SUPPORTING INFORMATION

A general protein *O*-glycosylation machinery conserved in *Burkholderia* species improves bacterial fitness and provides glycan immunogenicity in humans

Yasmine Fathy Mohamed, Nichollas E. Scott, Antonio Molinaro, Carole Creuzenet, Ximena Ortega, Ganjana Lertmemongkolchai, Michael M. Tunney, Heather Green, Andrew M. Jones, David DeShazer, Bart J. Currie, Leonard J. Foster, Rebecca Ingram, Cristina De Castro, and Miguel A. Valvano*

Content

Table 1. ^{1}H (plain) and ^{13}C (italic) chemical shifts of the NMR spectra acquired at 600 MHz, at 15°C in D_2O, for the glycopeptides from DsbA after Proteinase K digestion and chromatographic purification. The methyl signals of the acetyls of the two GalNAc residues were at 2.08 and 2.01 ppm.

	1	**2**	**3**	**4**	**5**	**6,6'**
A	5.07	4.38	3.82	4.23	3.85	3.76 x 2
3)-α-GalNAc	*94.5*	*51.0*	*79.1*	*69.4*	*72.2*	*61.7*
B	4.62	4.11	3.83	4.14	3.67	3.82 x 2
3)-β-GalNAc	*101.7*	*52.0*	*75.8*	*64.6*	*76.4*	*62.2*
C	4.45	3.51	3.64	3.96	3.90	4.31;4.23
t-β-Gal	*106.3*	*71.6*	*73.4*	*69.4*	*73.1*	*64.4*

Table S2

m/z of most intense ion	**1353.6985**	**1421.39**	**1489.086667**	**1464.74658**	**1157.57959**	**1182.58154**
Extracted ion chromotograms m/z range	**1353.69-1353.72**	**1421.38-1421.41**	**1489.07-1489.10**	**1464.72-1464.75**	**1157.56-1157.59**	**1182.57-1182.60**
Strain/Data file	**Unmodified peptide**	**Peptide+HexNAc**	**Peptide+HexNAc2**	**Peptide+QuiNAc-Rha**	**complete glycans form A**	**complete glycans form B**
Nsco_BC_k562_WT_DsbA1	5694171	378957	73850	78417	67106443	80854183
Nsco_BC_k562_△pglL_DsbA1	49794099	100	33659	53826	6911	100
Nsco_BC_k562_△ogc_DsbA1	227191060	9573	198576	44420514	100755	42680
Nsco_BC_k562_△ogcX_DsbA1	287304020	29810476	143444	2433825	32169	49652
Nsco_BC_k562_△ogcI_DsbA1	274272635	73564	285542	2119523	14443	100
Nsco_BC_k562_△ogcAB_DsbA1	246683625	128284289	11378	1744402	224497	20083
Nsco_BC_k562_△ogcE_DsbA1	17269199	100	100	265201	1063190	739204

*Note in cases where no ion of interest was detected a value of 100 has been recorded

as a %	**Unmodified peptide**	**Peptide+HexNAc**	**Peptide+HexNAc2**	**Peptide+QuiNAc-Rha**	**complete glycans form A**	**complete glycans form B**
Nsco_BC_k562_WT_DsbA1	3.69	0.25	0.05	0.05	43.52	52.44
Nsco_BC_k562_△pglL_DsbA1	99.81	0.00	0.07	0.11	0.01	0.00
Nsco_BC_k562_△ogc_DsbA1	83.54	0.00	0.07	16.33	0.04	0.02
Nsco_BC_k562_△ogcX_DsbA1	89.85	9.32	0.04	0.76	0.01	0.02
Nsco_BC_k562_△ogcI_DsbA1	99.10	0.03	0.10	0.77	0.01	0.00
Nsco_BC_k562_△ogcAB_DsbA1	65.44	34.03	0.00	0.46	0.06	0.01
Nsco_BC_k562_△ogcE_DsbA1	89.31	0.00	0.00	1.37	5.50	3.82

