



## OPEN

SUBJECT AREAS:  
MICROBIOLOGY  
BIOCHEMISTRYReceived  
24 November 2014Accepted  
24 February 2015Published  
17 March 2015Correspondence and  
requests for materials  
should be addressed to  
M.A.V. (m.valvano@  
qub.ac.uk)

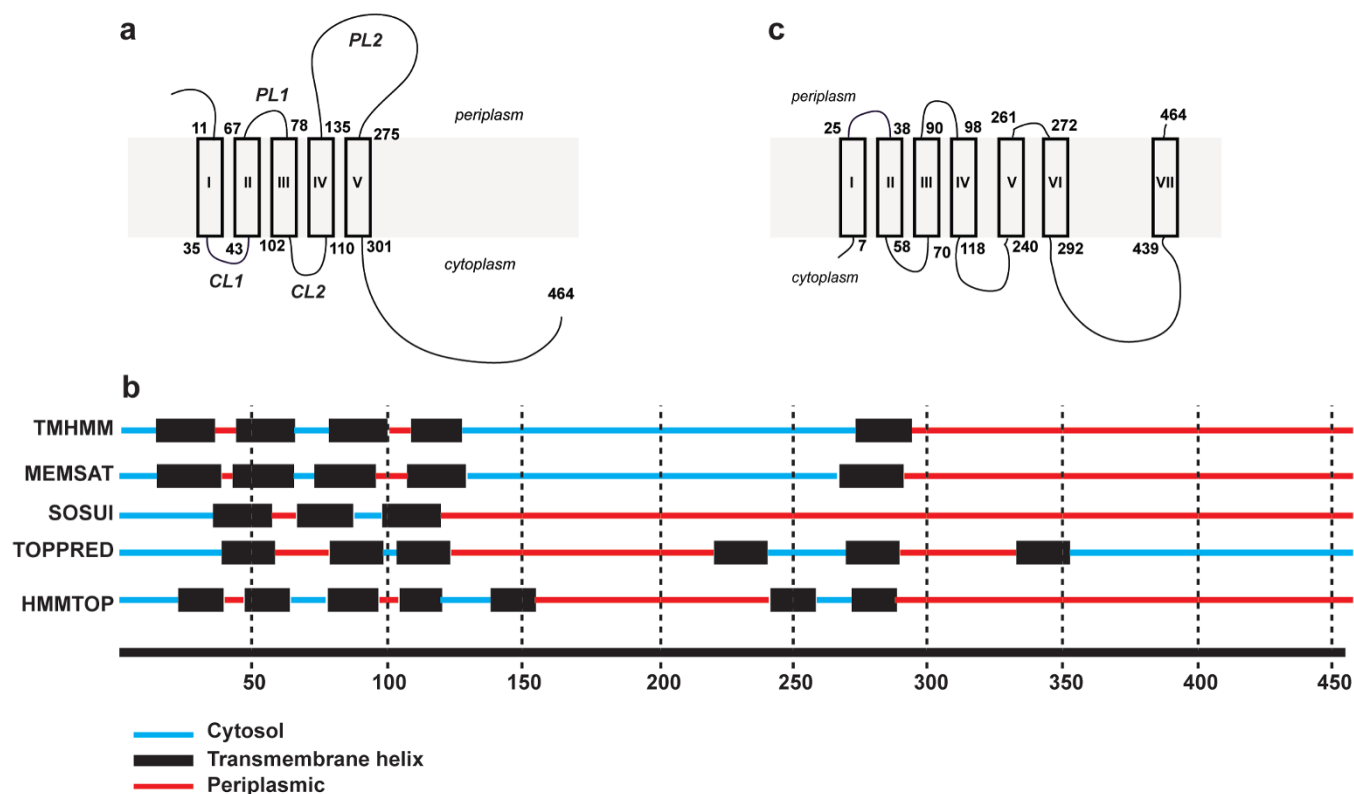
# Topological analysis of the *Escherichia coli* WcaJ protein reveals a new conserved configuration for the polyisoprenyl-phosphate hexose-1-phosphate transferase family

Sarah E. Furlong<sup>1</sup>, Amy Ford<sup>2</sup>, Lorena Albarnez-Rodriguez<sup>1</sup> & Miguel A. Valvano<sup>1,2</sup><sup>1</sup>Centre for Human Immunology, Department of Microbiology and Immunology, University of Western Ontario, London, Ontario, N6A 5C1, Canada, <sup>2</sup>Centre for Infection and Immunity, Queen's University Belfast, Belfast, United Kingdom, BT9 7AE.

WcaJ is an *Escherichia coli* membrane enzyme catalysing the biosynthesis of undecaprenyl-diphosphate-glucose, the first step in the assembly of colanic acid exopolysaccharide. WcaJ belongs to a large family of polyisoprenyl-phosphate hexose-1-phosphate transferases (PHPTs) sharing a similar predicted topology consisting of an N-terminal domain containing four transmembrane helices (TMHs), a large central periplasmic loop, and a C-terminal domain containing the fifth TMH (TMH-V) and a cytosolic tail. However, the topology of PHPTs has not been experimentally validated. Here, we investigated the topology of WcaJ using a combination of LacZ/PhoA reporter fusions and sulfhydryl labelling by PEGylation of novel cysteine residues introduced into a cysteine-less WcaJ. The results showed that the large central loop and the C-terminal tail both reside in the cytoplasm and are separated by TMH-V, which does not fully span the membrane, likely forming a "hairpin" structure. Modelling of TMH-V revealed that a highly conserved proline might contribute to a helix-break-helix structure in all PHPT members. Bioinformatic analyses show that all of these features are conserved in PHPT homologues from Gram-negative and Gram-positive bacteria. Our data demonstrate a novel topological configuration for PHPTs, which is proposed as a signature for all members of this enzyme family.

The biosynthesis of bacterial glycoconjugates such as cell wall peptidoglycan, exopolysaccharides, and the O antigen moiety of the lipopolysaccharide starts by the formation of a phosphoanhydride linkage between undecaprenyl pyrophosphate (Und-P) and a sugar-1-phosphate donated from a sugar nucleoside diphosphate<sup>1</sup>. Two distinct families of membrane enzymes, termed polyisoprenyl-phosphate N-acetylaminosugar-1-phosphate transferases and polyisoprenyl-phosphate hexose-1-phosphate transferases (PHPT), catalyse this reaction<sup>2,3</sup>. These enzymes are indispensable for lipid-linked glycan biosynthesis because they connect soluble cytosolic components (sugar phosphates donated as nucleotide sugars) with membrane-bound Und-P lipid carrier.

Unlike N-acetylaminosugar-1-phosphate transferases, PHPTs have no known eukaryotic homologues<sup>2</sup>. PHPTs initiate the biosynthesis of capsule<sup>4,5</sup>, O antigen<sup>6</sup>, and glycans attached to bacterial glycoproteins<sup>7</sup>. Most family members share similar predicted topology to that of *Salmonella enterica* WbaP, a prototypic PHPT that consists of five predicted transmembrane helices (TMHs), a large periplasmic loop, and an extended C-terminal region (Fig. 1a)<sup>6,8,9</sup>. While the predicted N-terminal and periplasmic domains are not always present in proteins of the PHPT family, the TMH-V and the C-terminal domain are conserved in all family members, and sufficient for activity in those proteins that naturally possess that "simpler" structure, as well as in truncated versions of the more complex family members<sup>6,8,10</sup>. The function of the N-terminal and large periplasmic loop domains is unclear. Deletion or overexpression of the predicted periplasmic loop of WbaP in *S. enterica* affects the distribution of O-antigen chain length and deletion mutant proteins show reduced *in vitro* transferase activity<sup>6</sup>. In the Gram-positive bacterium *S. pneumoniae*, the predicted extracellular loop of the glucosyltransferase Cps2E (corresponding to the periplasmic domain in PHPTs from Gram-negative bacteria) is important in modulating



**Figure 1 | Topologies of PHPT proteins based on bioinformatic predictions.** (a) Predicted model for WcaJ from *E. coli* W3110 that is similar to that of WbaP<sup>2,6</sup>, as explained in the Introduction. Numbers next to each predicted transmembrane helix (TMH) indicate the position of amino acid residues at the predicted boundaries between soluble and membrane embedded regions. TMHs are indicated by roman numerals. PL, periplasmic loop; CL, cytoplasmic loop. (b) Graphical representation of the topological predictions by various programs. The numbers indicate the position of the amino acid residues in WcaJ (464 amino acids). The location of soluble segments (cytosolic or periplasmic) and the positions of predicted TMHs are indicated. (c) Predicted model for BceB from *B. cenocepacia* IST432, a WcaJ homologue, which was derived from TMH predictions, experimental data, and refined based on hydrophobicity and conformational parameters for beta turns<sup>14</sup>. Numbers next to each TMH indicate the position of amino acid residues at the predicted boundaries between soluble and membrane embedded regions. TMHs are indicated by roman numerals.

polymer assembly and the rate of the initiation reaction<sup>4,11</sup>. Therefore, the large periplasmic/extracellular domain of PHPT proteins could regulate the catalytic region located on the C-terminal domain<sup>3</sup>.

