

SUPPLEMENTARY INFORMATION

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

Sarah E. Furlong¹, Amy Ford², Lorena Albarnez-Rodriguez¹, and Miguel A. Valvano^{1,2*}

¹Centre for Human Immunology, Department of Microbiology and Immunology, University of Western Ontario, London, Ontario, N6A 5C1, Canada

²Centre for Infection and Immunity, Queen's University Belfast, Belfast, United Kingdom, BT9 7AE
Correspondence and requests for materials should be addressed to M.A.V (m.valvano@qub.ac.uk)

Table S1 | PHPT homologues used for bioinformatic analyses

Sequence ID	Protein name	Number of amino acids	Species
P14186	EXOY_RHISN	226	<i>Rhizobium</i> sp. strain NGR234
Q48460	YC14_KLEPN	465	<i>Klebsiella pneumoniae</i>
Q57491	Y872_HAEIN	471	<i>Haemophilus influenzae</i>
O32274	TuaA protein	179	<i>Bacillus subtilis</i>
O34301		193	<i>Rhizobium etli</i>
O50321		442	<i>Erysipelothrix rhusiopathiae</i>
O85804	CpsA	476	<i>Erwinia stewartii</i>
O87932	CpsE	227	<i>Streptococcus thermophilus</i>
P71062	YvvC	202	<i>Bacillus subtilis</i>
P74057		478	<i>Synechocystis</i> sp. strain PCC 6803
P74347		243	<i>Synechocystis</i> sp. strain PCC 6803
Q44576		532	<i>Acetobacter xylinus</i>
Q45382	WlbG	197	<i>Bordetella pertussis</i>
Q54521		455	<i>Streptococcus pneumoniae</i>
Q55999		470	<i>Sphingomonas</i> sp.
Q56624		186	<i>Vibrio cholerae</i>
Q8D3V1		466	<i>Vibrio vulnificus</i>
Q8D9C8		213	<i>Vibrio vulnificus</i>
Q8DAA6		471	<i>Vibrio vulnificus</i>
Q8DL71	Tlr0626 protein	472	<i>Synechococcus elongatus</i>
Q8DNL5		243	<i>Streptococcus pneumoniae</i> strain ATCC BAA-255 / R6
Q8EMF5		215	<i>Oceanobacillus iheyensis</i>
Q8F399		473	<i>Leptospira interrogans</i>
Q8F913		462	<i>Leptospira interrogans</i>
Q8FG28		464	<i>Escherichia coli</i> O6
Q8G1D7		512	<i>Brucella suis</i>
Q8G7L5		573	<i>Bifidobacterium longum</i>
Q8G7L8	CpsD	514	<i>Bifidobacterium longum</i>
Q8GIC9	EpsB	457	<i>Methylobacillus</i> sp. 12S
Q8GPA7	Eps7E	462	<i>Streptococcus thermophilus</i>
Q8KH81		341	<i>Aeromonas hydrophila</i>
Q8KN16		492	<i>Streptomyces</i> sp. 139
Q8KW98	RB144	202	<i>Ruegeria</i> sp. PR1b
Q8NTF8		204	<i>Corynebacterium glutamicum</i>
Q8NYP1	Cap8M	185	<i>Staphylococcus aureus</i> strain MW2
Q8NYR9	MW0103 protein	230	<i>Staphylococcus aureus</i> strain MW2
Q8P7Z9	GumD	484	<i>Xanthomonas campestris</i> pv. <i>campestris</i>
Q8PRC0		469	<i>Xanthomonas axonopodis</i> pv. <i>citri</i>
Q8R6F8		230	<i>Fusobacterium nucleatum</i> subsp. <i>nucleatum</i>
Q8RBY9		201	<i>Thermoanaerobacter tengcongensis</i>
Q8RSV5		469	<i>Leptothrix cholodnii</i>
Q8UAP6		294	<i>Agrobacterium tumefaciens</i> strain C58 / ATCC 33970
Q8UG08		522	<i>Agrobacterium tumefaciens</i> strain C58 / ATCC 33970
Q8UJ36		252	<i>Agrobacterium tumefaciens</i> strain C58 / ATCC 33970
Q8VP01		199	<i>Bordetella avium</i>
Q8VST2	CpsE	449	<i>Streptococcus agalactiae</i>
Q8VTZ3		421	<i>Leptospira interrogans</i>
Q8VVI1		477	<i>Erwinia pyrifoliae</i>
Q8XMS4		220	<i>Clostridium perfringens</i>
Q8XN45		222	<i>Clostridium perfringens</i>
Q8XN65		205	<i>Clostridium perfringens</i>
Q8XR53		203	<i>Ralstonia solanacearum</i>
Q8YMV3	Alr4823	235	<i>Anabaena</i> sp. strain PCC 7120
Q8YN51	All4719	243	<i>Anabaena</i> sp. strain PCC 7120
Q8YNY9		445	<i>Anabaena</i> sp. strain PCC 7120
Q8YPN3	All4160	351	<i>Anabaena</i> sp. strain PCC 7120
Q8YT72	All2854	469	<i>Anabaena</i> sp. strain PCC 7120