Table S3. Bacterial strains and plasmids

Strain or plasmid	Description[a]	Source/reference
E. coli		
DH5α	F$^-$ ϕ 80 *lac*ZΔM15 *endA recA hsdR*17($r^-_K m^+_K$) *supE thi*-1Δ*gyrA* (Δ*lacZYA-arg*F)U169 *relA*1F$^-$	Lab stock
GT115	*mcrA* Δ*(mrr-hsdRMS-mcrBC)* ϕ *80lacZΔM15* Δ*lacX74 recA1 rpsL endA1* Δ*dcm uidA(ΔMluI)::pir-116* Δ*sbcC-sbcD*	Invivogen
B. cenocepacia		
K56-2	Clinical isolate, ET12 clone related to J2315	BCRRC[b]
MSS49	K56-2; Δ*pglL* (BCAL0960)	Lab stock
YFM35	K56-2; Δ*ogcE* (BCAM3117)	This study
YFM36	K56-2; Δ*ogcAB* (BCAL3115-BCAL3116)	This study
NS001	K56-2; Δ*ogcB* (BCAL3116)	This study
XOA25	K56-2; Δ*ogc* (BCAL3114-BCAL3118)	This study
XOA30	K56-2; Δ*ogcX* (BCAL3114)	This study
XOA31	K56-2; Δ*ogcI* (BCAL3118)	This study
MV4179	K56-2; ΔO-Antigen cluster (BCAL3119-BCAL3131)	Lab stock
Other *Burkholderia species*		
B. thailandensis e264	American Type Culture Collection isolate 700388	ATCC
B. gladioli MDU1	Clinical isolate AUSMDU00013928 from the Victorian Infectious Diseases Reference Laboratory	This study
B. gladioli MDU2	Clinical isolate AUSMDU00013929 from the Victorian Infectious Diseases Reference Laboratory	This study
B. pseudomallei NT1	Clinical isolate MSHR 831	This study
B. pseudomallei NT2	Clinical isolate MSHR 1967	This study
B. pseudomallei NT3	Clinical isolate MSHR 1824	This study
B. pseudomallei NT4	Clinical isolate MSHR 715	This study
B. pseudomallei NT5	Clinical isolate MSHR 8582	This study
B. pseudomallei NT6	Clinical isolate MSHR 9219	This study
B. pseudomallei NT7	Clinical isolate MSHR 9138	This study
B. pseudomallei NT8	Clinical isolate MSHR 9443	This study
Plasmids		
pGPISce-I	ori_{R6K}, *mob*$^+$, ΩTp^R , including an ISce-I restriction site	(52)
pDAI-SceI-SacB	ori_{pBBR1}, Tet^R, P_{dhfr}, *mob*$^+$, expressing ISce-I and the negative selection marker SacB	M. Hamad
pDA12	Cloning vector, ori_{pBBR1}, Tet^R, *mob*$^+$, P_{dhfr}	D. Aubert
pRK2013	ori_{colE1}, RK2 derivative, Kan^R, *mob*$^+$, *tra*$^+$	(52)
pYM4	pGPI-SceI with fragments flanking BCAL1907	This study
pYM36	pGPI-SceI with fragments flanking *ogcE*	This study
pYM37	pGPI-SceI with fragments flanking *ogcAB*	This study
pYM38	C- 6x His-tagged BCAL2640 cloned into pDA12, Tet^R	This study
pYM39	*ogcE* cloned in pET28a with N-6x His-Tag, Kn^R	This study
pXO23	*ogcE* cloned into pAP20, Cm^R	This study
pXO51	pGPI-SceI with fragments flanking *ogcX*	This study
pXO52	pGPI-SceI with fragments flanking *ogcI*	This study
pXO49	pGPI-SceI with fragments flanking *ogcX-ogcI*	This study
pNS01	pGPI-SceI with fragments flanking *ogcB*	This study
pAMF22	C-10× His-tagged DsbA1 from *N. meningitidis* MC58 cloned into pMLBAD, Tp^R	(28)

[a]Tp^R, trimethoprim resistance, Tet^R, tetracycline resistance, Cm^R, Chloramphenicol resistance and Kn^R, Kanamycin resistance. [b]BCRRC, *B. cepacia* Research and Referral Repository for Canadian CF Clinics.

