Supplementary Information File

"An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*"

Juan D. Valderrama-Rincon, Adam C. Fisher, Judith H. Merritt, Yao-Yun Fan, Craig A. Reading, Krishan Chhiba, Christian Heiss, Parastoo Azadi, Markus Aebi and Matthew P. DeLisa

Supplementary Methods

Bacterial strains and media. Antibiotic selection was maintained at: 100 μg/mL ampicillin (Amp), 25 μg/mL chloramphenicol (Cam), and 50 μg/mL kanamycin (Kan). Luria Bertani (LB) media was used for *E. coli*, supplemented with glucose at 0.2% as indicated. Protein expression was induced by adding L-arabinose and isopropyl β-d-thiogalactoside (IPTG) at 0.2% and 100 mM, respectively. Yeast FY834 was maintained on YPD media and synthetic-defined-Uracil media was used to select or maintain yeast plasmids. *E. coli* MC4100 (F^- *araD139* Δ(*argF-lac*)*U169 flbB5301 deoC1 ptsF25 relA1 rbsR22 rpsL150 thiA*) was used as the recipient strain for genetic manipulations. To delete the *gmd* and/or *waaL* genes in MC4100 cells, P1*vir* phage transduction was performed using standard methods and the Keio collection[1]. The Kan resistance cassette was removed from MC4100 *waaL*::kan cells using plasmid pCP20[2] prior to P1*vir* phage transduction to obtain the strain MC4100 *gmd*::kan Δ*waaL*.

Plasmid construction. Plasmid pTrc99A-*ALG13*::*ALG14* was used for co-expression of yeast Alg13 and Alg14. Plasmid pTrc99A-*ALG1* and pBAD(*ALG2*)-DEST49[3] were used for individual expression of Alg1 and Alg2, respectively. pBAD(*ALG2*)-DEST49 was a generous gift from Dr. Barbara Imperiali, while pTrc99A-*ALG13*::*ALG14* and pTrc99A-*ALG1* were constructed using standard techniques. *ALG13, ALG14* and *ALG1* genes were amplified from *Saccharomyces cerevisiae* genomic DNA. A C-terminal 6x-His tag was added to Alg1 and Alg13 while an N-terminal FLAG epitope tag was added to Alg14. An additional ribosomal binding site was introduced in front of *ALG14* to allow bicistronic expression. Plasmids pTrc99A-MBP$^{4x\text{-}DQNAT}$, pTrc99A-ssDsbA-scFv13-R4$^{4x\text{-}DQNAT}$ and pTrc-ssDsbA-Fc^{DQNAT} were used for expression of glycoproteins, as reported previously[4]. Plasmid pTrc99A-ssDsbA-scFv13-R4$^{1x\text{-}DQNAT}$ was constructed as described for pTrc99A-ssDsbA-scFv13-R4$^{4x\text{-}DQNAT}$. The

remaining plasmids were constructed using standard homologous recombination in *S. cerevisiae* as previously described[5]. Briefly, pYCG (see Fig. 1) was constructed by first PCR amplifying the genes *ALG13, ALG14, ALG1*, and *ALG2* from *S. cerevisiae* genomic DNA with primers containing appropriate regions of overlap. PCR products and the linearized vector pMQ70 were used to transform yeast strain FY834. Constructs assembled in yeast were electroporated into *E. coli* for verification via PCR, restriction enzyme digestion and/or sequencing. For glycosylation studies, vector pMW07 was similarly generated from pMQ70 by replacing the origin (*ori*) and antibiotic resistance cassette with the p15a *ori* and the *cat* gene for resistance to Cam. Plasmid pMW07 was the host vector for pYCG-PglB$_{Cj}$ encoding *C. jejuni* PglB amplified from (i) pACYC-pgl (see Fig. 1) or (ii) a synthesized, codon optimized version of *C. jejuni* PglB (Mr. Gene), (iii) pYCG-PglB$_{Cj}$mut amplified from pACYC-pgl *pglB*mut encoding an inactive PglB variant[6] and (iv) pYCG-PglB$_{Cl}$ encoding a codon optimized *C. lari* PglB synthesized for this work (Mr. Gene). The codon optimized version of PglB$_{Cj}$ showed a modest improvement over the wild-type enzyme (~15% increase in yield, data not shown) and was used to generate glycosylated scFv13-R4$^{1x\text{-}DQNAT}$ for Pronase E digestion, scFv13-R4$^{4x\text{-}DQNAT}$ for NMR analysis, RNAseA S32D for Western blot, and for the comparison of PglB$_{Cj}$ and PglB$_{Cl}$ (Fig. S4c). Plasmid pTrc99Y was generated from pTrc99A by adding the yeast origin and *URA* selection cassette from pMQ80[5]. The gene encoding the bovine RNAseA S32D mutant was amplified from pMIK81[7] for pTrc99Y-ssDsbA-RNAseA S32D. The gene encoding hGHv was synthesized (Mr. Gene) and amplified for pTrc99Y-MBP-hGHv^{DQNAT}. pYCGΔ*alg1* and pYCGΔ*alg2* were derived from pYCG using standard digestion and ligation.

Isolation of membranes and LLOs. Overnight cultures were diluted 1:100 into LB broth and incubated at 37^{o}C with shaking (250 rpm) until reaching A_{600} ~0.6. Protein expression was then

induced with 0.2% L-arabinose and cells were incubated at 30°C for 4h. At this point, cells where harvested and membranes where isolated similarly to the method of Marani *et al*[8]. For extraction of lipid-linked oligosaccharides (LLOs), a single colony was inoculated and cultured at 30°C overnight in screw-capped flasks filled with media. LLOs were extracted and partially purified as described previously[9]. Briefly, the LLO extraction protocol was based on a total lipid extraction protocol using chloroform:water (2:1) as described previously[10]. This solvent mixture extracts the most hydrophobic lipids, leaving LLOs in the precipitate. The precipitate was washed with water to remove hydrophilic molecules and then LLOs were extracted preferentially using chloroform:methanol:water (10:10:3). Contaminating oligosaccharides not attached to lipids were removed by adsorbing extracted LLOs on DEAE-cellulose. LLOs were then recovered from DEAE-cellulose using a 300 mM ammonium acetate solution. After releasing glycans from LLOs by acid hydrolysis (0.1 N HCl in 50% isopropanol at 60°C for 1 h), a butanol/water mixture was used to extract glycans to the water phase while lipids remained in the butanol phase. Finally, two ionic exchange resins (AG50W-X8 (Sigma) and AG1-X8 (BioRad)) were used for removing the remaining salts.

Flow cytometry and microscopy. *E. coli* cells were incubated at 30°C overnight in LB supplemented with 0.2% L-arabinose in filled, sealed culture tubes. Cells were pelleted, washed, resuspended in (PBS) and boiled for 10 min. 2.5 μg/mL *Canavalia ensiformis* concanavalin A (ConA)-AlexaFluor was added to the samples before incubation in the dark for 15 min at room temperature. 100 μL of each sample was analyzed using a FACSCalibur (Becton Dickinson). Median fluorescence was determined from 10,000 events. For microscopy, 5 μL of cells treated with ConA-AlexaFluor were imaged on a Zeiss Axioskop 40 fluorescent microscope.

Glycoprotein expression and purification. MC4100 *gmd*::kan Δ*waaL* cells were freshly transformed with pYCG-PglB$_{Cj}$, and a plasmid for glycoprotein expression. Overnight cultures were diluted 1:100 into LB broth with antibiotics and incubated at 30^{o}C with shaking (250 rpm) until reaching A_{600} ~3. Cultures were then induced with IPTG and L-arabinose, and returned to 30^{o}C for ~16 h. Cells were pelleted, frozen, resuspended in ConA column buffer (50 mM HEPES, 0.15 M NaCl, 0.1 mM $CaCl_2$, 0.01 mM $MnCl_2$, pH 7.0) with 1.0 mg/mL lysozyme and disrupted by either sonication (for Western blot analysis) or with a single pass through a microfluidizer (Microfluidics #110Y) at 20,000 psi (for MALDI-MS analysis). To isolate glycoproteins, clarified lysates (10 min at 18,000xg at 4^{o}C) were subject to affinity chromatography with a HiTrap ConA 4B column (GE Healthcare). Eluates were concentrated with 10-kDa MWCO columns (Sartorius Vivaspin20), diluted 40x in Ni-NTA buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), and again concentrated to 1 mL. His-tagged glycoproteins were then purified with a Qiagen Ni-NTA Kit per manufacturer's instructions. Fcs were then purified with a Nab Protein A/G Spin Kit (Thermo Scientific) per manufacturer's instructions. For MALDI-TOF MS, eluate was dialyzed in a Slyde-A-Lyzer 3500 MWCO cassette (Thermo Scientific) in 1.8 L deionized water (24 h, three volume replacements). Proteins from negative control cells expressing PglB$_{Cj}$mut were processed identically and no glycoprotein could be detected by ConA affinity, as expected (data not shown). Aglycosylated PglB$_{Cj}$mut control samples used in Western blot analysis were purified directly with a Qiagen Ni-NTA Kit or Nab Protein A/G Spin Kit per manufacturer's instructions.

