAGCGAAAGCGAAC
18_F_R126E_R	ATCCTGATGTTCCGATGACCACGGTCCCGCCCACATC
19_F_E10A_F	GGGCGACGCTCAAAAGCCAGCTGGCGATACTCTTCATC
19_F_E10A_R	TTTGAGCGTCGCCCGCACCAGATATCTTATGATTATCACTTTAAATACGCC

	C
19_F_E84A_F	GAAAGTGCGATTACGGTAATGCATGCCTGCGGCCTGAG
19_F_E84A_R	ATTACCGTAATCGCACTTTCGGTATACAGTTTGCCCAGCGTCATC
20_F_E58A_F	GCGATGGCGCAGCTTATCCTGCCATTAAGCCTGTTCC
20_F_E58A_R	GCTGCGCCATCGCCGGCACGCCCAACCCGAGAAG
20_F_E265A_F	CGCAGCGCTGAACTGGCGTATCACGTTGGTATTCACCG
20_F_E265A_R	GCTGCGCCATCGCCGGCACGCCCAACCCGAGAAG
20_F_E265A_F	CGCAGCGCTGAACTGGCGTATCACGTTGGTATTCACCG
20_F_E265A_R	AGTTCAGCGCTGCGCGAGCACGATCGGTGTCAGTG
21_F_D129A_F	ATCAGGCGGAAGTGTTAGCAGAAGCGAAAGCGAACCCTG
21_F_D129A_R	TAACACTTCCGCCTGATGACGCGATGACCACGGTCC
21_F_E265A_F	CGCAGCGCTGAACTGGCGTATCACGTTGGTATTCACCG
21_F_E265A_R	AGTTCAGCGCTGCGCGAGCACGATCGGTGTCAGTG
22_F_I18D_F	TGGCGGATCTCTTCATCTTGCTTTTGATCTTCTTCTGTCAAAAGTTAGTGAG
22_F_I18D_R	GATGAAGAGATCCGCCAGCTGGCTTTTGAGCGTCTCC
22_F_M74D_F	TGCTGGATACGCTGGGCaaaCTGTATACCGAAAGTgaaATTACGGTAATGC
22_F_M74D_R	CCAGCGTATCCAGCAGCCCGAGGAACAGGCTTAATGG
23_F_M74D_F	TGCTGGATACGCTGGGCaaaCTGTATACCGAAAGTgaaATTACGGTAATGC
23_F_M74D_R	CCAGCGTATCCAGCAGCCCGAGGAACAGGCTTAATGG
23_F_M299D_F	TACTGTCGGATCTGCCAGCCATGCTGCTGTATCTACTTTTCTTCC
23_F_M299D_R	TGGCAGATCCGACAGTACGCGTCCCTGACGTGGG
24_F_L66D_F	CTGCCAGATAGCCTGTTCCTCGGGCTGCTGATGACG
24_F_L66D_R	AACAGGCTATCTGGCAGGATAAGCTGCGCCATTTCCGGC
24_F_Y306D_F	GCTGCTGGATCTACTTTTCTTCCTGATCCAGACCTCCCTGAAATC
24_F_Y306D_R	GAAAAGTAGATCCAGCAGCATGGCTGGCAGCATCGACAG
25_F_W268E_F	AACTGAACGAACGTATCACGTTGGTATTCACCGTGTTTATGATGGCACTTATG
25_F_W268E_R	CGTGATACGTTCGTTCAGTTCTGCGCGAGCACGATCGG
25_F_F310E_F	CTACTTTTCGAACTGATCCAGACCTCCCTGAAATCGAACGGCG
25_F_F310E_R	CTGGATCAGTTCGAAAAGTAGATACAGCAGCATGGCTGGCAG
26_F_R212E_F	AGGGAACGGAATTCGAAGGCACTGCATTGTTACGTGATTTCC
26_F_R212E_R	TTCGAATTCCGTTCCCTGGTTGAGAGTGACGACCTGGG
26_F_Y230E_F	CAGGATGAACAGGCGATCATTGGTCACCAGGCGGTGG
26_F_Y230E_R	ATCGCCTGTTCATCCTGGAAGTCCGTAATGCGGAAATCACGTAAC
27_F_F149D_F	AAGGGCAAGATCAGCAAGCGACTAATGGCAGCTCGGTG
27_F_F149D_R	TTGCTGATCTTGCCCTTGCGCCAGCGCCGCCATGCC
27_F_P139D_F	AAAGCGAACGATGGCATGGCGGCGCTGGCGCAAGG
27_F_P139D_R	ATGCCATCGTTCGCTTTCGCTTCTGCTAACACTTCATCCTGATGACGC
28_G_K13E_F	CGCTATATCGGTGAAACTATTTTCACCACCATCATGATGACACTGTTCATGC
28_G_K13E_R	GAAAATAGTTTCACCGATATAGCGGTCAAGTACGCCAAAAGGTTGCAT
28_G_R86E_F	GCGCAGGAAAGCGAACTGGTGGTGATGCAGGCTTCTG

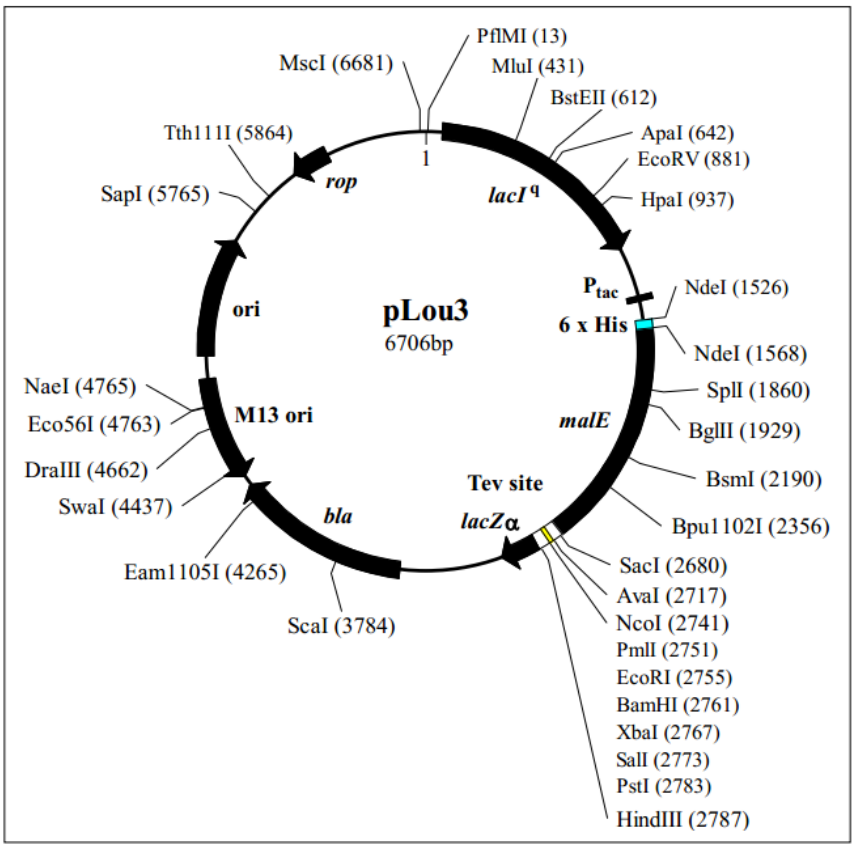
28_G_R86E_R	TCGCTTTCCTGCGCCAGCATCCCAAGACCAAGCAACG
29_G_R86E_F	GCGCAGGAAAGCGAACTGGTGGTGATGCAGGCTTCTG
29_G_R86E_R	TCGCTTTCCTGCGCCAGCATCCCAAGACCAAGCAACG
29_G_R301E_F	CACTGGAAAGCGTACCGATGGGCGTGCGTGTGG
29_G_R301E_R	GGTACGCTTTCCAGTGGGCCAAAGATGAACGACAGCGC
30_G_R301E_R308E_Fsm	GAAAGCGTACCGATGGGCGTGGAAGTGGTCACCGGTATCAGTTTCGGTTTTG
30_G_R301E_R308E_Rsm	TTCCACGCCCATCGGTACGCTTTCCAGTGGGCCAAAGATGAACGACAGC
31_G_K34E_F	GGCATTATCGAATTTGTCGATCAGCTGAAAAAAGCCGGGCAGG
31_G_K34E_R	ATCGACAAATTCGATAATGCCCGACAGCGACACCAGCATGAAC
31_G_R136E_F	GTAACTACGAAGCGCAGGCGATGTACGGCGGCTCG
31_G_R136E_R	CTGCGCTTCGTAGTTACGCGCCATCTGCTCGCC
32_G_K34E_F	GGCATTATCGAATTTGTCGATCAGCTGAAAAAAGCCGGGCAGG
32_G_K34E_R	ATCGACAAATTCGATAATGCCCGACAGCGACACCAGCATGAAC
32_G_K62E_F	CCGGAAGATGTGCAGATCTTCTTCCCGATGGCGGCTC
32_G_K62E_R	TCTGCACATCTTCCGGCACGCTCAGCAAGGTATACATTCCTGC
33_G_R133E_R136E_F	CGGAAAACTACGAAGCGCAGGCGATGTACGGCGGC
33_G_R133E_R136E_R	CGCTTCGTAGTTTTCCGCCATCTGCTCGCCCTGCG
34_G_R270E_R278E_Fsm	GAATATCAGCTCAACATGTGGAGCGAAATCTTCCAGCCGCTATCTGTGGC
34_G_R270E_R278E_Rsm	TTCGCTCCACATGTTGAGCTGATATTCTCCGGCATCCTGACCGCTC
35_G_K40E_F	GATCAGCTGGAAAAAGCCGGGCAGGGGAGTTACGACG
35_G_K40E_R	GGCTTTTTCCAGCTGATCGACAAACTTGATAATGCCCGACAGCGAC
35_G_R136E_F	GTAACTACGAAGCGCAGGCGATGTACGGCGGCTCG
35_G_R136E_R	CTGCGCTTCGTAGTTACGCGCCATCTGCTCGCC
36_G_T18D_F	TTTCACCGATATCATGATGACACTGTTCATGCTGGTGTCGCTGTC
36_G_T18D_R	TGTCATCATGATATCGGTGAAAATAGTTTTACCGATATAGCGGTCAAGTACGCCAAAAG
36_G_L78D_F	GCGTTGGATGGTCTTGGGATGCTGGCGCAGCGCAG
36_G_L78D_R	CAAGACCATCCAACGCCCCAAGCAGAGCCGCCATCG
37_G_V309D_F	CGTGATGTCACCGGTATCAGTTTCGGTTTTGTCTTCTACGTACTGGACC
37_G_V309D_R	TACCGGTGACATCACGCACGCCCATCGGTACGCTACG
37_G_L78D_F	GCGTTGGATGGTCTTGGGATGCTGGCGCAGCGCAG
37_G_L78D_R	CAAGACCATCCAACGCCCCAAGCAGAGCCGCCATCG
38_G_L26E_F	TTCATGGAAGTGTCGCTGTCGGGCATTATCAAGTTTGTCGATCAGC
38_G_L26E_R	AGCGACACTTCCATGAACAGTGTCATCATGATGGTGGTGAAAATAGTTTTACC
38_G_M70E_F	TTCTTCCCGGAAGCGGCTCTGCTTGGGGCGTTGC
38_G_M70E_R	CGCTTCCGGGAAGAAGATCTGCACATCTTTCGGCACGC
39_G_F67E_F	GCAGATCGAATTCCCGATGGCGGCTCTGCTTGGGGC
39_G_F67E_R	ATCGGGAATTCGATCTGCACATCTTTCGGCACGCTCAGCAAG
39_G_Y320E_F	TTGTCTTCGAAGTACTGGACCAGATCTTCGGCCCGCTGACG
39_G_Y320E_R	GTCCAGTACTTCGAAGACAAAACCGAAACTGATACCGGTGACCAC
40_G_W204D_F	CATAAAGTCGATCGTCTGTCGCAGGTTGATGAATCTGATCTGACCAATCC

