Escherichia coli and Pseudomonas aeruginosa lipopolysaccharide O-antigen ligases share similar membrane topology and biochemical properties

Ruan, X., Monjaras Feria, J., Hamad, M., & Valvano, M. A. (2018). Escherichia coli and Pseudomonas aeruginosa lipopolysaccharide O-antigen ligases share similar membrane topology and biochemical properties. *Molecular Microbiology*. https://doi.org/10.1111/mmi.14085

Published in:
Molecular Microbiology

Document Version:
Peer reviewed version

Escherichia coli and *Pseudomonas aeruginosa* lipopolysaccharide O-antigen ligases share similar membrane topology and biochemical properties

Xiang Ruan[1†], Julia Monjarás Feria[3†], Mohamad Hamad[2], and Miguel A. Valvano[1,3,*]

[1] *Department of Microbiology and Immunology, University of Western Ontario, London, Ontario, Canada N6A 5C1*

[2] *Department of Medical Laboratory Sciences, College of Health Sciences, University of Sharjah, Sharjah United Arab Emirates*

[3] *Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, United Kingdom BT9 7BL*

* For correspondence. E-mail: m.valvano@qub.ac.uk; Tel +44 (0) 28 9097 6025.

† Equal contribution

Running Title: Membrane topology of WaaL proteins

Summary

WaaL is an inner membrane glycosyltransferase that catalyzes the transfer of O-antigen polysaccharide from its lipid-linked intermediate to a terminal sugar of the lipid A-core oligosaccharide, a conserved step in lipopolysaccharide biosynthesis. Ligation occurs at the periplasmic side of the bacterial cell membrane, suggesting the catalytic region of WaaL faces the periplasm. Establishing the membrane topology of the WaaL protein family will enable understanding its mechanism and exploit it as a potential antimicrobial target. Applying oxidative labelling of native methionine/cysteine residues, we previously validated a topological model for *Escherichia coli* WaaL, which differs substantially from the reported topology of the *Pseudomonas aeruginosa* WaaL, derived from the analysis of truncated protein reporter fusions. Here, we examined the topology of intact *E. coli* and *P. aeruginosa* WaaL proteins by labeling engineered cysteine residues with the membrane-impermeable sulfhydryl reagent polyethylene glycol maleimide (PEG-Mal). The accessibility of PEG-Mal to targeted engineered cysteine residues in both *E. coli* and *P. aeruginosa* WaaL proteins demonstrates that both ligases share similar membrane topology. Further, we also demonstrate that *P. aeruginosa* WaaL shares similar functional properties with *E. coli* WaaL and that *E. coli* WaaL may adopt a functional dimer conformation.

Introduction

The Gram-negative bacterial cell envelope comprises an outer and an inner membrane separated by the periplasmic space. The lipopolysaccharide (LPS), a glycolipid made of lipid A, core oligosaccharide (core), and O-antigen (OAg) polysaccharide (Valvano, 2011), is a predominant component of the outer membrane. LPS helps maintain the stability and relative impermeability of the outer membrane. The lipid A provides a scaffold for the assembly of the core *via* sequential sugar transfer reactions. The lipid A-core is further remodelled by chemical modifications that provide bacteria with resistance to cationic antimicrobial peptides and hydrophobic antibiotics, contributing to bacterial evasion of innate immune recognition and intrinsic antimicrobial resistance (Needham and Trent, 2013).

OAgs are polymers of sugar repeating units that vary greatly in chemical composition, structure, and antigenicity across individual strains and bacterial species (Knirel, 2011). Bacteria expressing OAg resist innate immune attack by host macrophages and serum complement (Pluschke and Achtman, 1984, Joiner, 1988). OAgs are independently assembled as undecaprenyl-pyrophosphate (Und-PP)-linked saccharides and ligated to the lipid A-core (Valvano, 2011) resulting in the release of Und-PP, which is recycled into Und-P by a poorly characterized pathway. Und-P is essential for the biosynthesis of other surface polymers such as capsules and cell wall peptidoglycan (Valvano, 2011). All these processes occur in the inner membrane and require the activities of membrane-embedded proteins that participate in various key steps of lipid A-core and OAg assembly (Valvano, 2011). LPS export from the inner to the outer membrane and the proper insertion of LPS in the outer leaflet of the outer membrane require Lpt proteins, which form a trans-cell envelope multiprotein complex spaning the inner and outer membranes (Okuda *et al.*, 2016, Sperandeo *et al.*, 2017).

Und-PP-linked OAgs appearing on the periplasmic side of the membrane become ligated to preformed lipid A-core (Valvano, 2011). The membrane protein WaaL catalyzes the ligation step that joins the OAg to the rest of the LPS. WaaL can specifically recognize one or more terminal sugars of the lipid A-core acceptor (Feldman *et al.*, 1999), but shows no specificity for the structure of the lipid-linked saccharide substrate (Heinrichs *et al.*, 1998c). Therefore, despite different OAgs can be recombinantly expressed in *Escherichia coli* distinct WaaL proteins cannot cross-complement an *E. coli* K-12 *waaL* mutant (Valvano *et al.*, 2011). The periplasmic location of the ligation reaction suggests that potential inhibitory molecules need not enter into the bacterial cytosol and could evade membrane drug efflux mechanisms, which reduces the efficacy of conventional antibiotics. Elucidating the ligation mechanism is a critical step prior to the rational design of novel ligase inhibitors.

The best-studied ligases to date are the *E. coli* and *Pseudomonas aeruginosa* WaaL proteins (herein $WaaL_{Ec}$ and $WaaL_{Pa}$, respectively) (Abeyrathne *et al.*, 2005, Abeyrathne and Lam, 2007, Islam *et al.*, 2010, Pan *et al.*, 2012, Pérez *et al.*, 2008, Ruan *et al.*, 2012). Both proteins have 12 transmembrane helices (TMHs) and a characteristic large periplasmic loop (PL5). Conceivably, WaaL activity involves amino acids exposed to the periplasmic space where they could interact with donor and acceptor molecules (Pérez *et al.*, 2008, Ruan *et al.*, 2012). A conserved arginine residue critical for ligation, R215 in $WaaL_{Ec}$, is present in the preceding short periplasmic loop PL4 (Pérez *et al.*, 2008). We proposed a structural model for the $WaaL_{Ec}$ PL5 in which arginine-288 (R288) and histidine-338 (H338), two conserved residues that are needed for efficient ligation, are exposed to the solvent in a spatial arrangement that suggests direct interactions with substrate molecules (Pérez *et al.*, 2008). The critical H338 residue is part of a conserved motif

87 (H[NSQ]X_9GX_2[GTY]) spanning the C-terminal end of PL5 and the periplasm-membrane
88 boundary of the TMH-X (Ruan *et al.*, 2012).
89 Structural evidence from the analysis of the WaaL-catalyzed linkage between the acceptor
90 sugar residue at the lipid A-core and the OAg residue linked to Und-PP, together with
91 biochemical data, suggest WaaL proteins are metal-independent inverting glycosyltransferases
92 (Ruan *et al.*, 2012). Therefore, WaaL proteins should have common features, including a similar
93 topological location of functionally relevant amino acid residues. Both $WaaL_{Ec}$ and $WaaL_{Pa}$ have
94 relatively similar *in silico* predicted membrane topologies (Fig. 1A and B). However, the
95 experimental determination of the $WaaL_{Pa}$ topology using truncated protein derivatives arising
96 from a deletion/reporter fusion analysis revealed striking differences with $WaaL_{Ec}$ (Islam *et al.*,
97 2010). While experimental data and *in silico* predictions agree with the topology of the N- and
98 C-terminal regions of $WaaL_{Pa}$ (spanning TMH-I to TMH-III and TMH-IX to TMH-XII,
99 respectively), major differences appear in the central region of the protein spanning TMH-IV to
100 TMH-VIII (Fig. 1B). These TMHs in the experimental model had no concordance with the
101 predicted boundaries for each of the corresponding TMHs in the *in-silico* model (Fig. 1A and B).
102 A striking difference between the two models was the presence in the experimental model
103 ($WaaL_{Pa}$-Del/Fus in Fig. 1A and B) of a large cytosolic loop between amino acid residues at
104 positions 153-182 (CL3), which contained highly hydrophobic residues commonly present in
105 TMH segments. Further, the predicted conserved essential arginine residue R212 (corresponding
106 the R215 in $Waal_{Ec}$) was buried in TMH-VIII of the $WaaL_{Pa}$-Del/Fus model, but exposed to the
107 PL4 in the *in-silico* model (Fig. 1A and B). Intriguingly, the experimentally assigned TMH-VIII
108 in $WaaL_{Pa}$-Del/Fus also contained three positively charged residues in tandem (arginine-231,
109 lysine-232, and lysine-233) (Islam *et al.*, 2010), which would be highly unusual residues to be

embedded in the lipid bilayer. In this work, we have comparatively evaluated the topology of $WaaL_{Ec}$ and $WaaL_{Pa}$ by a scanning cysteine accessibility mutagenesis (SCAM) strategy using the membrane-impermeable sulfhydryl reagent polyethylene glycol maleimide (PEG-Mal). Our experiments demonstrate that both proteins share identical topology. Further, examining the *in vitro* ligation reaction using purified $WaaL_{Pa}$, we demonstrate that ligation occurs in the presence of EDTA, suggesting it cannot require ATP, as we previously observed for $WaaL_{Ec}$ (Ruan *et al.*, 2012), and we also provide evidence that $WaaL_{Ec}$ forms functional dimers. We conclude that $WaaL_{Ec}$ and $WaaL_{Pa}$ proteins are archetypes of a family of metal-independent, membrane embedded glycosyltransferases sharing similar topology and functional properties.

Results

E. coli *and* P. aeruginosa *WaaL proteins lacking native cysteines remain functional and both require a conserved arginine residue*

To investigate the topology of $WaaL_{Ec}$ and $WaaL_{Pa}$, we first expressed both proteins as C-terminal FLAG-10xHis-tag fusions under the control of exogenous L-arabinose from a pBAD24 vector derivative (Table 1). Protein expression was detected by Western blot using an anti-FLAG monoclonal antibody, and it was optimal in the presence of 0.2% L-arabinose (Fig. S1). A cysteine less version of $WaaL_{Ec}$ ($WaaL_{Ec}$C-less) was constructed by sequential site-directed replacements to alanine of each native cysteine residue in the native protein at positions 29, 54, 203, and 225. We also replaced by alanine a cysteine residue at position 428, which originated from the FLAG-tag fusion construction. Further, we generated $WaaL_{Ec}$ versions with a single

cysteine at position 29 ($WaaL_{Ec}C29$) and 428 ($WaaL_{Ec}C428$), whereby the other native cysteine residues were replaced by alanine. Additional single-cysteine replacements were made at selective sites.