Western blot analysis. Purified protein samples were separated using 12% SDS-PAGE gels and transferred to PVDF membranes. Proteins that harbored 6x-His affinity tags were detected with a monoclonal anti-polyhistidine-horse radish peroxidase (HRP) conjugate (Sigma), a monoclonal

anti-His (C-term) antibody (Invitrogen), or a monoclonal anti-FLAG® M2 antibodies (Stratagene), per manufacturers' instructions. Human Fcs were detected by anti-Human HRP conjugate antibodies (Promega). Protein-conjugated glycans were detected with 2.5 μg/mL ConA-HRP conjugate (Sigma). To remove *N*-linked glycans prior to Western blotting, purified glycoproteins were incubated with 5,000 units of PNGase F (New England BioLabs) at 37^{o}C for 1 hr. Samples that were not treated with PNGase F were handled identically except PNGase F was omitted from the solution.

Glycosylation characterization by mass spectrometry and NMR. Glycans were released from *E. coli* LLOs as described previously[9]. *N*-glycans were released from glycoproteins by treating 100 μg of purified scFv13-R4$^{4x\text{-}DQNAT}$ or scFv13-R4$^{1x\text{-}DQNAT}$ with trypsin followed by PNGase F digestion in 50 mM ammonium bicarbonate, pH 8. Glycans were collected as flow-through after SepPak C18 (Waters) separation. Some glycans were further digested with *C. ensiformis* α-exomannosidase in 50 mM ammonium acetate, pH 4.5, followed by further desalting steps using a carbon column (envi-carb SPE tube, Supelco). The mannosidase treatment was halted prior to completion to fully examine the glycan intermediates. Permethylation of glycans was performed as described[11]. Asparagine-linked glycans were analyzed as described elsewhere[12] with permethylation. MALDI-MS profiles were acquired using a 4800 MALDI-TOF/TOF (Applied Biosystems) in positive ion reflectron mode (500-4000 Da mass range). MALDI-MS/MS sequencing of permethylated glycans was performed in positive ion reflectron mode. Air was used as the collision gas. Structure assignment was according to the method of Yu *et al*[13]. All NMR analysis was performed at the Complex Carbohydrate Research Center (University of Georgia, Athens, GA). Briefly, 1-D Proton, TOCSY, NOESY and ROESY NMR spectra, run with water presaturation, and gradient enhanced COSY and HSQC spectra were acquired on a

Varian Inova-600 MHz spectrometer, equipped with a cryoprobe, at 21°C (for protein-released glycans) and at 25°C (for lipid-released glycans). Chemical shifts were measured relative to internal acetone (δ_H = 2.218 ppm, δ_C = 33.0 ppm).

Optimized PglB$_{Cj}$ gene.

5'_ATGCTGAAAAAAGAGTATCTGAAAAACCCTTATCTGGTGCTGTTCGCCATGATTG
TTCTGGCCTATGTGTTTAGCGTGTTCTGCCGCTTCTATTGGGTATGGTGGGCATCCGA
ATTCAACGAGTATTTCTTCAACAACCAGCTGATGATCATTAGCAATGACGGCTATGC
CTTTGCCGAAGGTGCTCGTGATATGATTGCCGGCTTTCACCAACCGAACGATCTGTC
CTATTACGGTAGCAGCCTGTCTACACTGACTTATTGGCTGTATAAAATCACCCCGTTC
TCATTTGAGAGCATTATCCTGTATATGTCCACGTTTCTGAGCAGTCTGGTAGTAATCC
CGATTATTCTGCTGGCCAACGAGTATAAACGCCCGCTGATGGGCTTTGTCGCCGCTC
TGCTGGCAAGTGTGGCAAATTCGTATTATAACCGTACCATGTCGGGCTATTACGATA
CCGATATGCTGGTCATCGTACTGCCGATGTTCATTCTGTTCTTTATGGTCCGTATGAT
TCTGAAAAAAGATTTCTTTAGCCTGATCGCTCTGCCTCTGTTTATTGGCATCTATCTG
TGGTGGTATCCGTCGTCGTATACCCTGAATGTTGCCCTGATTGGGCTGTTTCTGATCT
ATACGCTGATCTTCCACCGTAAAGAGAAAATTTTCTATATCGCCGTGATCCTGTCTA
GTCTGACCCTGAGCAATATTGCCTGGTTCTATCAGTCAGCCATCATCGTCATCCTGTT
TGCCCTGTTCGCTCTGGAACAAAAACGCCTGAACTTCATGATTATTGGCATCCTGGG
TAGCGCTACGCTGATCTTCCTGATTCTGTCTGGTGGTGTGGATCCTATTCTGTATCAA
CTGAAATTCTATATTTTCCGCTCCGATGAATCCGCTAACCTGACACAGGGGTTCATG
TATTTCAACGTCAACCAAACCATCCAAGAGGTCGAGAATGTCGATTTCTCCGAGTTT
ATGCGTCGCATTAGTGGCTCTGAGATTGTATTCCTGTTCTCACTGTTTGGGTTTGTGT
GGCTGCTGCGTAAACACAAATCAATGATTATGGCGCTGCCGATTCTGGTTCTGGGAT

TTCTGGCACTGAAAGGTGGTCTGCGCTTTACCATCTATAGCGTTCCGGTAATGGCAC

TGGGCTTTGGCTTTCTGCTGTCCGAGTTCAAAGCAATTCTGGTCAAAAAATATTCCC

AACTGACCTCGAATGTGTGTATTGTTTTCGCTACGATCCTGACGCTGGCACCTGTTTT

TATCCACATTTATAACTATAAAGCACCGACGGTCTTTTCCCAAAATGAAGCCTCACT

GCTGAATCAACTGAAAAACATTGCCAACCGTGAGGACTATGTGGTAACCTGGTGGG

ACTATGGTTATCCGGTTCGCTATTATTCCGACGTGAAAACCCTGGTTGATGGTGGTA

AACATCTGGGCAAAGACAACTTTTTCCCGAGCTTCTCACTGAGCAAAGATGAGCAG

GCAGCCGCTAACATGGCTCGTCTGAGCGTCGAGTATACCGAAAAAAGCTTCTATGCT

CCACAAAACGATATCCTGAAAAGTGACATCCTGCAGGCCATGATGAAAGACTATAA

CCAGAGCAACGTCGACCTGTTCCTGGCATCACTGAGTAAACCTGACTTCAAAATCGA

TACTCCAAAAACTCGTGACATTTATCTGTATATGCCGGCTCGTATGAGTCTGATCTTC

TCCACTGTTGCCTCGTTCTCGTTCATTAACCTGGATACGGGTGTTCTGGACAAACCGT

TCACCTTTTCAACCGCCTATCCGCTGGACGTGAAAAATGGCGAGATCTATCTGAGCA

ATGGCGTTGTGCTGTCGGACGATTTCCGTTCCTTCAAAATTGGGGACAACGTCGTGA

GCGTTAACAGTATTGTCGAGATTAACAGTATCAAACAAGGCGAGTACAAAATCACT

CCTATCGACGATAAAGCTCAATTCTATATCTTCTATCTGAAAGACTCCGCTATTCCGT

ATGCTCAATTCATTCTGATGGACAAAACGATGTTCAACTCCGCCTATGTCCAAATGT

TCTTCCTGGGCAACTATGACAAAAACCTGTTCGATCTGGTCATTAACAGTCGTGACG

CCAAAGTGTTCAAACTGAAAATCTGA_3’

Optimized PglB$_{Cl}$ gene.