40_G_W204D_R	GACAGACGATCGACTTTATGTTCCGGGTCAAACTTCGCAGTAGC
40_G_I163D_F	TTCGTCTACGATGAGCGGGTTAAAGGTGACGAAGAGTTAGGTGGC
40_G_I163D_R	CGCTCATCGTAGACGAAGTTGTTGCCATCTTTCGCCCATAAGC
41_G_L206D_V209D_F	TGATTCGCAGGATGATGAATCTGATCTGACCAATCCGAAACAGATTACCG
41_G_L206D_V209D_R	TCATCATCCTGCGAATCACGCCAGACTTTATGTTCCGGGTCAAAC
42_G_Y257E_Y271E_Fsm	AAGTCGAGCGGTCAGGATGCCGGAcgtGAACAGCTCAACATGTGGAGCAAAATCTTCC
42_G_Y257E_Y271E_Rsm	CTGACCGCTCGACTTCAGATACTTCACTTCGTTGTGCAAACCGCTGATAGAGAGTG
WaaB_V186A_F	ttctctatGCGgggcgtcttaaatttgaagggcagaaaagagttaaag
WaaB_V186A_R	gcccCGCatagagaaATACAGCGGGTTTATCGCGC
WaaB_Q194A_F	gaagggGCGaaaagagTTAAAGATTTATTTGATGGCTTAGCTCGTACG
WaaB_Q194A_R	ctcttttCGCcccttcaaatttaagacgcccTACatagagaaatacag
WaaB_K195A_F	agggCAGGCGagagTTAAAGATTTATTTGATGGCTTAGCTCGTACG
WaaB_K195A	ctctCGCCTGcccttcaaatttaagacgcccTACatagagaaatac
WaaB_I216A	atattGCGggtgatggCTCAGATTTTGAAAAGTGCCAGGC
WaaB_I216A_R	ccatcaccCGCaatatGTAGCTGCCATTCCCCTGTC
WaaB_W243A_F	atggtGCGcaaagcgcgccgtggcaagtcgtac
WaaB_W243A_R	gctttgCGCaccatACCAGATCACACGCTGCTCAATACC
WaaB_E268A_F	gcatttGCGggatttcCTATGACCCTGCTGGAAGCAATG
WaaB_E268A_R	gaaatccCGCaaatgcAGAGGTAAGTAGTAACGCGGTGAC
WaaB_T273A_F	ctatgGCGctgctgGAAGCAATGTCATATGGAATTCCGTG
WaaB_T273A_R	cagcagCGCcatagGAAATCCCTCAAATGCAGAGGTAAG
WaaB_E276A_F	gctgGCGgcaatgtcatatggaattccgtgtattagttctgattg
WaaB_E276A_R	cattgcCGCcagcAGGGTCATAGGAAATCCCTCAAATG

Table 11 Primers used in this thesis

The related primers used for molecular cloning and mutagenesis are listed here. The Tm temperatures are calculated using NEB Q5 TM CALCULATOR (https://tmcalculator.neb.com/). The annealing temperatures may vary from the Tm calculated for the websites and can be adjusted by cases.

pLou3 or pHIS_MalC2-Tev



pLou3 was derived from the NEB vector pMAL-c2X so has cytoplasmic expression of the fusion protein. New features are N-terminal 6xHis tag before MBP and a Tev protease cleavage site between MBP and the (modified) polylinker.
Unique restriction sites are shown on the map (+ NdeI which cuts twice).

pLOU3 Tev cleavage site and cloning sites:

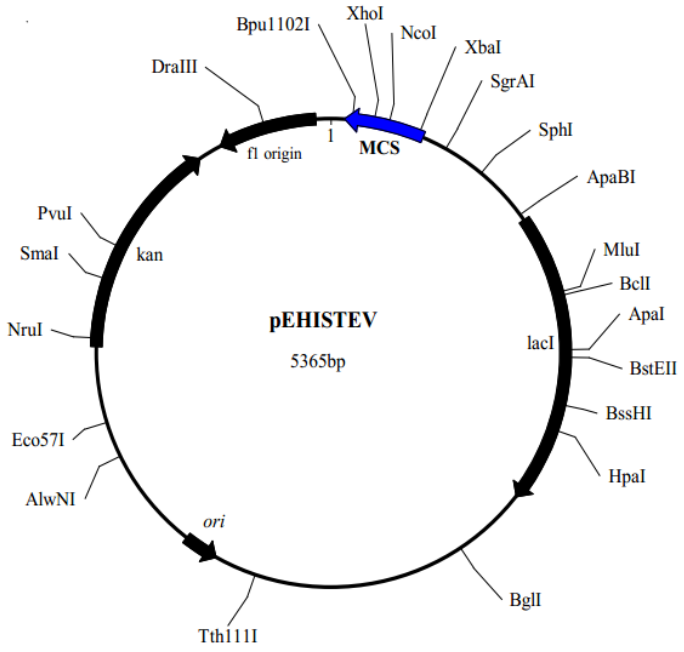
```
AvaI          TEV protease site          NcoI      PmlI    EcoRI   BamHI
CTC GGG GAA AAC CTG TAT TTT CAG GGC GCC ATG GAT CAC GTG GAA TTC GGA TCC
 L   G   E   N   L   Y   F   Q   G   A   M   D   H   V   E   F   G   S

 XbaI    SalI     PstI     HindIII
TCT AGA GTC GAC CTG CAG GCA AGC TTG GCA
 S   R   V   D   L   Q   A   S   L   A
```

Sequence for the 6xHis tag:

```
NdeI                    His-tag                         NdeI
CAT ATG AAA TAT TAC CAT CAC CAT CAC CAT CAC GAT TAC GAT CAT ATG
     M   K   Y   Y   H   H   H   H   H   H   S   Y   S   H   M
```

Based on pET28a(+)?: likely the BglII site in pEHISTEV



T7 Pro seq primer

T7 promoter

lac operator *Xba* I

ATCGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATA
ATTTTG

rbs *Nde* I 6xHis

TTTAACTTTAAGAAGGAGATATACAT ATG TCG TAC TAC CAT CAC CAT CAC CAT CAC GAT TAC GAT ATC CCA

M S Y Y H H H H H H D Y D I P

TEV protease site *Nco* I *Bam*H I *Eco*R I *Sac* I

ACG ACC GAA AAC CTG TAT TTT CAG GGC **GCC ATG GCT GAT ATC GGA TCC GAA TTC GAG CTC CGT**

T T E N L Y F Q* G A M A D I G S E F E L R

Sal I *Hind* III *Not* I *Xho* I

CGA CAA GCT TGC GGC CGC ACT CGA GCA CCA CCA CCA CCA CCA CTG AGA TCC GGC TGC TAA CAA

R Q A C G R T R A P P P P P L R S G C end

*Bpu*1102I T7 terminator

AGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGG

T7 Ter seq prime

GTCTTGAGGGGTTTTTTGCTGAA

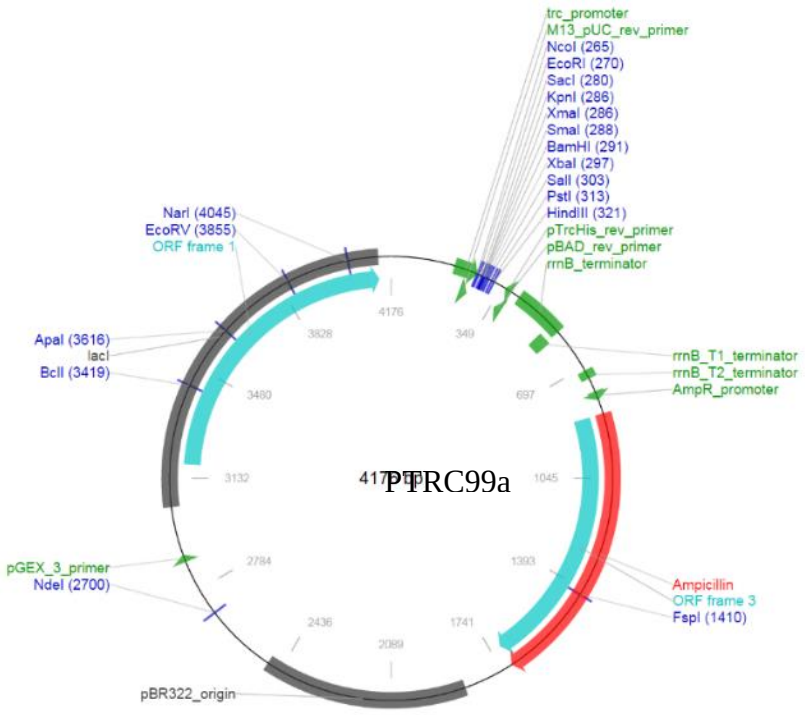


Table 12 Plasmids information in this thesis

The three main plasmids used in the thesis are pHISTEV, pLOU3, pTRC99a. pHISTEV (Kan^R) is derived from pET28 (Novagen) with a 6-histeine-TEV cleavage sequence inserted in front of the MCS, resulting in a plasmid that expresses the protein in an easy-tag-removal manner. pHISTEV uses T7 and IPTG inducible promoter. pLOU3 (Amp^R) is modified from pMalE plasmid (NEB). By inserting a 6-histeine in front of MBP and a TEV cleavage site between MBP and MCS, the pLOU3 is generated and can express the target protein in a MBP fusion mode and the removal of MBP and 6-His tag is also simple. MBP fusion method is a widely used method which can help improve the protein folding and can increase soluble expression level of some specific insoluble proteins, especially the proteins that form inclusion bodies during expression. pMalE uses tac IPTG inducible promoter. pTRC99a (Amp^R) is a high copy vector with strong IPTG inducible promoter trc. The high copy and the strong promoter properties of pTRC99a can maximize the protein expression level. The backbone of pTRC99a is quite simple even without RBS,

which allows the target gene insertion to be freer. This vector is most helpful when expression protein complexes in polycistron.