Similarly, we constructed a cysteine less version of $WaaL_{Pa}$ ($WaaL_{Pa}$C-less) by replacing the native cysteine residue 114 and the Cys-403 residue originated from the FLAG-tag fusion by alanine. By replacing each cysteine residue individually, we also generated versions of $WaaL_{Pa}$ with single cysteines at positions 114 ($WaaL_{Pa}C114$) and 403 ($WaaL_{Pa}C403$). Under induction with 0.2% L-arabinose, all the mutant derivatives of $WaaL_{Ec}$ or $WaaL_{Pa}$ were well expressed in *E. coli* DH5α (Fig. 2B and 3, upper panels, respectively) with apparent molecular masses of 50 kDa and 47 kDa, respectively. The parental FLAG-10xHis-tag fusions proteins $WaaL_{Ec}$ and $WaaL_{Pa}$ were also expressed with similar apparent molecular masses, respectively (Fig. S1).

The cysteine less version of $WaaL_{Ec}$ and all the single cysteine-alanine substitutions did not affect OAg ligation, as determined by the ability of each mutant protein to support O7 LPS surface expression in *E. coli* SCM3 (pJHCV32) (Fig. 4A). This *E. coli* K12 strain has a chromosomal deletion that eliminates the OAg and colanic capsule synthesis gene clusters and contains also a deletion of the *waaL* gene (Pérez *et al.*, 2008), while the cosmid pJHCV32 encodes the O7 lipopolysaccharide antigen gene cluster (Valvano and Crosa, 1989). The WaaL G286C mutant showed an apparently slight decrease in ligase activity, which could not be attributed to reduced protein expression and may be due to reduced efficiency of the enzyme.

Likewise, the cysteine less version of $WaaL_{Pa}$ and most of the single cysteine substitutions restored OAg LPS surface expression in the *P. aeruginosa* PAO1*waaL*::Gm^R mutant to levels comparable to those in the parental PAO1 strain (Fig. 4C). The replacements R147C, L155C, and S187C showed a small reduction in ligase function in comparison to the parental $WaaL_{Pa}$,

156 while the S169C and G282C replacements had a stronger effect since OAg in the LPS was not
157 detected by silver staining (Fig. 4C, upper panel). To determine if the reduced ligase activity
158 observed by OAg complementation was due to different levels of protein expression, we isolated
159 total membranes and detected the $WaaL_{Pa}$ proteins by immunoblot. This experiment
160 demonstrated that all the single Cys replacements are expressed at $WaaL_{Pa}$ parental levels, except
161 for the G282C in which the protein level is reduced (Fig. 4C, bottom panel). We used a more
162 sensitive approach to determine the ligase activity of the single Cys mutants based on
163 immunoblot with the anti-O5 monoclonal antibody (MF15-4). This experiment shows that only
164 the G282C replacement is affected in its ability to restore OAg ligation, while all the other
165 replacements supported OAg ligation (Fig. 4C, middle panel). A G282A replacement mutant
166 also exhibited the same phenotype as G282C (Fig. S2)., suggesting that the G282 residue in
167 $WaaL_{Pa}$ is important for ligase function either by being involved in the reaction or by
168 maintaining the local conformation of a putative catalytic fold. Together, we conclude that,
169 except for $WaaL_{Pa}$G282C the cysteine less versions of $WaaL_{Ec}$ and $WaaL_{Pa}$, as well as their
170 single cysteine replacement derivatives remain functional.
171 Because the Arg-212 residue in $WaaL_{Pa}$ is the counterpart of Arg-215 in $WaaL_{Ec}$, which is
172 required for ligase function (Pérez *et al.*, 2008, Ruan *et al.*, 2012), we generated a $WaaL_{Pa}$
173 derivative with an alanine replacement at this site ($WaaL_{Pa}$R212A). This mutant protein lost the
174 ability to complement the ligase function in the PAO1*waaL*::Gm^R ligase defective mutant (Fig.
175 4B). The loss of function phenotype is not due to defects in protein expression since the
176 $WaaL_{Pa}$R212A was produced at wild type levels (Fig. 4B, bottom panel). The results indicate
177 that similarly to Arg-215 in $WaaL_{Ec}$, the corresponding Arg-212 residue in $WaaL_{Pa}$ is also
178 critical for ligation.

Labeling of Cys residues with PEG-Mal supports the predicted topological model of $WaaL_{Ec}$

Our previous study using GFP as a topology probe and trypsin digestion determined a C-terminus inside, a large periplasmic loop PL5 and a small adjacent periplasmic loop PL4 for the *E. coli* K-12 WaaL, which is consistent with the predicted topology proposed by Perez et al. (Pérez *et al.*, 2008) and with a study using oxidative labeling of Met and Cys followed by mass spectroscopy (Pan *et al.*, 2012). In this study we used the substituted-cysteine accessibility method (SCAM) for mapping WaaL topology in their native state. This approach is based on the introduction of single Cys residues at various positions in a membrane protein followed by chemical modification with a membrane impermeable thiol-specific reagent, polyethylene glycol maleimide (PEG-Mal), either before or after disrupting cell membrane integrity by detergent treatment (Fig. 2A).

The results show that $WaaL_{Ec}$C-less could not be labelled by PEG-Mal in the absence or presence of SDS, or in the presence of N-ethylmaleimide (NEM), a low molecular weight, membrane permeable sulfhydryl-reactive compound that blocks PEGylation (see *Experimental Procedures* and Fig. 2A), and sodium dodecylsulfate SDS (Fig. 2B), providing a negative control for the PEGylation assay. In contrast, residue C29 neither was accessible to PEG-Mal in the absence of SDS nor could be blocked by NEM, but it was labelled by PEG-Mal in the presence of SDS (Fig. 2B), indicating that C29 is embedded within a TMH, as we previously demonstrated by oxidative labelling (Pan *et al.*, 2012). The residue C428 was inaccessible to PEG-Mal in the absence of SDS, but accessible to NEM and PEG-Mal in the presence of SDS (Fig. 2B), as expected for a residue facing the cytosol, which is consistent with the C-terminus inside topology of $WaaL_{Ec}$ (Pan *et al.*, 2012, Pérez *et al.*, 2008).

We then examined the accessibility of residues I162, T176 and T190, which were predicted to reside in the periplasmic PL3 loop, TMH-VI, and the cytoplasmic CL3 loop of $WaaL_{Ec}$, respectively (Fig. 1). The counterpart of this region in the model of $WaaL_{Pa}$ proposed by Islam *et al.* (2010) exhibits a very different topology (Fig. 1). The $WaaL_{Ec}$ I162C residue was accessible to PEG-Mal in the absence of SDS, but blocked by NEM and labeled by PEG-Mal in the presence of SDS (Fig. 2B), denoting a periplasmic location. Residue T176C was not accessible to PEG-Mal in the absence of SDS, could not be blocked by NEM and could be labeled by PEG-Mal in the presence of SDS (Fig. 2B), indicating the residue forms part of the TMH-VI. Residue T190C was consistently not labeled by PEG-Mal in the presence of SDS (Fig. 2B), so we could not precisely localize its location. This suggest the residue may be sequestered into a local secondary structure configuration that makes it inaccessible to PEG-Mal, as we have shown for other membrane proteins (Furlong *et al.*, 2015).

Residues T214 and A216 flank R215, a crucial residue for the $WaaL_{Ec}$ function (Pérez *et al.*, 2008, Ruan *et al.*, 2012). Both T214C and A216C were weakly PEGylated in the presence of SDS (Fig. 2B). Further, T214C was also labeled by PEG-Mal in the absence of SDS and blocked by NEM (Fig. 2B), indicating this residue faces the periplasmic space. In contrast, residue A216C was not accessible to PEG-Mal in the absence of SDS, but PEGylation was blocked by NEM (Fig. 2B), suggesting A216 is at the periplasmic border of TMH8. Together, these results support the prediction that R215 is located in a short loop exposed to the periplasm (Pérez *et al.*, 2008, Ruan *et al.*, 2012). Residue G286C was accessible to PEG-Mal, both in the absence and presence of SDS, and labeling was blocked by NEM (Fig. 2B), indicating that G286 is located in periplasm, consistent with the prediction that G286 is a residue in PL5. Together, the results

224 from PEGylation accessibility of the single-Cys mutants are consistent with the previously
225 predicted topology of $WaaL_{Ec}$ (Fig. 5A).
226
227 *PEG-Mal labelling of Cys residues in* P. aeruginosa *WaaL validates a topological model similar*
228 *to that of* E. coli *WaaL*
229 As shown in Fig. 3, $WaaL_{Pa}$C-less was not labeled at any condition providing a negative
230 control for the PEGylation assays. Residue C114 of $WaaL_{Pa}$ was not labeled by PEG-Mal in the
231 presence of SDS (Fig. 3), so we could not assign a location for this residue. Residue C403 was
232 inaccessible to PEG-Mal in the absence of SDS, weakly PEGylated with SDS, but PEGylation
233 was blocked by NEM in the presence of SDS (Fig. 3). This is consistent with a C-terminus inside
234 topology of $WaaL_{Pa}$. We then examined 10 residues predicted *in silico* to be located in TMH-V
235 (R147), the PL3 loop (L155, L157, T158 and G161), TMH-VI (N165, S169, A170 and A178),
236 and in the CL3 loop (S187) (Fig. 1). These same 10 residues were assigned to TMH-VI (R147),
237 CL3 (L155, L157, T158, G161, N165, S169, A170 and A178), and TMH-VII (S187) in the
238 model proposed by Islam *et al.* (2010) (Compare Fig. 5B and C). The single cysteine
239 substitutions of the 10 residues were labeled by PEG-Mal in the presence of SDS (Fig. 3).
240 Residue R147C was inaccessible to PEG-Mal in the absence of SDS and unblocked by NEM
241 (Fig. 3), indicating that R147 is embedded in a TMH. Residues L157C, T158C and G161C were
242 accessible to PEG-Mal in the absence of SDS and also blocked by NEM (Fig. 3), indicating they
243 are exposed to the periplasm. L155C was not accessible to PEG-Mal in the absence of SDS, but
244 accessible to PEG in the presence of SDS (Fig. 3), indicating a location near the periplasmic
245 border of the TMH. N165C, S169C, A170C and A178C were not accessible to PEG-Mal in the
246 absence of SDS and not blocked by NEM (Fig. 3), indicating they are part of the TMH. S187C

was not accessible to PEG-Mal in the absence of SDS, but PEGylation was blocked by NEM (Fig. 3), indicating that S187 is in cytoplasm. The results demonstrate that the topology of $WaaL_{Pa}$ resembles that of $WaaL_{Ec}$ (Fig. 5A and B).