Table S4. Primers

Primer number	Oligonucleotide sequence, 5'–3'	Restriction enzyme
Q539	TTTTTTCATATG ccgcgtcgagtacttcgtcgactg	NdeI
Q540	TTTTTTTCTAGAttaGTGGTGGTGGTGGTGGTG ccggttgccggccgtctccgacag	XbaI
Q653	TTTTTCTAGAgtcgtgctgccgtcgtac	XbaI
Q654	TTTTCTCGAGcttgctgttgacgaggttgt	XhoI
Q655	TTTTCTCGAGgagtgctatgcgaaccc	XhoI
Q656	TTTTGAATTCttggtcagatcctcgacgag	EcoRI
Q657	TTTTGAATTCcgtgttgctgtcgaaatgg	EcoRI
Q658	TTTTCTCGAGgtcgatacgggccgcgtag	XhoI
Q659	TTTTCTCGAGgaattcgacgaacagcagg	XhoI
Q660	TTTTTCTAGAgccgatgttgttctggtag	XbaI
Q729	TTTTCATATGaccgctaaaggcaccatcctcgtc	NdeI
Q728	TTTTGAATTCttatacaaaaccgcgcgggttcc	EcoRI
6030	TTTTGAATTCcgtccgactacgagtgcat	EcoRI
6031	TTTTCTCGAGgagatcgacggccgagtag	XhoI
6032	TTTTCTCGAGaacacggtcagcctgatg	XhoI
6033	TTTTTCTAGAtcaaccacaacatcgagacg	XbaI
Nsco_0182	gcatgcgatatcgagctctcccgtgcacgtgctgatcgtc	Gibson
Nsco_0183	cgatgcttccatcagcgtCTAaggtggcggacatgaaaa	Gibson
Nsco_0184	ttttcatgtccgccacctTAGacgctgatggaagcatcg	Gibson
Nsco_0185	cggataacaatttgtggaattcccgcgatcagcttcgactgg	Gibson

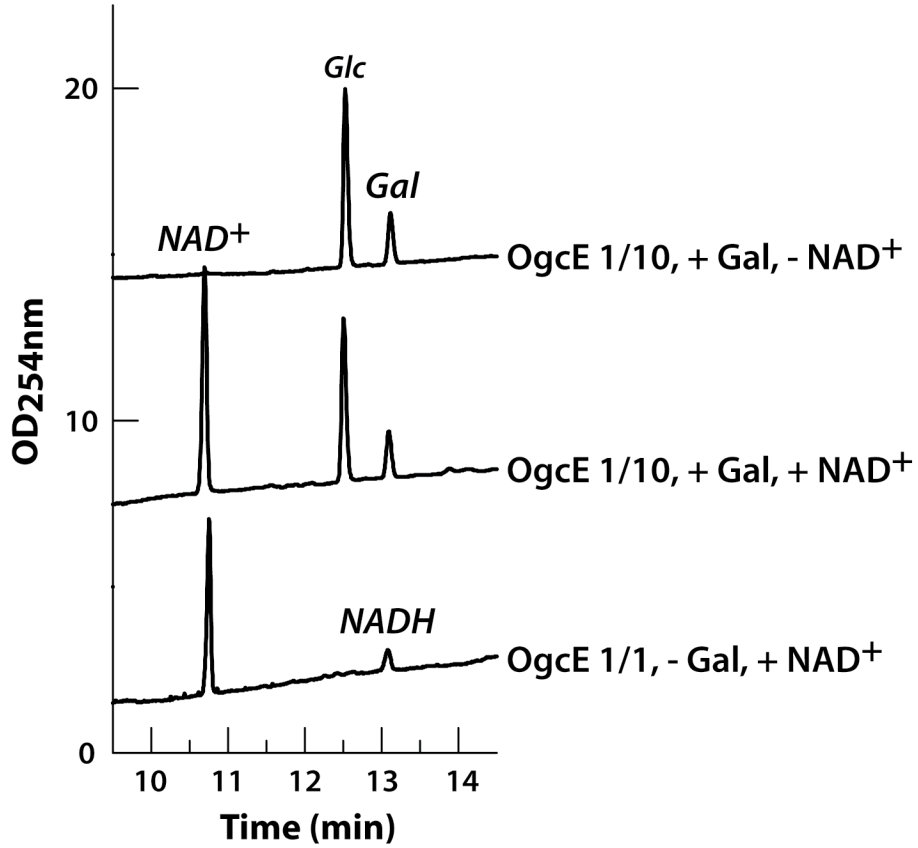

Fig. S1. Capillary electrophoresis electropherograms showing that OgcE does not need any exogenous cofactor for full activity and showing the release of enzyme-bound cofactor (seen as NADH). The reactions were of the same composition as on (Fig. 5) except that the enzyme concentration and the presence of cofactor or of substrate varied as indicated on the figure. The enzyme dilution 1/1 is the same dilution as in Figure 5. The reactions were incubated for 4 h at 37°C. Data shown for 1 reaction per condition optimized for enzyme to substrate ratio and for reaction time after initial result was observed on 4 reactions per condition.