References:

1. Okuda, S., Sherman, D. J., Silhavy, T. J., Ruiz, N., and Kahne, D. (2016) Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model. *Nature reviews. Microbiology* **14**, 337-345
2. Gan, L., Chen, S., and Jensen, G. J. (2008) Molecular organization of Gram-negative peptidoglycan. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 18953-18957
3. Vollmer, W., and Holtje, J. V. (2004) The architecture of the murein (peptidoglycan) in gram-negative bacteria: vertical scaffold or horizontal layer(s)? *Journal of bacteriology* **186**, 5978-5987
4. Nikaido, H. (2003) Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiology and Molecular Biology Reviews* **67**, 593-656
5. Raetz, C. R., and Whitfield, C. (2002) Lipopolysaccharide endotoxins. *Annual review of biochemistry* **71**, 635-700
6. Berg, H. C. (2003) The rotary motor of bacterial flagella. *Annual review of biochemistry* **72**, 19-54
7. Nan, B., McBride, M. J., Chen, J., Zusman, D. R., and Oster, G. (2014) Bacteria that glide with helical tracks. *Current biology : CB* **24**, R169-173
8. Lai, Y., Rosenshine, I., Leong, J. M., and Frankel, G. (2013) Intimate host attachment: enteropathogenic and enterohaemorrhagic Escherichia coli. *Cellular microbiology* **15**, 1796-1808
9. Laverty, G., Gorman, S. P., and Gilmore, B. F. (2014) Biomolecular Mechanisms of Pseudomonas aeruginosa and Escherichia coli Biofilm Formation. *Pathogens* **3**, 596-632
10. Koebnik, R., Locher, K. P., and Van Gelder, P. (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Molecular microbiology* **37**, 239-253
11. Bos, M. P., Robert, V., and Tommassen, J. (2007) Biogenesis of the gram-negative bacterial outer membrane. *Annual review of microbiology* **61**, 191-214
12. Nikaido, H. (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiology And Molecular Biology Reviews* **67**, 593
13. de Keyzer, J., van der Does, C., and Driessen, A. J. M. (2003) The bacterial translocase: a dynamic protein channel complex. *Cell Mol Life Sci* **60**, 2034-2052
14. Eppens, E. F., Nouwen, N., and Tommassen, J. (1997) Folding of a bacterial outer membrane protein during passage through the periplasm. *Embo J* **16**, 4295-4301
15. Noinaj, N., Kuszak, A. J., Gumbart, J. C., Lukacik, P., Chang, H. S., Easley, N. C., Lithgow, T., and Buchanan, S. K. (2013) Structural insight into the biogenesis of beta-barrel membrane proteins. *Nature* **501**, 385

16. Iadanza, M. G., Higgins, A. J., Schiffrin, B., Calabrese, A. N., Brockwell, D. J., Ashcroft, A. E., Radford, S. E., and Ranson, N. A. (2016) Lateral opening in the intact beta-barrel assembly machinery captured by cryo-EM. *Nat Commun* **7**
17. Gu, Y. H., Li, H. Y., Dong, H. H., Zeng, Y., Zhang, Z. Y., Paterson, N. G., Stansfeld, P. J., Wang, Z. S., Zhang, Y. Z., Wang, W. J., and Dong, C. J. (2016) Structural basis of outer membrane protein insertion by the BAM complex. *Nature* **531**, 64
18. Han, L., Zheng, J. G., Wang, Y., Yang, X., Liu, Y. Q., Sun, C. Q., Cao, B. H., Zhou, H. Z., Ni, D. C., Lou, J. Z., Zhao, Y. F., and Huang, Y. H. (2016) Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. *Nat Struct Mol Biol* **23**, 192-196
19. Sankaran, K., and Wu, H. C. (1994) Lipid Modification Of Bacterial Prolipoprotein - Transfer Of Diacylglyceryl Moiety From Phosphatidylglycerol. *Journal Of Biological Chemistry* **269**, 19701-19706
20. Dev, I. K., and Ray, P. H. (1984) Rapid Assay And Purification Of a Unique Signal Peptidase That Processes the Prolipoprotein From Escherichia-Coli-B. *Journal Of Biological Chemistry* **259**, 1114-1120
21. Gupta, S. D., Gan, K., Schmid, M. B., and Wu, H. C. (1993) Characterization Of a Temperature-Sensitive Mutant Of Salmonella-Typhimurium Defective In Apolipoprotein N-Acyltransferase. *Journal Of Biological Chemistry* **268**, 16551-16556
22. Terada, M., Kuroda, T., Matsuyama, S., and Tokuda, H. (2001) Lipoprotein sorting signals evaluated as the Lo1A-dependent release of lipoproteins from the cytoplasmic membrane of Escherichia coli. *Journal Of Biological Chemistry* **276**, 47690-47694
23. Ito, Y., Kanamaru, K., Taniguchi, N., Miyamoto, S., and Tokuda, H. (2006) A novel ligand bound ABC transporter, LolCDE, provides insights into the molecular mechanisms underlying membrane detachment of bacterial lipoproteins. *Molecular microbiology* **62**, 1064-1075
24. Missiakas, D., Betton, J. M., and Raina, S. (1996) New components of protein folding in extracytoplasmic compartments of Escherichia coli SurA, FkpA and Skp/OmpH. *Molecular microbiology* **21**, 871-884
25. Cronan, J. E. (2003) Bacterial membrane lipids: Where do we stand? *Annual review of microbiology* **57**, 203-224
26. Huijbregts, R. P. H., de Kroon, A. I. P. M., and de Kruijff, B. (2000) Topology and transport of membrane lipids in bacteria. *Bba-Rev Biomembranes* **1469**, 43-61
27. Mi, W., Li, Y., Yoon, S. H., Ernst, R. K., Walz, T., and Liao, M. (2017) Structural basis of MsbA-mediated lipopolysaccharide transport. *Nature* **549**, 233-237
28. Okuda, S., Freinkman, E., and Kahne, D. (2012) Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in E. coli. *Science* **338**, 1214-1217
29. Dong, H., Xiang, Q., Gu, Y., Wang, Z., Paterson, N. G., Stansfeld, P. J., He, C., Zhang, Y., Wang, W., and Dong, C. (2014) Structural basis for outer membrane

lipopolysaccharide insertion. *Nature* **511**, 52-56
30. Noinaj, N., Gumbart, J. C., and Buchanan, S. K. (2017) The beta-barrel assembly machinery in motion. *Nature Reviews Microbiology* **15**
31. Mori, H., and Ito, K. (2001) The Sec protein-translocation pathway. *Trends Microbiol* **9**, 494-500
32. Okuda, S., and Tokuda, H. (2011) Lipoprotein Sorting in Bacteria. *Annual Review Of Microbiology, Vol 65* **65**, 239-259
33. Palmer, T., and Berks, B. C. (2012) The twin-arginine translocation (Tat) protein export pathway. *Nature Reviews Microbiology* **10**, 483-496
34. Galloway, S. M., and Raetz, C. R. (1990) A mutant of Escherichia coli defective in the first step of endotoxin biosynthesis. *Journal of Biological Chemistry* **265**, 6394-6402
35. Raetz, C. R. H. (1990) Biochemistry Of Endotoxins. *Annual review of biochemistry* **59**, 129-170
36. Schnaitman, C. A., and Klena, J. D. (1993) Genetics Of Lipopolysaccharide Biosynthesis In Enteric Bacteria. *Microbiological Reviews* **57**, 655-682
37. Brabetz, W., MullerLoennies, S., Holst, O., and Brade, H. (1997) Deletion of the heptosyltransferase genes rfaC and rfaF in Escherichia coli K-12 results in an Re-type lipopolysaccharide with a high degree of 2-aminoethanol phosphate substitution. *Eur J Biochem* **247**, 716-724
38. Raetz, C. R. H. (1993) Bacterial-Endotoxins - Extraordinary Lipids That Activate Eukaryotic Signal-Transduction. *Journal of bacteriology* **175**, 5745-5753
39. Wyckoff, T. J. O., Raetz, C. R. H., and Jackman, J. E. (1998) Antibacterial and anti-inflammatory agents that target endotoxin. *Trends Microbiol* **6**, 154-159
40. Kelly, T. M., Stachula, S. A., Raetz, C. R. H., and Anderson, M. S. (1993) The Fira Gene Of Escherichia-Coli Encodes Udp-3-O-(R-3-Hydroxymyristoyl)-Glucosamine N-Acyltransferase - the 3rd Step Of Endotoxin Biosynthesis. *Journal Of Biological Chemistry* **268**, 19866-19874
41. Babinski, K. J., and Raetz, C. R. H. (1998) Identification of a gene encoding a novel Escherichia coli UDP-2,3-diacylglucosamine hydrolase. *Faseb J* **12**, A1288-A1288
42. Ray, B. L., and Raetz, C. R. H. (1987) The Biosynthesis Of Gram-Negative Endotoxin - a Novel Kinase In Escherichia-Coli Membranes That Incorporates the 4'-Phosphate Of Lipid-A. *Journal Of Biological Chemistry* **262**, 1122-1128
43. Clementz, T., and Raetz, C. R. H. (1991) A Gene Coding for 3-Deoxy-D-Manno-Octulosonic-Acid Transferase In Escherichia-Coli - Identification, Mapping, Cloning, And Sequencing. *Journal Of Biological Chemistry* **266**, 9687-9696
44. Brozek, K. A., and Raetz, C. R. H. (1990) Biosynthesis Of Lipid-a In Escherichia-Coli - Acyl Carrier Protein-Dependent Incorporation Of Laurate And Myristate. *Journal Of Biological Chemistry* **265**, 15410-15417
45. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., ColladoVides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis,