Further, residues S211 and T213 flank R212, a counterpart of the functional residue R215 in the PL4 of $WaaL_{Ec}$. Both $WaaL_{Pa}$ S211C and T213C residues were labeled by PEG-Mal in the presence of SDS (Fig. 3). S211C was also PEGylated in the absence of SDS and accessible to NEM (Fig. 3), indicating this residue faces the periplasm. On the contrary, T213C was inaccessible to PEG-Mal in the absence of SDS but accessible to NEM (Fig. 3), suggesting a location in the periplasmic borders closer to TMH. These data demonstrate that R212 is also part of short periplasmic loop PL4, in contrast to the model by Islam *et al.* (2010) (Fig. 5C), in which S211, R212 and T213 are located in TMH-VIII.

We also examined six residues spanning the segment between L251 and V309, which are predicted to be located in PL5 and its borders (Fig. 5B). All the single-Cys substitutions of these residues were accessible to PEG-Mal in the presence of SDS (Fig. 3). Residue V255C was also labeled by PEG-Mal in the absence of SDS and labeling blocked by NEM (Fig. 3), indicating a periplasmic location (Fig. 5B). L251C was not accessible to PEG-Mal in the absence of SDS, but PEGylation in SDS was blocked by NEM (Fig. 3), indicating it is located at the periplasmic borders closer to the C-terminus of TMH-IX. P253C was not accessible to PEG-Mal in the absence of SDS and not blocked by NEM (Fig. 3), also suggesting this residue may be located near the border of TMH-IX (Fig. 5C). Residue G282C was accessible to PEG-Mal in the absence of SDS and labeling blocked by NEM (Fig. 3), indicating a periplasmic location within PL5 (Fig. 5B). Both L307C and V309C were not accessible to PEG-Mal in the absence of SDS and not blocked by NEM (Fig. 3), indicating they are embedded within TMH-X. Therefore, the

experimental results obtained by PEG-Mal labeling of these six residues agree with both the *in silico* model for $WaaL_{Pa}$ (Fig. 1B) and the topology proposed by Islam *et al.* (2010) (Fig. 1C). Together, our experimental data demonstrate that the topology predictions by Islam *et al.* (2010), which were based on deletion-fusion experiments are not accurate to delineate the topology of the central portion of $WaaL_{Pa}$ (Figs. 1 and 5C) since they are not supported by PEG-Mal labeling of intact protein derivatives. Based on our results, we conclude that both $WaaL_{Pa}$ and $WaaL_{Ec}$ share a similar membrane topology, as shown in Fig. 5A and B.

$WaaL_{Pa}$ exhibits similar functionality as $WaaL_{Ec}$

Previous work reported that $WaaL_{Pa}$ required ATP hydrolysis for ligation, while this was not the case for the $WaaL_{Ec}$ protein (Han *et al.*, 2012, Ruan *et al.*, 2012). To assess the $WaaL_{Pa}$ functionality we purified this protein and performed an *in vitro* ligase assay, as previously described (Ruan *et al.*, 2012). Initial experiments using LPS ligation substrates containing a mixture of Und-PP-linked O-polysaccharides (including A-band homopolysaccharide and B-band heteropolysaccharide) and LPS core prepared from PAO1*waaL::*Gm^R and mixed them with purified $WaaL_{Pa}$ did not result in ligation products that were detectable by silver staining. *P. aeruginosa* produces two forms of lipid A-core. One of them consists of a capped core in which an L-rhamnose is attached to a terminal glucose (glucose-I) by WapR, while the other (uncapped core) contains L-rhamnose attached to glucose-II by MigA (Kocincova and Lam, 2011). From these two core forms, only the capped core is a substrate for ligation with O-antigen polysaccharide. To increase the proportion of capped core we constructed pXR93, overexpressing WapR under the control of the *lac* promoter. The strategy of preparing the ligation substrates from the *P. aeruginosa* strain PAO1*waaL*::Gm^R(pXR93) expressing the

WapR protein allowed us to detect LPS products by silver staining (Fig. 6). The results indicated that 37°C was a suitable temperature for the ligation reaction in contrast to 22°C and 30°C (Fig. 6, Lanes 3-5). The addition of IPTG showed increased detection of ligation products (Fig. 6, Lanes 5-8). However, 10 mM ethylenediaminetetraacetic acid (EDTA), a concentration that could chelate divalent cations including Mg^{2+} and disrupt the ATPase hydrolysis, did not inhibit the reaction (Fig. 6, Lane 9). This indicated that, as in the *in vitro* ligase assay for $WaaL_{Ec}$, $WaaL_{Pa}$ is also an ATP hydrolysis-independent, metal ion-independent glycosyltransferase (Ruan *et al.*, 2012).

Ligation results in inversion and is specific for the lipid A core substrate

Structural data on LPS of many different species revealed a β-linkage of the sugar from the OAg unit attached to the terminal lipid A-core acceptor site. Interestingly, this sugar is also the one proximal to Und-PP, which arises from the diphosphate nucleotide sugar precursors in the α-configuration employed in the initiation reaction by enzymes of the WecA and WbaP families (Ruan *et al.*, 2012, Valvano, 2011), indicating WaaL is an inverting glycosyltransferase (Ruan *et al.*, 2012). To gather additional evidence for this mechanism we examined the reactivity of chemically characterized *E. coli* lipid A-core forms with a monoclonal antibody specific to β-O-linked *N*-acetylglucosamine (β-O-GlcNAc). For this experiment, we employed LPS from different *E. coli* strains (Fig. 7). W3110 has a lipid A-core with a terminal β-linked-GlcNAc, which results from the synthesis of a truncated O-antigen unit consisting of Und-PP-GlcNAc (Feldman *et al.*, 1999). CLM35, a strain with *wecA* and *wzx* double deletion, produces lipid A-core devoid of GlcNAc (Marolda *et al.*, 2006b). The LPS from CLM24, a strain with *waaL* deletion contains a mixture of lipid A-core and Und-PP-GlcNAc (Feldman *et al.*, 2005), while

316 CLM24(pXR1) contains the plasmid pXR1 expressing *E. coli* WaaL and therefore producing
317 lipid A-core with a terminal GlcNAc residue (Fig. 7A). The strain F632 is an *E. coli* R2
318 chemotype bacterium producing a lipid A-core with a terminal α-linked-GlcNAc (Leipold *et al.*,
319 2007). While the LPS preparations looked similar by silver staining, only those from W3110 and
320 CLM24(pXR1) reacted with the anti-β-O-GlcNAc antibody (Fig. 7A), confirming that the
321 terminal GlcNAc attached to K-12 LPS core is in β-configuration. The absence of a positive
322 signal with the antiserum in the F632 LPS shows that the antiserum cannot detect GlcNAc in the
323 α-configuration. These results provide additional support to the notion that WaaL is an inverting
324 glycosyltransferase.
325 Another property of the ligase is its specificity for the terminal core acceptor site, which was
326 determined by genetic experiments (Heinrichs *et al.*, 1998a, Heinrichs *et al.*, 1998b, Heinrichs *et*
327 *al.*, 1998c). We took advantage of the *in vitro* ligation assay to determine *in vitro* if the $WaaL_{Ec}$
328 can ligate O-antigen substrates from different bacterial species. We used four LPS preparations
329 containing lipid A-core and endogenously made Und-PP-linked O-antigen substrates (Fig. 7B).
330 They included LPS from strain CLM35 ($Core_{Ec}$) containing lipid A-core plus Und-PP-GlcNAc,
331 LPS from strain CLM24(pMF19) (LPS_{Ec}) containing lipid A-core plus Und-PP-OAg, LPS from
332 strain PAO1waaL::Gm^R(pXR93) (LPS_{Pa}) containing lipid A-core core plus Und-PP-OAg (from
333 *P. aeruginosa*), and lipid A-core from strain K56-2*waaL*$_{Bc}$::pGPΩTp (LPS_{Bc}) containing lipid A-
334 core plus Und-PP-OAg (from *Burkholderia cenocepacia*). We mixed purified $WaaL_{Ec}$ with
335 different combination of the LPS preparations as described in *Experimental Procedures*. Our
336 results show that LPS preparations from *P. aeruginosa* (LPS_{Pa}) and *B. cenocepacia* (LPS_{Bc})
337 enriched for Und-PP-linked O antigen result in the production of *E. coli* lipid A-core with the
338 ligated O-polysaccharides from the respective species (Fig. 7B, Lane 5, and 7C, Lane 6,

respectively), as efficient as LPS preparation from *E. coli* (LPS_{Ec}) (Fig . 7B, Lane 6). From these experiments, we conclude WaaL lacks specificity of the Und-PP-linked OAgs.

E. coli *WaaL forms functional oligomers*

Purified *E. coli* and *P. aeruginosa* WaaL preparations typically show variable amounts of larger products in SDS-polyacrylamide gel electrophoresis (PAGE) that react with the anti-FLAG antibody suggesting the presence of multimeric forms of the protein samples (Fig. S1) likely due to incomplete denaturation since heating at 100°C in SDS precludes detection of polytopic membrane proteins in SDS-PAGE (Abeyrathne and Lam, 2007, Lehrer *et al.*, 2007, Marolda *et al.*, 2004, Pérez *et al.*, 2008, Ruan *et al.*, 2012, Vinés *et al.*, 2005, Tatar *et al.*, 2007). However, these larger forms could also occur by dimerization. To investigate if WaaL can form functional dimers, we reasoned that constructs lacking ligase function if expressed singly could revert to functional enzyme forms when expressed jointly. To explore this possibility, we used $WaaL_{Ec}$ functionally defective constructs that exhibited observable protein expression, but could potentially restore ligation if jointly expressed. These included WaaLΔN2TM, a derivative of $WaaL_{Ec}$ with a deletion removing amino acid residues from Leu-2 to Asn-71 and eliminating TMHs I and II, WaaL-R288A and WaaL-R215A (Ruan *et al.*, 2012). All these recombinant proteins, expressed from pBAD24 in the presence of L-arabinose were detected by immunoblot at comparable levels (Fig. 8A, left panel). Similarly, WaaLΔN2TM expressed from pBAD33, a plasmid that is compatible with the pBAD24 (Guzman *et al.*, 1995), was also readily detected by immunoblot (Fig. 8A, right panel). We first confirmed that the recombinant WaaL proteins function *in vivo* as expected. For this we transformed the *E. coli* SCM3(pJHCV32) with various plasmids encoding each of the WaaL proteins and examined the production of O7