Q8YUQ0		333	<i>Anabaena</i> sp. strain PCC 7120
Q8Z5J2	WbaP	476	<i>Salmonella typhi</i>
Q92QS6		520	<i>Rhizobium meliloti</i>
Q92VP9		216	<i>Rhizobium meliloti</i>
Q936H1		185	<i>Staphylococcus aureus</i>
Q93CQ5	PssA	263	<i>Rhizobium leguminosarum</i> biovar <i>trifolii</i>
Q93I52		469	<i>Sphaerotilus natans</i>
Q93QV6		198	<i>Bacteroides fragilis</i>
Q93TW2	PglB2	525	<i>Neisseria meningitidis</i>
Q97EN5		222	<i>Clostridium acetobutylicum</i>
Q97EP9		223	<i>Clostridium acetobutylicum</i>
Q97GN6		461	<i>Clostridium acetobutylicum</i>
Q982Y9		462	<i>Rhizobium loti</i>
Q985Q8		184	<i>Rhizobium loti</i>
Q988F1		205	<i>Rhizobium loti</i>
Q98C89	ExoY	225	<i>Rhizobium loti</i>
Q9A5M2	CC2425	512	<i>Caulobacter crescentus</i>
Q9A5R3		188	<i>Caulobacter crescentus</i>
Q9A878		186	<i>Caulobacter crescentus</i>
Q9ABR0		267	<i>Caulobacter crescentus</i>
Q9AEE9		421	<i>Leptospira interrogans</i>
Q9AHD1	WciI/Cps4E	211	<i>Streptococcus pneumoniae</i>
Q9CM33	RfbP	200	<i>Pasteurella multocida</i>
Q9EYG6		471	<i>Actinobacillus pleuropneumoniae</i>
Q9F0C3	EpsE	219	<i>Lactobacillus delbrueckii</i> subsp. <i>bulgaricus</i>
Q9F742	WbgY	196	<i>Plesiomonas shigelloides</i>
Q9FDJ5		466	<i>Lactobacillus rhamnosus</i>
Q9HSZ7	Vng0011c	477	<i>Halobacterium</i> sp. strain NRC-1
Q9I1N8		478	<i>Pseudomonas aeruginosa</i>
Q9JRN2		472	<i>Actinobacillus actinomycetemcomitans</i>
Q9JRR6		400	<i>Actinobacillus actinomycetemcomitans</i>
Q9K6L3		207	<i>Bacillus halodurans</i>
Q9KTG8		465	<i>Vibrio cholerae</i>
Q9KV93		184	<i>Vibrio cholerae</i>
Q9KYG5		484	<i>Streptomyces coelicolor</i>
Q9PAX8	GumD	484	<i>Xylella fastidiosa</i>
Q9PNH0		200	<i>Campylobacter jejuni</i>
Q9R927	Cps23fE	455	<i>Streptococcus pneumoniae</i>
Q9RFX2	Cps7F	232	<i>Streptococcus suis</i>
Q9RG41	Cps9F	200	<i>Streptococcus suis</i>
Q9RGI8		202	<i>Bacteroides fragilis</i>
Q9RJK7		491	<i>Streptomyces coelicolor</i>
Q9RMC7		203	<i>Acinetobacter lwoffii</i>
Q9RR58	PglB	413	<i>Neisseria meningitidis</i> serogroup B
Q9RZC0		481	<i>Deinococcus radiodurans</i>
Q9S4G7		421	<i>Leptospira interrogans</i>
Q9WZA0		408	<i>Thermotoga maritima</i>
Q9X484		251	<i>Lactococcus lactis</i> subsp. <i>cremoris</i>
Q9X490		228	<i>Lactococcus lactis</i> subsp. <i>cremoris</i>)
Q9X4C0	WbaP	476	<i>Escherichia coli</i>
Q9X4U9	Cps2E	459	<i>Streptococcus suis</i>
Q9X648		485	<i>Aeromonas hydrophila</i>
Q9XDP9		452	<i>Actinobacillus actinomycetemcomitans</i>
Q9ZAY8		253	<i>Anabaena</i> sp. strain CA / ATCC 33047

The dataset is available from the authors as a FASTA file upon request.