5’_ATGAAACTGCAACAGAACTTTACCGATAACAACTCCATCAAATATACCTGTATCC

TGATCCTGATCGCCTTTGCCTTTAGTGTGCTGTGTCGCCTGTATTGGGTAGCATGGGC

ATCCGAATTCTATGAGTTTTTCTTCAACGACCAGCTGATGATTACCACCAACGATGG

TTATGCCTTCGCTGAGGGTGCTCGTGATATGATTGCCGGCTTCCACCAACCGAATGA

TCTGTCCTATTTCGGCAGCAGTCTGAGTACACTGACATATTGGCTGTATAGCATCCTG

CCTTTCTCGTTTGAAAGCATTATCCTGTATATGAGCGCCTTCTTTGCTTCTCTGATTGT

TGTCCCGATTATTCTGATCGCTCGTGAGTATAAACTGACCACCTATGGCTTCATTGCC

GCTCTGCTGGGTTCAATTGCTAACTCGTATTATAACCGTACCATGTCGGGCTACTATG

ACACTGATATGCTGGTTCTGGTTCTGCCAATGCTGATTCTGCTGACCTTTATTCGTCT

GACTATTAACAAAGACATCTTCACCCTGCTGCTGTCACCGGTTTTTATCATGATTTAT

CTGTGGTGGTATCCGTCCTCTTATAGCCTGAATTTCGCCATGATCGGGCTGTTTGGTC

TGTATACCCTGGTGTTCCACCGTAAAGAGAAAATCTTCTATCTGACGATCGCCCTGA

TGATTATTGCCCTGTCTATGCTGGCCTGGCAGTATAAACTGGCCCTGATTGTTCTGCT

GTTTGCCATCTTCGCCTTTAAAGAGGAAAAAATCAACTTCTATATGATCTGGGCACT

GATCTTCATTAGCATCCTGATCCTGCATCTGTCTGGAGGTCTGGATCCAGTACTGTAT

CAACTGAAATTCTATGTGTTCAAAGCCTCCGATGTTCAAAATCTGAAAGACGCCGCC

TTCATGTATTTTAACGTGAACGAAACCATTATGGAGGTGAATACCATTGATCCGGAA

GTCTTTATGCAGCGCATTAGTAGCAGTGTTCTGGTCTTTATCCTGAGCTTCATCGGGT

TCATTCTGCTGTGTAAAGATCACAAAAGCATGCTGCTGGCTCTGCCTATGCTGGCAC

TGGGTTTTATGGCTCTGCGTGCGGGTCTGCGTTTTACCATTTATGCCGTGCCTGTTAT

GGCTCTGGGTTTTGGCTATTTCCTGTATGCCTTTTTCAACTTCCTGGAAAAAAAACAA

ATCAAACTGAGCCTGCGCAACAAAAACATTCTGCTGATTCTGATTGCCTTCTTTAGC

ATTTCTCCGGCACTGATGCACATCTATTATTATAAAAGCTCCACCGTGTTTACCAGCT

ATGAGGCGTCAATTCTGAATGACCTGAAAAACAAAGCCCAGCGTGAGGATTATGTT

GTGGCATGGTGGGACTATGGATATCCGATTCGCTATTATTCCGACGTGAAAACCCTG

ATCGATGGCGGAAAACATCTGGGTAAAGACAACTTCTTCAGCAGCTTTGTGCTGAGT

AAAGAACAAATTCCGGCAGCGAATATGGCTCGTCTGAGCGTGGAGTATACCGAAAA
ATCTTTTAAAGAAAACTATCCGGACGTGCTGAAAGCCATGGTCAAAGACTATAACA
AAACCTCGGCGAAAGACTTTCTGGAGTCCCTGAACGACAAAGATTTCAAATTCGAC
ACCAACAAAACACGTGACGTGTATATCTATATGCCGTATCGTATGCTGCGTATCATG
CCTGTAGTAGCCCAATTTGCCAACACGAACCCTGACAATGGGGAGCAAGAGAAATC
GCTGTTCTTCTCACAGGCAAACGCCATTGCCCAAGACAAAACGACCGGTAGCGTTAT
GCTGGATAACGGTGTGGAGATTATCAACGACTTTCGTGCCCTGAAAGTGGAGGGAG
CCTCTATTCCACTGAAAGCGTTCGTCGACATTGAGAGTATCACCAATGGAAAATTCT
ATTATAACGAAATCGACTCGAAAGCCCAAATCTATCTGCTGTTCCTGCGCGAGTATA
AAAGTTTCGTGATCCTGGATGAAAGCCTGTATAACAGCTCGTATATTCAAATGTTCC
TGCTGAACCAGTATGACCAGGACCTGTTTGAGCAAATCACTAACGACACCCGTGCC
AAAATCTATCGCCTGAAACGCTGA_3’

References Cited

1. Baba, T. et al. *Mol Syst Biol* **2**, 2006 0008 (2006).
2. Datsenko, K.A. & Wanner, B.L. *Proc Natl Acad Sci U S A* **97**, 6640-5 (2000).
3. O'Reilly, M.K., Zhang, G. & Imperiali, B. *Biochemistry* **45**, 9593-603 (2006).
4. Fisher, A.C. et al. *Appl Environ Microbiol* **77**, 871-81 (2011).
5. Shanks, R.M., Caiazza, N.C., Hinsa, S.M., Toutain, C.M. & O'Toole, G.A. *Appl Environ Microbiol* **72**, 5027-36 (2006).
6. Wacker, M. et al. *Science* **298**, 1790-3 (2002).
7. Kowarik, M. et al. *Science* **314**, 1148-50 (2006).
8. Marani, P. et al. *Protein Sci* **15**, 884-9 (2006).

9. Gao, N. & Lehrman, M.A. *Methods Enzymol* **415**, 3-20 (2006).

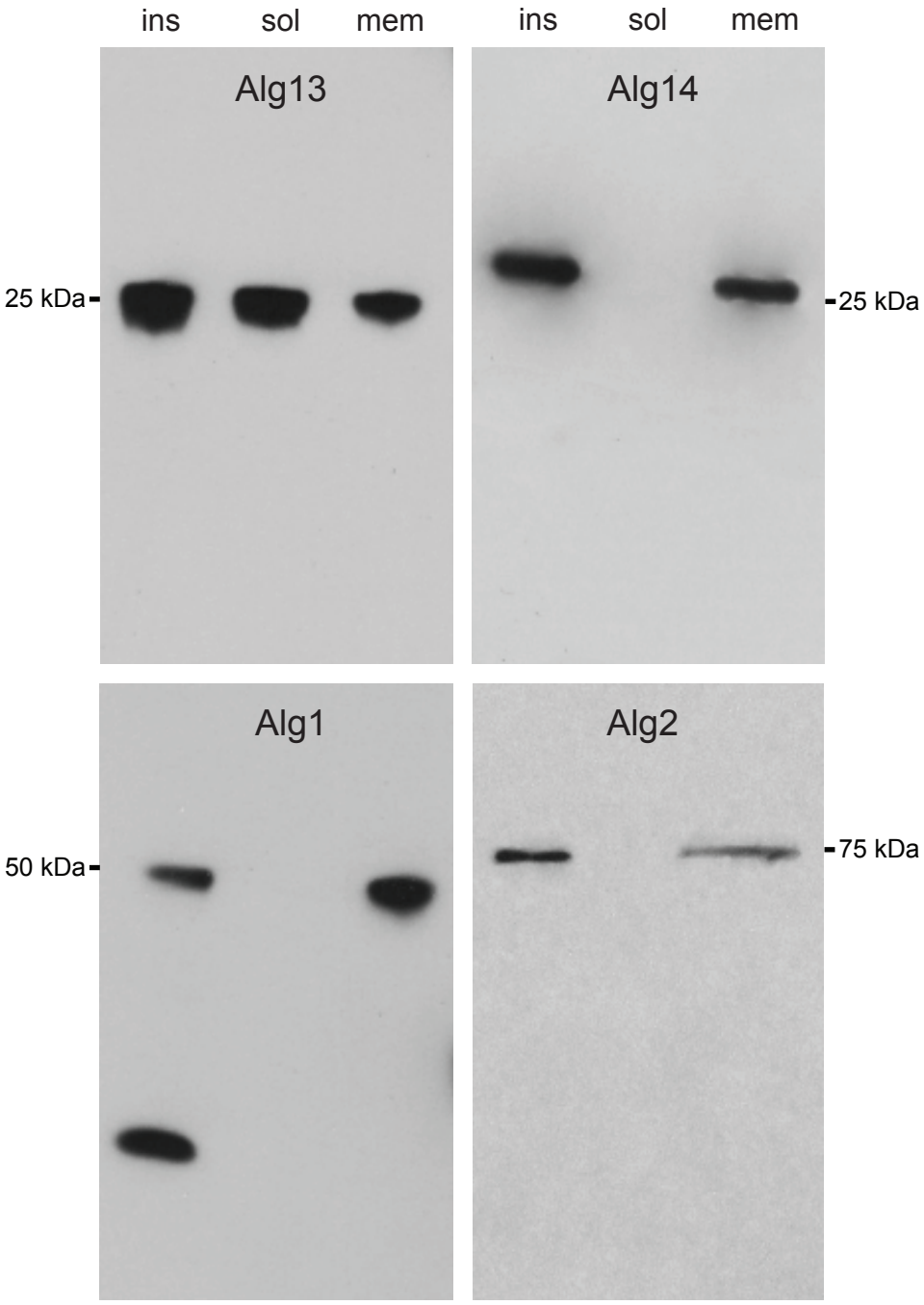
10. Folch, J., Lees, M. & Stanley, G.H.S. *J Biol Chem* **226**, 497-509 (1957).