Unfortunately, much of the current understanding of PHPT function relies on a predicted topological model of the protein that has not been experimentally validated. The predicted topology of WbaP was first proposed by Valvano<sup>2</sup> based on combined bioinformatic data from DAS and TMHMM programs<sup>12,13</sup>, which indicated two large soluble regions separated by one TMH. The DAS program does not provide topological information for the location of soluble domains, while TMHMM suggested that the C-terminal soluble domain was in the periplasmic space (Fig. 1b). Since this prediction did not agree with experimental data indicating that the C-terminal half of WbaP is required for the transferase function<sup>9</sup>, the C-terminal region of WbaP was assigned to the cytosolic space, where sugar nucleoside diphosphate precursors become available. This assignment was subsequently validated experimentally by protease accessibility experiments and microscopy using recombinant WbaP C-terminally fused to the green fluorescent protein, a reporter for cytoplasmic location of membrane protein segments<sup>8</sup>. However, *in silico* predictions of PHPT proteins give inconsistent results depending on the program used (Fig. 1b). Major areas of disagreement in the predictions are the number of TMHs (5 to 7) and the subcellular localization of the periplasmic/extracellular region and the C-terminal tails (Fig. 1b). Protease accessibility experiments in WbaP showed no cleavage in the predicted large periplasmic region despite the presence of multiple trypsin cleavage sites, suggesting that

this domain may not be located in the periplasm<sup>8</sup>. In contrast, using LacZ and PhoA fusions combined with secondary structure predictions, Videira *et al.*<sup>14</sup> proposed that BceB, a PHPT enzyme that initiates the biosynthesis of the cepacian exopolysaccharide in *Burkholderia cenocepacia*, consists of seven TMHs. This model also predicts a novel TMH adjacent to TMH-V and another TMH near the C-terminus that would place the C-terminal end of the protein on the periplasmic space (Fig. 1c)<sup>14</sup>. The inconsistencies between *in silico* predictions and limited experimental data<sup>8,14</sup> underscore the need to carefully address the topology of PHPTs.

In this work, we utilized WcaJ, a model PHPT that shares similar features as WbaP<sup>5</sup>. WcaJ is an integral membrane protein in *Escherichia coli* K-12 and other *Enterobacteriaceae*, which catalyses the transfer of glucose-1-phosphate from UDP-glucose (UDP-Glc) to Und-P resulting in the formation of Und-P-P-Glc. This is the first step in the biosynthesis of colanic acid (CA) capsule or M antigen<sup>5,15</sup>. CA is a high-molecular weight, extracellular polysaccharide required for biofilm formation, resistance to desiccation, and cell stress responses<sup>16,17</sup>. CA is loosely connected with the outer membrane, but under specific conditions can also be attached to the lipopolysaccharide (LPS) core, resulting in M<sub>LPS</sub><sup>18</sup>. The production of CA is regulated by the *rsc* (regulator of capsule biosynthesis) genes *rscA*, *rscB*, and *rscC*. RcsA and RcsB proteins are positive transcriptional regulators and are required for maximal CA production<sup>19</sup>.

Here, we report a detailed topological analysis of WcaJ using two methods: reporter LacZ/PhoA fusions and sulfhydryl labelling by PEGylation of novel cysteine residues that are introduced into a cysteine-less (Cys-less) WcaJ protein. Our results reveal that WcaJ



comprises 4 canonical TMHs and a TMH-V that does not fully span the membrane, likely forming a hairpin. This topological arrangement places both the adjacent central loop and C-terminal domain towards the cytoplasm. The new conserved topological configuration for WcaJ is proposed as a signature for all PHPTs, and provides a framework to better understand the structure-function of this important membrane enzyme family.

## Results and Discussion

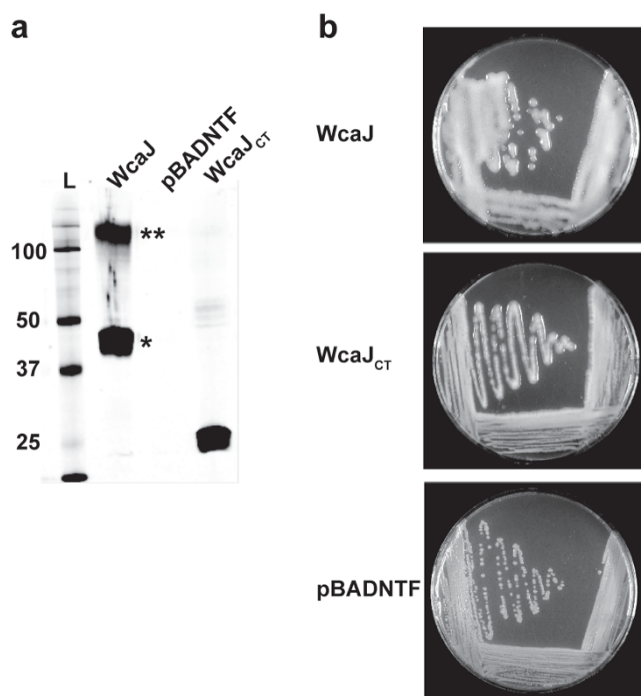
**The C-terminus of the *E. coli* WcaJ is sufficient for enzymatic activity *in vivo*.** Before using WcaJ as a model PHPT for topological analysis, we determined if its C-terminal domain was functionally comparable to the WbaP C-terminus<sup>10</sup>. Full-length <sub>FLAG</sub>WcaJ (53 kDa) and <sub>FLAG</sub>WcaJ<sub>CT</sub> (25 kDa), a truncated version comprising residues E252-Y464, which is analogous to WbaP<sub>CT</sub> (E258-Y476)<sup>6,8</sup>, were constructed and expressed from an arabinose inducible promoter (Fig. 2a). <sub>FLAG</sub>WcaJ migrates by SDS-PAGE with an apparent mass of ~48 kDa. A larger band of ~120 kDa (Fig. 2a, double asterisk) reacting with the anti-FLAG antibody was interpreted as an oligomeric form of WcaJ. Anomalous gel migration and oligomerization are commonly seen with membrane proteins<sup>6,20,21</sup>, particularly in those with high isoelectric points<sup>22</sup> (the theoretical pI for WcaJ is 9.2), since the mild denaturing conditions used for the preparation of the cell lysates is insufficient to disperse protein oligomers. Plasmids encoding <sub>FLAG</sub>WcaJ and <sub>FLAG</sub>WcaJ<sub>CT</sub> (pLA3 and pLA5, respectively) were introduced by transformation

into *E. coli*  $\Delta$ wcaJ also containing a plasmid expressing RcsA (pWQ499), the positive transcriptional regulator of the CA biosynthesis gene cluster<sup>23</sup>. In excess of RcsA, CA overproduction generates a highly mucoid colony phenotype<sup>23</sup>. Expression of <sub>FLAG</sub>WcaJ<sub>CT</sub> and <sub>FLAG</sub>WcaJ in  $\Delta$ wcaJ(pWQ499) restored the mucoid phenotype to the same levels as the parental strain, while no CA was observed in bacteria transformed with the vector pBADNTF (Fig. 2b), demonstrating that the WcaJ C-terminal domain is sufficient for enzymatic activity, as in the other PHPT family members<sup>5,10</sup>.