Table S2 | List of strains and plasmids

Strain or Plasmid	Relevant Properties	Source or Reference
Strains		
CC118	$\Delta(ara\ leu)\ \Delta lac\ phoA\ galE\ galK\ thi\ rpsL\ rpsB\ argE\ recA$	Laboratory stock
DH5 α	F ⁻ $\phi 80lacZ\Delta M15\ endA\ recA\ hsdR(\tau_K^- m_K^-)\ nupG\ thi\ glnV\ deoR\ gyrA\ relA1\ \Delta(lacZYA-argF)U169$	Laboratory stock
W3110	<i>rph-1</i> IN(<i>rrnD-rrnE</i>)1, <i>wbbL::IS5</i>	Laboratory stock
XBF1	W3110, $\Delta wcaJ::aph$, KnR	¹
Plasmids		
pAF06	pBAD expressing _{FLAG} WcaJ-Cys-Less-V39C	This study
pAF07	pBAD expressing _{FLAG} WcaJ-Cys-Less-G74C	This study
pAF08	pBAD expressing _{FLAG} WcaJ-Cys-Less-D106C	This study
pAF09	pBAD expressing _{FLAG} WcaJ-Cys-Less-A134C	This study
pAF10	pBAD expressing _{FLAG} WcaJ-Cys-Less-P180C	This study
pAF11	pBAD expressing _{FLAG} WcaJ-Cys-Less-L223C	This study
pAF12	pBAD expressing _{FLAG} WcaJ-Cys-Less-N254C	This study
pAF13	pBAD expressing _{FLAG} WcaJ-Cys-Less-L272C	This study
pAF14	pBAD expressing _{FLAG} WcaJ-Cys-Less-K301	This study
pAF15	pBAD expressing _{FLAG} WcaJ-Cys-Less-G316C	This study
pAF16	pBAD expressing _{FLAG} WcaJ-Cys-Less-S360C	This study
pAF31	pBAD expressing _{FLAG} WcaJ-Cys-Less-S120C	This study
pAF32	pBAD expressing _{FLAG} WcaJ-Cys-Less-N143C	This study
pAF33	pBAD expressing _{FLAG} WcaJ-Cys-Less-M160C	This study
pAH18	pBAD vector expressing _{FLAG} PhoA	²
pBAD24	Cloning vector, inducible by arabinose, Amp ^R	³
pBADNTF	pBAD24 vector for N-terminal FLAG fusions, Amp ^R	⁴
pLA1	pBAD expressing His5-WcaJ (full length)	This study
pLA3	pBAD expressing _{FLAG} WcaJ (full length)	This study
pLA5	pBAD expressing _{FLAG} WcaJ _{E252-Y464} (C-terminal)	This study
pLA6	pBAD expressing _{FLAG} WcaJ-A134-PhoA	This study
pLA7	pBAD expressing _{FLAG} WcaJ-D106-PhoA	This study
pLA8	pBAD expressing _{FLAG} WcaJ-V39-PhoA	This study
pLA9	pBAD expressing _{FLAG} WcaJ-G74-PhoA	This study
pLA11	pBAD expressing _{FLAG} WcaJ-S360-PhoA	This study
pLA18	pBAD expressing _{FLAG} WcaJ-L223-PhoA	This study
pLA19	pBAD expressing _{FLAG} WcaJ-N254-PhoA	This study
pLA20	pBAD expressing _{FLAG} WcaJ-L272-PhoA	This study
pLA23	pBAD expressing _{FLAG} WcaJ-G316-LacZ	This study
pLA24	pBAD expressing _{FLAG} WcaJ-G316-PhoA	This study
pLA30	pBAD expressing _{FLAG} WcaJ-P180-LacZ	This study
pLA31	pBAD expressing _{FLAG} WcaJ-A134-LacZ	This study
pLA32	pBAD expressing _{FLAG} WcaJ-D106-LacZ	This study
pLA33	pBAD expressing _{FLAG} WcaJ-V39-LacZ	This study
pSEF62	pBAD expressing _{FLAG} WcaJ-G74-LacZ	This study
pSEF63	pBAD expressing _{FLAG} WcaJ-L272-LacZ	This study
pSEF64	pBAD expressing _{FLAG} WcaJ-N254-LacZ	This study
pSEF65	pBAD expressing _{FLAG} WcaJ-L223-LacZ	This study
pSEF66	pBAD expressing _{FLAG} WcaJ-S120-LacZ	This study
pSEF67	pBAD expressing _{FLAG} WcaJ-N143-LacZ	This study
pSEF68	pBAD expressing _{FLAG} WcaJ-M160-LacZ	This study
pSEF69	pBAD expressing _{FLAG} WcaJ-K301-LacZ	This study
pSEF71	pBAD expressing _{FLAG} WcaJ-S120-PhoA	This study
pSEF72	pBAD expressing _{FLAG} WcaJ-N143-PhoA	This study
pSEF73	pBAD expressing _{FLAG} WcaJ-M160-PhoA	This study
pSEF74	pBAD expressing _{FLAG} WcaJ-K301-PhoA	This study
pSEF75	pBAD expressing _{FLAG} WcaJ-P180-PhoA	This study
pSEF102	pBAD expressing _{FLAG} WcaJ-Cys-Less	This study
pSEF103	pBAD expressing _{FLAG} WcaJ-Cys-Less-P291C	This study
pSEF104	pBAD expressing _{FLAG} WcaJ-Cys-Less-A282C	This study
pSEF105	pBAD expressing _{FLAG} WcaJ-Cys-Less-A276C	This study
pSEF106	pBAD expressing _{FLAG} WcaJ-Cys-Less-A294C	This study

pSEF107	pBAD expressing _{FLAG} WcaJ-Cys-Less-G316C	This study
pSEF108	pBAD expressing _{FLAG} WcaJ-Cys-Less-L272C	This study
pSEF109	pBAD expressing _{FLAG} WcaJ-Cys-Less-S303C	This study
pSEF110	pBAD expressing _{FLAG} WcaJ-Cys-Less-K301C	This study
pWQ499	pKV102 containing <i>resAK30</i> ; TcR	C. Whitfield
pXF1	pBAD24 encoding W3110 WcaJ	⁵