11. Jang-Lee, J. et al. *Methods Enzymol* **415**, 59-86 (2006).

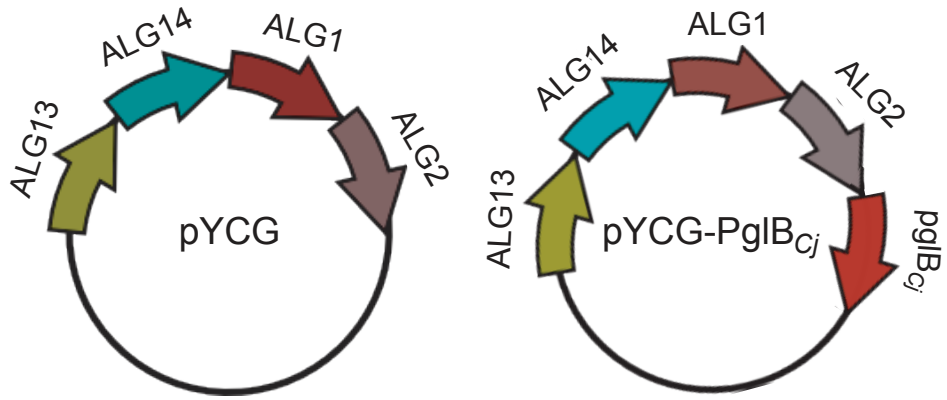
12. Liu, X. et al. *Anal Chem* **78**, 6081-7 (2006).

13. Yu, S.Y., Wu, S.W. & Khoo, K.H. *Glycoconj J* **23**, 355-69 (2006).

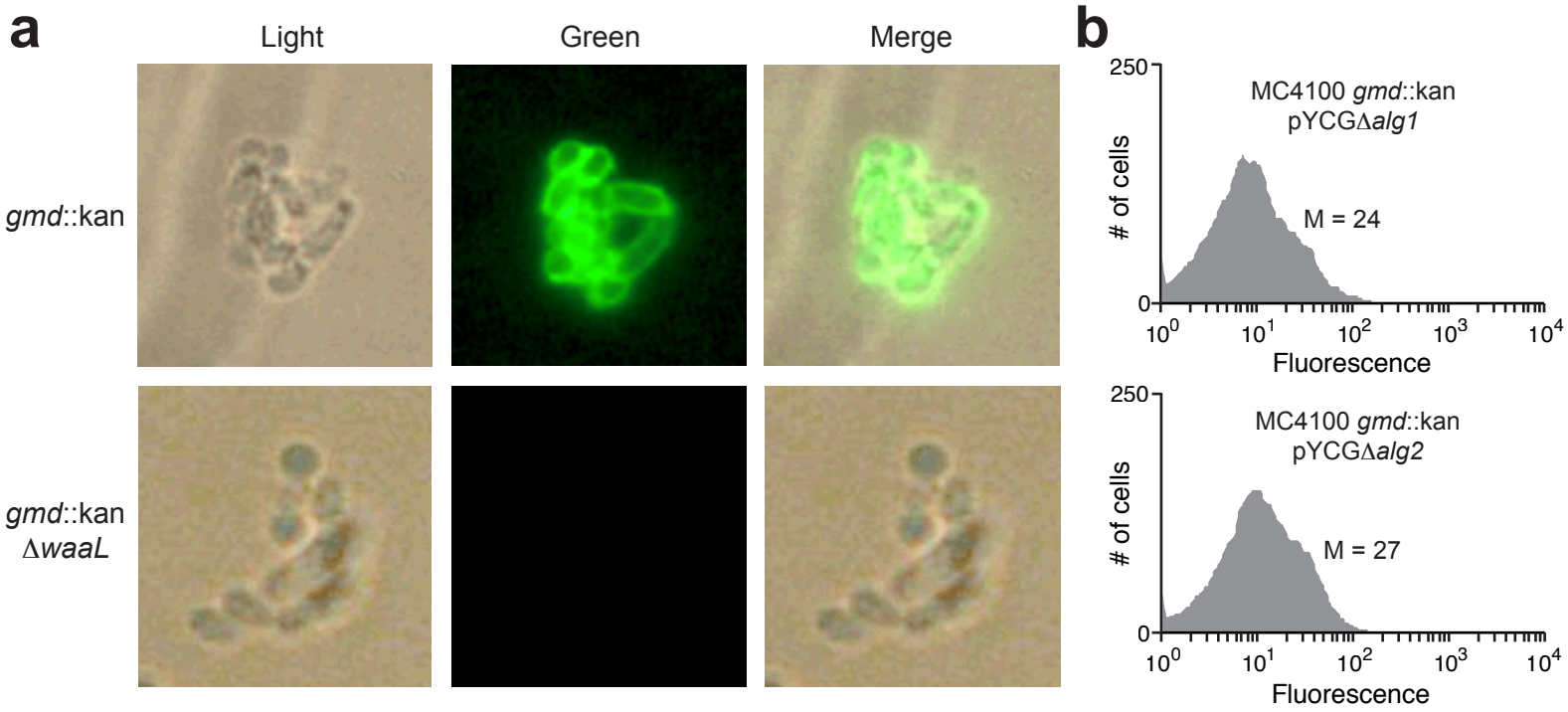
Supplementary Results



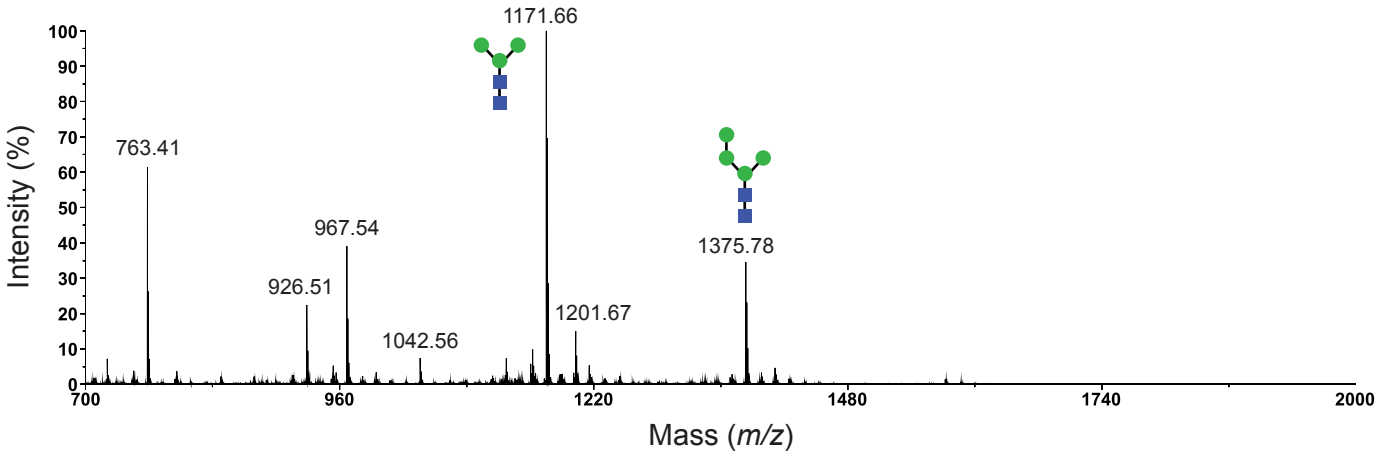
Supplementary Figure 1. Western blot analysis of insoluble (ins), soluble (sol) and membrane (mem) fractions isolated from MC4100 *gmd*::kan cells carrying pYCG. All proteins migrated according to their predicted molecular weights. Alg14 was detected using anti-FLAG antibodies; the rest were detected using anti-His antibodies as described in the Supplementary Methods.