## Topological analysis of WcaJ using LacZ and PhoA reporter fusions reveals an unexpected membrane topology.

To experimentally assess the membrane topology of WcaJ (Fig. 1) we employed two complementary approaches: a dual reporter strategy and the substituted cysteine accessibility method (SCAM). The dual reporter fusion strategy was based on PhoA and LacZ reporters that were C-terminally fused to <sub>FLAG</sub>WcaJ truncations and expressed in *E. coli* strains CC118 ( $\Delta$ phoA) and DH5 $\alpha$  (*lacZ* $\Delta$ M15), respectively. Colonies were examined for blue or white phenotypes on LB-agar plates supplemented with the chromogenic substrates 5-bromo-4-chloro-3-indolyl-phosphate (XP) and 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (XGal), as appropriate and the *in vitro* alkaline phosphatase and  $\beta$ -galactosidase activities were measured as Miller units<sup>24</sup> (Table 1). PhoA is only active in the periplasm, as it requires an oxidizing environment to form disulphide bonds<sup>25,26</sup>. Conversely, LacZ functions in the cytosol. The results were consistent with the topological model shown in Fig. 3. <sub>FLAG</sub>WcaJ-LacZ fusions at V39 and D106, which reside in predicted cytoplasmic loops 1 and 2 (Fig. 1a), did not exhibit  $\beta$ -galactosidase activity and the colonies containing these constructs remained white (Table 1). On the contrary, the corresponding <sub>FLAG</sub>WcaJ-PhoA fusions at the same positions yielded strongly blue colonies in XP and positive alkaline phosphatase activity (2112  $\pm$  95 and 938  $\pm$  70 Miller units, respectively; Table 1), indicating that these residues are exposed to the periplasmic space. The LacZ fusion to G74, a residue predicted to be in periplasmic loop 1 (Fig. 1a), resulted in low  $\beta$ -galactosidase activity, but yielded blue colonies on XGal plates while the corresponding PhoA fusion gave no colour on XP plates. This result supports the conclusion that G74 is exposed to the cytosolic space (see below). Furthermore, LacZ fusions at residues in the predicted periplasmic loop II (N143, M160, P180, L223, N254, and L272) gave blue colonies on X-Gal plates and half of them showed strong *in vitro*  $\beta$ -galactosidase activity (Fig. 3 and Table 1). None of the corresponding PhoA fusions demonstrated blue colour in XP plates and had very low or no detectable alkaline phosphatase activity (Table 1). These results demonstrate that the large soluble domain of WcaJ predicted to be exposed to the periplasm (Fig. 1a) faces instead the cytoplasmic compartment (Fig. 3).

LacZ and PhoA fusions at S120 and K301 gave inconclusive results, as bacteria expressing these constructs produced blue colonies when plated on XGal or XP plates (Table 1). It has been reported that LacZ and PhoA fusions to amino acid residues located on TMHs can both be functional, failing to provide accurate topological information<sup>27,28</sup>. Therefore, we tentatively assigned S120 within TMH-IV and K301 to TMH-V (Fig. 3, see text below). In contrast, the fusions at G316, which is expected to be part of the functional C-terminal domain of WcaJ, gave results consistent with its predicted cytosolic location (Fig. 3 and Table 1). Immunoblotting was performed on total membrane preparations of CC118 cells expressing <sub>FLAG</sub>WcaJ-PhoA or DH5 $\alpha$  cells expressing <sub>FLAG</sub>WcaJ-LacZ fusion proteins. All of the fusion proteins were detectable by immunoblot, albeit at different expression levels, suggesting differences in protein stability depending on the fusion joint (Supplemental Figures S1 and S2 online). Together, the reporter fusion data suggest the large soluble loop between TMH-IV and -V is cytoplasmic.



**Figure 2** | (a) Membrane protein expression of WcaJ and WcaJ<sub>CT</sub>. Both proteins were engineered as N-terminal FLAG-tagged proteins and expressed from the arabinose-inducible vector pBADNTF (see Supplementary Table S2). Total membranes were isolated from *E. coli* DH5 $\alpha$  expressing the constructs, separated by electrophoresis in 14% SDS-polyacrylamide gels, and detected by immunoblotting using 1 : 10,000 dilution of anti-FLAG M2 monoclonal antibody as indicated in Methods. \* and \*\* indicate the position of monomeric WcaJ and oligomeric forms of WcaJ, respectively. L, molecular mass protein ladder (b) Complementation of colanic acid production in *E. coli* XBF1(pWQ499). Bacteria expressing the WcaJ constructs or containing the vector control were plated on LB agar supplemented with antibiotics, as needed, and 0.2% (w/v) arabinose. Colonies were evaluated for mucoid phenotype after 24 h at room temperature.



Table 1 | Topological location of selected WcaJ amino acid residues based on reporter fusions and PEGylation

Amino Acid position	Reporter fusion	Plate assay	Miller Units	PEGylation	Location
V39	PhoA	+	2112 ± 95	OUT	PL1
V39	LacZ	–	UD		
G74	PhoA	–	4.0 ± 2	IN	CL1
G74	LacZ	+	0.2 ± 0.1		
D106	PhoA	+	938 ± 70	OUT	PL2
D106	LacZ	–	UD		
S120	PhoA	+	97 ± 2	TM	TMIV
S120	LacZ	+	UD		
A134	PhoA	–	15 ± 1	IN	CL2
A134	LacZ	–	UD		
N143	PhoA	–	15 ± 4	IN	CL2
N143	LacZ	+	1 ± 0.2		
M160	PhoA	–	13 ± 3	IN	CL2
M160	LacZ	+	1 ± 0.2		
P180	PhoA	–	UD	IN	CL2
P180	LacZ	–	0.3 ± 0.1		
L223	PhoA	–	UD	IN	CL2
L223	LacZ	+	65.8 ± 3.2		
N254	PhoA	–	U.D.	IN	CL2
N254	LacZ	+	15 ± 0.6		
L272	PhoA	–	UD	BORDER	CL2
L272	LacZ	+	22 ± 0.4		
A276	ND	ND	ND	TM	TMV
A282	ND	ND	ND	TM	TMV
P291	ND	ND	ND	UD	?
C294	ND	ND	ND	TM	TMV
K301	PhoA	+	611 ± 44	IN	CT
K301	LacZ	+	5 ± 0.5		
S303	ND	ND	ND	BORDER IN	CT
G316	PhoA	–	UD	IN	CT
G316	LacZ	+	13 ± 1		
S360	ND	ND	ND	IN	CT

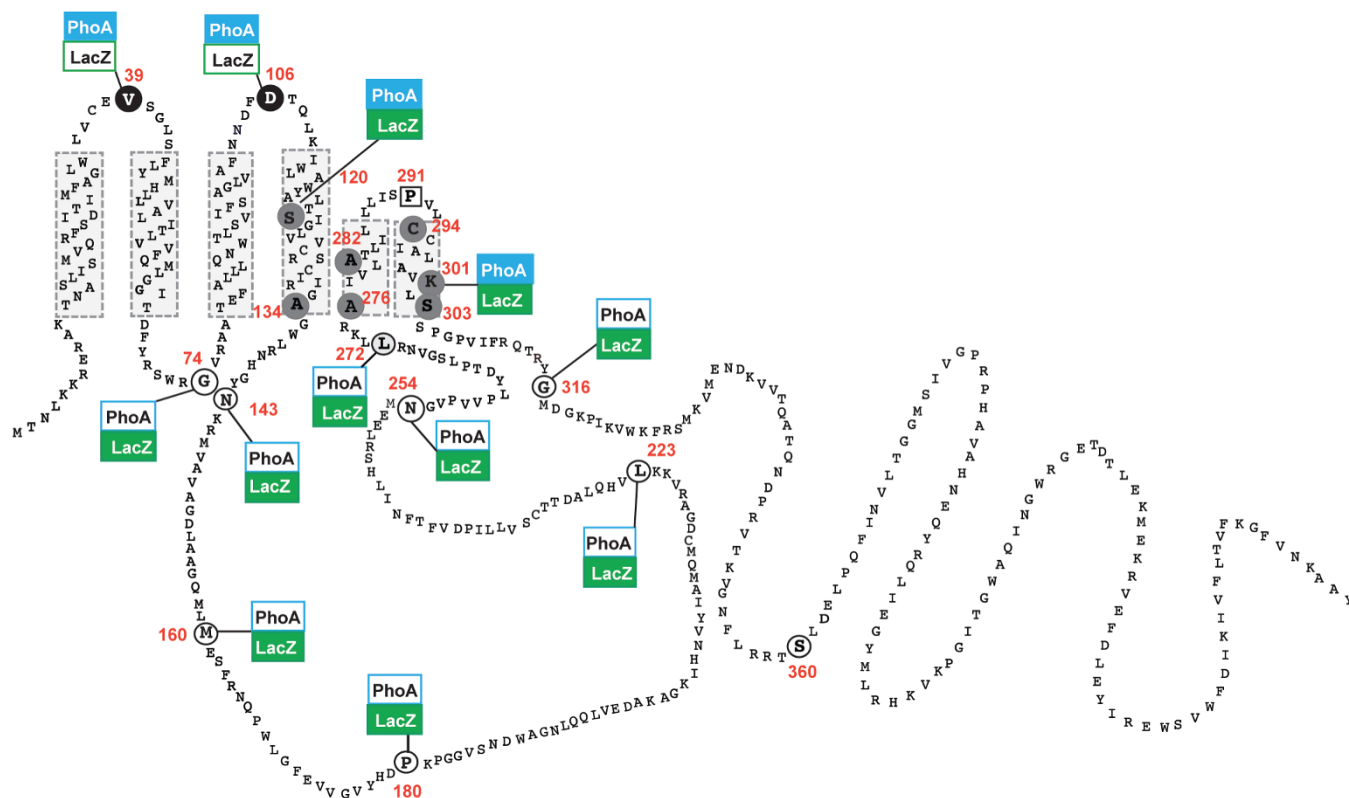
WcaJ-LacZ/PhoA fusion proteins were expressed in *E. coli* DH5 $\alpha$  or CC118 (see Methods). Blue colonies were assessed on plates containing XGal or XP. Alkaline phosphatase and  $\beta$ -galactosidase activities were determined (in triplicate) by measuring the hydrolysis of *p*-nitrophenyl phosphate and ortho-nitrophenyl- $\beta$ -galactoside, respectively, and the results calculated as Miller units (mean  $\pm$  standard deviation). PEGylation of WcaJ<sub>Cysless</sub> with ectopic cysteine residues at the indicated positions was done as described in Methods (see also Fig. 5). UD, undetectable; ND, not done; ?, location could not be established. PL, periplasmic loop; CL, cytoplasmic loop; CT, C-terminal domain.