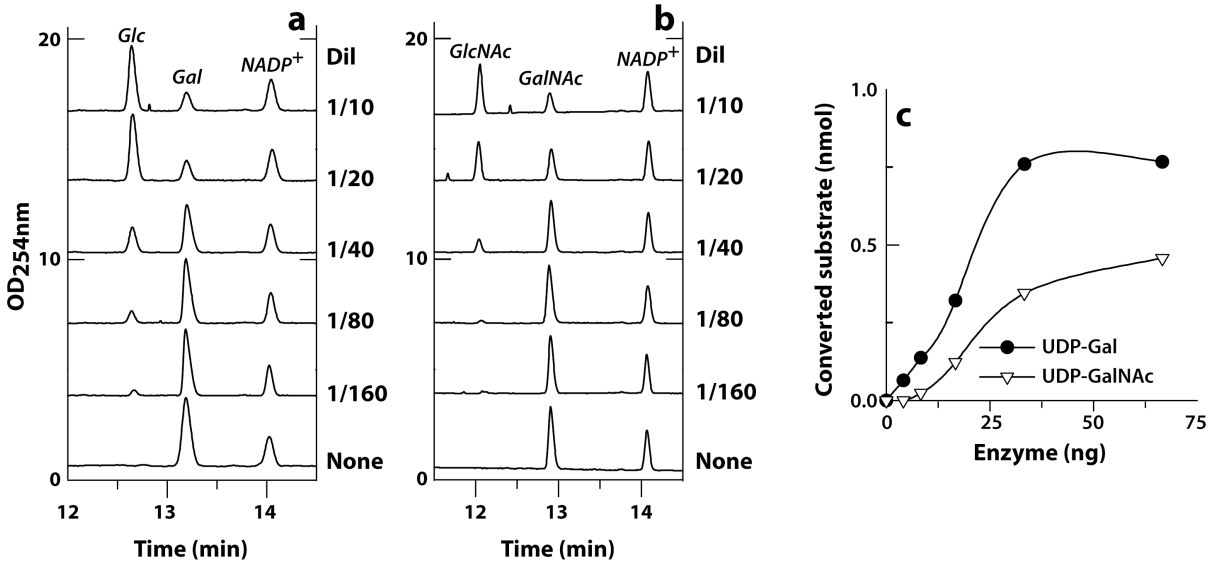

Fig. S2. Enhanced catalysis of UDP-Gal versus UDP-GalNAc at all enzyme dilutions tested. The reactions were incubated for 10 min. Data are representative of two independent series showing the same trend. One sample per enzyme concentration in each series.

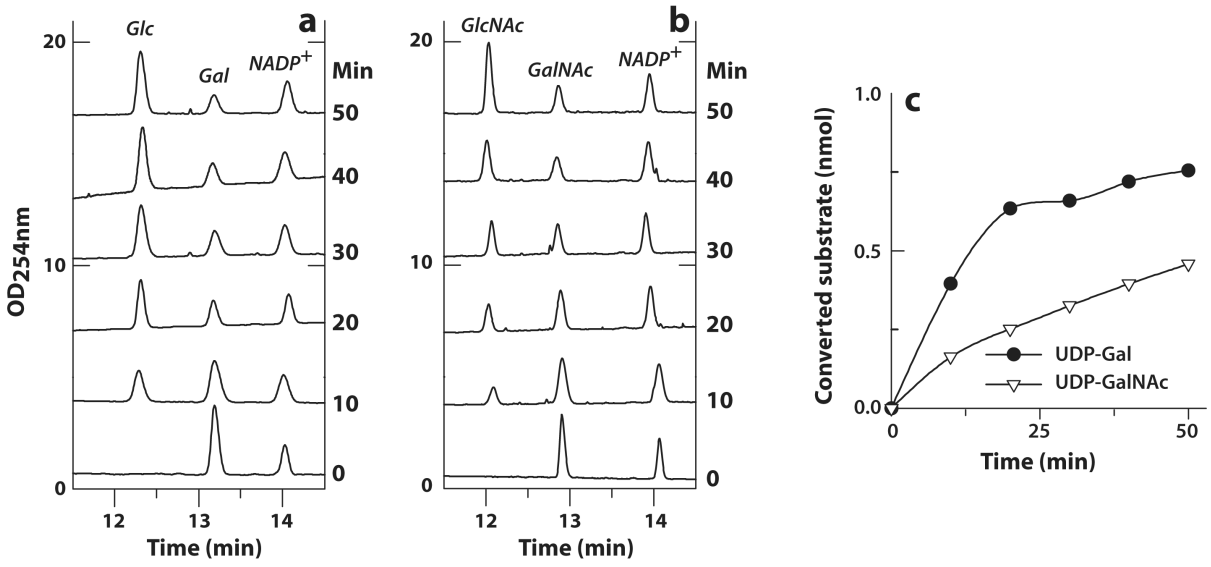

Fig. S3. Enhanced catalysis of UDP-Gal versus UDP-GalNAc at all time points tested. The enzyme was used in a 1/160 dilution. Data are representative of two independent series showing the same trend. One sample per time point in each series.