Supplementary Figure 2. Plasmid pYCG for the biosynthesis of a trimannosyl core glycan (left) and pYCG-PglB$_{Cj}$ for producing a trimannosyl core glycan with the *C. jejuni* OTase PglB (right) were constructed in plasmids pMQ70 and pMW07, respectively.



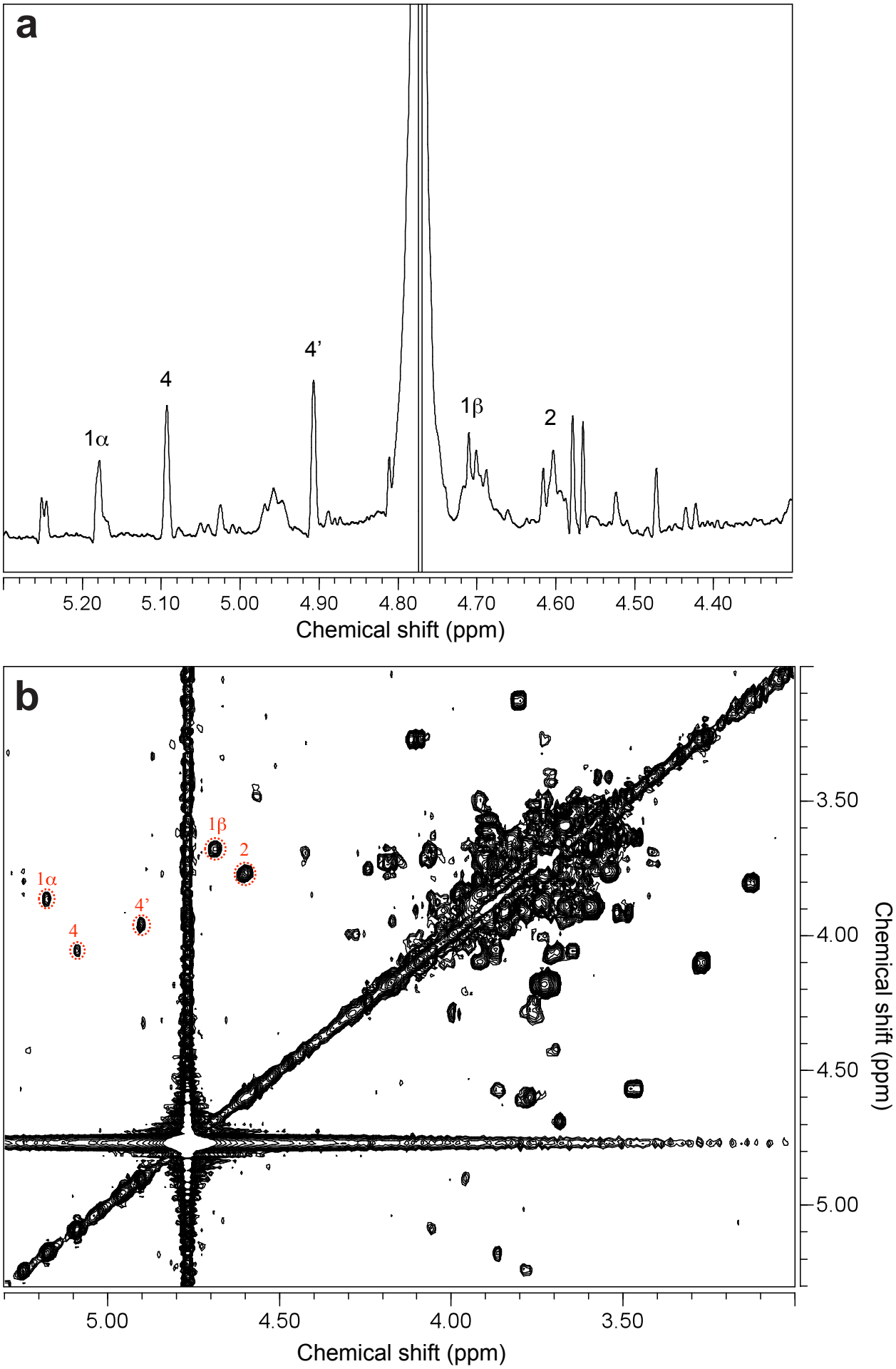
Supplementary Figure 3. (a) ConA-AlexaFluor labeling of MC4100 *gmd*::kan (top panels) or MC4100 *gmd*::kan Δ*waaL* (bottom panels) cells each carrying plasmid pYCG. Cells were visualized by light and fluorescence microscopy as described in the Supplementary Methods. Panels show phase contrast microscopy (left), fluorescence microscopy using green emission filter (center) and merge (right). **(b)** Flow cytometric analysis of *E. coli* MC4100 *gmd*::kan cells carrying pYCGΔ*alg1* or pYCGΔ*alg2* plasmids as indicated. Cells were labeled with ConA-AlexaFluor prior to flow cytometry. Median cell fluorescence (M) values are given for each histogram.



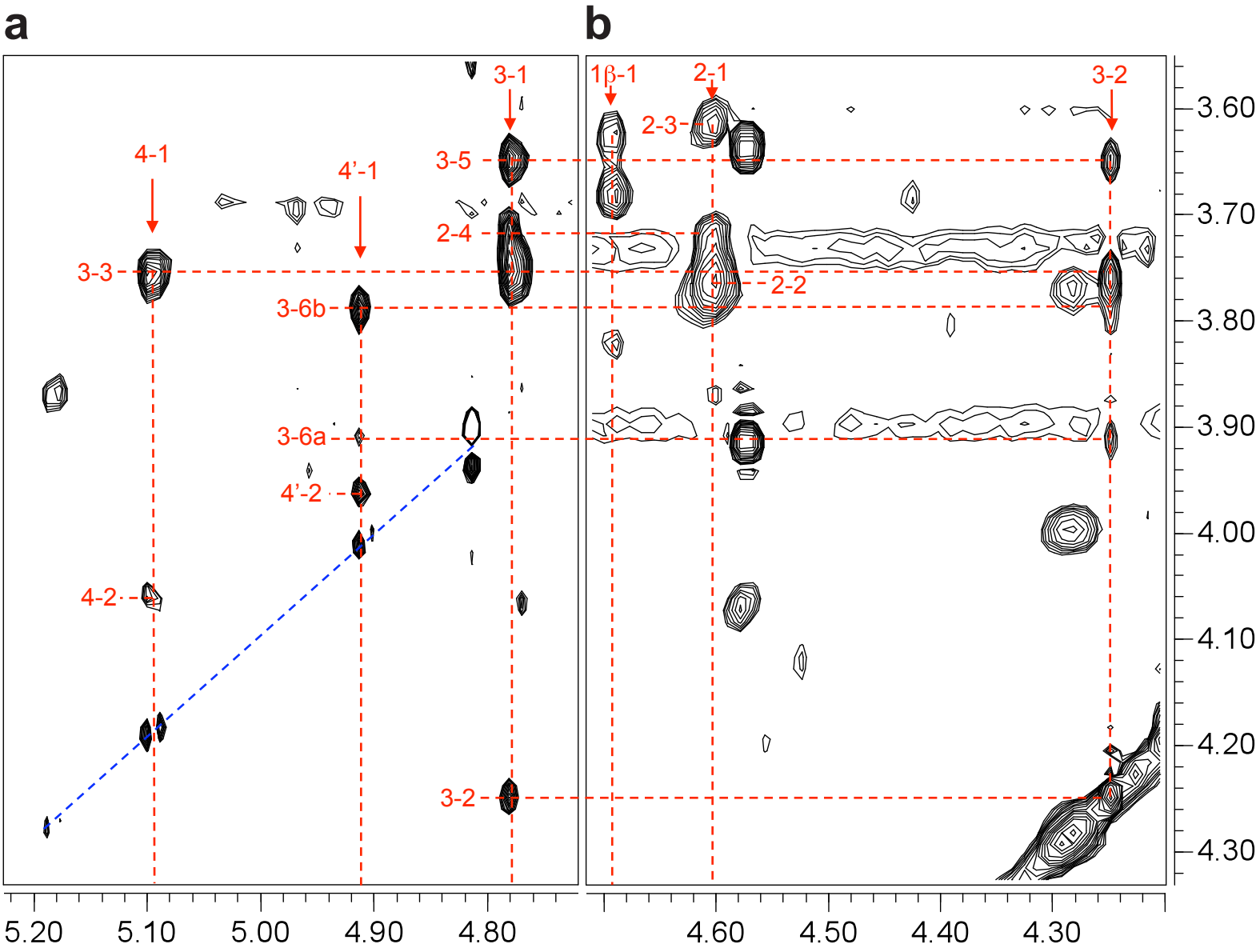
Supplementary Figure 4. MALDI-MS profile of permethylated glycans released from LLOs by acid hydrolysis. LLOs were extracted from *E. coli* MC4100 *gmd*::kan Δ*waaL* cells carrying plasmid pYCG. The major signal at *m/z* 1171 corresponds to $[M+Na]^+$ of $Hex_3HexNAc_2$.