The topological information derived from the dual fusions was refined by SCAM using methoxypolyethylene glycol maleimide (PEG-Mal) labelling<sup>29,30</sup>. WcaJ has 7 cysteine residues at positions 37, 126, 129, 232, 294 and 295, which were replaced by alanine to generate a cysteine-less derivative (WcaJ<sub>Cysless</sub>). WcaJ<sub>Cysless</sub> was expressed in membrane fractions (Fig. 4a) and restored CA production in strain XBF1(pWQ499) (Fig. 4b), demonstrating that the native cysteine residues in the protein are dispensable for function. PEG-Mal forms irreversible thioester bonds with the sulfhydryl groups on cysteine residues in the presence of a water molecule. Reaction of a cysteine with PEG-Mal results in a mass addition of ~5–10 kDa to the protein, which appears as a band shift by immunoblotting.

To determine the topology of the novel cysteine residues in WcaJ, whole cells were pre-treated with thiol-blocking reagents 2-sulfonatoethyl methanethiosulfonate (MTSES; permeable only to the outer membrane, and N-ethylmaleimide (NEM; permeable to both inner and outer membranes). After lysis and sulfhydryl-labelling with PEG-Mal, the topology was assessed by the ability of the thiol-blocking reagents to bind to the introduced novel cysteine residues and prevent them from becoming accessible to PEG-Mal and subsequently causing a band shift. Treatment with both MTSES and NEM prevented labelling of PEG-Mal to residues V39C and D106C indicating that these residues are exposed to the periplasm, corresponding with Pho/LacZ fusion results (Fig. 5). Similarly, residues G74C, A134C, N143C, M160C, P180C, L223C, N254C, K301C, S303C, G316C and S360C could only be labelled in samples pre-treated with MTSES but not with NEM, suggesting they face the cytoplasm (Fig. 5). Labelling of PEG-Mal was observed after treat-

ment with both MTSES and NEM in the reintroduced native cysteine C294 suggesting this residue is in a transmembrane domain. Residue L272C was partially PEGylated after treatment with both MTSES and NEM. This residue is predicted to be on the interface between the cytoplasm and a transmembrane domain, which could explain the partial PEGylation under all treatment conditions. Together, the combined results from reporter fusions and SCAM using PEGylation support a topological model for WcaJ whereby the protein contains two prominent cytosolic regions separated by a single TMH (Fig. 3).

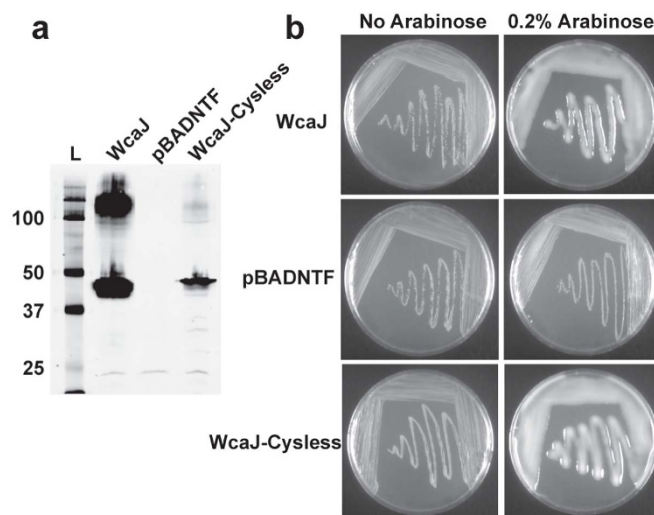
**Analysis of the borders of TMH-V suggests a helix-break-helix structure.** Our experimental results indicating that TMH-V separates two cytosolic soluble domains, suggest this TMH has an unusual structure. P291C could not be labelled with PEG-Mal under conditions that allowed PEGylation of nearby residues (Fig. 5). To better examine the behaviour of this P291C replacement mutant, we also performed PEGylation experiments in isolated total membrane protein fractions before and after SDS treatment. This experiment showed that similar to the PEGylation conditions using whole bacteria (Fig. 5), the protein with the P291C replacement could not be labelled, while replacements in the neighbouring residues A276C, A282C and A294C were PEGylated in the presence of SDS (Fig. 6). Together, the results suggest that the cysteine at position 291 is located in a region of the protein adopting a conformation that prevents labelling even after SDS treatment. WcaJ<sub>P291C</sub> was also poorly expressed and migrated abnormally compared to the other proteins (Figs. 5 and 6), suggesting an alteration of the secondary structure in this mutant protein. However, expression of WcaJ<sub>P291C</sub>



**Figure 3 | Refined topological model of the *E. coli* WcaJ based on reporter fusions and substituted cysteine labelling with PEG-Mal.** Blue coloured boxes indicate positive PhoA results and green coloured boxes indicate positive LacZ results. White coloured boxes indicate negative reporter fusion results. Residues indicated in circles were replaced by cysteine in the WcaJ<sub>CysLess</sub> version of the protein. Residues that were labelled with PEG-Mal denoting a periplasmic location are depicted with black background. Residues indicated in grey are those only detected by PEG-Mal after SDS solubilization, indicating they are buried in the membrane bilayer. Residues exposed to the cytosol are depicted in white background. The square indicates the position of the conserved P291, which could not be labelled with PEG-Mal under any of the treatments employed including SDS solubilization of membrane fractions.

in  $\Delta wcaJ$ (pWQ499) restored colanic acid biosynthesis (Supplemental Figure S3 online), indicating that the replacement of proline by cysteine at this position does not compromise the functionality of WcaJ.

To gain additional insight on the potential structure of this TMH, we analysed its secondary structure using PSIPRED<sup>31</sup>, which predicted a helical conformation between L273 and L302. This prediction was in agreement with the experimental boundaries determined for TMH-V. The region encompassing residues R270–G306 was modelled using TMKink, a program specifically tailored to identify kinks within TMHs. Meruelo *et al.*<sup>32</sup> demonstrated that there are residue preferences in TMHs that contribute to kinks and that a score above the designated threshold indicates a kink in the helix. Using TMKink, we found that R270–G306 contains a predicted kink in TMH-V with a value of 0.719 that is as above the 0.67 cut-off threshold<sup>32</sup>. The predicted kink spans the sequence I284–LLLISPV–L292, which includes P291 (Fig. 7a). Similar results were obtained from TMKink predictions using a dataset of 112 PHPT homologues (Supplementary Table S1 and Supplemental Figure S4 online), obtaining in all cases predicted values higher than the cut-off threshold within a region of the protein that includes a conserved proline at a position corresponding to the P291 residue in WcaJ. We also modelled this region in WcaJ with PEP-FOLD<sup>33</sup>, which revealed a helix-break-helix predicted structure (Fig. 7b). The break region corresponded to S290 and P291. Further analysis using WcaJ residues E252–Y464 with ConSeq<sup>34</sup> also predicted that this proline is buried (Fig. 7c). PHPT homologues typically fall into two different classes: proteins containing the three canonical domains, (e.g. WbaP and WcaJ), and proteins that have the C-terminal domain including TMH-V (e.g. PssY<sup>5</sup>). Fewer PHPT homologues only have the soluble