362 polysaccharide. Fig. 8B, left panel, shows that no O7 polysaccharide was produced in the
363 presence of WaaLΔN2TM, WaaL-R288A, and WaaL-R215A. In contrast, O7 polysaccharide
364 was detected in SCM3(pJHCV32) bacteria expressing the parental WaaL and $WaaL_{R215K}$, as
365 expected from previous results showing that a positive charged amino acid is required at position
366 215 of *E. coli* WaaL (Pérez *et al.*, 2008, Ruan *et al.*, 2012). Next, we co-expressed each of these
367 proteins with WaaLΔN2TM, which was expressed from pBAD33. The results revealed that O7
368 polysaccharide was produced by co-expression of WaaLΔN2TM/WaaL-R288A or
369 WaaLΔN2TM/WaaL-R215A (Fig. 8, right panel), indicating that the reversal of the ligase
370 function could originate from protein-protein interactions between WaaLΔN2TM and WaaL-
371 R288A, or between WaaLΔN2TM and WaaL-R215A.
372 We also investigated WaaL oligomerization by chemical crosslinking. The *E. coli* WaaL and
373 WaaLC-less proteins with FLAG and 10xHis tags were purified from *E. coli* JM105v, as
374 described previously (Ruan *et al.*, 2012). The membrane-permeable crosslinker dithiobis
375 (succinimidylpropionate) (DSP) was used, which crosslink molecules whose respective lysine
376 residues come within 12.0 Å, the spacer length of the crosslinker (Smith *et al.*, 2011). Extensive
377 crosslinking of WaaL was apparent by anti-His immunoblot. A portion of each sample was
378 treated with the reducing agent dithiothreitol (DTT) to cleave the crosslinking. The results show
379 that $WaaL_{Ec}$ forms oligomeric species of higher apparent mass in the presence of DSP, while the
380 monomeric protein is efficiently recovered by addition of DTT (Fig. 8C,). To investigate if
381 dimerization involves disulfide bond formation, we also performed crosslinking experiments
382 with the WaaLC-less protein. The results show that native cysteines are partially involved in
383 dimerization, as crosslinking was less efficient with the cysteine less protein (Fig. 8C).

Therefore, the genetic reconstitution experiments and biochemical crosslinking data indicate WaaL can form oligomers and is likely a functional dimer *in vivo*.

Discussion

Using a SCAM strategy based on PEGylation, we show here the topologies of *E. coli* and *P. aeruginosa* WaaL proteins are highly similar. The $WaaL_{Pa}$ topology model we constructed differs from that previously reported in the literature, which was based on a dual-reporter fusion approach (Alexeyev and Winkler, 1999) and resulted in a library of truncated versions of $WaaL_{Pa}$ C-terminally fused to reporter β-galactosidase and alkaline phosphatase enzymes for determining the cytoplasmic or periplasmic location of each fusion end point, respectively (Islam *et al.*, 2010). Reporter fusions are commonly used to map the topology on membrane proteins, but the strategy is not devoid of pitfalls, especially considering that the topogenic signals for membrane protein insertion depend from both amino acid and membrane lipid compositions (Bogdanov and Dowhan, 1999). Indeed, the dual-reporter fusion strategy for $WaaL_{Pa}$ did not accurately predict TMHs and loops boundaries in the central region of the protein. This could most likely be caused by disruption of TMH-TMH interactions leading to alternative topologies of the truncated WaaL segment included in the fusion. Similar observations were made for TMH-I of the *E. coli* α-ketoglutarate permease, which was not detected by a PhoA fusion owing to the presence of positively charged residues in other TMH domains that were required to neutralize the negatively charged residues in TMH-I, facilitating its membrane insertion (Seol and Shatkin, 1993). Alternatively, truncated TMH domains may form helical hairpins that result in the mislocalization of the reporter protein (Cassel *et al.*, 2008).

SCAM is also not free of drawbacks, as residues in certain locations may become inaccessible to sulfhydryl labeling (Bogdanov *et al.*, 2010). We noted that the cysteine replacement at position 190 in $WaaL_{Ec}$ (T190C) could not be labeled under any of the conditions tested. This residue, presumably located in the CL3, is flanked by a tyrosine residue at one end and two lysine residues at the other end, suggesting the possibility of a local secondary structure stabilized by salt bridges between the tyrosine and the lysine residues, which would prevent the access of PEG-Mal by steric hindrance. Local secondary structure preventing access of PEG-Mal was noted before in the TMH-V of membrane protein WcaJ, which does span the membrane and forms a hairpin helix around a proline and a serine residue (Furlong *et al.*, 2015). Similarly, the native cysteine residue at position 114 of $WaaL_{Pa}$ could not be labeled. This residue is predicted to be embedded in TMH-IV, and lack of labeling with PEG-Mal in this case could arise from the absence of ionization of the thiol group within a strong hydrophobic environment (Bogdanov *et al.*, 2010). However, despite the difficulties with the localization of these specific residues, the overall topology of $WaaL_{Pa}$ derived from PEG-Mal SCAM agrees with that of $WaaL_{Ec}$. Indeed, the critical residues for ligase function in both proteins, Arg-215 in $WaaL_{Ec}$ and Arg-212 in $WaaL_{Pa}$ are orientated towards the periplasmic space within a short PL4 region. This is consistent for a protein that requires access to substrates in the periplasmic space.

Further, our *in vitro* ligation experiments demonstrated that the $WaaL_{Pa}$-catalyzed ligation occurs in the presence of high concentration of the metal divalent ion chelator EDTA, indicating that the reaction is metal cation independent, which also rules out an involvement of ATP hydrolysis, as previously reported (Abeyrathne and Lam, 2007). Therefore, we conclude that $WaaL_{Pa}$ has the same general properties of previously characterized ligases such as those from *E.*

coli K-12 (Ruan *et al.*, 2012), *E. coli* O86 (Han *et al.*, 2012), and the *Helicobacter pylori* (Hug *et al.*, 2010), displaying similar membrane topology and biochemical function.

In a previous study, we provided evidence suggesting the $WaaL_{Ec}$ is an inverting glycosyltransferase (Ruan *et al.*, 2012). This means the glycosyl bond formed by the ligation reaction has an inverted anomeric configuration with respect to the donor substrate (Lairson *et al.*, 2008). The substrate for the ligase reaction is the Und-PP-linked proximal sugar, which in turn is the product of the initiation reaction resulting in the formation of a phosphoanhydride bond without a change in the anomerization of the sugar backbone structure (Valvano, 2011). Indeed, the diphosphate nucleotide sugar precursors for the initiation reaction, such as UDP-Gal, UDP-Glc and UDP-GlcNAc are all in α-configuration (Murazumi *et al.*, 1979, Weisgerber and Jann, 1982, Olsthoorn *et al.*, 2000). In this work, we provide additional evidence for inversion of anomerization upon ligation by showing that a single GlcNAc residue ligated to *E. coli* K-12 lipid A-core reacted with an anti-β-O-GlcNAc antibody, confirming that this residue is in the β-configuration.

Several cytoplasmic membrane proteins implicated in the assembly of cell surface polysaccharides form oligomeric structures (Collins *et al.*, 2017, Liston *et al.*, 2015, Whitfield and Trent, 2014). Further, the MraY protein that initiates cell wall peptidoglycan synthesis forms dimers both in detergent micelles and in the membrane (Chung *et al.*, 2013). To probe whether WaaL can form oligomers *in vivo*, we employed a genetic reconstitution strategy by co-expressing an N-terminally truncated $WaaL_{Ec}$ version that contains intact functional residues in PL4 and PL5 regions with an intact $WaaL_{Ec}$ with single substitutions in functional amino acids. This experiment resulted in the restoration of O-antigen synthesis, which was not possible with any of these proteins when expressed individually. Oligomerization of $WaaL_{Ec}$ was

independently supported by chemical cross-linking analyses, suggesting this protein adopts a functional oligomeric form in the membrane.

Together, this study demonstrates that $WaaL_{Ec}$ and $WaaL_{Pa}$ proteins share similar topology and functionality, indicating that O-antigen ligases have conserved features consistent with their function at the periplasmic side of the membrane and pointing out to a common glycosyl transfer mechanism.

Experimental Procedures

Bacterial strains, plasmids and growth conditions

Strains and plasmids used in this study are listed in Table 1. Bacteria were cultured at 37°C in a Luria-Bertani (LB) medium supplemented with ampicillin (100 μg ml^{-1}), tetracycline (20 μg ml^{-1}), chloramphenicol (30 μg ml^{-1}) for *E. coli*, or with tetracycline (100 μg ml^{-1}), gentamicin (30 μg ml^{-1}) for *P. aeruginosa*, or with 0.2% (w/v) L-arabinose, when appropriate.

Construction of strains and plasmids

P. aeruginosa genomic knockout mutants PAO1*waaL*::Gm^R was constructed for complementation. The *waaL* gene of the *P. aeruginosa* strain PAO1 was amplified from genomic DNA by PCR using forward primer 5368 and reverse primer 5369 (Table S1), and cloned into pEX18Tc vector, followed by inserting a gentamicin-resistance cassette into *Sal*I site within the *waaL* gene, producing the insertional mutant PAO1*waaL*::Gm^R. This construct was transformed into PAO1 by electroporation, and transformants plated on LB + 100 μg ml^{-1} of

tetracycline. *waaL* knockout mutants were selected by subsequent plating on 5% sucrose agar plates containing 30 μg ml^{-1} of gentamicin. For complementation experiments, the PAO1 *waaL* gene was subcloned from pXR14, pXR23, and all pXR23 and pJM plasmids with single cysteine substitution mutants by digesting with *Eco*RI and *Hind*III, and ligating it into pUCP26, also digested with the same restriction enzymes. The *wapR* gene of PAO1 was amplified from the genomic DNA by PCR using forward primer 7049 and reverse primer 7050 (Table S1), and the amplicon digested with *Eco*RI and *Pst*I and ligated to pUCP26, which was also digested with the same restriction enzymes, yielding plasmid pXR93.

The *E. coli waaL*$_{\Delta N2TM}$, which carries a deletion eliminating amino acids Leu-2 to Asn-71, was amplified from pXR1 using primers 5995 and 5730 (Table S1), and the amplicon digested with *Eco*RI and *Pst*I was ligated to pBADFLAG, which was also digested with the same restriction enzymes, yielding plasmid pXR39. This strategy deletes the complete WaaL encoding sequence and allows replacement with the truncated one. To generate pXR76, encoding *E. coli* waaLΔN2TM in pBAD33, both pXR39 and pBAD33 were digested with *Cla*I and *Hind*III, and the 2.7 Kb insert was ligated to the 4-Kb fragment from pBAD33. All replacements were verified by DNA sequencing.