No.	Residue	Position	Chemical Shift (ppm)						
			1	2	3	4	5	6a/6b	NOE
I	4-α-GlcNAc	1	5.180	3.87	3.64	n.d.	n.d.	n.d.	
			93.1	56.2	n.d.	n.d.	n.d.	n.d.	
II	α-Man	4	5.093	4.05	3.87	3.79	n.d.	n.d.	VI-3
			105.2	72.5	71.7	n.d.	n.d.	n.d.	
V	α-Man	4’	4.907	3.97	3.85	3.64	3.79	n.d.	VI-6a,6b
			102.2	72.5	71.7	n.d.	n.d.	n.d.	
VI	3,6-β-Man	3	4.773	4.25	3.76	3.75	3.65	3.92/3.80	VIII-4
			102.6	72.5	82.8	68.2	76.8	68.7	
VII	4-β-GlcNAc	1	4.70	3.68	3.63	3.65	3.51	n.d.	
			n.d.	n.d.	75.1	n.d.	n.d.	n.d.	
VIII	4-β-GlcNAc	2	4.60	3.77	3.62	3.72	n.d.	n.d.	VII-4
			n.d.	57.5	n.d.	n.d.	n.d.	n.d.	

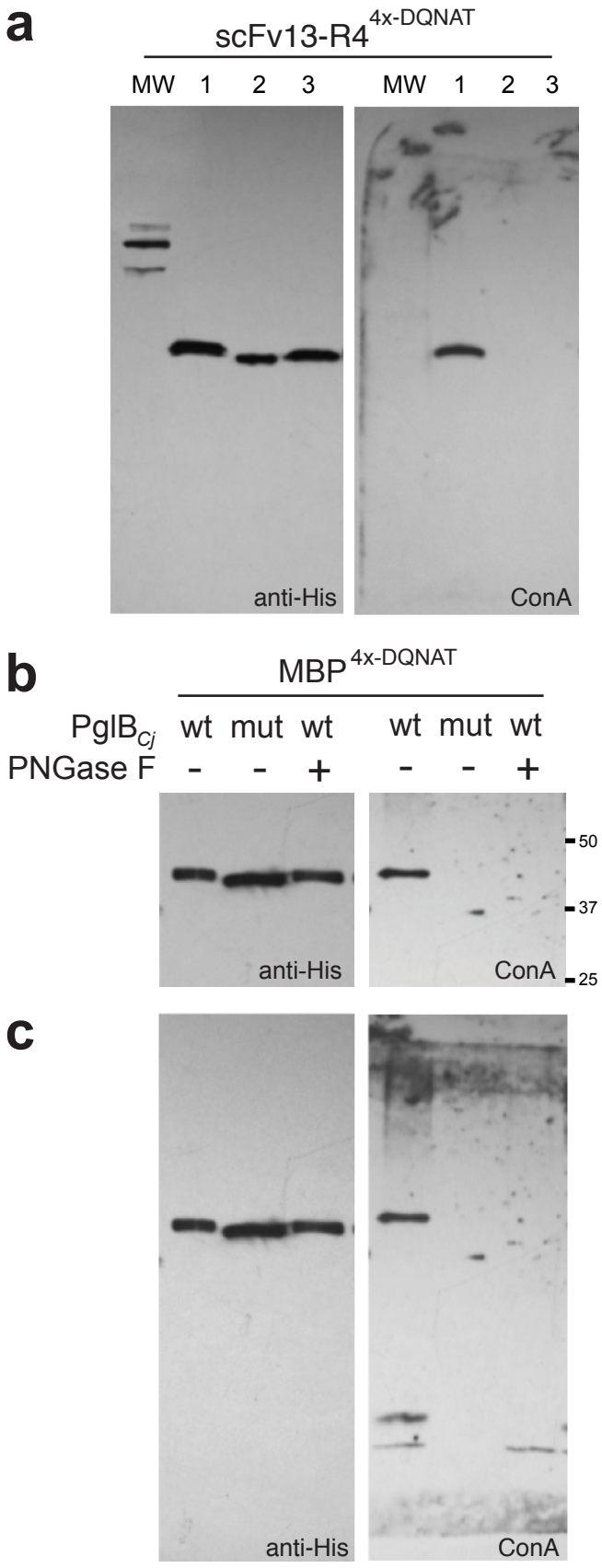
Supplementary Figure 5. Partial assignment of the NMR signals belonging to $Man_3GlcNAc_2$ from lipid-released glycans. The chemical shifts obtained from the 2-D spectra were consistent with the main component of the sample being $Man_3GlcNAc_2$. This sample was available in sufficient quantity to allow acquisition of a 2-D HSQC spectrum, which confirmed the proton peak assignments.