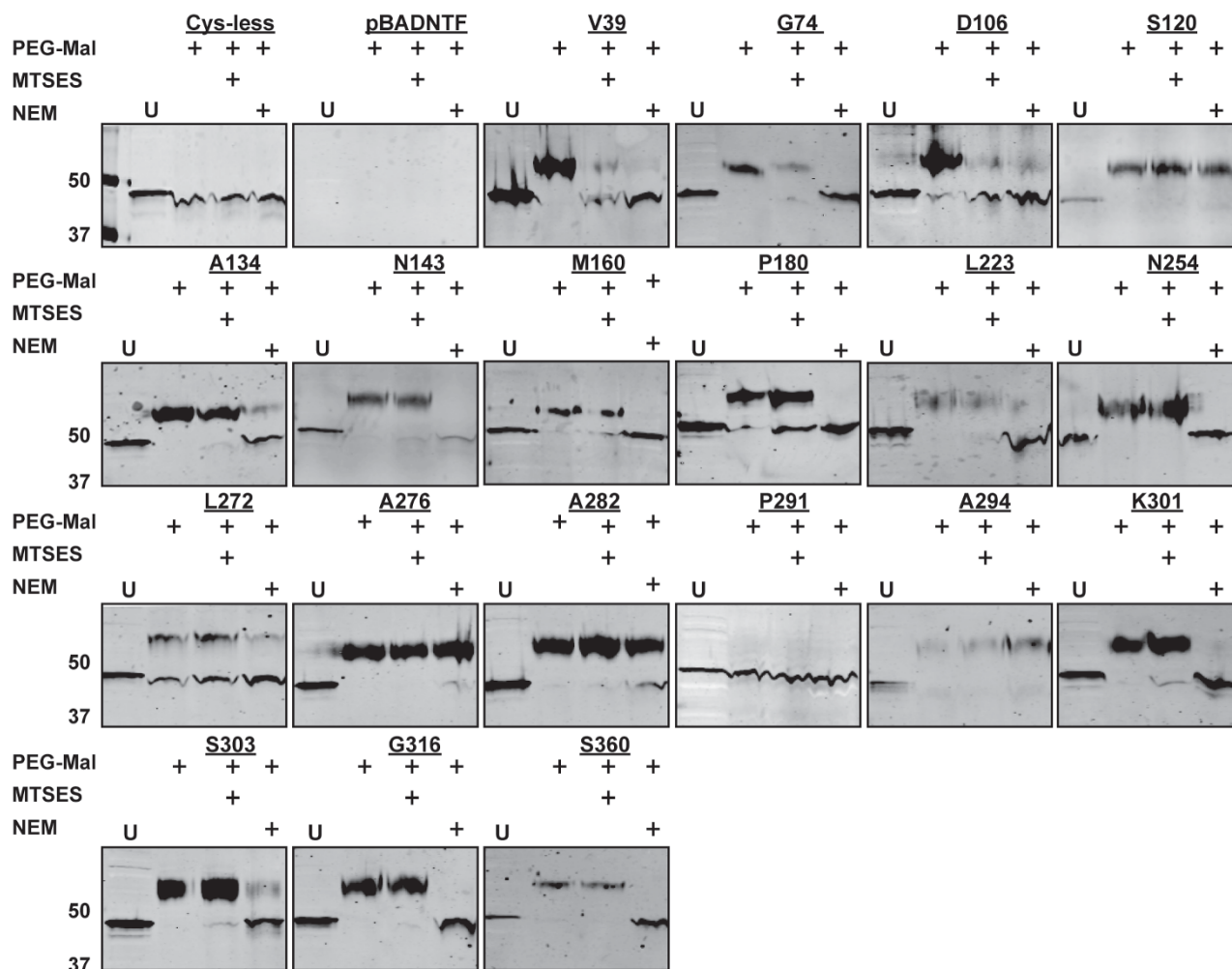
**Figure 4 | (a) Membrane protein expression of WcaJ<sub>CysLess</sub>.** Ten  $\mu$ g of total membrane preparations from *E. coli* DH5 $\alpha$  cells expressing WcaJ (pLA3) and WcaJ<sub>CysLess</sub> (pSEF102), both engineered with an N-terminal FLAG tag, were separated by 14% SDS-PAGE and detected by immunoblotting using 1:10,000 dilution of anti-FLAG M2 monoclonal antibody. L, molecular mass protein ladder. (b) Complementation of colanic acid production in *E. coli* XBF1(pWQ499). Bacteria expressing full length WcaJ (pLA3), WcaJ<sub>CysLess</sub> (pSEF102) or containing the vector control (pBADNTF) were plated on LB agar supplemented with antibiotics, as needed, and 0.2% (w/v) arabinose. Colonies were evaluated for mucoid phenotype after 24 h.



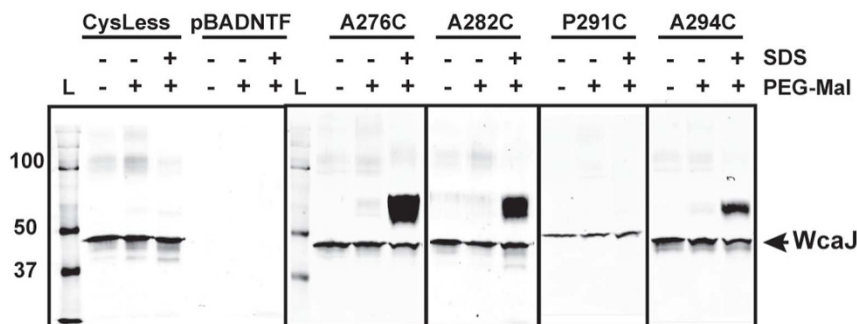
C-terminal domain, but it is unknown if these proteins are functional. We examined a dataset of 112 PHPT homologues (Supplementary Table S1 online) by Clustal Omega. In most cases, we found a conserved proline residue at the same position corresponding to P291 and within a region in these proteins that is predicted as a transmembrane domain (Supplemental Figure S4 online). Together, our results strongly suggest that TMH-V has a subdomain adopting a helix-break-helix structure and this feature is conserved in PHPT family members.

**Concluding Remarks.** This study experimentally elucidates the topology of WcaJ, and also demonstrates that the predicted TMH-V is flanked by two large cytosolic regions, one of which corresponds to the C-terminal domain of the protein. The availability of a cysteine-less WcaJ allowed us to investigate the topology of TMH-V and its neighbouring soluble regions utilizing a substituted cysteine replacement and a sulphhydryl labelling technique with PEG-Mal. These experiments conclusively established that the N- and C-terminal borders of TMH-V (L272 and S304, respectively) face the cytosol while the other intervening residues are buried in the lipid bilayer. P291 was critical to maintain the secondary structure of TMH-V, and alignment of this region with other PHPT members reveals that the proline residue at the equivalent

position as P291 is highly conserved. Also, an aspartic acid residue corresponding to that at position 278 in WcaJ that delineates the N-terminal border of TMH-V (Fig. 3) is highly conserved in PHPT proteins (Supplementary Figure S4), suggesting a possible motif including P291 common to the members of this family. Therefore, the results obtained for WcaJ extend to other members of the PHPT family. Studies on the membrane protein caveolin-1 have identified a helix-break-helix structure causing the N- and C-terminal domains of this helix to reside on the same side of the membrane forming an “intra-membrane horseshoe conformation” that does not fully span the membrane<sup>35</sup>. A proline and an isoleucine were among the key residues contributing to the “break” in helix, and their replacement by alanine dramatically disrupted the helix-break-helix structure<sup>35</sup>. Interestingly, the TMH-V is of similar size to the helix-break-helix structure (~28 residues) and has a highly conserved proline and isoleucine residues at positions 291 and 289, respectively. Therefore, the altered migration of WcaJ<sub>P291C</sub> on SDS-PAGE, its reduced membrane expression or stability, and inaccessibility of the cysteine residue replacing P291 to PEGylation, even after SDS treatment, suggests that TMH-V contains a helix-break-helix structure and cannot span the membrane bilayer, thus explaining why a single predicted TM is found between two regions of the proteins that are facing the cytosol.