Site-directed mutagenesis

pXR1, encoding a C-terminal FLAG followed by 10xHis-tagged *E. coli* WaaL, was used as a DNA template for site-directed mutagenesis to construct pXR36, encoding cysteine-less *E. coli* WaaL. pXR36 was then used as a DNA template for site-directed mutagenesis to construct *E. coli* WaaL mutants with single cysteine replacements. pXR14, encoding a C-terminal FLAG followed by 10xHis-tagged *P. aeruginosa* WaaL, was used as a DNA template for site-directed

497 mutagenesis to construct pXR23, encoding cysteine-less *P. aeruginosa* WaaL. pXR23 was then
498 used as a DNA template for site-directed mutagenesis to construct *P. aeruginosa* WaaL mutants
499 with single cysteine replacements. The Cys-428 codon in *E. coli* WaaL and Cys-403 codon in *P.*
500 *aeruginosa* WaaL do not exist in the native gene, but from the construction of FLAG-tag. pXR14
501 was also used as a DNA template for site-directed mutagenesis to construct *P. aeruginosa*
502 WaaL-R212A. Table S1 of supplementary data shows the DNA sequences of all primers used for
503 mutagenesis. *Pfu* polymerase was used in PCR amplification for site-directed mutagenesis. The
504 PCR products were digested overnight with 1 U *Dpn*I at 37°C, and then introduced into *E. coli*
505 DH5α competent cells by transformation. Ampicillin-resistant colonies were screened. All
506 substitutions were confirmed by DNA sequencing.
507
508 *Sulfhydryl Labeling*
509 Overnight cultures of DH5α cells containing the appropriate plasmids (Table 1) supplemented
510 with 100 μg ampicillin ml^{-1} were diluted in LB medium (5 ml) to an OD_{600} of 0.02. When the
511 OD_{600} reached around 0.6, protein expression was induced with 0.2% (w/v) L-arabinose. After a
512 2.5-h induction, the OD_{600} for each culture was measured and adjusted to 0.5 with LB medium.
513 Then 2 ml of each adjusted culture was centrifuged at 16,000 x *g* for 1 min, washed twice with 4-
514 [2-hydroxyethyl]-1-piperazineethanesulfonic acid (HEPES) buffer (50 mM HEPES [pH6.8], 5
515 mM $MgCl_2$) and re-suspended in 0.2 ml of the same buffer, which is referred to as '0.2 ml cell
516 suspension' as below.
517 For protein expression controls, the 0.2 ml cell suspension were added 0.1 mg ml^{-1} lysozyme
518 and 25 μg ml^{-1} DNaseI, incubated at 37°C for 20 min, and then mixed with protein loading dye
519 for SDS-PAGE on 14% acrylamide gels. For sulfhydryl labeling with Polyethyleneglycol

maleimide (PEG-Mal, 5 KDa, Sigma), each of the 0.2 ml cell suspension was incubated with 5 mM EDTA at room temperature for 10 min (+PEG-Mal) or 0.2% (w/v) SDS at 37°C for 5 min (+SDS/+PEG-Mal), and then incubated with 1 mM PEG-Mal at room temperature for 1 h. Samples were quenched with 45 mM DTT for 10 min. Then the samples were treated with lysozyme and DNaseI before SDS-PAGE as mentioned above. For PEG-Mal labeling in the presence of N-ethylmaleimide (NEM, Sigma), each of the 0.2 ml cell suspension was incubated with 5 mM NEM at room temperature for 30 min (+NEM/+SDS/+PEG-Mal), and centrifuged at 16,000 x *g* for 1 min. Pellets were washed twice with the HEPES buffer, and re-suspended in 0.2 ml of the same buffer. Samples were then incubated with 2% (w/v) SDS at 37°C for 5 min and with 1 mM PEG-Mal at room temperature for 1 h, and quenched with 45 mM DTT for 10 min. Then the samples were treated as mentioned above for SDS-PAGE.

Purification of WaaL proteins and in vitro ligase assay

The *E. coli* and *P. aeruginosa* WaaL proteins with FLAG and 10xHis tags were purified from *E. coli* JM105v as described previously (Ruan et al. 2012). LPS was prepared as described previously (Marolda *et al.*, 2006a). In the *in vitro* ligase assay for $WaaL_{Pa}$, the LPS from PAO1*waaL::*Gm^R(pXR93) expressing WapR in the absence or presence of IPTG (0-1 mM) contains both substrates for the reaction. The *in vitro* ligase assay for $WaaL_{Pa}$ was performed using 6 µl of LPS and purified *P. aeruginosa* WaaL (1 µM) in 50 mM phosphate buffer at pH 7.0 in 60 µl reaction volume, and the mixture was incubated at 22°C, 30°C or 37°C for 25 h, and terminated by adding 30 µg ml^{-1} proteinase K at 60°C for 20 min. The sample was deproteinized by adding 60 µl of hot phenol at 70°C for 15 min and centrifuged to separate the phenol phase.

542 The upper aqueous phase was collected and analyzed by Tricine-SDS-PAGE. The *in vitro* ligase
543 assay for $WaaL_{Ec}$ was processed as described previously (Ruan *et al.*, 2012).
544
545 *LPS and protein analysis*
546 LPS samples were separated on 14% (w/v) Tricine-SDS-PAGE and the gels stained with silver
547 nitrate (Marolda *et al.*, 2006a). For Western blot to detect β-O-linked N-acetylglucosamine,
548 samples in the polyacrylamide gel were transferred to nitrocellulose membranes and reacted with
549 O-GlcNAc Monoclonal Antibody (IgM) (Covance) at a 1:1000 dilution. For O-antigen detection
550 anti-O5 antigen monoclonal antibody MF15-4 (Antibodies-online) was used at a 1:10,000
551 dilution. Alexa Fluor® 680 Goat Anti-Mouse IgM (Invitrogen) or IRDye® 800CW Goat anti-
552 Mouse IgG (LI-COR) were used as secondary antibodies. For protein analysis, polyacrylamide
553 gels were transferred onto nitrocellulose membranes, which were blocked with 10% Western
554 Blocking Solution (Roche Diagnostics). Membranes were incubated overnight at 4°C with anti-
555 FLAG M2 monoclonal Antibody (Sigma) at a 1:5000 dilution or anti-His (Sigma) at a 1:10,000
556 dilution. IRDye® 800CW Goat anti-Mouse IgG (LI-COR) or Alexa Fluor® 680 Goat anti-mouse
557 immunoglobulin G (IgG) (Invitrogen) were used as secondary antibodies. Reacting bands were
558 detected by fluorescence with an Odyssey infrared imaging system (Li-cor Bioscience, Lincoln,
559 NE).
560
561 *Topology prediction*
562 The membrane topology of WaaL proteins was assessed by TMHMM (Sonnhammer *et al.*,
563 1998) and graphically displayed using Textopo (Beitz, 2000).
564

Chemical cross-linking

The *E. coli* WaaL and WaaLC-less proteins with FLAG-His tags were purified from *E. coli* as described previously (Ruan *et al.*, 2012). Cross-linking was performed using dithiobis[succinimidyl propionate] (DSP) following manufacturer´s instructions (Thermo Scientific). DSP was prepared as a 25 mM solution in DMSO and mixed with purified protein (in PBS) to a 5 mM final concentration. The mix was incubated for 45 min at room temperature. The reaction was stopped by adding the quenching solution (50 mM Tris-HCl pH 7.5) and incubated for 15 min. To cleave DSP, DTT was added at 50 mM final concentration and incubated at 37°C for 30 min. Samples were separated on 12.5% SDS-PAGE, using a loading buffer without DTT. Samples in the polyacrylamide gel were transferred to nitrocellulose membranes and incubated overnight with anti-His (Sigma) monoclonal antibody and IRDye® 800CW Goat anti-Mouse IgG (LI-COR) was used as secondary antibody. Reacting bands were detected by an Odyssey infrared imaging system (Li-cor Bioscience, Lincoln, NE).

Acknowledgements

This research was supported by grants from the Canadian Institutes of Health Research to M.A.V. The authors declare no conflict of interest.

References

Abeyrathne, P. D., Daniels, C., Poon, K. K., 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. *J Bacteriol* **187**: 3002-3012.

Abeyrathne, P. D., and Lam, J. S. (2007) WaaL of *Pseudomonas aeruginosa* utilizes ATP in *in vitro* ligation of O antigen onto lipid A-core. *Mol Microbiol* **65**: 1345-1359.

Alexeyev, M. F., and Winkler, H. H. (1999) Membrane topology of the *Rickettsia prowazekii* ATP/ADP translocase revealed by novel dual *pho-lac* reporters. *J Mol Biol* **285**: 1503-1513.

Beitz, E. (2000) TEXtopo: shaded membrane protein topology plots in LATEX 2ε. *Bioinformatics* **16**: 1050-1051.

Bogdanov, M., and Dowhan, W. (1999) Lipid-assisted protein folding. *J Biol Chem* **274**: 36827-36830.

Bogdanov, M., Heacock, P. N., and Dowhan, W. (2010) Study of polytopic membrane protein topological organization as a function of membrane lipid composition. *Methods Mol Biol* **619**: 79-101.

Cassel, M., Seppala, S., and von Heijne, G. (2008) Confronting fusion protein-based membrane protein topology mapping with reality: the *Escherichia coli* ClcA H^+/Cl^- exchange transporter. *J Mol Biol* **381**: 860-866.

Chung, B. C., Zhao, J., Gillespie, R. A., Kwon, D. Y., Guan, Z., Hong, J., Zhou, P., and Lee, S. Y. (2013) Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. *Science* **341**: 1012-1016.

Collins, R. F., Kargas, V., Clarke, B. R., Siebert, C. A., Clare, D. K., Bond, P. J., Whitfield, C., and Ford, R. C. (2017) Full-length, Oligomeric Structure of Wzz Determined by Cryoelectron Microscopy Reveals Insights into Membrane-Bound States. *Structure* **25**: 806-815 e803.

Feldman, M. F., Marolda, C. L., Monteiro, M. A., Perry, M. B., Parodi, A. J., and Valvano, M. A. (1999) The activity of a putative polyisoprenol-linked sugar translocase(Wzx) involved in *Escherichia coli* O antigen assembly is independent of the chemical structure of the O repeat. *J Biol Chem* **274**: 35129-35138.

Feldman, M. F., Wacker, M., Hernandez, M., Hitchen, P. G., Marolda, C. L., Kowarik, M., Morris, H. R., Dell, A., Valvano, M. A., and Aebi, M. (2005) Engineering *N*-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc Natl Acad Sci USA* **102**: 3016-3021.