Table S3 | List of primers used for cloning

Primer Name	Primer Number	Sequence (5'-3')
pBAD forward	252	GATTAGCGGATCCTACCTGA
pBAD reverse	258	GACCGCTTCTGCGTTCTGAT
R-His-WcaJ	4779	ATGGTGGTGGATGGTGCATCCTTGTTCCTCCATGGTG
F- His-WcaJ	4780	CACCATCATCACCATACAAATCTAAAAAAGCGC
F-SmaI-WcaJ	4911	CTAGCCCGGGACAAATCTAAAAAAGCGC
R-HindIII-WcaJ	4912	CTAGAAGCTTTC AATATGCCGCTTTGTTA
F-PstI-WcaJCT	5094	CCTGCTGCAGGAGATGAACGGCGTACCG
R-PstI-WcaJCT	5095	TCGACTGCAGCTTGTTCATCGTCATCCTTG
F-XbaI-PhoA	5257	CCCGTCTAGAGTCGACCTGCAGCCTGTTCTG
R-XbaI-V39	5364	CGCATCTAGAGACTTCGCAAACCAGCCATAG
R-XbaI-G74	5363	CGCATCTAGAACC GCGCCATGAGCGATAAAA
R-XbaI-D106	5259	CGCATCTAGAGTCGAAATCATTGTTGAACGC
R-XbaI-S120	5841	CGCATCTAGAGCTGGTCAGCGCATAACCACGCCA
R-XbaI-A134	5258	CGCATCTAGACGCCCAATGCGAATACACGA
R-XbaI-N143	5840	CGCATCTAGAGTTATAGCCATGATTACGCAGCC
R-XbaI-M160	5839	CGCATCTAGACATCAGCATTTGCCCGGCGGCTA
R-XbaI-P180	5093	CGCATCTAGACGGGTCGTGGTAAACGCCAC
R-XbaI-L223	5394	CGCATCTAGAGACCAGTTTTTTTCACTCGCGCGC
R-XbaI-N254	5393	CGCATCTAGAGCCGTTTCATCTCTTCGAGGCGTGA
R-XbaI-L272	5392	CGCATCTAGACAGGCGGTTAACCCCGGAAAG
R-XbaI-K301	5867	CGCATCTAGATTTTACCACCAGCGCAATACAGC
R-XbaI-G316	5580	CGCATCTAGAGCCGTAGCGAGTCTGGCGG

Table S4 | List of primer pairs and plasmid templates used for plasmid constructions.

Plasmid	Primer Pair	Template Plasmid
pLA1	4779/4780	pXF1
pLA3	4911/4912	pLA1
pLA5	5094/5095	pLA3
pLA8 (V39-PhoA)	5257/5364	pLA6
pLA9 (G74-PhoA)	5257/5363	pLA6
pLA7 (D106-PhoA)	5257/5259	pLA7
pSEF71 (S120-PhoA)	4911/5841	pLA3
pLA6 (A134-PhoA)	4911/5258	pLA3
pSEF72 (N143-PhoA)	4911/5840	pLA3
pSEF73 (M160-PhoA)	4911/5839	pLA3
pSEF75 (P180-PhoA)	4911/5093	pLA3
pLA18 (L223-PhoA)	5257/5394	pLA11
pLA19 (N254-PhoA)	5257/5393	pLA11
pLA20 (L272-PhoA)	5257/5392	pLA11
pSEF74 (K301-PhoA)	4911/5867	pLA3
pLA24 (G316-PhoA)	5257/5580	pLA11
pLA33 (V39-LacZ)	5640/5364	pLA23
pSEF62 (G74-LacZ)	5640/5363	pLA23
pLA32 (D106-LacZ)	5640/5259	pLA23
pSEF66 (S120-LacZ)	5640/5841	pLA23
pLA31 (A134-LacZ)	5640/5258	pLA23
pSEF67 (N143-LacZ)	5640/5840	pLA23
pSEF68 (M160-LacZ)	5640/5839	pLA23
pLA30 (P180-LacZ)	5640/5093	pLA23
pSEF65 (L223-LacZ)	5640/5394	pLA23
pSEF64 (N254-LacZ)	5640/5393	pLA23
pSEF63 (L272-LacZ)	5640/5392	pLA23
pSEF69 (K301-LacZ)	5640/5867	pLA23
pLA23 (G316-LacZ)	5640/5580	pLA23