**Figure 5 | Substituted cysteine labelling of WcaJ<sub>Cysless</sub> derivatives containing cysteine replacements.** Whole bacteria were treated with thiol-blocking reagents MTSES and NEM prior to lysis and labelling with PEG-Mal. The figure is a composite of multiple experiments each involving three experimental samples plus untreated control and the molecular mass markers. All samples were separated on a 14% polyacrylamide gel and detected by 1 : 10,000 dilution of anti-FLAG M2 monoclonal antibody. The topology of the cysteine residue was determined by the ability of MTSES and NEM to prevent PEG-Mal labelling. PEG-Mal labelling is indicated by a band shift. \* and \*\* indicate the position of WcaJ and PEGylated WcaJ, respectively. U, untreated control.



**Figure 6 | Substituted cysteine labelling in the presence of SDS using PEG-Mal in total membrane preparations expressing FLAG-WcaJ cysteine replacement proteins.** Total membranes were either untreated, incubated with 1 mM PEG-Mal to reveal exposed residues, or 1 mM PEG-Mal with 1% SDS to reveal TMH residues. Samples were examined by SDS-PAGE in 14% polyacrylamide and detected by immunoblotting with 1 : 10,000 dilution of anti-FLAG M2 monoclonal antibody. Each experiment indicated in the figure was run under identical conditions. Band shift denotes PEGylated polypeptides. Ten  $\mu\text{g}$  of protein were labelled, except for WcaJ<sub>P291C</sub>, which required 30  $\mu\text{g}$ . L, molecular mass protein ladder.

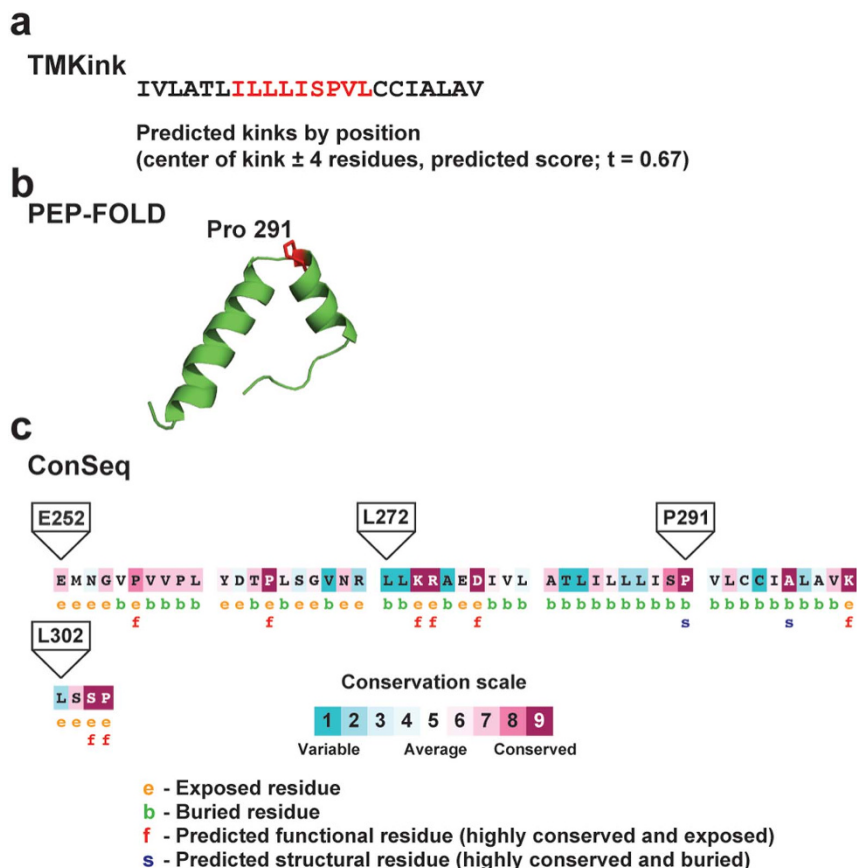
In summary, this study reveals an unexpected membrane topology for WcaJ, which is proposed as a signature for all members of the PHPT family. Further investigations are underway in our laboratory to elucidate the function of the central cytosolic domain, previously thought to be located on the periplasmic space.

## Methods

**Bacterial strains, plasmids, media and growth conditions.** Bacterial strains and plasmids used in this study are listed in Supplementary Table S2 online. Bacteria were grown at 37°C in Luria-Bertani (LB) medium (Difco) (10 mg tryptone  $\text{ml}^{-1}$ ; 5 mg yeast extract  $\text{ml}^{-1}$ ; 5 mg NaCl  $\text{ml}^{-1}$ ). Growth medium was supplemented with 100  $\mu\text{g}$

ampicillin  $\text{ml}^{-1}$ , 20  $\mu\text{g}$  tetracycline  $\text{ml}^{-1}$ , 40  $\mu\text{g}$  kanamycin  $\text{ml}^{-1}$  or 30  $\mu\text{g}$  chloramphenicol  $\text{ml}^{-1}$ , as required. The expression of hybrid proteins with  $\beta$ -galactosidase (LacZ) or alkaline phosphatase (PhoA) activities was qualitatively assessed by examining the blue-colony phenotypes on LB plates containing 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (XGal) and 5-bromo-4-chloro-3-indolyl-phosphate (XP), respectively, both at a final concentration of 40  $\mu\text{g}$   $\text{ml}^{-1}$ . XGal and XP were from Boehringer Mannheim. *E. coli* DH5 $\alpha$  was transformed with plasmids by the calcium chloride method<sup>36</sup>. *E. coli* CC118 and XBF1 were transformed by electroporation<sup>37</sup>.

**Construction of recombinant plasmids.** Primers are listed in Supplementary Table S3 online. DNA ligations were performed with T4 DNA ligase. All constructs were verified by DNA sequencing, which was performed at York University in Toronto,



**Figure 7 | *In silico* analysis of the WcaJ TMH-V.** (a) A potential kink (highlighted in red) was identified in the  $\alpha$  helix of TMH-V using TMKink. (b) Structural modelling of TMH-V using PEP-FOLD, which predicts a helix-break-helix structure involving serine-290 and proline-291. (c) ConSeq prediction from multiple PHPT homologues identifying conservation scores for each residue and predicting if they are buried or exposed. Proline-291 is a predicted as a highly conserved and buried structural residue. A proline at this position is also conserved in a dataset of PHPT homologues (Supplementary Figure S4 online). The positions of the neighbouring Cys substitutions are also indicated.



Ontario, Canada, Eurofins MWG Operon, Huntsville, Alabama, USA, or GATC Biotech AG, Constance, Germany. Primers 4779 and 4780 were used to introduce novel histidine residues into the coding region of *wcaJ* gene in pXF1<sup>5</sup> using *Pfu* AD polymerase (Stratagene). *DpnI* was added to PCR reactions for overnight digestion of parental plasmid DNA at 37°C. The resulting DNA was introduced into *E. coli* DH5 $\alpha$  by transformation and transformants selected on LB-agar containing 100- $\mu$ g ampicillin ml<sup>-1</sup>. This resulted in pLA1 (Supplementary Table S2 online), which was used as a template to construct pLA3, expressing an N-terminal FLAG fusion to WcaJ (FLAGWcaJ), using primers 4911 (containing a *SmaI* restriction site) and 4912 (containing a *HindIII* restriction site; Supplementary Table S3 online). The PCR product and the vector pBADNTF were digested with *SmaI* and *HindIII*. Digested pBADNTF was also treated with alkaline phosphatase to prevent recircularisation prior to ligation with digested PCR product. After transformation, colony PCR with pBADNTF external primers 252 and 258 (Supplementary Table S3 online) was used to identify the correct construct. Primers 5094 and 5095 served to amplify a region of the *wcaJ* gene encoding the C-terminal domain (WcaJ<sub>E252-Y464</sub>), giving rise to pLA5 (Supplementary Table S2 online).

PhoA fusions to residues S120, N143, M160, P180, and K301 (expressed in plasmids pSEF71, pSEF72, pSEF73, pSEF75, and pSEF74, respectively; Supplementary Table S2 online) were constructed by amplifying the *wcaJ* gene from pLA3 using primer 4911 with primers 5841, 5840, 5839, 5093, 5867, respectively (Supplementary Tables S3 and S4 online). Each PCR product and pAH18 were digested with *SmaI* and *XbaI*, purified and ligated together using T4 ligase. Construction of the A134-PhoA fusion (pLA6, Supplementary Table S2 online) involved amplifying *wcaJ* from pLA3 with primers 4911 and 5258 (Supplementary Tables S3 and S4 online); the PCR product and pSEF74 were digested with *SmaI* and *XbaI* and ligated. For the constructions of the fusions indicated above, PCR amplifications were done with *Pwo* polymerase. *DpnI* was added to each PCR product to digest the plasmid template and the PCR product was purified using the PCR purification kit (Qiagen) prior to digestion with the restrictions enzymes.