Furlong, S. E., Ford, A., Albarnez-Rodriguez, L., and Valvano, M. A. (2015) Topological analysis of the *Escherichia coli* WcaJ protein reveals a new conserved configuration for the polyisoprenyl-phosphate hexose-1-phosphate transferase family. *Sci Rep* **5**: 9178.

Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. *J Bacteriol* **177**: 4121-4130.

Han, W., Wu, B., Li, L., Zhao, G., Woodward, R., Pettit, N., Cai, L., Thon, V., and Wang, P. G. (2012) Defining function of lipopolysaccharide O-antigen ligase WaaL using chemoenzymatically synthesized substrates. *J Biol Chem* **287**: 5357-5365.

Heinrichs, D. E., Monteiro, M. A., Perry, M. B., and Whitfield, C. (1998a) 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. *J Biol Chem* **273**: 8849-8859.

Heinrichs, D. E., Yethon, J. A., Amor, P. A., and Whitfield, C. (1998b) The assembly system for the outer core portion of R1- and R4-type lipopolysaccharides of *Escherichia coli*. The R1 core-specific β-glucosyltransferase provides a novel attachment site for O-polysaccharides. *J Biol Chem* **273**: 29497-29505.

Heinrichs, D. E., Yethon, J. A., and Whitfield, C. (1998c) Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. *Mol Microbiol* **30**: 221-232.

Hoang, T. T., Karkhoff-Schweizer, R. R., Kutchma, A. J., and Schweizer, H. P. (1998) A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. *Gene* **212**: 77-86.

Holloway, B. W., Krishnapillai, V., and Morgan, A. F. (1979) Chromosomal genetics of *Pseudomonas*. *Microbiol Rev* **43**: 73-102.

Hug, I., Couturier, M. R., Rooker, M. M., Taylor, D. E., Stein, M., and Feldman, M. F. (2010) *Helicobacter pylori* lipopolysaccharide is synthesized via a novel pathway with an evolutionary connection to protein *N*-glycosylation. *PLoS Path* **6**: e1000819.

Islam, S. T., Taylor, V. L., Qi, M., and Lam, J. S. (2010) Membrane topology mapping of the O-antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from *Pseudomonas aeruginosa* PAO1 reveals novel domain architectures. *mBio* **1**: e00189-00110-e00189-00119.

Joiner, K. A. (1988) Complement evasion by bacteria and parasites. *Ann Rev Microbiol* **42**: 201-230.

Knirel, Y. A. (2011) Structure of O-antigens. In: Bacterial lipopolysaccharides: Structure, chemical synthesis, biogenesis and interaction with host cells. Knirel, Y. A., and Valvano, M. A. (eds). Wein: Springer Verlag Publishing Inc., pp. 41-115.

Kocincova, D., and Lam, J. S. (2011) Structural diversity of the core oligosaccharide domain of *Pseudomonas aeruginosa* lipopolysaccharide. *Biochemistry (Mosc)* **76**: 755-760.

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu Rev Biochem* **77**: 521-555.

Lehrer, J., Vigeant, K. A., Tatar, L. D., and Valvano, M. A. (2007) Functional characterization and membrane topology of *Escherichia coli* WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O antigen lipopolysaccharide. *J Bacteriol* **189**: 2618-2628.

Leipold, M. D., Vinogradov, E., and Whitfield, C. (2007) Glycosyltransferases involved in biosynthesis of the outer core region of *Escherichia coli* lipopolysaccharides exhibit broader substrate specificities than is predicted from lipopolysaccharide structures. *J Biol Chem* **282**: 26786-26792.

Liston, S. D., Clarke, B. R., Greenfield, L. K., Richards, M. R., Lowary, T. L., and Whitfield, C. (2015) Domain interactions control complex formation and polymerase specificity in the biosynthesis of the *Escherichia coli* O9a antigen. *J Biol Chem* **290**: 1075-1085.

Marolda, C. L., Lahiry, P., Vinés, E., Saldías, S., and Valvano, M. A. (2006a) Micromethods for the characterization of lipid A-core and O-antigen lipopolysaccharide *Methods Mol Biol* **347**: 237-252.

Marolda, C. L., Tatar, L. D., Alaimo, C., Aebi, M., and Valvano, M. A. (2006b) Interplay of the wzx translocase and the corresponding polymerase and chain length regulator proteins in the

translocation and periplasmic assembly of lipopolysaccharide O antigen. *J Bacteriol* **188**: 5124-5135.

Marolda, C. L., Vicarioli, J., and Valvano, M. A. (2004) Wzx proteins involved in O antigen biosynthesis function in association with the first sugar of the O-specific lipopolysaccharide subunit. *Microbiology* **150**: 4095-4105.

Murazumi, N., Yamamori, S., Araki, Y., and Ito, E. (1979) Anomeric configuration of *N*-acetylglucosaminyl phosphorylundecaprenols formed in *Bacillus cereus* membranes. *J Biol Chem* **254**: 11791-11793.

Needham, B. D., and Trent, M. S. (2013) Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nat Rev Microbiol* **11**: 467-481.

Neuhard, J., and Thomassen, E. (1976) Altered deoxyribonucleic pools in P2 eductants of *Escherichia coli* K-12 due to deletion of the *dcd* gene. *J Bacteriol* **126**: 999-1001.

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. *Nat Rev Microbiol* **14**: 337-345.

Olsthoorn, M. M., Petersen, B. O., Duus, J., Haverkamp, J., Thomas-Oates, J. E., Bock, K., and Holst, O. (2000) The structure of the linkage between the O-specific polysaccharide and the core region of the lipopolysaccharide from *Salmonella enterica* serovar Typhimurium revisited. *Eur J Biochem* **267**: 2014-2027.

Ortega, X., Silipo, A., Saldias, M. S., Bates, C. C., Molinaro, A., and Valvano, M. A. (2009) Biosynthesis and structure of the *Burkholderia cenocepacia* K56-2 lipopolysaccharide core oligosaccharide: truncation of the core oligosaccharide leads to increased binding and sensitivity to polymyxin B. *J Biol Chem* **284**: 21738-21751.

Pan, Y., Ruan, X., Valvano, M. A., and Konermann, L. (2012) Validation of protein topology models by oxidative labeling and mass spectrometry. *J Am Soc Mass Spectr* **23**: 889-898.

Pérez, J. M., McGarry, M. A., Marolda, C. L., and Valvano, M. A. (2008) Functional analysis of the large periplasmic loop of the *Escherichia coli* K-12 WaaL O-antigen ligase. *Mol Microbiol* **70**: 1424-1440.

Pluschke, G., and Achtman, M. (1984) Degree of antibody-independent activation of the classical complement pathway by K1 *Escherichia coli* differs with O antigen type and correlates with virulence of meningitis in newborns. *Infect Immun* **43**: 684-692.

Ruan, X., Pérez, J. M., Marolda, C. L., 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.

Saldías, M. S., Patel, K., Marolda, C. L., Bittner, M., Contreras, I., and Valvano, M. A. (2008) Distinct functional domains of the *Salmonella enterica* WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit. *Microbiology* **154**: 440-453.

Seol, W., and Shatkin, A. J. (1993) Membrane topology model of *Escherichia coli* α-ketoglutarate permease by phoA fusion analysis. *J Bacteriol* **175**: 565-567.

Smith, A. L., Friedman, D. B., Yu, H., Carnahan, R. H., and Reynolds, A. B. (2011) ReCLIP (reversible cross-link immuno-precipitation): an efficient method for interrogation of labile protein complexes. *PLoS One* **6**: e16206.

Sonnhammer, E. L. L., von Heijne, G., and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In: Proceedings of Sixth International Conference on Intelligent Systems for Molecular Biology. Glasgow, J., Littlejohn, T., Major, F., Lathrop, R., Sankoff, D., and Sensen, C. (eds). Menlo Park, Ca: AAAI Press, pp. 175-182.

Sperandeo, P., Martorana, A. M., and Polissi, A. (2017) The lipopolysaccharide transport (Lpt) machinery: a non-conventional transporter for lipopolysaccharide assembly at the outer membrane of Gram-negative bacteria. *J Biol Chem*.

Tatar, L. D., Marolda, C. L., Polischuk, A. N., van Leeuwen, D., and Valvano, M. A. (2007) An Escherichia coli undecaprenyl-pyrophosphate phosphatase implicated in undecaprenyl phosphate recycling. *Microbiology* **153**: 2518-2529.

Valvano, M. A. (2011) Common Themes in Glycoconjugate Assembly Using the Biogenesis of O-Antigen Lipopolysaccharide as a Model System. *Biochemistry-Moscow* **76**: 729-735.

Valvano, M. A., and Crosa, J. H. (1989) Molecular cloning and expression in *Escherichia coli* K-12 of chromosomal genes determining the O7 lipopolysaccharide antigen of a human invasive strain of *E. coli* O7:K1. *Infect Immun* **57**: 937-943.

Valvano, M. A., Patel, K. B., and Furlong, S. E. (2011) Genetics, biosynthesis and assembly of O antigen. In: Bacterial lipopolysaccharides: Structure, chemical synthesis, biogenesis and interaction with host cells. Knirel, Y., and Valvano, M. A. (eds). Wein: Springer Verlag Publishing Inc., pp. 275-310.

Vinés, E., Marolda, C. L., Balachandran, A., and Valvano, M. A. (2005) Defective O antigen polymerization in *tolA* and *pal* mutants of *Escherichia coli* in response to extracytoplasmic stress. *J Bacteriol* **187**: 3359-3368.

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.

West, S. E., Schweizer, H. P., Dall, C., Sample, A. K., and Runyen-Janecky, L. J. (1994) Construction of improved *Escherichia-Pseudomonas* shuttle vectors derived from pUC18/19 and sequence of the region required for their replication in *Pseudomonas aeruginosa*. *Gene* **148**: 81-86.