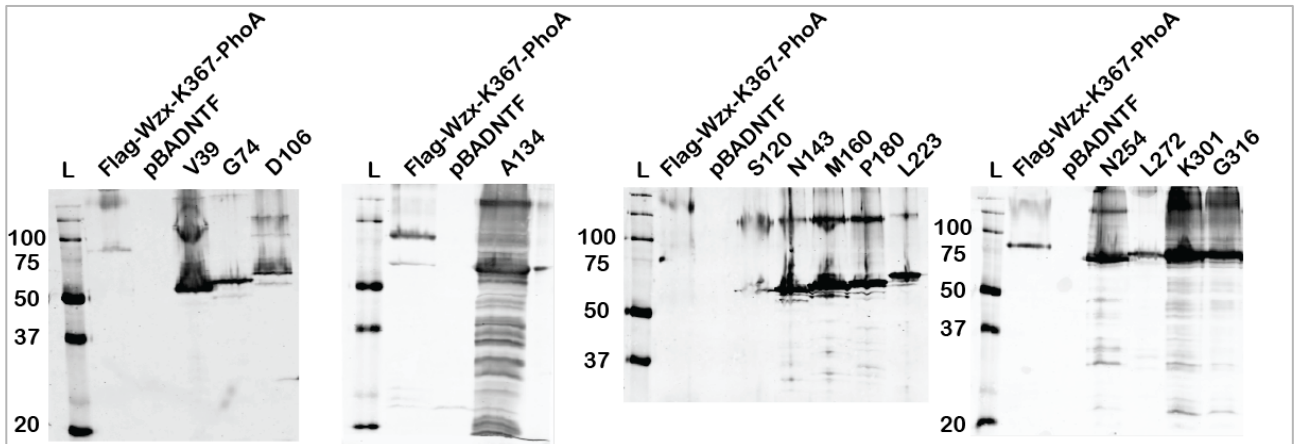


Figure S1. Membrane expression of $_{FLAG}WcaJ$ -PhoA fusion proteins in CC118 cells. 30 μ g of total membrane preparations from CC118 expressing $_{FLAG}WcaJ$ -PhoA fusions were separated by 12% SDS-PAGE and immunoblotting was performed using anti-Flag antibodies. $_{FLAG}Wzx$ -K367-PhoA² was used as positive control for PhoA western blots and experiments.