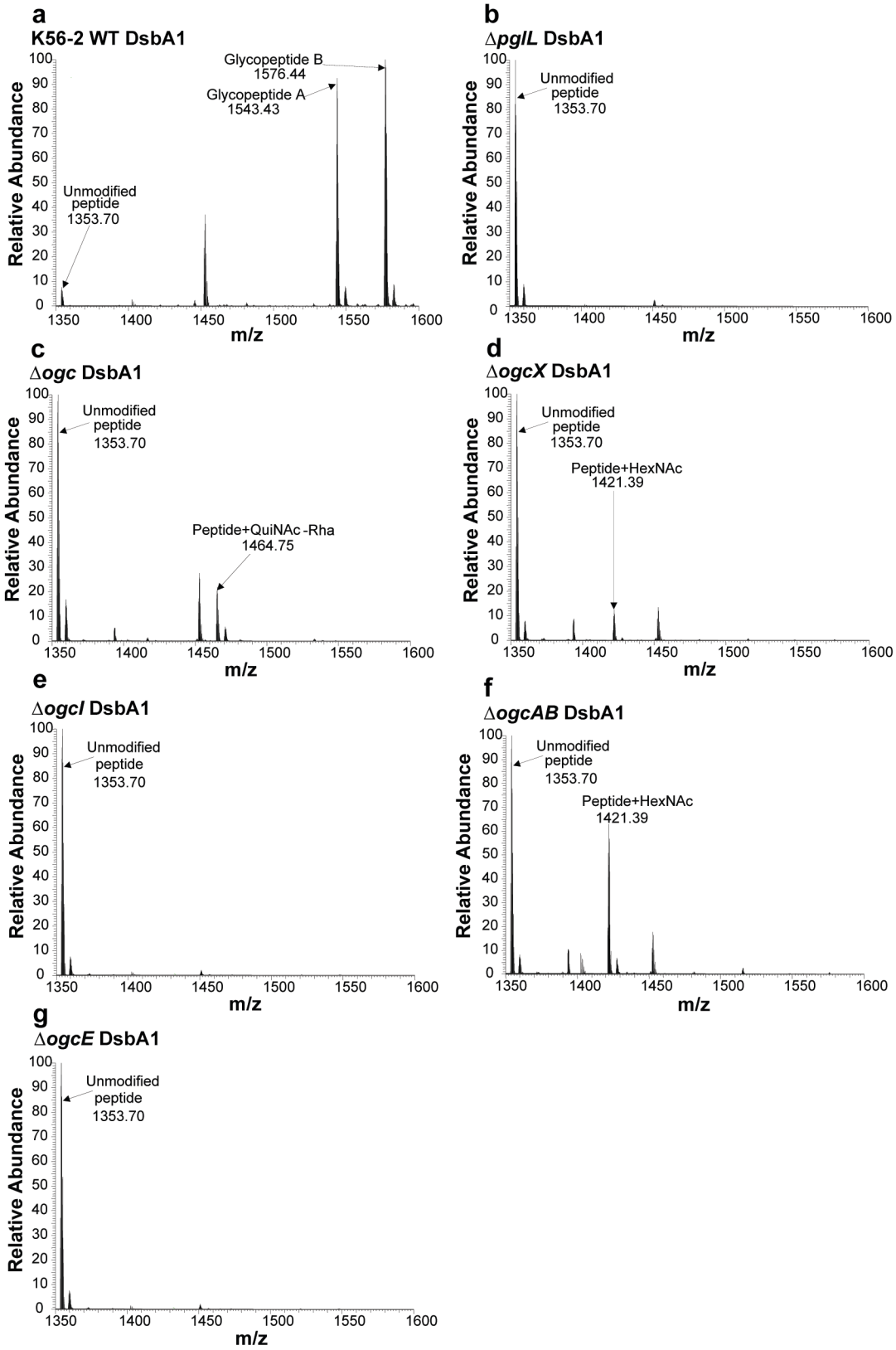

Fig. S4. MS spectrum of the 6xHis-tagged DsbA1 derived peptides ^{23}VQTSVPADSAPAASAAAAPAGLVEGQNYTVLANPIPQQQAGK^{64}. Peptide forms within each strains **a** K56-2 wildtype (WT) **b** Δ*pglL* **c** Δ*ogI* **d** Δ*ogcX* **e** Δ*ogcI* **f** Δ*ogcAB* **g** Δ*ogcE*