N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453

46. Ward, A., Reyes, C. L., Yu, J., Roth, C. B., and Chang, G. (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19005-19010
47. Amor, K., Heinrichs, D. E., Frirdich, E., Ziebell, K., Johnson, R. P., and Whitfield, C. (2000) Distribution of core oligosaccharide types in lipopolysaccharides from Escherichia coli. *Infection and immunity* **68**, 1116-1124
48. Appelmelk, B. J., An, Y. Q., Hekker, T. A., Thijs, L. G., MacLaren, D. M., and de Graaf, J. (1994) Frequencies of lipopolysaccharide core types in Escherichia coli strains from bacteraemic patients. *Microbiology* **140 (Pt 5)**, 1119-1124
49. Gibb, A. P., Barclay, G. R., Poxton, I. R., and di Padova, F. (1992) Frequencies of lipopolysaccharide core types among clinical isolates of Escherichia coli defined with monoclonal antibodies. *J Infect Dis* **166**, 1051-1057
50. Currie, C. G., and Poxton, I. R. (1999) The lipopolysaccharide core type of Escherichia coli O157:H7 and other non-O157 verotoxin-producing E. coli. *FEMS Immunol Med Microbiol* **24**, 57-62
51. Di Padova, F. E., Mikol, V., Barclay, G. R., Poxton, I. R., Brade, H., and Rietschel, E. T. (1994) Anti-lipopolysaccharide core antibodies. *Progress in clinical and biological research* **388**, 85-94
52. Bennett-Guerrero, E., McIntosh, T. J., Barclay, G. R., Snyder, D. S., Gibbs, R. J., Mythen, M. G., and Poxton, I. R. (2000) Preparation and preclinical evaluation of a novel liposomal complete-core lipopolysaccharide vaccine. *Infection and immunity* **68**, 6202-6208
53. Di Padova, F. E., Brade, H., Barclay, G. R., Poxton, I. R., Liehl, E., Schuetze, E., Kocher, H. P., Ramsay, G., Schreier, M. H., McClelland, D. B., and et al. (1993) A broadly cross-protective monoclonal antibody binding to Escherichia coli and Salmonella lipopolysaccharides. *Infection and immunity* **61**, 3863-3872
54. Stanislavsky, E. S., and Lam, J. S. (1997) Pseudomonas aeruginosa antigens as potential vaccines. *Fems Microbiol Rev* **21**, 243-277
55. O'Toole, G., Kaplan, H. B., and Kolter, R. (2000) Biofilm formation as microbial development. *Annual review of microbiology* **54**, 49-79
56. Liebau, J., Pettersson, P., Szpryngiel, S., and Maler, L. (2015) Membrane Interaction of the Glycosyltransferase WaaG. *Biophysical journal* **109**, 552-563
57. Yethon, J. A., Heinrichs, D. E., Monteiro, M. A., Perry, M. B., and Whitfield, C. (1998) Involvement of waaY, waaQ, and waaP in the modification of Escherichia coli lipopolysaccharide and their role in the formation of a stable outer membrane. *Journal Of Biological Chemistry* **273**, 26310-26316
58. Walsh, A. G., Matewish, M. J., Burrows, L. L., Monteiro, M. A., Perry, M. B., and Lam, J. S. (2000) Lipopolysaccharide core phosphates are required for viability and intrinsic drug resistance in Pseudomonas aeruginosa. *Molecular*

microbiology **35**, 718-727
59. Creeger, E. S., and Rothfield, L. I. (1979) Cloning of genes for bacterial glycosyltransferases. I. Selection of hybrid plasmids carrying genes for two glucosyltransferases. *The Journal of biological chemistry* **254**, 804-810
60. Heinrichs, D. E., Yethon, J. A., and Whitfield, C. (1998) Molecular basis for structural diversity in the core regions of the lipopolysaccharides of Escherichia coli and Salmonella enterica. *Molecular microbiology* **30**, 221-232
61. Olsthoorn, M. M. A., Petersen, B. O., Schlecht, S., Haverkamp, J., Bock, K., Thomas-Oates, J. E., and Holst, O. (1998) Identification of a Novel Core Type in SalmonellaLipopolysaccharide COMPLETE STRUCTURAL ANALYSIS OF THE CORE REGION OF THE LIPOPOLYSACCHARIDE FROM SALMONELLA ENTERICA sv. ARIZONAE O62. *Journal of Biological Chemistry* **273**, 3817-3829
62. Brade, H. (1999) *Endotoxin in health and disease*, CRC Press
63. Vinogradov, E., Cedzynski, M., Ziolkowski, A., and Swierzko, A. (2001) The structure of the core region of the lipopolysaccharide from Klebsiella pneumoniae O3. 3-deoxy-alpha-D-manno-octulosonic acid (alpha-Kdo) residue in the outer part of the core, a common structural element of Klebsiella pneumoniae O1, O2, O3, O4, O5, O8, and O12 lipopolysaccharides. *Eur J Biochem* **268**, 1722-1729
64. Sadovskaya, I., Brisson, J. R., Thibault, P., Richards, J. C., Lam, J. S., and Altman, E. (2000) Structural characterization of the outer core and the O-chain linkage region of lipopolysaccharide from Pseudomonas aeruginosa serotype O5. *Eur J Biochem* **267**, 1640-1650
65. Knirel' Iu, A., and Kochetkov, N. K. (1994) [Structure of lipopolysaccharides from gram-negative bacteria. III. Structure of O-specific polysaccharides]. *Biokhimiia* **59**, 1784-1851
66. Popoff, M. Y., and Leminor, L. (1985) Expression Of Antigenic Factor-O-54 Is Associated with the Presence Of a Plasmid In Salmonella. *Ann Inst Pasteur Mic* **136B**, 169
67. Keenleyside, W. J., Perry, M., Maclean, L., Poppe, C., and Whitfield, C. (1994) A Plasmid-Encoded Rfb(O/54) Gene-Cluster Is Required for Biosynthesis Of the O/54 Antigen In Salmonella-Enterica Serovar Borreze. *Molecular microbiology* **11**, 437-448
68. Burns, S. M., and Hull, S. I. (1998) Comparison of loss of serum resistance by defined lipopolysaccharide mutants and an acapsular mutant of uropathogenic Escherichia coli O75 : K5. *Infection and immunity* **66**, 4244-4253
69. Joiner, K. A. (1988) Complement Evasion by Bacteria And Parasites. *Annual review of microbiology* **42**, 201-230
70. Mcgrath, B. C., and Osborn, M. J. (1991) Localization Of the Terminal Steps Of O-Antigen Synthesis In Salmonella-Typhimurium. *Journal of bacteriology* **173**, 649-654
71. Liu, D., and Reeves, P. R. (1994) Escherichia-Coli K12 Regains Its O-Antigen. *Microbiol-Uk* **140**, 49-57

72. Liu, D., Haase, A. M., Lindqvist, L., Linberg, A. A., and Reeves, P. R. (1993) Glycosyl Transferases Of O-Antigen Biosynthesis In Salmonella-Enterica - Identification And Characterization Of Transferase Genes Of Group-B, Group-C2, And Group-E1. *Journal of bacteriology* **175**, 3408-3413
73. Susskind, M., Brade, L., Brade, H., and Holst, O. (1998) Identification of a novel heptoglycan of alpha 1 -> 2-linked D-glycero-D-manno-heptopyranose - Chemical and antigenic structure of lipopolysaccharides from Klebsiella pneumoniae ssp. Pneumoniae rough strain R20 (O1(-): K20(-)). *Journal Of Biological Chemistry* **273**, 7006-7017
74. Clarke, B. R., Bronner, D., Keenleyside, W. J., Severn, W. B., Richards, J. C., and Whitfield, C. (1995) Role Of Rfe And Rfbf In the Initiation Of Biosynthesis Of D-Galactan-I, the Lipopolysaccharide O-Antigen From Klebsiella-Pneumoniae Serotype-O1. *Journal of bacteriology* **177**, 5411-5418
75. Guan, S., Clarke, A. J., and Whitfield, C. (2001) Functional analysis of the galactosyltransferases required for biosynthesis of D-galactan I, a component of the lipopolysaccharide O1 antigen of Klebsiella pneumoniae. *Journal of bacteriology* **183**, 3318-3327
76. Kido, N., Torgov, V. I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T., Kato, N., and Jann, K. (1995) Expression Of the O9 Polysaccharide Of Escherichia-Coli - Sequencing Of the Escherichia-Coli O9 Rfb Gene-Cluster, Characterization Of Mannosyl Transferases, And Evidence for an Atp-Binding Cassette Transport-System. *Journal of bacteriology* **177**, 2178-2187
77. Weisgerber, C., and Jann, K. (1982) Glucosyldiphosphoundecaprenol, the Mannose Acceptor In the Synthesis Of the O9 Antigen Of Escherichia-Coli - Biosynthesis And Characterization. *Eur J Biochem* **127**, 165-168
78. Whitfield, C., Amor, P. A., and Koplin, R. (1997) Modulation of the surface architecture of Gram-negative bacteria by the action of surface polymer: Lipid A-core ligase and by determinants of polymer chain length. *Molecular microbiology* **23**, 629-638
79. Keenleyside, W. J., and Whitfield, C. (1996) A novel pathway for O-polysaccharide biosynthesis in Salmonella enterica serovar Borreze. *Journal Of Biological Chemistry* **271**, 28581-28592
80. Keenleyside, W. J., and Whitfield, C. (1995) Lateral Transfer Of Rfb Genes - a Mobilizable Cole1-Type Plasmid Carries the Rfb(O-54) (O-54 Antigen Biosynthesis) Gene-Cluster From Salmonella-Enterica Serovar Borreze. *Journal of bacteriology* **177**, 5247-5253
81. Keenleyside, W. J., Clarke, A. J., and Whitfield, C. (2001) Identification of residues involved in catalytic activity of the inverting glycosyl transferase WbbE from Salmonella enterica serovar borreze. *Journal of bacteriology* **183**, 77-85
82. Slauch, J. M., Lee, A. A., Mahan, M. J., and Meklanos, J. J. (1996) Molecular characterization of the oafA locus responsible for acetylation of Salmonella typhimurium O-antigen: OafA is a member of a family of integral membrane trans-acylases. *Journal of bacteriology* **178**, 5904-5909