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

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Table 1. Strains and plasmids

Strain or plasmid	Relevant properties[a]	Source/reference
E. coli		
CLM24	W3110, *ΔwaaL*	(Feldman *et al.*, 2005)
CLM35	W3110, *ΔwecA Δwzx*	(Marolda *et al.*, 2006b)
DH5α	F^- Φ80 *lacZM15 endA recA hsdR*(r_k^- m_k^+) *supE thi gyrA relA* Δ*(lacZYA-argF) U169*	Laboratory stock
F632	*E. coli* R2, R-LPS mutant of O100:K? (B):H2	(Heinrichs *et al.*, 1998a)
SCM3	SØ874, Δ*waaL*	(Pérez *et al.*, 2008)
SØ874	*lacZ trp*Δ*(sbcB-rfb)upp rel rspL*	(Neuhard and Thomassen, 1976)
W3110	*E. coli* K-12 *rph-1* IN*(rrnD-rrnE)1*	Laboratory stock
JM105v	JM105 derivative with a spontaneous mutation resulting in a truncated lipid A-core OS	Laboratory stock
P. aeruginosa		
PAO1	Prototypical laboratory strain (from L. Pasador)	(Holloway *et al.*, 1979)
PAO1*waaL*::Gm^R	PAO1, Δ*waaL*, Gm^R	This study
B. cenocepacia		
K56-2	Cystic fibrosis clinical isolate	Laboratory stock
K56-2*waaL*::pGPΩTp	K56-2, *waaL*::pGPΩTp	(Ortega *et al.*, 2009)
Plasmids		
pBADFLAG	pBAD24; for FLAG-tagged C-terminal fusions; arabinose inducible, Ap^R	(Saldías *et al.*, 2008)
pBAD33	pBAD18 containing the pBR322-compatible replication origin from pACYC 184, Cm^R	(Guzman *et al.*, 1995)
pEX18Tc	Tc^R*, oriT*$^+$, *sacB*$^+$, gene replacement vector with multi cloning site from pUC18	(Hoang *et al.*, 1998)
pJHCV32	*wb(rfb)*$_{EcO7}$ cosmid, Tc^R $O7^+$	(Valvano and Crosa, 1989)
pUCP26	pUC18-derived broad-host-range vector, Tc^R	(West *et al.*, 1994)
pXR1	pBADFLAG, *E. coli waaL*$_{FLAG\text{-}10xHis}$, Ap^R	(Ruan *et al.*, 2012)
pXR4	pXR1, *E. coli waaL*$_{R215A}$, Ap^R	(Ruan *et al.*, 2012)
pXR5	pXR1, *E. coli waaL*$_{R215K}$, Ap^R	(Ruan *et al.*, 2012)
pXR7	pXR1, *E. coli waaL*$_{R288A}$, Ap^R	(Ruan *et al.*, 2012)
pXR14	pBADFLAG, *P. aeruginosa waaL*$_{FLAG\text{-}10xHis}$, Ap^R	(Ruan *et al.*, 2012)
pXR39	pXR1, *E. coli waaL*$_{ΔN2TM}$; *E. coli waaL*$_{(L2\text{-}N71)}$ deleted; Ap^R	This study
pXR76	pBAD33, *E. coli waaL*$_{ΔN2TM}$, Ap^R	This study
pXR36	pXR1, *E. coli waaL*$_{Cys\text{-}less}$, Ap^R	This study
pXR33	pXR1, *E. coli waaL*$_{C29}$, from *E. coli waaL*$_{C54A/C203A/C225A/C428A}$, Ap^R	This study
pXR38	pXR1, *E. coli waaL*$_{C428}$, from *E. coli waaL*$_{C29A/C54A/C203A/C225A}$, Ap^R	This study
pXR52	pXR36, *E. coli waaL*$_{I162C}$, Ap^R	This study
pXR53	pXR36, *E. coli waaL*$_{T176C}$, Ap^R	This study
pXR54	pXR36, *E. coli waaL*$_{T190C}$, Ap^R	This study

pXR55	pXR36, *E. coli waaL*$_{T214C}$, Ap^R	This study
pXR56	pXR36, *E. coli waaL*$_{A216C}$, Ap^R	This study
pXR43	pXR36, *E. coli waaL*$_{G286C}$, Ap^R	This study
pXR23	pXR14, *P. aeruginosa waaL*$_{Cys\text{-}less}$, from *P. aeruginosa waaL*$_{C114A/C403A}$, Ap^R	This study
pXR22	pXR14, *P. aeruginosa waaL*$_{C114}$, from *P. aeruginosa waaL*$_{C403A}$, Ap^R	This study
pXR21	pXR14, *P. aeruginosa waaL*$_{C403}$, from *P. aeruginosa waaL*$_{C114A}$, Ap^R	This study
pXR65	pXR23, *P. aeruginosa waaL*$_{R147C}$, Ap^R	This study
pXR45	pXR23, *P. aeruginosa waaL*$_{L155C}$, Ap^R	This study
pXR46	pXR23, *P. aeruginosa waaL*$_{L157C}$, Ap^R	This study
pXR66	pXR23, *P. aeruginosa waaL*$_{T158C}$, Ap^R	This study
pXR67	pXR23, *P. aeruginosa waaL*$_{G161C}$, Ap^R	This study
pXR68	pXR23, *P. aeruginosa waaL*$_{N165C}$, Ap^R	This study
pXR47	pXR23, *P. aeruginosa waaL*$_{S169C}$, Ap^R	This study
pXR69	pXR23, *P. aeruginosa waaL*$_{A170C}$, Ap^R	This study
pXR48	pXR23, *P. aeruginosa waaL*$_{A178C}$, Ap^R	This study
pXR49	pXR23, *P. aeruginosa waaL*$_{S187C}$, Ap^R	This study
pXR50	pXR23, *P. aeruginosa waaL*$_{S211C}$, Ap^R	This study
pXR51	pXR23, *P. aeruginosa waaL*$_{T213C}$, Ap^R	This study
pXR70	pXR23, *P. aeruginosa waaL*$_{L251C}$, Ap^R	This study
pXR71	pXR23, *P. aeruginosa waaL*$_{P253C}$, Ap^R	This study
pXR72	pXR23, *P. aeruginosa waaL*$_{V255C}$, Ap^R	This study
pXR44	pXR23, *P. aeruginosa waaL*$_{G282C}$, Ap^R	This study
pXR73	pXR23, *P. aeruginosa waaL*$_{L307C}$, Ap^R	This study
pXR74	pXR23, *P. aeruginosa waaL*$_{V309C}$, Ap^R	This study
pXR88	pXR14, *P. aeruginosa waaL*$_{R212A}$, Ap^R	This study
pXR75	pUCP26, *P. aeruginosa waaL*, Tc^R	This study
pXR79	pUCP26, *P. aeruginosa waaL*$_{Cys\text{-}less}$, from *P. aeruginosa waaL*$_{C114A/C403A}$, Tc^R	This study
pXR89	pUCP26, *P. aeruginosa waaL*$_{R212A}$, Tc^R	This study
pXR93	pUCP26, *P. aeruginosa wapR,* Tc^R	This study
pXR95	pUCP26, *P. aeruginosa waaL*$_{N165C}$, Tc^R	This study
pJM7	pUCP26, *P. aeruginosa waaL*$_{R147C}$, Tc^R	This study
pJM8	pUCP26, *P. aeruginosa waaL*$_{L155C}$, Tc^R	This study
pJM9	pUCP26, *P. aeruginosa waaL*$_{L157C}$, Tc^R	This study
pJM10	pUCP26, *P. aeruginosa waaL*$_{T158C}$, Tc^R	This study
pJM11	pUCP26, *P. aeruginosa waaL*$_{G161C}$, Tc^R	This study
pJM12	pUCP26, *P. aeruginosa waaL*$_{N165C}$, Tc^R	This study
pJM13	pUCP26, *P. aeruginosa waaL*$_{S169C}$, Tc^R	This study
pJM14	pUCP26, *P. aeruginosa waaL*$_{A170C}$, Tc^R	This study
pJM15	pUCP26, *P. aeruginosa waaL*$_{A178C}$, Tc^R	This study
pJM16	pUCP26, *P. aeruginosa waaL*$_{S187C}$, Tc^R	This study
pJM17	pUCP26, *P. aeruginosa waaL*$_{S211C}$, Tc^R	This study
pJM18	pUCP26, *P. aeruginosa waaL*$_{T213C}$, Tc^R	This study
pJM19	pUCP26, *P. aeruginosa waaL*$_{L251C}$, Tc^R	This study
pJM20	pUCP26, *P. aeruginosa waaL*$_{P253C}$, Tc^R	This study
pJM21	pUCP26, *P. aeruginosa waaL*$_{V255C}$, Tc^R	This study
pJM22	pUCP26, *P. aeruginosa waaL*$_{G282C}$, Tc^R	This study
pJM23	pUCP26, *P. aeruginosa waaL*$_{L307C}$, Tc^R	This study
pJM24	pUCP26, *P. aeruginosa waaL*$_{V309C}$, Tc^R	This study
pJM25	pXR79, *P. aeruginosa waaL*$_{G282A}$, Tc^R	This study

a. Ap^R, ampicillin resistance; Cm^R, chloramphenicol resistance; Tc^R, tetracycline resistance; Gm^R, gentamicin resistance.

Figure legends

Fig. 1. Topology of WaaL proteins from *E. coli* and *P. aeruginosa*.

A. Cartoon representation of *in silico* topology models of WaaL proteins from *E. coli* and *P. aeruginosa* ($WaaL_{Ec}$-TMHMM and $WaaL_{Pa}$-TMHMM, respectively), which were obtained with TMHMM (Sonnhammer *et al.*, 1998). Also shown is a cartoon representation of the topology model of *P. aeruginosa* WaaL ($WaaL_{Pa}$-Del/Fus) obtained from the analysis of *phoA-lacZα* dual-reporter fusions (Islam *et al.*, 2010). Numbers next to each predicted transmembrane helix indicate the position of amino acid residues at the predicted boundaries between soluble and membrane (TMH) embedded regions. TMHs are indicated by roman numerals. Shaded TMHs in $WaaL_{Pa}$-Del/Fus denote helices that show major differences with in the $WaaL_{Pa}$-TMHMM model. PL, periplasmic loop; CL, cytoplasmic loop. Purple stars symbolize the location of known functional residues in $WaaL_{Ec}$ (Pérez *et al.*, 2008, Ruan *et al.*, 2012) and $WaaL_{Pa}$ (Abeyrathne and Lam, 2007), which were identified previously and also in this study (R212 in $WaaL_{Pa}$).

B. Linear representation of the topological models in A. The numbers indicate the position of the amino acid residues in $WaaL_{Ec}$ (419 amino acids) and $WaaL_{Pa}$ (401 amino acids). Concordant TMH segments in the three models are indicated in black. TMH segments IV, V, VI, VII and VIII in $WaaL_{Pa}$-Del/Fus (light gray) differ from those predicted *in silico* for both *E. coli* and *P. aeruginosa* WaaL proteins.

887 **Fig. 2.** Sulfhydryl labeling of *E. coli* WaaL cysteine-replacement mutants.

888 A. Schematic description of the sulfhydryl labeling test by PEGylation. Predicted western blot
889 profiles of WaaL polypeptides with single cysteine replacements based on the accessibility of the
890 residue depending on its location and the treatment with NEM, SDS or both. The expected
891 profiles denote residues located in the periplasm, in periplasmic/TMH borders, embedded in the
892 TMH, in TMH/cytoplasmic borders, or exposed to the cytoplasm. The structures of PEG-Mal
893 and NEM are also indicated.