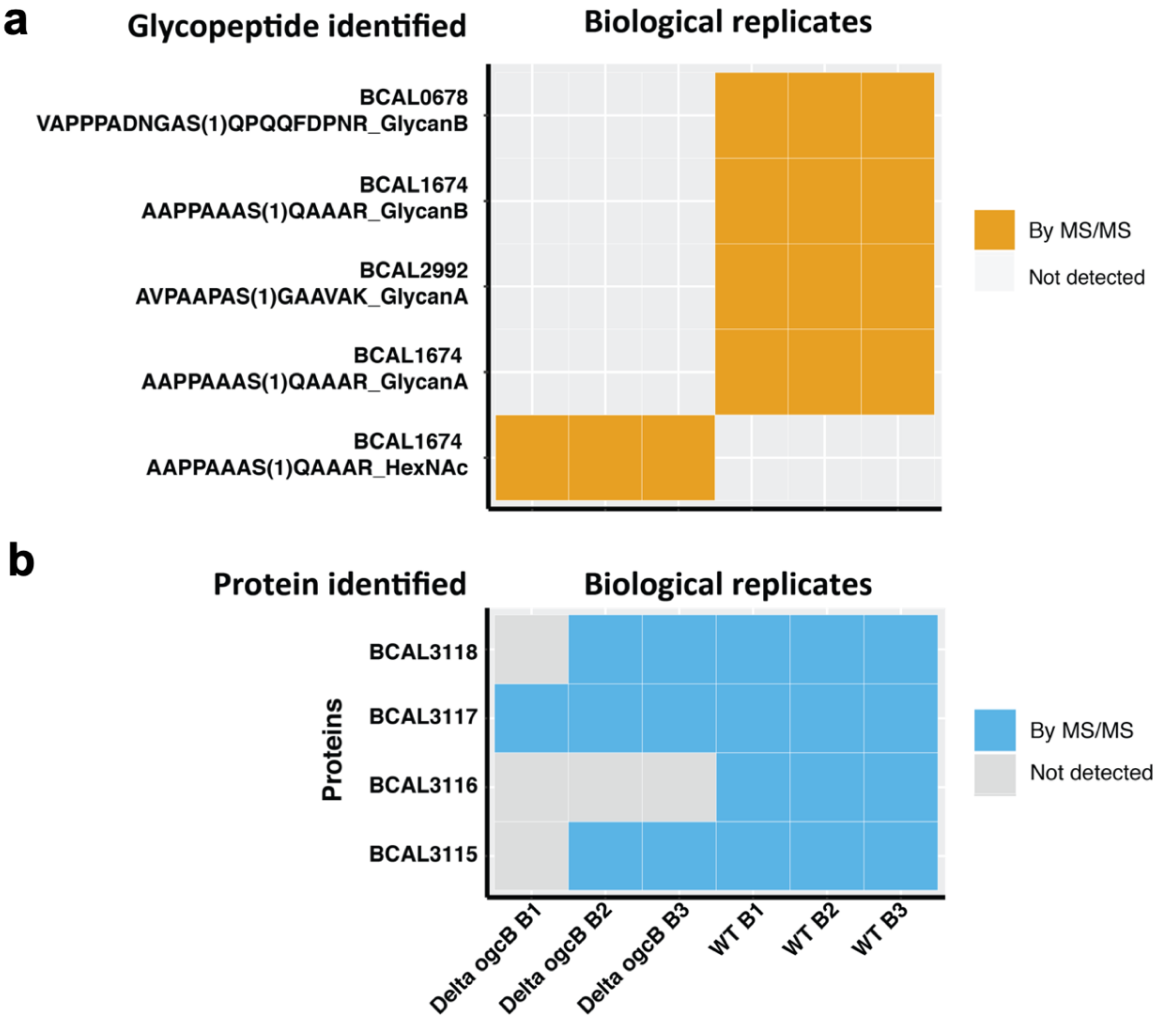

Fig. S5. Proteomic analysis of K56-2 Δ*ogcB* compared to Wild type K56-2." By MS/MS" denotes samples in which the glycopeptide/protein was identified by MS/MS using Maxquant based analysis **a** Disruption of *ogcB* results in the loss glycan A- and B-decorated glycopeptides observed within wildtype replicates. Examples shown include the putative membrane protein BCAL0678, the putative lipoprotein BCAL2992, and the multidrug efflux system AmrA protein BCAL1674 (uniport accessions: B4E9M8, B4EB92, and B4E8U6, respectively). Also, the AmrA protein BCAL1674 glycopeptide ^{397}AAPPAAASQAAAR^{409} is modified with a single HexNAc, which is not observed within WT replicates. **b** Proteome analysis confirms the disruption of *ogcB* does not affect flanking genes since the proteins BCAL3115 (OgcA), BCAL3116 (OgcB), BCAL3117 (OgcE) and BCAL3118 (OgcI) encoded by the *ogc* cluster are readily detected in whole cell proteome samples, while OgcB was undetectable.