83. Verma, N. K., Brandt, J. M., Verma, D. J., and Lindberg, A. A. (1991) Molecular Characterization Of the O-Acetyl Transferase Gene Of Converting Bacteriophage Sf6 That Adds Group Antigen-6 To Shigella-Flexneri. *Molecular microbiology* **5**, 71-75
84. Newton, G. J., Daniels, C., Burrows, L. L., Kropinski, A. M., Clarke, A. J., and Lam, J. S. (2001) Three-component-mediated serotype conversion in Pseudomonas aeruginosa by bacteriophage D3. *Molecular microbiology* **39**, 1237-1247
85. Whitfield, C., Richards, J. C., Perry, M. B., Clarke, B. R., and Maclean, L. L. (1991) Expression Of 2 Structurally Distinct D-Galactan O-Antigens In the Lipopolysaccharide Of Klebsiella-Pneumoniae Serotype-O1. *Journal of bacteriology* **173**, 1420-1431
86. Whitfield, C., Perry, M. B., Maclean, L. L., and Yu, S. H. (1992) Structural-Analysis Of the O-Antigen Side-Chain Polysaccharides In the Lipopolysaccharides Of Klebsiella Serotypes O2(2a), O2(2a,2b), And O2(2a,2c). *Journal of bacteriology* **174**, 4913-4919
87. Kelly, J., Masoud, H., Perry, M. B., Richards, J. C., and Thibault, P. (1996) Separation and characterization of O-deacylated lipooligosaccharides and glycans derived from Moraxella catarrhalis using capillary electrophoresis electrospray mass spectrometry and tandem mass spectrometry. *Analytical biochemistry* **233**, 15-30
88. Kelly, R. F., Severn, W. B., Richards, J. C., Perry, M. B., Maclean, L. L., Tomas, J. M., Merino, S., and Whitfield, C. (1993) Structural Variation In the O-Specific Polysaccharides Of Klebsiella-Pneumoniae Serotype-O1 And Serotype-O8 Lipopolysaccharide - Evidence for Clonal Diversity In Rfb Genes. *Molecular microbiology* **10**, 615-625
89. Maclean, L. L., Whitfield, C., and Perry, M. B. (1993) Characterization Of the Polysaccharide Antigen Of Klebsiella-Pneumoniae O-9 Lipopolysaccharide. *Carbohydrate research* **239**, 325-328
90. Qiao, S., Luo, Q., Zhao, Y., Zhang, X. C., and Huang, Y. (2014) Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. *Nature* **511**, 108-111
91. Sperandeo, P., Deho, G., and Polissi, A. (2009) The lipopolysaccharide transport system of Gram-negative bacteria. *Biochimica et biophysica acta* **1791**, 594-602
92. Simpson, B. W., May, J. M., Sherman, D. J., Kahne, D., and Ruiz, N. (2015) Lipopolysaccharide transport to the cell surface: biosynthesis and extraction from the inner membrane. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**
93. May, J. M., Sherman, D. J., Simpson, B. W., Ruiz, N., and Kahne, D. (2015) Lipopolysaccharide transport to the cell surface: periplasmic transport and assembly into the outer membrane. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**
94. Sperandeo, P., Lau, F. K., Carpentieri, A., De Castro, C., Molinaro, A., Deho,

G., Silhavy, T. J., and Polissi, A. (2008) Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of Escherichia coli. *Journal of bacteriology* **190**, 4460-4469

95. Osborn, M. J., Rick, P. D., and Rasmussen, N. S. (1980) Mechanism Of Assembly Of the Outer-Membrane Of Salmonella-Typhimurium - Translocation And Integration Of an Incomplete Mutant Lipid a into the Outer-Membrane. *Journal Of Biological Chemistry* **255**, 4246-4251
96. Sherman, D. J., Lazarus, M. B., Murphy, L., Liu, C., Walker, S., Ruiz, N., and Kahne, D. (2014) Decoupling catalytic activity from biological function of the ATPase that powers lipopolysaccharide transport. *Proceedings of the National Academy of Sciences* **111**, 4982-4987
97. Sperandeo, P., Villa, R., Martorana, A. M., Samalikova, M., Grandori, R., Deho, G., and Polissi, A. (2011) New insights into the Lpt machinery for lipopolysaccharide transport to the cell surface: LptA-LptC interaction and LptA stability as sensors of a properly assembled transenvelope complex. *Journal of bacteriology* **193**, 1042-1053
98. Sherman, D. J., Okuda, S., Denny, W. A., and Kahne, D. (2013) Validation of inhibitors of an ABC transporter required to transport lipopolysaccharide to the cell surface in Escherichia coli. *Bioorganic & medicinal chemistry* **21**, 4846-4851
99. Villa, R., Martorana, A. M., Okuda, S., Gourlay, L. J., Nardini, M., Sperandeo, P., Deh ò, G., Bolognesi, M., Kahne, D., and Polissi, A. (2013) The Escherichia coli Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains. *Journal of bacteriology* **195**, 1100-1108
100. Suits, M. D., Sperandeo, P., Deho, G., Polissi, A., and Jia, Z. (2008) Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *Journal of molecular biology* **380**, 476-488
101. Okuda, S., Watanabe, S., and Tokuda, H. (2008) A short helix in the C-terminal region of LolA is important for the specific membrane localization of lipoproteins. *FEBS letters* **582**, 2247-2251
102. Tefsen, B., Geurtsen, J., Beckers, F., Tommassen, J., and de Cock, H. (2005) Lipopolysaccharide transport to the bacterial outer membrane in spheroplasts. *The Journal of biological chemistry* **280**, 4504-4509
103. Chng, S. S., Gronenberg, L. S., and Kahne, D. (2010) Proteins required for lipopolysaccharide assembly in Escherichia coli form a transenvelope complex. *Biochemistry* **49**, 4565-4567
104. Botos, I., Majdalani, N., Mayclin, S. J., McCarthy, J. G., Lundquist, K., Wojtowicz, D., Barnard, T. J., Gumbart, J. C., and Buchanan, S. K. (2016) Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens. *Structure* **24**, 965-976
105. Bos, M. P., Tefsen, B., Geurtsen, J., and Tommassen, J. (2004) Identification of an outer membrane protein required for the transport of lipopolysaccharide to the bacterial cell surface. *Proceedings of the National Academy of Sciences of*

the United States of America **101**, 9417-9422
106. Wu, T., McCandlish, A. C., Gronenberg, L. S., Chng, S. S., Silhavy, T. J., and Kahne, D. (2006) Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 11754-11759
107. Hagan, C. L., Silhavy, T. J., and Kahne, D. (2011) beta-Barrel membrane protein assembly by the Bam complex. *Annual review of biochemistry* **80**, 189-210
108. Freinkman, E., Chng, S.-S., and Kahne, D. (2011) The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. *Proceedings of the National Academy of Sciences* **108**, 2486-2491
109. Tran, A. X., Dong, C., and Whitfield, C. (2010) Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in Escherichia coli. *The Journal of biological chemistry* **285**, 33529-33539
110. Bos, M. P., Robert, V., and Tommassen, J. (2007) Biogenesis of the gram-negative bacterial outer membrane. *Annual review of microbiology* **61**, 191-214
111. Freinkman, E., Okuda, S., Ruiz, N., and Kahne, D. (2012) Regulated assembly of the transenvelope protein complex required for lipopolysaccharide export. *Biochemistry* **51**, 4800-4806
112. Bowyer, A., Baardsnes, J., Ajamian, E., Zhang, L., and Cygler, M. (2011) Characterization of interactions between LPS transport proteins of the Lpt system. *Biochemical and biophysical research communications* **404**, 1093-1098
113. Whitfield, C., and Trent, M. S. (2014) Biosynthesis and export of bacterial lipopolysaccharides. *Annual review of biochemistry* **83**, 99-128
114. Ruiz, N., Gronenberg, L. S., Kahne, D., and Silhavy, T. J. (2008) Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5537-5542
115. Narita, S., and Tokuda, H. (2009) Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides. *FEBS letters* **583**, 2160-2164
116. Anthony Ivetac, J. D. C., and Mark S. P. Sansom*. (2007) Dynamics and Function in a Bacterial ABC Transporter: Simulation Studies of the BtuCDF System and Its Components†. *Biochemistry* **46**, 2767-2778
117. Higgins, C. F. (2001) ABC tansporters: physiology, structure and mechanism - an overview *Res Microbiol*, **152**, 205-210
118. Blight, I. B. H. a. M. A. (1991) ABC-ATPases, Adaptable Energy Generators Fuelling Transmembrane Movement of a Variety of Molecules in Organisms from Bacteria to Humans. *J. Mol. Biol.* **293**, 381-399
119. Davidson, A. L., and Chen, J. (2004) ATP-binding cassette transporters in bacteria. *Annual review of biochemistry* **73**, 241-268
120. Delcour, A. H. (2003) Solute uptake through general porins. *Front Biosci* **8**, D1055-D1071

121. Postle, K., and Kadner, R. J. (2003) Touch and go: tying TonB to transport. *Molecular microbiology* **49**, 869-882
122. Covitz, K. M. Y., Panagiotidis, C. H., Hor, L. I., Reyes, M., Treptow, N. A., and Shuman, H. A. (1994) Mutations That Alter the Transmembrane Signaling Pathway In an Atp Binding Cassette (Abc) Transporter. *Embo J* **13**, 1752-1759
123. Dassa, E., and Bouige, P. (2001) The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Res Microbiol* **152**, 211-229
124. Dinh, T., Paulsen, I. T., and Saier, M. H. (1994) A Family Of Extracytoplasmic Proteins That Allow Transport Of Large Molecules across the Outer Membranes Of Gram-Negative Bacteria. *Journal of bacteriology* **176**, 3825-3831
125. Paulsen, I. T., Park, J. H., Choi, P. S., and Saier, M. H. (1997) A family of Gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from Gram-negative bacteria. *Fems Microbiol Lett* **156**, 1-8
126. V, K., Sharff, A., Koronakis, E., Luisi, B., and Hughes, C. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914-919
127. Andersen, C., Koronakis, E., Bokma, E., Eswaran, J., Humphreys, D., Hughes, C., and Koronakis, V. (2002) Transition to the open state of the TolC periplasmic tunnel entrance. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11103-11108
128. Delepelaire, P., and Wandersman, C. (1998) The SecB chaperone is involved in the secretion of the Serratia marcescens HasA protein through an ABC transporter. *Embo J* **17**, 936-944
129. Palacios, J. L., Zaror, I., Martinez, P., Uribe, F., Opazo, P., Socias, T., Gidekel, M., and Venegas, A. (2001) Subset of hybrid eukaryotic proteins is exported by the type I secretion system of Erwinia chrysanthemi. *Journal of bacteriology* **183**, 1346-1358
130. Fernandez, L. A., and de Lorenzo, V. (2001) Formation of disulphide bonds during secretion of proteins through the periplasmic-independent type I pathway. *Molecular microbiology* **40**, 332-346
131. Sapriel, G., Wandersman, C., and Delepelaire, P. (2002) The N terminus of the HasA protein and the SecB chaperone cooperate in the efficient targeting and secretion of HasA via the ATP-binding cassette transporter. *Journal Of Biological Chemistry* **277**, 6726-6732
132. Yuan, Y. R., Blecker, S., Martsinkevich, O., Millen, L., Thomas, P. J., and Hunt, J. F. (2001) The crystal structure of the MJ0796 ATP-binding cassette - Implications for the structural consequences of ATP hydrolysis in the active site of an ABC transporter. *Journal Of Biological Chemistry* **276**, 32313-32321
133. Karpowich, N., Martsinkevich, O., Millen, L., Yuan, Y. R., Dai, P. L., MacVey, K., Thomas, P. J., and Hunt, J. F. (2001) Crystal structures of the MJ1267 ATP binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter. *Structure* **9**, 571-586