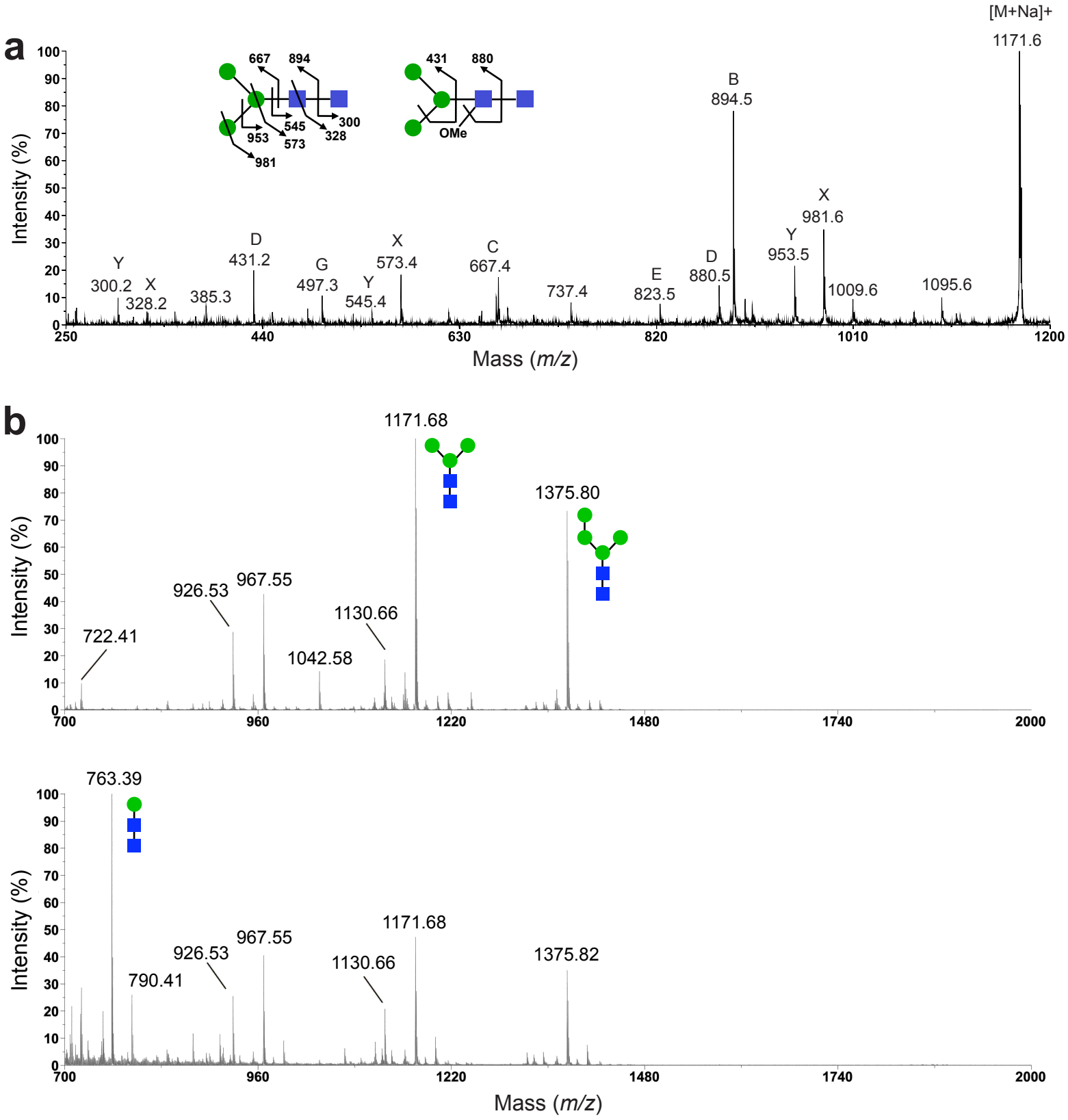
Supplementary Figure 6. NMR analysis of glycans released from extracted LLOs. **(a)** Anomeric region of the resolution-enhanced 1-D proton spectrum acquired at 25°C. **(b)** Partial 2-D gCOSY spectrum acquired at 25°C.



Supplementary Figure 7. NMR analysis of glycans released from extracted LLOs. **(a)** Partial 2-D ROESY spectrum acquired at 25°C. **(b)** Partial 2-D TOCSY spectrum acquired at 25°C. The dotted red lines indicate the relevant correlations from the anomeric proton, except for the β-mannose residue, which, due to the small H1-H2 coupling constant shows only one cross peak from H-1. Therefore the correlations from H-2 were used. The NOE correlations demonstrate the linkages between the three mannose residues in $Man_3GlcNAc_2$. The signals marked by the blue line are a cancellation artifact from the ROESY experiment.



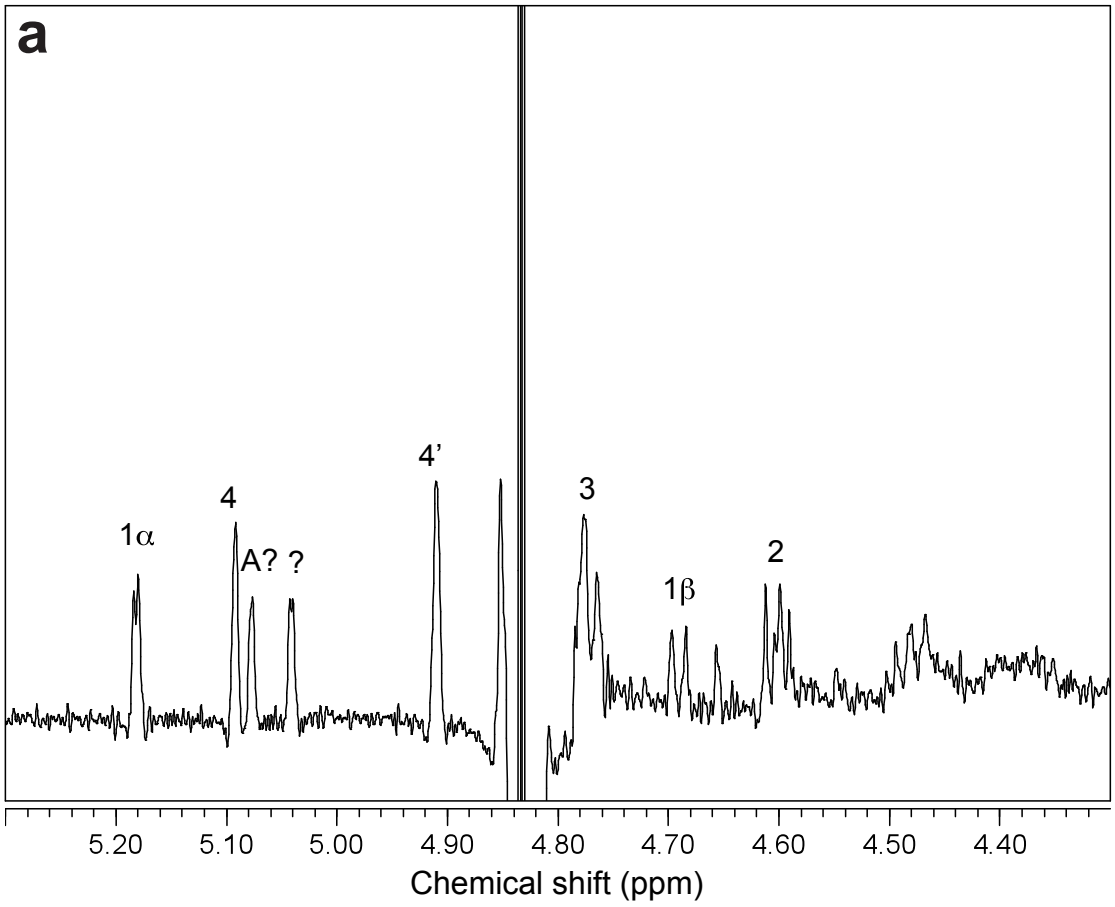
Supplementary Figure 8. (a) Full, uncut image of Western blot shown in Fig. 3a in the main text. Lanes 1-3 correspond exactly to the same lanes in Fig. 3a. MW, molecular weight standard. **(b)** Western blot analysis of MBP$^{4x\text{-}DQNAT}$ affinity purified from *E. coli* MC4100 *gmd*::kan Δ*waaL* cells carrying pYCG-PglB$_{Cj}$ or pYCG-PglB$_{Cj}$mut as indicated. Proteins isolated from cells expressing wild-type PglB$_{Cj}$ were further treated with PNGase F for removal of *N*-linked glycans. Polyhistidine tags on the proteins were detected using anti-His antibodies while mannose glycans on the proteins were detected using ConA. **(c)** Full, uncut image of Western blot shown in (b).