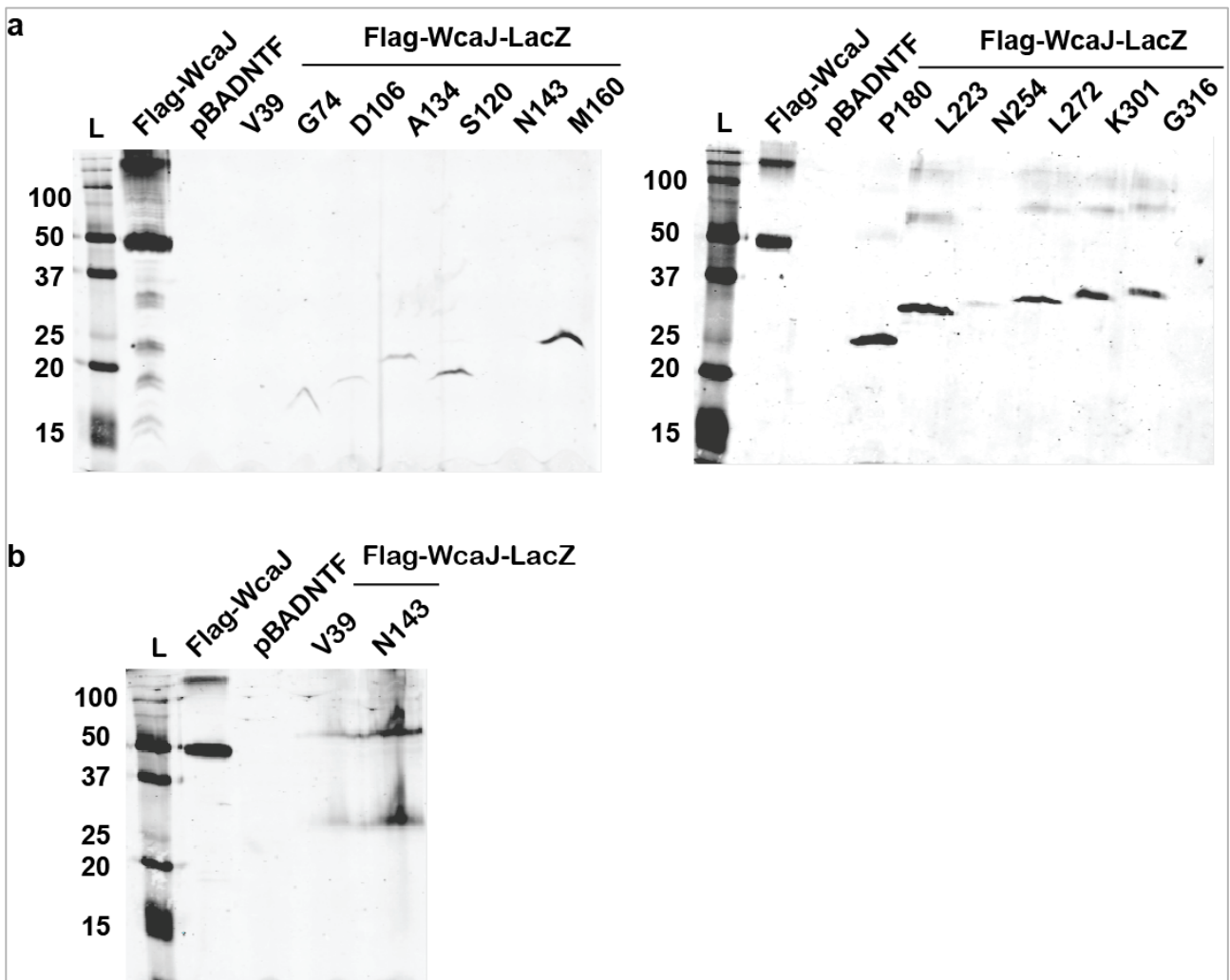


Figure S2. Membrane expression of $_{FLAG}WcaJ$ -LacZ fusion proteins in DH5 α cells. Total membrane preparations were separated by 18% SDS-PAGE and immunoblotting was performed using anti-FLAG antibodies. (a) 10 μ g of total membranes from DH5 α expressing $_{FLAG}WcaJ$ were used as a positive control. 30 μ g of total membranes from DH5 α expressing $_{FLAG}WcaJ$ -LacZ fusions were separated by SDS-PAGE. (b) 10 μ g of total membranes from DH5 α expressing $_{FLAG}WcaJ$ were used as a positive control. 50 μ g of total membranes from DH5 α expressing $_{FLAG}WcaJ$ -LacZ fusions were separated by SDS-PAGE.

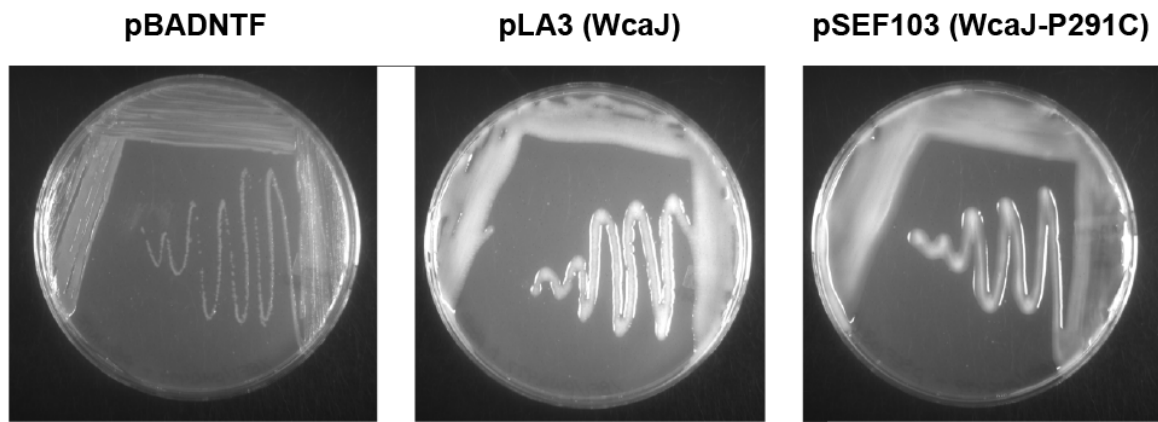


Figure S3. Complementation of colanic acid production by WcaJ_{p291C}. The *E. coli* strain $\Delta wcaJ$ (pWQ499) (XBF1; Table S2) was transformed with plasmids expressing WcaJ, WcaJ-P291C, and the vector control pBADNTF and plates were incubated in LB containing 0.2% arabinose (w/v).

"TMH-V"