134. Verdon, G., Albers, S. V., van Oosterwijk, N., Dijkstra, B. W., Driessen, A. J. M., and Thunnissen, A. M. W. H. (2003) Formation of the productive ATP-Mg2+-bound dimer of GlcV, an ABC-ATPase from Sulfolobus solfataricus. *Journal of molecular biology* **334**, 255-267
135. Smith, P. C., Karpowich, N., Millen, L., Moody, J. E., Rosen, J., Thomas, P. J., and Hunt, J. F. (2002) ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell* **10**, 139-149
136. Moody, J. E., Millen, L., Binns, D., Hunt, J. F., and Thomas, P. J. (2002) Cooperative, ATP-dependent association of the nucleotide binding cassettes during the catalytic cycle of ATP-binding cassette transporters. *Journal Of Biological Chemistry* **277**, 21111-21114
137. Diederichs, K., Diez, J., Greller, G., Muller, C., Breed, J., Schnell, C., Vonrhein, C., Boos, W., and Welte, W. (2000) Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon Thermococcus litoralis. *Embo J* **19**, 5951-5961
138. Chen, J., Lu, G., Lin, J., Davidson, A. L., and Quiocho, F. A. (2003) A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol Cell* **12**, 651-661
139. Hung, L. W., Wang, I. X. Y., Nikaido, K., Liu, P. Q., Ames, G. F. L., and Kim, S. H. (1998) Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature* **396**, 703-707
140. Gaudet, R., and Wiley, D. C. (2001) Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. *Embo J* **20**, 4964-4972
141. Verdon, G., Albers, S. V., Dijkstra, B. W., Driessen, A. J. M., and Thunnissen, A. M. W. H. (2003) Crystal structures of the ATPase subunit of the glucose ABC transporter from Sulfolobus solfataricus: Nucleotide-free and nucleotide-bound conformations. *Journal of molecular biology* **330**, 343-358
142. Locher, K. P., Lee, A. T., and Rees, D. C. (2002) The E-coli BtuCD structure: A framework for ABC transporter architecture and mechanism. *Science* **296**, 1091-1098
143. Liu, H., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC biotechnology* **8**, 91
144. Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **43**, 186-190
145. Waterman, D. G., Winter, G., Parkhurst, J. M., Fuentes-Montero, L., Hattne, J., Brewster, A., Sauter, N. K., and Evans, G. (2013) The DIALS framework for integration software.
146. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., and McCoy, A. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235-242

147. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr D* **69**, 1204-1214
148. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D* **66**, 479-485
149. Taylor, G. L. (2010) Introduction to phasing. *Acta Crystallogr D Biol Crystallogr* **66**, 325-338
150. Terwilliger, T. (2004) SOLVE and RESOLVE: automated structure solution, density modification, and model building. *J Synchrotron Radiat* **11**, 49-52
151. Cowtan, K. D., and Zhang, K. Y. J. (1999) Density modification for macromolecular phase improvement. *Prog Biophys Mol Bio* **72**, 245-270
152. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* **54**, 905-921
153. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature protocols* **3**, 1171-1179
154. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D* **62**, 1002-1011
155. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., and Grosse-Kunstleve, R. W. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213-221
156. Prince, C., and Jia, Z. (2015) An Unexpected Duo: Rubredoxin Binds Nine TPR Motifs to Form LapB, an Essential Regulator of Lipopolysaccharide Synthesis. *Structure* **23**, 1500-1506
157. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126-2132
158. Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G. N. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2184-2195
159. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D* **67**, 355-367
160. Weng, J. W., Gu, S., Gao, X., Huang, X. H., and Wang, W. N. (2017) Maltose-binding protein effectively stabilizes the partially closed conformation of the ATP-binding cassette transporter MalFGK(2). *Phys Chem Chem Phys* **19**, 9366-9373
161. Ivetac, A., Campbell, J. D., and Sansom, M. S. P. (2007) Dynamics and function

in a bacterial ABC transporter: Simulation studies of the BtuCDF system and its components. *Biochemistry* **46**, 2767-2778

162. Perez, C., Kohler, M., Janser, D., Pardon, E., Steyaert, J., Zenobi, R., and Locher, K. P. (2017) Structural basis of inhibition of lipid-linked oligosaccharide flippase PglK by a conformational nanobody. *Sci Rep-Uk* **7**
163. Chang, G., and Roth, C. B. (2001) Structure of MsbA from E-coli: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters (Retracted Article. See vol 314, pg 1875, 2006). *Science* **293**, 1793-1800
164. Narita, S., Tanaka, K., Matsuyama, S., and Tokuda, H. (2002) Disruption of lolCDE, encoding an ATP-binding cassette transporter, is lethal for Escherichia coli and prevents release of lipoproteins from the inner membrane. *Journal of bacteriology* **184**, 1417-1422
165. Taniguchi, N., and Tokuda, H. (2008) Molecular events involved in a single cycle of ligand transfer from an ATP binding cassette transporter, LolCDE, to a molecular chaperone, LolA. *The Journal of biological chemistry* **283**, 8538-8544
166. Sherman, D. J., Lazarus, M. B., Murphy, L., Liu, C., Walker, S., Ruiz, N., and Kahne, D. (2014) Decoupling catalytic activity from biological function of the ATPase that powers lipopolysaccharide transport. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 4982-4987
167. Simpson, B. W., Owens, T. W., Orabella, M. J., Davis, R. M., May, J. M., Trauger, S. A., Kahne, D., and Ruiz, N. (2016) Identification of Residues in the Lipopolysaccharide ABC Transporter That Coordinate ATPase Activity with Extractor Function. *Mbio* **7**
168. Luo, Q. S., Yang, X., Yu, S., Shi, H. G., Wang, K., Xiao, L., Zhu, G. Y., Sun, C. Q., Li, T. T., Li, D. F., Zhang, X. Z., Zhou, M., and Huang, Y. H. (2017) Structural basis for lipopolysaccharide extraction by ABC transporter LptB(2)FG. *Nat Struct Mol Biol* **24**, 469
169. Korkhov, V. M., Mireku, S. A., and Locher, K. P. (2012) Structure of AMP-PNP-bound vitamin B12 transporter BtuCD-F. *Nature* **490**, 367-372
170. Control, C. f. D., and Prevention. (2007) Multistate outbreak of Salmonella serotype Tennessee infections associated with peanut butter--United States, 2006-2007. *MMWR. Morbidity and mortality weekly report* **56**, 521
171. Trevejo, R., Courtney, J., Starr, M., and Vugia, D. (2003) Epidemiology of salmonellosis in California, 1990–1999: morbidity, mortality, and hospitalization costs. *American Journal of Epidemiology* **157**, 48-57
172. Control, C. f. D., and Prevention. (2003) Outbreaks of Salmonella serotype enteritidis infection associated with eating shell eggs--United States, 1999-2001. *MMWR. Morbidity and mortality weekly report* **51**, 1149
173. Murray, G. L., Attridge, S. R., and Morona, R. (2003) Regulation of Salmonella typhimurium lipopolysaccharide O antigen chain length is required for virulence; identification of FepE as a second Wzz. *Molecular microbiology* **47**, 1395-1406
174. Klena, J., Zhang, P., Schwartz, O., Hull, S., and Chen, T. (2005) The core

lipopolysaccharide of Escherichia coli is a ligand for the dendritic-cell-specific intercellular adhesion molecule nonintegrin CD209 receptor. *Journal of bacteriology* **187**, 1710-1715

175. Valtonen, M., Plosila, M., Valtonen, V., and Mäkelä, P. (1975) Effect of the quality of the lipopolysaccharide on mouse virulence of Salmonella enteritidis. *Infection and immunity* **12**, 828-832
176. Yethon, J. A., Gunn, J. S., Ernst, R. K., Miller, S. I., Laroche, L., Malo, D., and Whitfield, C. (2000) Salmonella enterica Serovar TyphimuriumwaaP Mutants Show Increased Susceptibility to Polymyxin and Loss of Virulence In Vivo. *Infection and immunity* **68**, 4485-4491
177. Hoare, A., Bittner, M., Carter, J., Alvarez, S., Zaldívar, M., Bravo, D., Valvano, M. A., and Contreras, I. (2006) The outer core lipopolysaccharide of Salmonella enterica serovar Typhi is required for bacterial entry into epithelial cells. *Infection and immunity* **74**, 1555-1564
178. Devyatyarova-Johnson, M., Rees, I. H., Robertson, B. D., Turner, M. W., Klein, N. J., and Jack, D. L. (2000) The Lipopolysaccharide Structures of Salmonella enterica Serovar Typhimurium and Neisseria gonorrhoeaeDetermine the Attachment of Human Mannose-Binding Lectin to Intact Organisms. *Infection and immunity* **68**, 3894-3899
179. Morgan, E., Campbell, J. D., Rowe, S. C., Bispham, J., Stevens, M. P., Bowen, A. J., Barrow, P. A., Maskell, D. J., and Wallis, T. S. (2004) Identification of host‐specific colonization factors of Salmonella enterica serovar Typhimurium. *Molecular microbiology* **54**, 994-1010
180. Qian, J., Garrett, T. A., and Raetz, C. R. (2014) In vitro assembly of the outer core of the lipopolysaccharide from Escherichia coli K-12 and Salmonella typhimurium. *Biochemistry* **53**, 1250-1262
181. Kadam, S., Rehemtulla, A., and Sanderson, K. (1985) Cloning of rfaG, B, I, and J genes for glycosyltransferase enzymes for synthesis of the lipopolysaccharide core of Salmonella typhimurium. *Journal of bacteriology* **161**, 277-284
182. Pradel, E., Parker, C. T., and Schnaitman, C. A. (1992) Structures of the rfaB, rfaI, rfaJ, and rfaS genes of Escherichia coli K-12 and their roles in assembly of the lipopolysaccharide core. *Journal of bacteriology* **174**, 4736-4745
183. Ruan, X., Loyola, D. E., Marolda, C. L., Perez-Donoso, J. M., and Valvano, M. A. (2012) The WaaL O-antigen lipopolysaccharide ligase has features in common with metal ion-independent inverting glycosyltransferases. *Glycobiology* **22**, 288-299
184. Su, J., Timbely, D., Zhu, M., Hua, X., Liu, B., Pang, Y., Shen, H., Qi, J., and Yang, Y. (2009) RfaB, a galactosyltransferase, contributes to the resistance to detergent and the virulence of Salmonella enterica serovar Enteritidis. *Medical microbiology and immunology* **198**, 185-194
185. Martinez-Fleites, C., Proctor, M., Roberts, S., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2006) Insights into the synthesis of lipopolysaccharide and antibiotics through the structures of two retaining glycosyltransferases from family GT4. *Chemistry & biology* **13**, 1143-1152