894 B. *E. coli* DH5α producing plasmid-encoded Cys-substituted variants of $WaaL_{Ec}$ were grown at
895 37ºC in 5 ml of cultures to mid-exponential phase and processed as explained in *Experimental*
896 *Procedures*. 15-µl aliquots were separated by SDS-PAGE and immunoblotted using the anti-
897 FLAG monoclonal antibody. Modification of *E. coli* WaaL by PEG-Mal causes a shift in the
898 WaaL band of ~20 kDa. M, Dual Color Precision Plus Protein standards (Bio-Rad).

899 No PEGMal. Control experiment in HEPES buffer only. A sample from *E. coli* expressing the
900 parental WaaL ($WaaL_{Ec}$) was loaded as an expression control.

901 PEGMal. Bacterial cell suspensions were treated with 1 mM PEG-Mal and no SDS.

902 NEM/SDS/PEGMal. Cell suspensions were first treated with 5 mM NEM and then with 1 mM
903 PEG-Mal in the presence of 2% SDS.

904 SDS/PEGMal. Cell suspensions were treated with 1 mM PEG-Mal in the presence of 2% SDS.

905

906 **Fig. 3.** Sulfhydryl labeling of *P. aeruginosa* WaaL cysteine-replacement mutants.

907 *E. coli* DH5α producing plasmid-encoded Cys-substituted variants of $WaaL_{Pa}$ were grown and
908 processed as indicated in the legend to Fig. 2.

No PEGMal. Control experiment in HEPES buffer only. A sample from *E. coli* expressing the parental WaaL ($WaaL_{Pa}$) was loaded as an expression control.

PEGMal. Bacterial cell suspensions were treated with 1 mM PEG-Mal and no SDS.

NEM/SDS/PEGMal. Cell suspensions were first treated with 5 mM NEM and further treated with 1 mM PEG-Mal in the presence of 2% SDS

SDS/PEGMal. Cell suspensions were treated with 1 mM PEG-Mal in the presence of 2% SDS.

Fig. 4. Cysteine less and single-cysteine substitutions of $WaaL_{Ec}$ and $WaaL_{Pa}$ do not affect ligase function.

A. Silver-stained SDS-polyacrylamide gel showing the LPS profile of *E. coli* SCM3(pJHCV32) cells expressing parental $WaaL_{Ec}$, pBADFLAG plasmid control, $WaaL_{Ec}$C-less, and various single cysteine replacement $WaaL_{Ec}$ mutants, as indicated.

B. Silver-stained SDS-PAGE showing the LPS profile of *P. aeruginosa* PAO1 as a positive control, and the profiles of *P. aeruginosa* PAO1*waaL*::Gm^R with no plasmid, pUCP26 vector control, and plasmids expressing $WaaL_{Pa}$, WaaL-R212A and $WaaL_{Pa}$C-less. Lower panel shows immunoblot using anti-His antibody of total membrane fraction (Mem) of of *P. aeruginosa* PAO1*waaL*::Gm^R with no plasmid, pUCP26 vector control, and plasmids expressing $WaaL_{Pa}$, WaaL-R212A and $WaaL_{Pa}$C-less.

C. Silver-stained SDS-PAGE showing the LPS profile of *P. aeruginosa* PAO1*waaL*::Gm^R expressing parental $WaaL_{Pa}$, no plasmid, pUCP26 vector control, and plasmids expressing $WaaL_{Pa}$C-less and $WaaL_{Pa}$N165C. Middle panel shows the immunoblot of extracted LPS using anti-Oantigen antibody. Protein expression was tested in total membrane (Mem) fractions using anti-His antibody (lower panel).

Fig. 5. Topological models of $WaaL_{Ec}$ and $WaaL_{Pa}$ displayed with Textopo (Beitz, 2000). The *E. coli* WaaL and B. *P. aeruginosa* WaaL topologies deduced from PEGylation experiments in this study were compared to the *P. aeruginosa* WaaL topology based on dual-reporter fusions (Islam *et al.*, 2010). PL, periplasmic loop; CL, cytoplasmic loop. TMHs are indicated in roman numerals. Residues in black squares denote TMH location. Purple diamonds indicate functional residues that when replaced by alanine abolish ligation. Methionine residues accessible by oxidation in $WaaL_{Ec}$ are indicated by blue squares with black letters. Residues in blue circles are exposed to the periplasmic space, while residues in red circles are exposed to the cytosol. White residues in blue square background were experimentally positioned at the TMH-periplasmic border. The C-terminal residues of $WaaL_{Ec}$ and $WaaL_{Pa}$ protein (N419 and P401, respectively), as well as the additional residues of the FLAG-10xHis tag peptides are also indicated. To better emphasize the differences between the two $WaaL_{Pa}$ models the location of the residues determined by PEGylation in *P. aeruginosa* WaaL (revised topology) has been maintained in the *P. aeruginosa* WaaL model.

A. Topology of $WaaL_{Ec}$.

B. Revised topology of $WaaL_{Pa}$ based on cysteine accessibility experiments using the intact protein.

C. $WaaL_{Pa}$ topology based on dual-reporter fusions using truncated proteins (Islam *et al.*, 2010).

Fig. 6. *In vitro* ligase assay for *P. aeruginosa* WaaL. LPS produced *in vivo* by the PAO1*waaL::*Gm^R expressing *P. aeruginosa* WaaL (Lane 1) or WapR (Lane 2). The *in vitro* ligation assay using purified $WaaL_{Pa}$ mixed with extracted LPS from PAO1*waaL::*Gm^R was

955 tested at 22°C (Lane 3), 30°C (Lane 4), or 37°C (Lane 5) . Product of *in vitro* ligation incubated
956 at 37°C using purified *P. aeruginosa* WaaL mixed with LPS from PAO1*waaL::*Gm^R expressing
957 WapR at different IPTG concentrations (Lane 6-8). Product of *in vitro* ligation in the prescence
958 of 10 mM EDTA mixed with LPS from PAO1*waaL::*Gm^R expressing WapR.(Lane 9). Samples
959 were analyzed by SDS-PAGE followed by silver staining. The resolved O-antigen products are
960 indicated by the dashed rectangle.
961
962 **Fig. 7.** Properties of the ligation reaction
963 A. The chemical structures of the core oligosaccharide of the various LPS types used in these
964 experiments are indicated in the respective boxes. LPS samples were examined by silver staining
965 and by immunoblot with a monoclonal antibody against β-O-linked N-acetylglucosamine (anti-
966 β-O-GlcNAc).
967 B. The *in vitro* ligation of K-12 core with *P. aeruginosa* PAO1 O-antigen by purified $WaaL_{Ec}$
968 was performed as described in *Experimental Procedures*. Lanes: 1, PAO1, LPS produced *in vivo*
969 by the PAO1 *waaL::*Gm^R(pXR75) strain expressing $WaaL_{Pa}$; 2, $Core_{Ec}$+ LPS_{Pa}, negative control
970 for the *in vitro* ligation assay consisting of a mixture of LPS from CLM35 ($Core_{Ec}$) and from
971 PAO1*waaL::*Gm^R(pXR93) prepared in the presence of 0.2 mM IPTG (LPS_{Pa}); 3,
972 $Core_{Ec}$+$WaaL_{Ec}$, negative control for the *in vitro* ligation assay with purified $WaaL_{Ec}$ and *E. coli*
973 LPS ($Core_{Ec}$); 4, LPS_{Pa}+$WaaL_{Ec}$, *in vitro* ligation assay with $WaaL_{Ec}$ and LPS_{Pa}; 5,
974 $Core_{Ec}$+LPS_{Pa}+$WaaL_{Ec}$, *in vitro* ligation assay with $WaaL_{Ec}$ mixed with $Core_{Ec}$ and LPS_{Pa}; 6,
975 LPS_{Ec}+$WaaL_{Ec}$, *in vitro* ligation assay with $WaaL_{Ec}$ and LPS from CLM24(pMF19) (LPS_{Ec})
976 expressing K-12 O-antigen; 7, CLM24(pMF19/pXR1) O-antigen production *in vivo*.

977 C. *In vitro* ligation of K-12 core with *B. cenocepacia* K56-2 O-antigen by purified $WaaL_{Ec}$. *In*
978 *vitro* ligation reactions were carried out as described in panel B. Lanes: 1, K56-2, LPS produced
979 *in vivo* by parental *B. cenocepacia* K56-2; 2, $waaL_{Bc}$::pGPΩTp, LPS produced *in vivo* by the *B.*
980 *cenocepacia* K56-2 $waaL_{Bc}$::pGPΩTp ligase deficient mutant; 3, $Core_{Ec}$+LPS_{Bc}, negative control
981 for the *in vitro* ligation assay consisting of a mixture of LPS from CLM35 ($Core_{Ec}$) and from
982 K56-2 $waaL_{Bc}$::pGPΩTp (LPS_{Bc}); 4, $Core_{Ec}$+$WaaL_{Ec}$, negative control for the *in vitro* ligation
983 assay with purified $WaaL_{Ec}$ and *E. coli* LPS ($Core_{Ec}$); 5, LPS_{Bc}+$WaaL_{Ec}$, *in vitro* ligation assay
984 with $WaaL_{Ec}$ and LPS_{Bc}; 6, $Core_{Ec}$+LPS_{Bc}+$WaaL_{Ec}$, *in vitro* ligation assay with $WaaL_{Ec}$, $Core_{Ec}$,
985 and LPS_{Bc}.
986
987 **Fig. 8.** Reconstitution of *E. coli* ligase function from two inactive WaaL ligases co-expressed in
988 the same bacterial strain and WaaL oligomerization.
989 A. Expression of *E. coli* WaaL constructs in intact cells detected by immunoblot with the anti-
990 FLAG antibody. The constructs were expressed in the presence of 0.2% (w/v) L-arabinose, as
991 described in *Experimental Procedures.*
992 B. O7 antigen production by *E. coli* SCM3(pJHCV32) expressing single (left panel) or double
993 (right panel) combinations of WaaL constructs, expressed from, pBAD24, pBAD33 or both, as
994 indicated by the cartoons above each lane.
995 C. Oligomerization analysis of purified WaaL and WaaLC-less after treatment with the DSP
996 crosslinker (5mM) and under reducing conditions (DTT). Samples were processed for SDS-
997 PAGE and Western blotting with anti-His antibody. WaaL forms protein complexes that are not
998 present when incubated with DTT or weakly formed in the C-less version. In non-reducing
999 conditions, a monomeric, a dimeric form (**) and multimeric forms (****) were visible.