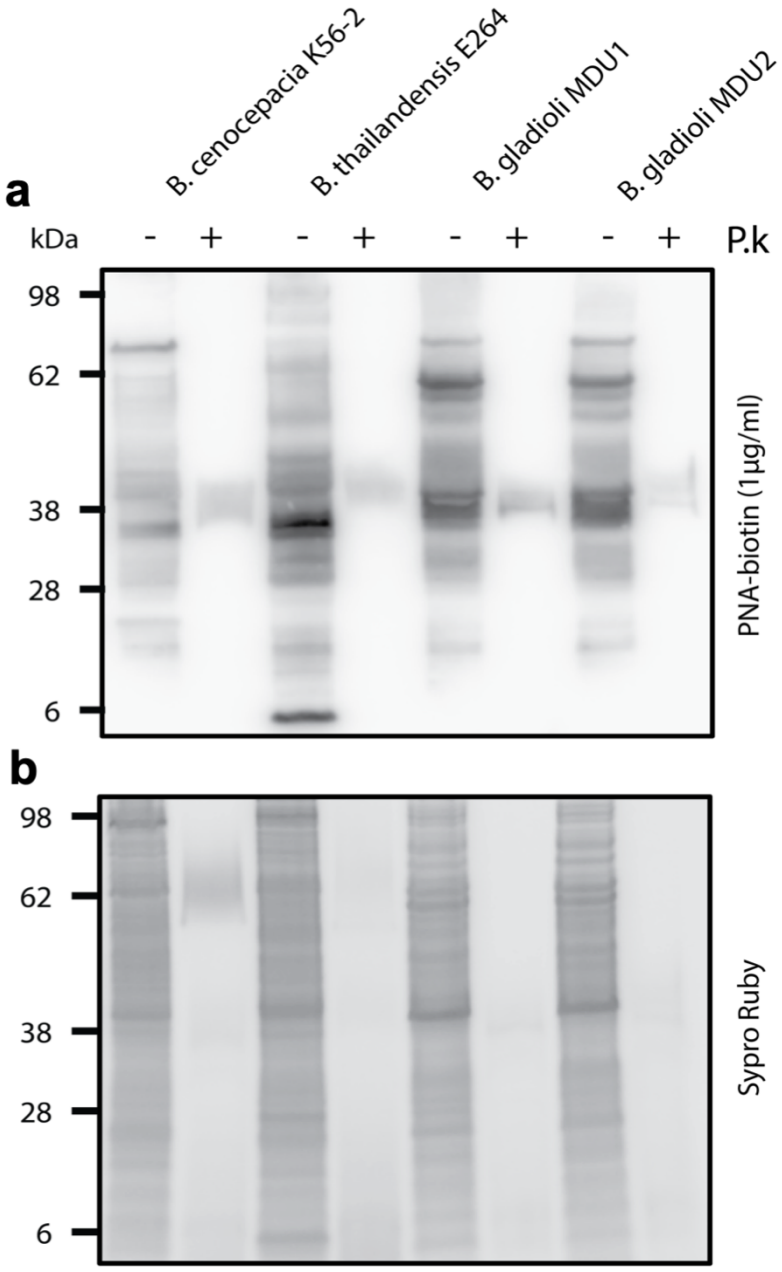

Fig. S6. Peanut agglutinin (PNA) lectin blotting to demonstrate the presence of Gal-β(1-3)-GalNAc modified glycoproteins in *Burkholderia* species. **a** lectin blotting with the Gal-β(1-3)-GalNAc specific lectin PNA conjugated to biotin and probed with fluorescent streptavidin demonstrates the presence of multiple reactive bands in *B. cenocepacia, B. thailandensis* and *B. gladioli.* Treatment with proteinase K (p.K,) abolishes lectin detection. **b** Protein staining of *Burkholderia* lysates demonstrates that the observed banding pattern within PNA lectin blotting is unique to that of the total proteome and confirm the removal of protein upon proteinase K treatment.