186. Chang, A., Singh, S., Phillips, G. N., and Thorson, J. S. (2011) Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr Opin Biotech* **22**, 800-808
187. Schuman, B., Evans, S. V., and Fyles, T. M. (2013) Geometric Attributes of Retaining Glycosyltransferase Enzymes Favor an Orthogonal Mechanism. *Plos One* **8**
188. Ness, S. R., de Graaff, R. A. G., Abrahams, J. P., and Pannu, N. S. (2004) CRANK: New methods for automated macromolecular crystal structure solution. *Structure* **12**, 1753-1761
189. de Graaff, R. A. G., Hilge, M., van der Plas, J. L., and Abrahams, J. P. (2001) Matrix methods for solving protein substructures of chlorine and sulfur from anomalous data. *Acta Crystallogr D* **57**, 1857-1862
190. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr D* **58**, 1772-1779
191. Pannu, N. S., and Read, R. J. (2004) The application of multivariate statistical techniques improves single-wavelength anomalous diffraction phasing. *Acta Crystallogr D* **60**, 22-27
192. Abrahams, J. P., and Leslie, A. G. W. (1996) Methods used in the structure determination of bovine mitochondrial F-1 ATPase. *Acta Crystallogr D* **52**, 30-42
193. Cowtan, K. (1994) 'DM': An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Prorein Crystallogr.* **31**, 34-38
194. Sheldrick, G. M. (2002) Macromolecular phasing with SHELXE. *Z Kristallogr* **217**, 644-650
195. Cowtan, K. (2000) General quadratic functions in real and reciprocal space and their application to likelihood phasing. *Acta Crystallogr D* **56**, 1612-1621
196. Terwilliger, T. C. (2000) Maximum-likelihood density modification. *Acta Crystallogr D* **56**, 965-972
197. Terwilliger, T. C. (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D* **59**, 38-44
198. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**, 458-463
199. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240-255
200. French, S., and Wilson, K. (1978) On the treatment of negative intensity observations. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **34**, 517-525
201. Pannu, N. S., McCoy, A. J., and Read, R. J. (2003) Application of the complex multivariate normal distribution to crystallographic methods with insights into multiple isomorphous replacement phasing. *Acta Crystallogr D* **59**, 1801-1808
202. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L.

C., and Read, R. J. (2007) Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674
203. Dong, A. P., Xu, X. H., and Edward, A. M. (2007) In situ proteolysis for protein crystallization and structure determination. *Nat Methods* **4**, 1019-1021
204. Schrodinger, LLC. (2015) The PyMOL Molecular Graphics System, Version 1.8.
205. Celniker, G., Nimrod, G., Ashkenazy, H., Glaser, F., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. (2013) ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Israel Journal of Chemistry* **53**, 199-206
206. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003) ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* **19**, 163-164
207. Silhavy, T. J., Kahne, D., and Walker, S. (2010) The bacterial cell envelope. *Cold Spring Harbor perspectives in biology* **2**, a000414
208. Dong, J. H., Yang, G. Y., and Mchaourab, H. S. (2005) Structural basis of energy transduction in the transport cycle of MsbA. *Science* **308**, 1023-1028
209. Liu, D., Cole, R. A., and Reeves, P. R. (1996) An O-antigen processing function for Wzx (RfbX): A promising candidate for O-unit flippase. *Journal of bacteriology* **178**, 2102-2107
210. Bengoechea, J. A., Pinta, E., Salminen, T., Oertelt, C., Holst, O., Radziejewska-Lebrecht, J., Piotrowska-Seget, Z., Venho, R., and Skurnik, M. (2002) Functional characterization of gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of Yersinia enterocolitica serotype O : 8. *Journal of bacteriology* **184**, 4277-4287
211. Heinrichs, D. E., Monteiro, M. A., Perry, M. B., and Whitfield, C. (1998) The assembly system for the lipopolysaccharide R2 core-type of Escherichia coli is a hybrid of those found in Escherichia coli K-12 and Salmonella enterica - Structure and function of the R2 WaaK and WaaL homologs. *Journal Of Biological Chemistry* **273**, 8849-8859
212. Kaniuk, N. A., Vinogradov, E., and Whitfield, C. (2004) Investigation of the structural requirements in the lipopolysaccharide core acceptor for ligation of O antigens in the genus Salmonella - WaaL "ligase" is not the sole determinant of acceptor specificity. *Journal Of Biological Chemistry* **279**, 36470-36480
213. Abeyrathne, P. D., Daniels, C., Poon, K. K. H., Matewish, M. J., and Lam, J. S. (2005) Functional characterization of WaaL, a ligase associated with linking O-antigen polysaccharide to the core of Pseudomonas aeruginosa lipopolysaccharide. *Journal of bacteriology* **187**, 3002-3012
214. Berry, M. C., McGhee, G. C., Zhao, Y. F., and Sundin, G. W. (2009) Effect of a waaL mutation on lipopolysaccharide composition, oxidative stress survival, and virulence in Erwinia amylovora. *Fems Microbiol Lett* **291**, 80-87
215. Xu, L. L., Wang, Q. Y., Xiao, J. F., Liu, Q., Wang, X., Chen, T., and Zhang, Y. X. (2010) Characterization of Edwardsiella tarda waaL: roles in lipopolysaccharide biosynthesis, stress adaptation, and virulence toward fish.

Arch Microbiol **192**, 1039-1047
216. Jerry D. King, D. K. n., Erin L. Westman, Joseph S. Lam. (2009) Lipopolysaccharide biosynthesis in Pseudomonas aeruginosa. *Innate Immunity* **15**, 261-312
217. Mahalakshmi, S., Sunayana, M. R., SaiSree, L., and Reddy, M. (2014) yciM is an essential gene required for regulation of lipopolysaccharide synthesis in Escherichia coli. *Molecular microbiology* **91**, 145-157
218. Anderson, M. S., and Raetz, C. R. H. (1987) Biosynthesis Of Lipid-a Precursors In Escherichia-Coli - a Cytoplasmic Acyltransferase That Converts Udp-N-Acetylglucosamine To Udp-3-0-(R-3-Hydroxymyristoyl)-N-Acetylglucosamine. *Journal Of Biological Chemistry* **262**, 5159-5169
219. Wyckoff, T. J. O., Lin, S. H., Cotter, R. J., Dotson, G. D., and Raetz, C. R. H. (1998) Hydrocarbon rulers in UDP-N-acetylglucosamine acyltransferases. *Journal Of Biological Chemistry* **273**, 32369-32372
220. Wyckoff, T. J. O., and Raetz, C. R. H. (1999) The active site of Escherichia coli UDP-N-acetylglucosamine acyltransferase - Chemical modification and site-directed mutagenesis. *Journal Of Biological Chemistry* **274**, 27047-27055
221. Anderson, M. S., Bull, H. G., Galloway, S. M., Kelly, T. M., Mohan, S., Radika, K., and Raetz, C. R. H. (1993) Udp-N-Acetylglucosamine Acyltransferase Of Escherichia-Coli - the 1st Step Of Endotoxin Biosynthesis Is Thermodynamically Unfavorable. *Journal Of Biological Chemistry* **268**, 19858-19865
222. Anderson, M. S., Robertson, A. D., Macher, I., and Raetz, C. R. H. (1988) Biosynthesis Of Lipid-a In Escherichia-Coli - Identification Of Udp-3-O-[(R)-3-Hydroxymyristoyl]-Alpha-D-Glucosamine as a Precursor Of Udp-N,2,O-3-Bis[(R)-3-Hydroxymyristoyl]-Alpha-D-Glucosamine. *Biochemistry* **27**, 1908-1917
223. Katherine Young, L. L. S., David Bramhill§, Patricia Cameron, Suzanne S. Eveland§Christian R. H. Raetzi, Sheryl A. Hyland§, and Matt S. Anderson. (1995) The envA Permeability/Cell Division Gene of Escherichia coli Encodes the Second Enzyme of Lipid A Biosynthesis. *The Journal of biological chemistry* **270**, 30384-30391
224. Sorensen, P. G., Lutkenhaus, J., Young, K., Eveland, S. S., Anderson, M. S., and Raetz, C. R. H. (1996) Regulation of UDP-3-0-[R-3-hydroxymyristoyl]-N-acetylglucosamine deacetylase in Escherichia coli - The second enzymatic step of lipid a biosynthesis. *Journal Of Biological Chemistry* **271**, 25898-25905
225. Jackman, J. E., Raetz, C. R. H., and Fierke, C. A. (1999) UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase of Escherichia coli is a zinc metalloenzyme. *Biochemistry* **38**, 1902-1911
226. Jackman, J. E., Raetz, C. R. H., and Fierke, C. A. (2001) Site-directed mutagenesis of the bacterial metalloamidase UDP-(3-O-acyl)-N-acetylglucosamine deacetylase (LpxC). Identification of the zinc binding site. *Biochemistry* **40**, 514-523
227. Bulawa, C. E., and Raetz, C. R. H. (1984) The Biosynthesis Of Gram-Negative

Endotoxin - Identification And Function Of Udp-2,3-Diacylglucosamine In Escherichia-Coli. *Journal Of Biological Chemistry* **259**, 4846-4851