The V39-, G74-, and D106-PhoA fusions (pLA8, pLA9, and pLA7, respectively; Supplementary Table S2 online) were constructed using the direct primer 5257 that amplifies from the *phoA* gene in pLA6 onwards and reverse primers 5364, 5363, and 5259 that amplify from various positions in the *wcaJ* gene (Supplementary Tables S3 and S4 online). The L223-, -N254, and L272-PhoA fusions (pLA18, pLA19, and pLA20, respectively; Supplementary Table S2 online) were also built using the pLA11 DNA template, the direct primer 5257, and reverse primers 5394, 5393, and 5392 (Supplementary Table S3 and S4 online). The WcaJ-LacZ fusions were constructed using the direct primer 5640, which amplifies from the *lacZ* gene in pLA23 onwards, and reverse primers 5364, 5363, 5259, 5841, 5258, 5840, 5839, 5093, 5394, 5393, 5392, 5867, and 5580 (Supplementary Tables S3 and S4 online). Each primer pair mentioned above contained an *XbaI* site. PCR amplifications for these constructs were done with *Pfu* AD polymerase, which extends around the entire plasmid template. *DpnI* was added to each PCR product to digest the plasmid template and the PCR product was purified as indicated above and digested with *XbaI*. In all of these experiments, ligation products were introduced into *E. coli* DH5 $\alpha$  cells by transformation and selected on LB agar supplemented with 100  $\mu$ g/ml ampicillin.

The cysteine-less FLAGWcaJ was constructed by replacing the native cysteine residues with alanine using site directed mutagenesis. Replacements were introduced by PCR using primers containing the desired mutations, and pLA3 as the DNA template (Supplementary Tables S2 and S3). Amplification was performed using *Pfu* Turbo AD polymerase (Stratagene). *DpnI* was added to PCR reactions for overnight digestion of parental plasmid DNA at 37°C. The resulting DNA was introduced into *E. coli* DH5 $\alpha$  by transformation. This resulted in pSEF102 encoding FLAGWcaJ<sub>Cys-Less</sub>. Novel cysteine residues were introduced into FLAGWcaJ<sub>Cys-Less</sub> by site directed mutagenesis.

**Complementation of colanic acid biosynthesis in *E. coli*  $\Delta$ wcaJ.** Plasmids expressing full length WcaJ (pLA3), WcaJ<sub>Cys-Less</sub> (pSEF102), and WcaJ-P291C (pSEF103) were introduced into XBF1/pWQ499 by electroporation and selected on LB-agar supplemented with ampicillin, tetracycline, and kanamycin. Cells were replated onto LB-agar with antibiotics with 0.2% arabinose to promote expression of WcaJ. Plates were incubated at 37°C overnight before incubation at room temperature for an additional 24–48 h to observe mucoidity. CA is optimally produced at 20°C<sup>19</sup>.

**Alkaline phosphatase and  $\beta$ -galactosidase assays.** The  $\beta$ -galactosidase and alkaline phosphatase activities were quantitated as previously described<sup>38,39</sup>.

**Protein analyses.** Bacterial cultures containing arabinose-inducible plasmids were grown overnight in LB containing antibiotics, as needed. The next day, cultures were diluted to an optical density measured at 600 nm (OD<sub>600</sub>) of 0.2 in LB with the respective antibiotics. Cultures (100 ml) were grown at 37°C until an OD<sub>600</sub> of 0.5–0.7 at which point arabinose was added to a final concentration of 0.2% (w/v) and incubation continued for 3 h. Bacteria were harvested by centrifugation 8 000 g and cell pellets were frozen at -20°C until needed. Cell pellets were resuspended in 5 ml of 50 mM Tris-HCl, pH 8, with protease inhibitor cocktail (Roche Diagnostics), and lysed at 10 000 PSI with a cell disruptor (Constant Systems Ltd). Cell debris was pelleted at 27 216 g. Total membranes were isolated by centrifugation in microfuge tubes at 39 191 g and resuspended in 50  $\mu$ l of 50 mM Tris-HCl, pH 8. Protein concentration was determined by the Bradford protein assay (Bio-Rad). Total membrane preparations were used for immunoblotting and PEGylation. Membrane

proteins were separated by 14% SDS-PAGE (For detection of PhoA and LacZ protein fusions 12% and 18% polyacrylamide gels were used; see Supplemental Figures S1 and S2 online). Proteins were transferred to nitrocellulose membrane. Membranes were blocked overnight in 5% Western blocking reagent (Roche Diagnostics) dissolved in TBS pH 7.5 (50 mM Tris-HCl, 150 mM NaCl). The primary antibody, 4.6 mg anti-FLAG M2 monoclonal antibody ml<sup>-1</sup> (Sigma), was diluted to 1 : 10 000 and applied for 1.5 h, and the secondary antibody, 2 mg goat anti-mouse Alexa Fluor 680 IgG antibody ml<sup>-1</sup> (Invitrogen Molecular Probes) was diluted to 1 : 20 000 and applied for 20 min. Immunoblots were developed using LI-COR Odyssey infrared imaging system. Bio-Rad Precision Plus Protein Standards were used as protein mass standards.

#### Protein labelling with methoxy polyethylene glycol maleimide (PEG-Mal).

Plasmids containing FLAGWcaJ cysteine substitutions were introduced into *E. coli* XBF1 by CaCl<sub>2</sub> transformation. Cultures were grown overnight in LB containing antibiotics, as needed. The next day, cultures were diluted to an OD<sub>600</sub> of 0.2 in LB with the respective antibiotics and grown at 30°C for 1.5 h before induction with 0.2% (w/v) arabinose for 1 h. Cells were adjusted to OD<sub>600</sub> of 2 in 200  $\mu$ l of 1 $\times$  Phosphate Buffered Saline (PBS) (8 mg NaCl ml<sup>-1</sup>, 0.2 mg KCl ml<sup>-1</sup>, 1.44 mg Na<sub>2</sub>HPO<sub>4</sub> ml<sup>-1</sup>, 0.24 mg KH<sub>2</sub>KPO<sub>4</sub> ml<sup>-1</sup>, pH7.4) and divided into 50  $\mu$ l aliquots. Five  $\mu$ l of 55 mM MTSES or NEM in doubled distilled water (ddH<sub>2</sub>O) were added to the samples and incubated at room temperature for 1 h in the dark. Cells were harvested by centrifugation 13,000 rpm for 1 min at 4°C and cell pellets washed with 100  $\mu$ l of 1 $\times$  PBS. Cells were resuspended in 22.5  $\mu$ l of lysis buffer (15  $\mu$ l of 1 M Tris-HCl pH7.6, 100  $\mu$ l 10% SDS, 0.376 g urea and up to 1 ml ddH<sub>2</sub>O) and 7.5  $\mu$ l of 25 mM PEG-Mal solubilised in DMSO was added. Samples were incubated at room temperature for 1 h in the dark before addition of 30  $\mu$ l of 2 $\times$  AB buffer (6.84 mM Na<sub>2</sub>HPO<sub>4</sub>, 3.16 mM NaH<sub>2</sub>PO<sub>4</sub>, 50 mM Tris-HCl pH 6.8, 6 M Urea, 1%  $\beta$ -mercaptoethanol, 3% SDS, 10% glycerol, 0.1% Bromophenol Blue) and frozen at -20°C until required. Samples were separated by 14% SDS-PAGE and transferred onto nitrocellulose membranes and immunoblotting performed as indicated above.

**Bioinformatic analyses.** Amino acid sequences of PHPT proteins were downloaded from the SYSTERS database (Cluster O 141559; <http://systems.molgen.mpg.de/cgi-bin/nph-fetchcluster.pl?CLNR=141559&SIZE1=1&SIZE2=1&LINES=38&v=4>)<sup>40</sup>. All these proteins contain PFAM PF02397<sup>41</sup> corresponding to bacterial sugar transferases. The collection was manually curated to eliminate incomplete sequences, and sequences corresponding to acetyltransferases. The remaining sequences were examined for TMH predictions with TMHMM<sup>13</sup> and those sequences lacking TMHs were removed from the dataset, which ultimately contained 112 unique sequences (Supplementary Table S1). This dataset is available from the authors as a FASTA file upon request. The PHPT dataset was aligned using Clustal Omega<sup>42</sup> with default parameters and displayed by Jalview<sup>43</sup> (Supplementary Figure S4).