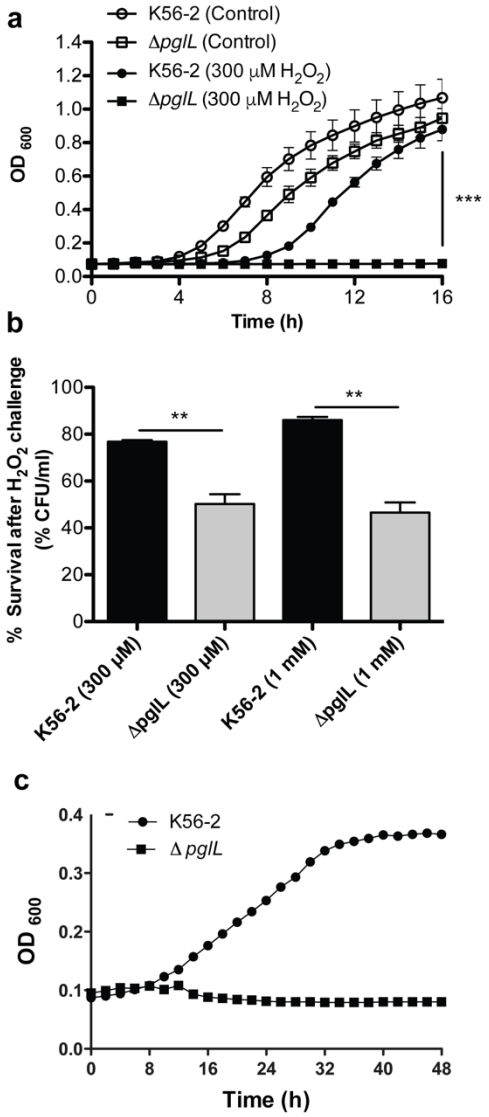

Fig. S7. a Δ*pglL* mutant bacteria are more sensitive to oxidative stress. Sensitivity of K56-2 and Δ*pglL* to 300 μg/ml H_2O_2 as determined turbidimetrically using a Bioscreen C. **b** sensitivity to oxidation determined by challenging either low or high bacterial inoculum by 300 μg/ml and 1 mM of H_2O_2, respectively in LB for 1 h at 37°C. $n = 6$ from 3 independent experiments each done in duplicate. **c** Δ*pglL* mutant bacteria are more sensitive to osmotic stress. Growth curve depending on OD_{600} of K56-2 and Δ*pglL* in M9 minimal media in the presence of 4% NaCl.

Supplementary Data S1. Identified *Burkholderia thailandensis E264* Glycopeptides. Eight unique glycopeptides were observed within whole cell proteome samples of *B. thailandensis* E264. For assigned glycopeptides the gene, protein name, uniprot accession (mapped to *B. thailandensis* E264), observed m/z, charge state, MH+ mass, peptide mass, glycan, peptide sequence, MASCOT ion score and containing annotated CID and HCD data are provided.

Supplementary Data S2. Identified *Burkholderia gladioli* MDU1 (AUSMDU00013928) Glycopeptides. Twenty-nine unique glycopeptides were observed within whole cell proteome samples of B. gladioli MDU1. For assigned glycopeptides the gene, protein name, uniprot accession (mapped to *B. gladioli* BSR3), observed m/z, charge state, MH+ mass, peptide mass, glycan, peptide sequence, MASCOT ion score and containing annotated CID and HCD data are provided.

Supplementary Data S3. Identified *Burkholderia gladioli* MDU2 (AUSMDU00013929) Glycopeptides. Twenty-four unique glycopeptides were observed within whole cell proteome samples of *B. gladioli* MDU2. For assigned glycopeptides the gene, protein name, uniprot accession (mapped to *B. gladioli* BSR3), observed m/z, charge state, MH+ mass, peptide mass, glycan, peptide sequence, MASCOT ion score and containing annotated CID and HCD data are provided.

Supplementary Data S4. Identified *Burkholderia pseudomallei* Glycopeptides observed across strains. Ten unique glycopeptides were observed within whole cell proteome samples of *B. pseudomallei* strains. For assigned glycopeptides the *B. pseudomallei* strain, gene, protein name, uniprot accession (mapped to *B. pseudomallei* K96243), observed m/z, charge state, MH+ mass, peptide mass, glycan, peptide sequence, MASCOT ion score and containing annotated CID and HCD data are provided.

Supplementary Data S5. Phenotypic array analysis. Excel file with the raw Area under the Curve (AUC) values used for the calculations of the relative growth defects of *pglL* in the various carbon sources.

Supplementary Data S6. qRT-PCR data. Excel file with the qRT-PCR data used for the calculation of the normalized fold expression of target transcripts in *Galleria mellonella*.