	10	20	30	40	50	60
SPR P14186 EXOY_RHISN/1-55	DVLI	AI	LALI	ALS	P	-LFLLV-MGLVKF-S
SPR Q48460 YC14_KLEPN/1-55	DII	IVSS	LIL	LIS	P	-ILLV-IATAVKT-T
SPR Q57491 Y872_HAEIN/1-54	DIV	VGSL	AI	IIF	S	-VLLYL-YFAVK-K
TRE Q32274 Q32274/1-55	DIF	WAL	I	GLAI	L	P-MIAV-FSILICL
TRE Q34301 Q34301/1-55	D	FLAL	IASV	LIVL	V	-PILV-VAFSVRL
TRE Q50321 Q50321/1-55	DIF	VAV	VGA	LSS	P	-IWI-I-AAIAIKL
TRE Q85804 Q85804/1-54	D	IVG	ALS	I	VM	L-LALLV-GFLVG
TRE Q87932 Q87932/1-56	D	ILVSS	I	GLI	L	P-LFLIVALIMKCS
TRE P71062 P71062/1-55	D	LTA	AIF	L	LCT	S-V-I-LFTI
TRE P74057 P74057/1-55	D	YGG	AL	F	IV	V-LFLA-I-VIAIKL
TRE P74347 P74347/1-55	D	ICF	S	L	SV	L-LYLA-I-ALMIAV
TRE Q44576 Q44576/1-55	D	LTA	S	V	L	L-LTTL-VALAIKL
TRE Q45382 Q45382/1-55	D	VVC	S	GL	L	L-LLAL-IAIAIKL
TRE Q54521 Q54521/1-54	D	ICG	A	I	L	L-ASLVL-VP-LIR
TRE Q55999 Q55999/1-55	D	L	L	I	T	P-LMIV-VAI
TRE Q56624 Q56624/1-54	D	F	L	A	A	L-LFLI-VT
TRE Q803V1 Q803V1/1-55	D	I	V	F	S	L-LMLIS
TRE Q8D9C8 Q8D9C8/1-59	D	L	V	F	A	L-LFPL-I
TRE Q8DA6 Q8DA6/1-55	D	L	V	I	G	S-LVLA-V
TRE Q8DL71 Q8DL71/1-55	D	W	M	L	A	L-LFVVV-A
TRE Q8DNL5 Q8DNL5/1-55	D	D	I	L	S	L-LI
TRE Q8EMF5 Q8EMF5/1-55	D	L	T	F	S	L-L
TRE Q8F399 Q8F399/1-55	D	E	I	V	L	S-L
TRE Q8F913 Q8F913/1-55	D	I	V	L	A	T-L
TRE Q8FG28 Q8FG28/1-55	D	I	V	F	S	L-L
TRE Q8G1D7 Q8G1D7/1-55	D	I	V	L	S	L-L
TRE Q8G7L5 Q8G7L5/1-55	D	I	V	L	S	L-L
TRE Q8G7L8 Q8G7L8/1-55	D	I	V	L	S	L-L
TRE Q8GIC9 Q8GIC9/1-55	D	I	A	F	S	L-L
TRE Q8GPA7 Q8GPA7/1-54	D	I	C	G	A	L-L
TRE Q8KH81 Q8KH81/1-55	D	I	L	A	A	L-L
TRE Q8KN16 Q8KN16/1-55	D	R	L	G	A	L-L
TRE Q8KW98 Q8KW98/1-55	D	L	F	F	A	L-L
TRE Q8NTR8 Q8NTR8/1-55	D	I	G	V	S	L-L
TRE Q8NYP1 Q8NYP1/1-55	D	V	V	S	I	L-L
TRE Q8NYP9 Q8NYP9/1-55	D	L	I	L	S	L-L
TRE Q8P729 Q8P729/1-55	D	K	I	L	A	L-L
TRE Q8PRC0 Q8PRC0/1-55	D	E	R	V	L	A
TRE Q8RF68 Q8RF68/1-55	D	I	V	A	S	L-L
TRE Q8RB9 Q8RB9/1-55	D	I	F	V	S	L-L
TRE Q8RSV5 Q8RSV5/1-55	D	I	V	S	A	L-L
TRE Q8UA6 Q8UA6/1-55	D	M	T	S	A	L-L
TRE Q8UG08 Q8UG08/1-55	D	I	T	F	L	M
TRE Q8UJ36 Q8UJ36/1-55	D	I	A	G	A	L-L
TRE Q8VP01 Q8VP01/1-55	D	I	A	V	S	L-L
TRE Q8VST2 Q8VST2/1-54	D	I	M	G	A	L-L
TRE Q8VTZ3 Q8VTZ3/1-55	D	F	L	A	A	L-L
TRE Q8VV11 Q8VV11/1-54	D	L	V	G	A	L-L
TRE Q8XMS4 Q8XMS4/1-55	D	I	L	C	S	L-L
TRE Q8XN45 Q8XN45/1-56	D	I	L	G	S	L-L
TRE Q8XN65 Q8XN65/1-55	D	I	C	V	A	L-L
TRE Q8XR53 Q8XR53/1-55	D	I	V	V	A	L-L
TRE Q8YMV3 Q8YMV3/1-55	D	I	I	G	A	L-L
TRE Q8YN51 Q8YN51/1-55	D	I	L	G	S	L-L
TRE Q8YNY9 Q8YNY9/1-55	D	I	L	C	F	S
TRE Q8YPN3 Q8YPN3/1-55	D	I	F	G	A	L-L
TRE Q8YT72 Q8YT72/1-55	D	I	M	L	T	C
TRE Q8YUQ0 Q8YUQ0/1-55	D	I	V	G	A	L-L
TRE Q8Z52 Q8Z52/1-54	D	I	F	S	L-L	
TRE Q92Q56 Q92Q56/1-55	D	L	A	V	A	L-L
TRE Q92VP9 Q92VP9/1-55	D	I	F	S	L-L	
TRE Q936H1 Q936H1/1-55	D	I	F	S	L-L	
TRE Q93C5 Q93C5/1-55	D	I	V	S	L-L	
TRE Q93I52 Q93I52/1-55	D	I	V	S	L-L	
TRE Q93QV6 Q93QV6/1-55	D	I	I	A	S	L-L
TRE Q93TW2 Q93TW2/1-55	D	V	F	G	S	L-L
TRE Q97EN5 Q97EN5/1-55	D	I	V	F	S	L-L