- Valvano, M. A., Patel, K. B. & Furlong, S. E. in *Bacterial lipopolysaccharides: Structure, chemical synthesis, biogenesis and interaction with host cells* (eds Knirel, Y. & Valvano, M. A.) 275–310 (Springer Verlag Publishing Inc., 2011).
- Valvano, M. A. Export of O-specific lipopolysaccharide. *Front Biosci* **8**, s452–471 (2003).
- Valvano, M. A. Common Themes in Glycoconjugate Assembly Using the Biogenesis of O-Antigen Lipopolysaccharide as a Model System. *Biochemistry-Moscow* **76**, 729–735 (2011).
- Xayarath, B. & Yother, J. Mutations blocking side chain assembly, polymerization, or transport of a Wzy-dependent *Streptococcus pneumoniae* capsule are lethal in the absence of suppressor mutations and can affect polymer transfer to the cell wall. *J Bacteriol* **189**, 3369–3381 (2007).
- Patel, K. B. *et al.* Functional characterization of UDP-glucose:undecaprenyl-phosphate glucose-1-phosphate transferases of *Escherichia coli* and *Caulobacter crescentus*. *J Bacteriol* **194**, 2646–2657 (2012).
- Saldias, M. S. *et al.* 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 (2008).
- Steiner, K. *et al.* Functional characterization of the initiation enzyme of S-layer glycoprotein glycan biosynthesis in *Geobacillus stearothermophilus* NRS 2004/3a. *J Bacteriol* **189**, 2590–2598 (2007).
- Patel, K. B., Furlong, S. E. & Valvano, M. A. Functional analysis of the C-terminal domain of the WbaP protein that mediates initiation of O antigen synthesis in *Salmonella enterica*. *Glycobiology* **20**, 1389–1401 (2010).
- Wang, L., Liu, D. & Reeves, P. R. C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *J Bacteriol* **178**, 2598–2604 (1996).
- Patel, K. B., Ciepichal, E., Swiezewska, E. & Valvano, M. A. The C-terminal domain of the *Salmonella enterica* WbaP (UDP-galactose:Und-P galactose-1-phosphate transferase) is sufficient for catalytic activity and specificity for undecaprenyl monophosphate. *Glycobiology* **22**, 116–122 (2012).
- James, D. B., Gupta, K., Hauser, J. R. & Yother, J. Biochemical activities of *Streptococcus pneumoniae* serotype 2 capsular glycosyltransferases and significance of suppressor mutations affecting the initiating glycosyltransferase Cps2E. *J Bacteriol* **195**, 5469–5478 (2013).





12. Cserzo, M., Wallin, E., Simon, I., von Heijne, G. & Elofsson, A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the Dense Alignment Surface method. *Prot. Eng.* **10**, 673–676 (1997).
13. Sonnhammer, E. L. L., von Heijne, G. & Krogh, A. in *Proceedings of Sixth International Conference on Intelligent Systems for Molecular Biology* (eds Glasgow, J. et al.) 175–182. (AAAI Press, 1998).
14. Videira, P. A., Garcia, A. P. & Sa-Correia, I. Functional and topological analysis of the *Burkholderia cenocepacia* priming glucosyltransferase BceB, involved in the biosynthesis of the cepacian exopolysaccharide. *J. Bacteriol.* **187**, 5013–5018 (2005).
15. Stevenson, G., Andrianopoulos, K., Hobbs, M. & Reeves, P. R. Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *J. Bacteriol.* **178**, 4885–4893 (1996).
16. Whitfield, C. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* **75**, 39–68 (2006).
17. Danese, P. N., Pratt, L. A. & Kolter, R. Exopolysaccharide production is required for development of *Escherichia coli* K-12 biofilm architecture. *J. Bacteriol.* **182**, 3593–3596 (2000).
18. Meredith, T. C. et al. Modification of lipopolysaccharide with colanic acid (M-antigen) repeats in *Escherichia coli*. *J. Biol. Chem.* **282**, 7790–7798 (2007).
19. Gottesman, S. & Stout, V. Regulation of capsule polysaccharide synthesis in *Escherichia coli* K-12. *Molecular microbiology* **5**, 1599–1606 (1991).
20. Amer, A. O. & Valvano, M. A. The N-terminal region of the *Escherichia coli* WecA (Rfe) protein containing three predicted transmembrane helices is required for function but not for membrane insertion. *J. Bacteriol.* **182**, 498–503 (2000).
21. Lehrer, J., Vigeant, K. A., Tatar, L. D. & Valvano, M. A. 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 (2007).
22. Kashino, Y. Separation methods in the analysis of protein membrane complexes. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **797**, 191–216 (2003).
23. Stout, V., Torres-Cabassa, A., Maurizi, M. R., Gutnick, D. & Gottesman, S. RcsA, an unstable positive regulator of capsular polysaccharide synthesis. *J. Bacteriol.* **173**, 1738–1747 (1991).
24. Manoel, C. Analysis of membrane protein topology using alkaline phosphatase and beta-galactosidase gene fusions. *Methods Cell. Biol.* **34**, 61–75 (1991).
25. Derman, A. I. & Beckwith, J. *Escherichia coli* alkaline phosphatase fails to acquire disulfide bonds when retained in the cytoplasm. *J. Bacteriol.* **173**, 7719–7722 (1991).
26. Derman, A. I. & Beckwith, J. *Escherichia coli* alkaline phosphatase localized to the cytoplasm slowly acquires enzymatic activity in cells whose growth has been suspended: a caution for gene fusion studies. *J. Bacteriol.* **177**, 3764–3770 (1995).
27. Calamia, J. & Manoel, C. Lac permease of *Escherichia coli*: topology and sequence elements promoting membrane insertion. *Proc Natl Acad Sci U S A* **87**, 4937–4941 (1990).
28. Gött, P. & Boos, W. The transmembrane topology of the *sn*-glycerol-3-phosphate permease of *Escherichia coli* analysed by *phoA* and *lacZ* protein fusions. *Mol. Microbiol.* **2**, 655–663 (1988).
29. Koch, S., Fritsch, M. J., Buchanan, G. & Palmer, T. *Escherichia coli* TatA and TatB proteins have N-out, C-in topology in intact cells. *J. Biol. Chem.* **287**, 14420–14431 (2012).
30. Nasie, I., Steiner-Mordoch, S. & Schuldiner, S. Topology determination of untagged membrane proteins. *Methods Mol Biol.* **1033**, 121–130 (2013).
31. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
32. Meruelo, A. D., Samish, I. & Bowie, J. U. TMKink: a method to predict transmembrane helix kinks. *Protein Sci.* **20**, 1256–1264 (2011).
33. Thevenet, P. et al. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.* **40**, W288–293 (2012).
34. Berezin, C. et al. ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* **20**, 1322–1324 (2004).
35. Lee, J. & Glover, K. J. The transmembrane domain of caveolin-1 exhibits a helix-break-helix structure. *Biochim Biophys Acta* **1818**, 1158–1164 (2012).
36. Cohen, S. N., Chang, A. C. & Hsu, L. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc. Natl. Acad. Sci. U S A* **69**, 2110–2114 (1972).
37. Dower, W. J., Miller, J. F. & Ragsdale, C. W. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127–6145 (1988).
38. Tatar, L. D., Marolda, C. L., Polischuk, A. N., van Leeuwen, D. & Valvano, M. A. An *Escherichia coli* undecaprenyl-pyrophosphate phosphatase implicated in undecaprenyl phosphate recycling. *Microbiology* **153**, 2518–2529 (2007).
39. Marolda, C. L. et al. Membrane topology and identification of critical amino acid residues in the Wzx O-antigen translocase from *Escherichia coli* O157:H4. *J. Bacteriol.* **192**, 6160–6171 (2010).
40. Krause, A., Stoye, J. & Vingron, M. Large scale hierarchical clustering of protein sequences. *BMC bioinformatics* **6**, 15 (2005).
41. Finn, R. D. et al. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–230 (2014).
42. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
43. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

## Acknowledgments

This work was supported by a grant from the Canadian Institutes of Health Research and an Internal Research grant from the School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast (to M.A.V.). S.E.F. was supported by an Ontario Graduate Scholarship in Science and Technology. A.F. was supported by a Doctoral Research Scholarship from the Department of Employment and Learning, Northern Ireland.

## Author contributions

M.A.V. conceived the study. M.A.V. and S.E.F. wrote the manuscript with contributions from A.F., S.E.F., A.F. and L.A.-R. carried out experimental work. M.A.V. performed the bioinformatics analyses.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Furlong, S.E., Ford, A., Albarnez-Rodriguez, L. & Valvano, M.A. 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; DOI:10.1038/srep09178 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>