228. Ray, B. L., Painter, G., and Raetz, C. R. H. (1984) The Biosynthesis Of Gram-Negative Endotoxin - Formation Of Lipid a Disaccharides From Monosaccharide Precursors In Extracts Of Escherichia-Coli. *Journal Of Biological Chemistry* **259**, 4852-4859
229. Carty, S. M., Sreekumar, K. R., and Raetz, C. R. H. (1999) Effect of cold shock on lipid A biosynthesis in Escherichia coli - Induction at 12 degrees C of an acyltransferase specific for palmitoleoyl-acyl carrier protein. *Journal Of Biological Chemistry* **274**, 9677-9685
230. Zhou, Z. M., Lin, S. H., Cotter, R. J., and Raetz, C. R. H. (1999) Lipid a modifications characteristic of Salmonella typhimurium are induced by NH4VO3 in Escherichia coli K12 - Detection of 4-amino-4-deoxy-L-arabinose, phosphoethanolamine and palmitate. *Journal Of Biological Chemistry* **274**, 18503-18514
231. Lin Guo, K. B. L., John S. Gunn, Brian Bainbridge Richard P. Darveau, Murray Hackett, Samuel I. Miller†. (1997) Regulation of Lipid A Modifications by Salmonella typhimurium Virulence Genes phoP-phoQ. *Science* **276-11**, 250
232. Gunn, J. S., Lim, K. B., Krueger, J., Kim, K., Guo, L., Hackett, M., and Miller, S. I. (1998) PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Molecular microbiology* **27**, 1171-1182
233. Helander, I. M., Kilpelainen, I., and Vaara, M. (1994) Increased Substitution Of Phosphate Groups In Lipopolysaccharides And Lipid-a Of the Polymyxin-Resistant Pmra Mutants Of Salmonella-Typhimurium - a P-31-Nmr Study. *Molecular microbiology* **11**, 481-487
234. Nummila, K., Kilpelainen, I., Zahringer, U., Vaara, M., and Helander, I. M. (1995) Lipopolysaccharides Of Polymyxin-B-Resistant Mutants Of Escherichia-Coli Are Extensively Substituted by 2-Aminoethyl Pyrophosphate And Contain Aminoarabinose In Lipid-A. *Molecular microbiology* **16**, 271-278
235. Trent, M. S., Ribeiro, A. A., Doerrler, W. T., Lin, S. H., Cotter, R. J., and Raetz, C. R. H. (2001) Accumulation of a polyisoprene-linked amino sugar in polymyxin-resistant Salmonella typhimurium and Escherichia coli - Structural characterization and transfer to lipid A in the periplasm. *Journal Of Biological Chemistry* **276**, 43132-43144
236. Guo, L., Lim, K. B., Poduje, C. M., Daniel, M., Gunn, J. S., Hackett, M., and Miller, S. I. (1998) Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell* **95**, 189-198
237. Bishop, R. E., Gibbons, H. S., Guina, T., Trent, M. S., Miller, S. I., and Raetz, C. R. H. (2000) Transfer of palmitate from phospholipids to lipid A in outer membranes of Gram-negative bacteria. *Embo J* **19**, 5071-5080
238. Li, C. J., Guan, Z. Q., Liu, D., and Raetz, C. R. H. (2011) Pathway for lipid A biosynthesis in Arabidopsis thaliana resembling that of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of*

America **108**, 11387-11392
239. Patricia G. Sorensen, J. L., Katherine Youngi, Suzanne S. Eveland,Matt S. Anderson, and Christian R. H. Raetz. (1996) Regulation of UDP-3-O-[R-3-hydroxymyristoyl]-N-acetylglucosamine Deacetylase in Escherichia coli. *The Journal of biological chemistry* **271**, 25898 –25905
240. Fuhrer, F., Langklotz, S., and Narberhaus, F. (2006) The C-terminal end of LpxC is required for degradation by the FtsH protease. *Molecular microbiology* **59**, 1025-1036
241. Raetz, C. R., Reynolds, C. M., Trent, M. S., and Bishop, R. E. (2007) Lipid A modification systems in gram-negative bacteria. *Annual review of biochemistry* **76**, 295-329
242. Klein, G., Kobylak, N., Lindner, B., Stupak, A., and Raina, S. (2014) Assembly of lipopolysaccharide in Escherichia coli requires the essential LapB heat shock protein. *The Journal of biological chemistry* **289**, 14829-14853
243. Klein, G., Kobylak, N., Lindner, B., Stupak, A., and Raina, S. (2014) Assembly of Lipopolysaccharide in Escherichia coli Requires the Essential LapB Heat Shock Protein. *Journal Of Biological Chemistry* **289**, 14829-14853
244. Nicolaes, V., El Hajjaji, H., Davis, R. M., Van der Henst, C., Depuydt, M., Leverrier, P., Aertsen, A., Haufroid, V., de Choudens, S. O., De Bolle, X., Ruiz, N., and Collet, J. F. (2014) Insights into the Function of YciM, a Heat Shock Membrane Protein Required To Maintain Envelope Integrity in Escherichia coli. *Journal of bacteriology* **196**, 300-309
245. Mahalakshmi, S., Sunayana, M. R., SaiSree, L., and Reddy, M. (2014) yciM is an essential gene required for regulation of lipopolysaccharide synthesis in Escherichia coli. *Molecular microbiology* **91**, 145-157
246. Graham, A. I., Hunt, S., Stokes, S. L., Bramall, N., Bunch, J., Cox, A. G., McLeod, C. W., and Poole, R. K. (2009) Severe Zinc Depletion of Escherichia coli ROLES FOR HIGH AFFINITY ZINC BINDING BY ZinT, ZINC TRANSPORT AND ZINC-INDEPENDENT PROTEINS. *Journal Of Biological Chemistry* **284**, 18377-18389
247. Vanduyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) Atomic Structures Of the Human Immunophilin Fkbp-12 Complexes with Fk506 And Rapamycin. *Journal of molecular biology* **229**, 105-124
248. Lin, K., Simossis, V. A., Taylor, W. R., and Heringa, J. (2005) A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics* **21**, 152-159
249. Prince, C., and Jia, Z. C. (2015) An Unexpected Duo: Rubredoxin Binds Nine TPR Motifs to Form LapB, an Essential Regulator of Lipopolysaccharide Synthesis. *Structure* **23**, 1500-1506
250. Wang, X. Y., and Quinn, P. J. (2010) Lipopolysaccharide: Biosynthetic pathway and structure modification. *Prog Lipid Res* **49**, 97-107
251. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K‐12 in‐frame, single‐gene knockout mutants: the Keio

collection. *Molecular systems biology* **2**

252. Ruiz, N., Kahne, D., and Silhavy, T. J. (2009) Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nature reviews. Microbiology* **7**, 677-683
253. Tefsen, B., Bos, M. P., Beckers, F., Tommassen, J., and de Cock, H. (2005) MsbA is not required for phospholipid transport in Neisseria meningitidis. *Journal Of Biological Chemistry* **280**, 35961-35966
254. Kol, M. A., van Dalen, A., de Kroon, A. I. P. M., and de Kruijff, B. (2003) Translocation of phospholipids is facilitated by a subset of membrane-spanning proteins of the bacterial cytoplasmic membrane. *Journal Of Biological Chemistry* **278**, 24586-24593
255. Dalebroux, Z. D., Matamouros, S., Whittington, D., Bishop, R. E., and Miller, S. I. (2014) PhoPQ regulates acidic glycerophospholipid content of the Salmonella Typhimurium outer membrane. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1963-1968
256. Dalebroux, Z. D., Edrozo, M. B., Pfuetzner, R. A., Ressl, S., Kulasekara, B. R., Blanc, M. P., and Miller, S. I. (2015) Delivery of cardiolipins to the Salmonella outer membrane is necessary for survival within host tissues and virulence. *Cell host & microbe* **17**, 441-451
257. Lev, S. (2010) Non-vesicular lipid transport by lipid transfer proteins and beyond. *Nature Reviews Molecular Cell Biology* **11**, 739-750
258. Schlattner, U., Tokarska-Schlattner, M., Rousseau, D., Boissan, M., Mannella, C., Epand, R., and Lacombe, M. L. (2014) Mitochondrial cardiolipin/phospholipid trafficking: the role of membrane contact site complexes and lipid transfer proteins. *Chemistry and physics of lipids* **179**, 32-41
259. Schlame, M. (2008) Thematic review series: Glycerolipids - Cardiolipin synthesis for the assembly of bacterial and mitochondrial membranes. *J Lipid Res* **49**, 1607-1620
260. Tuson, H. H., Auer, G. K., Renner, L. D., Hasebe, M., Tropini, C., Salick, M., Crone, W. C., Gopinathan, A., Huang, K. C., and Weibel, D. B. (2012) Measuring the stiffness of bacterial cells from growth rates in hydrogels of tunable elasticity. *Molecular microbiology* **84**, 874-891
261. Osman, C., Voelker, D. R., and Langer, T. (2011) Making heads or tails of phospholipids in mitochondria. *The Journal of cell biology* **192**, 7-16
262. Paradies, G., Paradies, V., De Benedictis, V., Ruggiero, F. M., and Petrosillo, G. (2014) Functional role of cardiolipin in mitochondrial bioenergetics. *Biochimica et biophysica acta* **1837**, 408-417
263. Yu, F., He, F., Yao, H., Wang, C., Wang, J., Li, J., Qi, X., Xue, H., Ding, J., and Zhang, P. (2015) Structural basis of intramitochondrial phosphatidic acid transport mediated by Ups1-Mdm35 complex. *EMBO reports* **16**, 813-823
264. Miliara, X., Garnett, J. A., Tatsuta, T., Abid Ali, F., Baldie, H., Perez-Dorado, I., Simpson, P., Yague, E., Langer, T., and Matthews, S. (2015) Structural insight into the TRIAP1/PRELI-like domain family of mitochondrial phospholipid

transfer complexes. *EMBO reports* **16**, 824-835
265. Malinverni, J. C., and Silhavy, T. J. (2009) An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 8009-8014
266. Murakami, S., Nakashima, R., Yamashita, E., and Yamaguchi, A. (2002) Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587-593
267. Ramraj, V., Evans, G., Diprose, J., and Esnouf, R. (2012) Nearest-cell: a fast and easy tool for locating crystal matches in the PDB. *Acta Crystallographica Section D: Biological Crystallography* **68**, 1697-1700
268. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* **66**, 12-21
269. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic acids research* **38**, W545-W549
270. Boltes, I., Czapinska, H., Kahnert, A., von Bulow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M. A., and Uson, I. (2001) 1.3 angstrom structure of arylsulfatase from Pseudomonas aeruginosa establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family. *Structure* **9**, 483-491
271. Campeotto, I., Percy, M. G., MacDonald, J. T., Forster, A., Freemont, P. S., and Grundling, A. (2014) Structural and Mechanistic Insight into the Listeria monocytogenes Two-enzyme Lipoteichoic Acid Synthesis System. *Journal Of Biological Chemistry* **289**, 28054-28069
272. Chu, C. T., Ji, J., Dagda, R. K., Jiang, J. F., Tyurina, Y. Y., Kapralov, A. A., Tyurin, V. A., Yanamala, N., Shrivastava, I. H., Mohammadyani, D., Qiang Wang, K. Z., Zhu, J., Klein-Seetharaman, J., Balasubramanian, K., Amoscato, A. A., Borisenko, G., Huang, Z., Gusdon, A. M., Cheikhi, A., Steer, E. K., Wang, R., Baty, C., Watkins, S., Bahar, I., Bayir, H., and Kagan, V. E. (2013) Cardiolipin externalization to the outer mitochondrial membrane acts as an elimination signal for mitophagy in neuronal cells. *Nature cell biology* **15**, 1197-1205
273. Watanabe, Y., Tamura, Y., Kawano, S., and Endo, T. (2015) Structural and mechanistic insights into phospholipid transfer by Ups1-Mdm35 in mitochondria. *Nat Commun* **6**
274. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *Journal of molecular biology* **229**, 105-124