TRE Q97EP9 Q97EP9/1-55	D	I	V	V	A	L-L
TRE Q97GN6 Q97GN6/1-55	D	I	V	F	S	L-L
TRE Q982Y9 Q982Y9/1-55	D	L	V	A	T	A
TRE Q985Q8 Q985Q8/1-55	D	L	A	V	A	L-L
TRE Q988F1 Q988F1/1-55	D	I	V	G	S	L-L
TRE Q98C89 Q98C89/1-55	D	L	V	V	A	L-L
TRE Q9A5M2 Q9A5M2/1-55	D	I	V	V	A	L-L
TRE Q9A5R3 Q9A5R3/1-55	D	V	V	A	L-L	
TRE Q9A878 Q9A878/1-55	D	V	V	A	L-L	
TRE Q9ABR0 Q9ABR0/1-55	D	I	F	A	L-L	
TRE Q9AEE9 Q9AEE9/1-55	D	I	L	L	S	L-L
TRE Q9AHD1 Q9AHD1/1-55	D	I	V	I	S	L-L
TRE Q9CM33 Q9CM33/1-55	D	I	L	G	S	L-L
TRE Q9EY66 Q9EY66/1-54	D	F	L	A	S	L-L
TRE Q9FOC3 Q9FOC3/1-55	D	V	I	V	A	L-L
TRE Q9FDJ5 Q9FDJ5/1-55	D	I	V	V	L	L
TRE Q9HSZ7 Q9HSZ7/1-55	D	I	A	F	S	L-L
TRE Q9I1N8 Q9I1N8/1-55	D	L	V	L	A	L-L
TRE Q9JRN2 Q9JRN2/1-54	D	I	A	M	S	L-L
TRE Q9JRR6 Q9JRR6/1-56	D	Y	L	I	T	L
TRE Q9K6L3 Q9K6L3/1-55	D	L	V	L	S	L-L
TRE Q9KTG8 Q9KTG8/1-55	D	L	V	L	S	L-L
TRE Q9KV93 Q9KV93/1-54	D	F	I	F	A	L-L
TRE Q9KYG5 Q9KYG5/1-55	D	V	T	V	S	L-L
TRE Q9PAX8 Q9PAX8/1-55	D	K	V	L	A	L-L
TRE Q9PNH0 Q9PNH0/1-54	D	F	I	L	A	L-L
TRE Q9R927 Q9R927/1-54	D	I	V	G	A	L-L
TRE Q9RFX2 Q9RFX2/1-55	D	K	L	L	A	L-L
TRE Q9RC41 Q9RC41/1-55	D	A	I	I	S	L-L
TRE Q9RCI8 Q9RCI8/1-56	D	F	V	V	F	L-L
TRE Q9RJ7 Q9RJ7/1-55	D	R	L	G	A	L-L
TRE Q9RMC7 Q9RMC7/1-55	D	I	V	I	A	L-L
TRE Q9RR58 Q9RR58/1-55	D	I	V	A	S	L-L
TRE Q9RZC0 Q9RZC0/1-55	D	L	M	L	S	L-L
TRE Q9S4C7 Q9S4C7/1-55	D	F	L	M	A	L-L
TRE Q9WZAO Q9WZAO/1-55	D	I	A	V	S	L-L
TRE Q9X484 Q9X484/1-56	D	I	C	S	L-L	
TRE Q9X490 Q9X490/1-54	D	I	L	G	L	L
TRE Q9X4C0 Q9X4C0/1-54	D	I	V	G	S	L-L
TRE Q9X4U9 Q9X4U9/1-54	D	I	L	G	A	L-L
TRE Q9X648 Q9X648/1-55	D	R	V	F	A	L-L
TRE Q9XDP9 Q9XDP9/1-54	D	I	V	S	L-L	
TRE Q9ZAY8 Q9ZAY8/1-55	D	I	V	F	S	L-L

Figure S4. Alignment of the TMH-V and C terminal region of PHPT homologues. The dataset is described in Table S3. Initial alignments using Clustal Omega ⁶ revealed that an aspartic acid residue (in green) corresponding to D-278 in WcaJ is highly conserved (see Fig. 3 in the main manuscript). To facilitate the alignment, all sequences were trimmed to remove all amino acids N-terminally located with the respect to D-278, which serve to mark the predicted border of TMH-V. The highly conserved proline corresponding to P-291 in WcaJ is indicated in red.

Supplementary References

- 1 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).
- 2 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).
- 3 Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. *J. Bacteriol.* **177**, 4121-4130. (1995).
- 4 Marolda, C. L., Vicarioli, J. & Valvano, M. A. Wzx proteins involved in O antigen biosynthesis function in association with the first sugar of the O-specific lipopolysaccharide subunit. *Microbiology* **150**, 4095-4105 (2004).
- 5 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).
- 6 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7:539**, doi:10.1038/msb.2011.75 (2011).