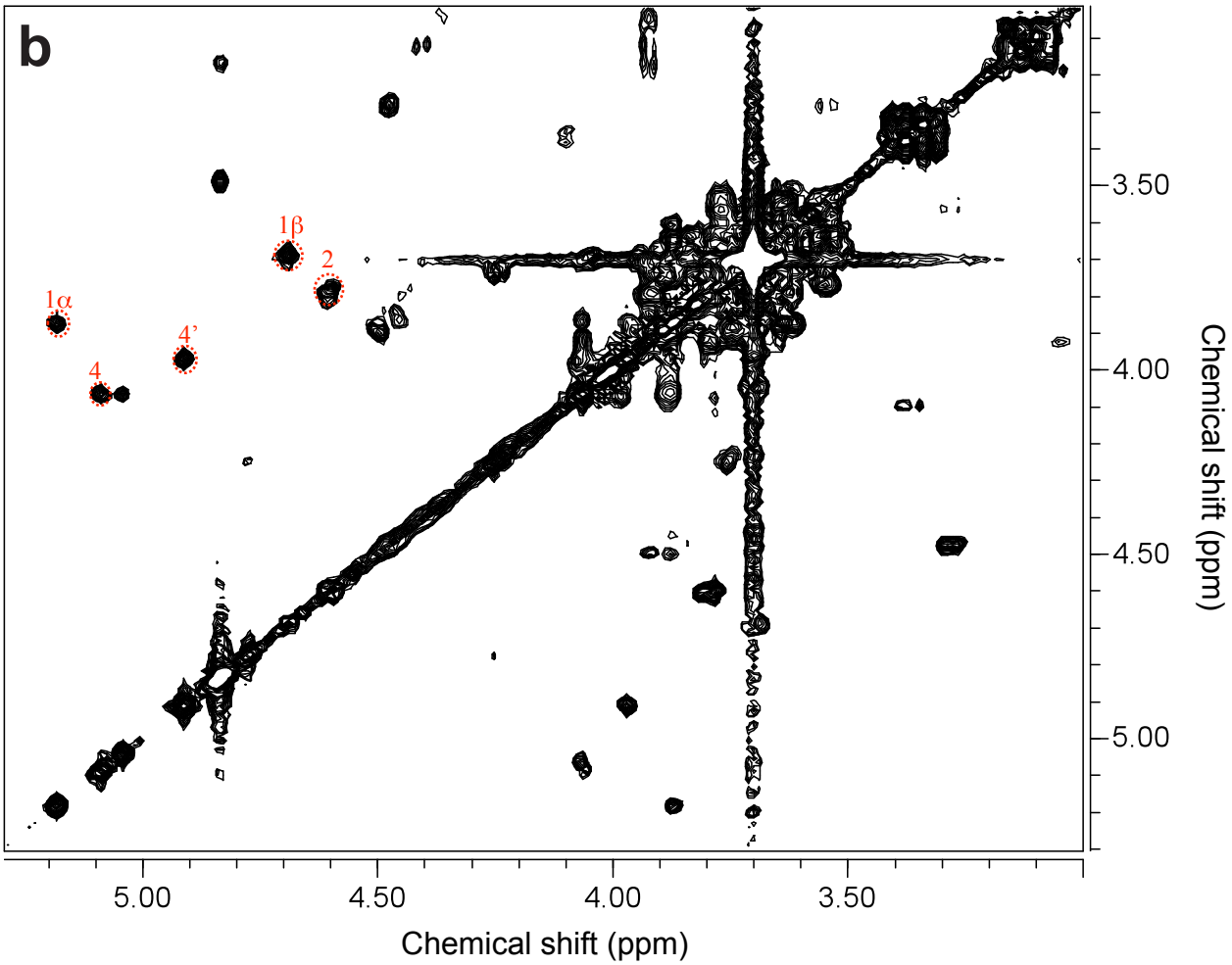


Supplementary Figure 9. (a) MALDI-TOF/TOF MS^2 sequencing of *N*-glycans at *m/z* 1171. X/Y ion series at *m/z* 300/328, 545/573, 953/981, coupled with B/Y ion pair at *m/z* 894/300, showed the presence of a biantennary trihexosyl-core *N*-glycan structure. D ions at *m/z* 431 and 880 further confirmed the structure. **(b)** MALDI-MS profile of glycans released from scFv13-$R4^{1x\text{-}DQNAT}$ by PNGase F (top panel) and partially digested with *C. ensiformis* α-exomannosidase (bottom panel) to determine the identity of terminal hexose units.

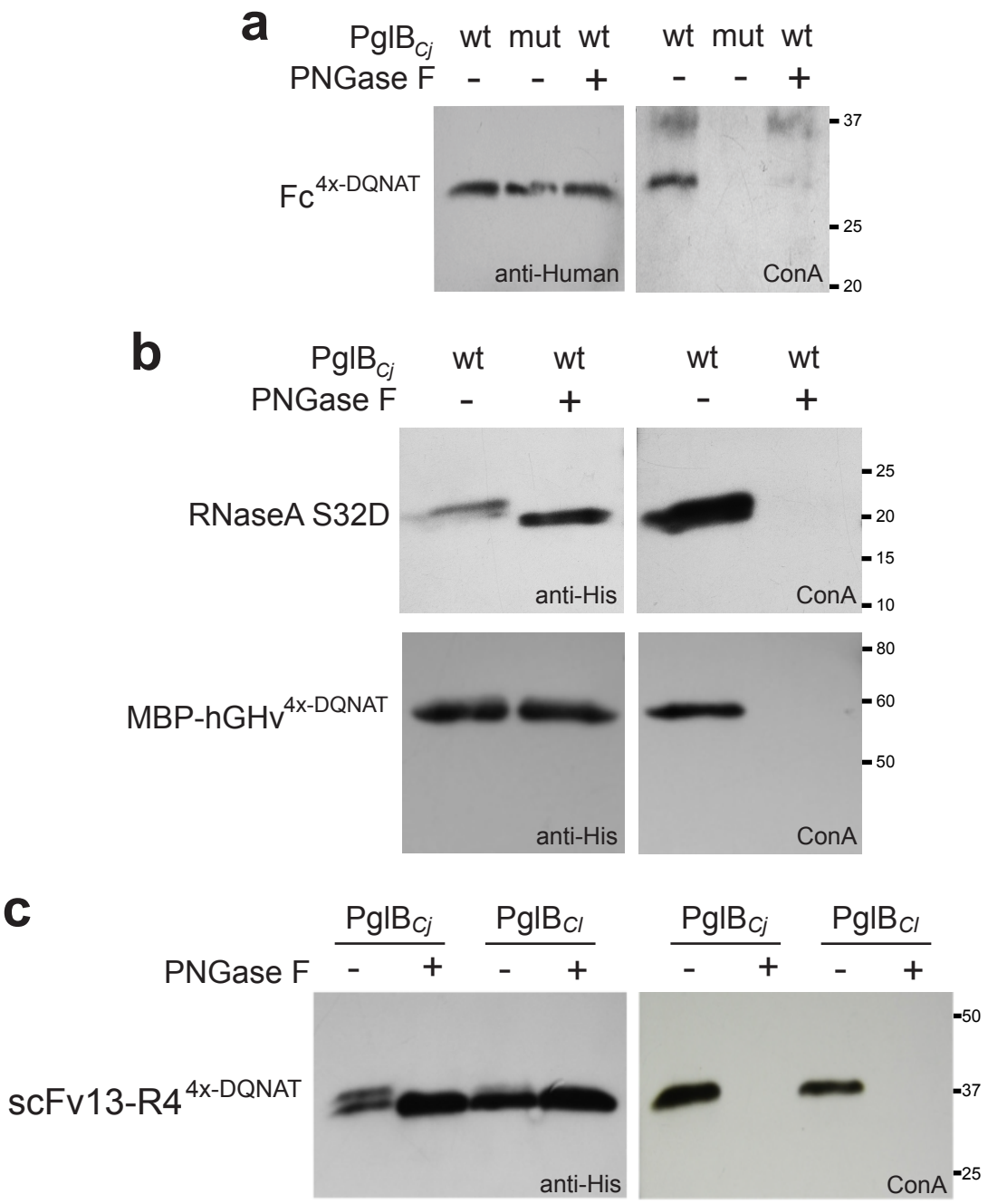
No.	Residue	Position	Chemical Shift (ppm)						NOE
			1	2	3	4	5	6	
I	4-α-GlcNAc	1	5.183	3.88	3.65	n.d.	n.d.	n.d.	
II	α-Man	4	5.094	4.059	3.89	n.d.	n.d.	n.d.	VI-3
III	?	?	5.080	4.065	n.d.	n.d.	n.d.	n.d.	
IV	?	?	5.042	4.065	n.d.	n.d.	n.d.	n.d.	
V	α-Man	4’	4.910	3.97	3.89	n.d.	n.d.	n.d.	
VI	3,6-β-Man	3	4.777	4.25	3.75	n.d.	n.d.	n.d.	
VII	4-β-GlcNAc	1	4.692	3.69	n.d.	n.d.	n.d.	n.d.	
VIII	4-β-GlcNAc	2	4.602	3.79	3.62	n.d.	n.d.	n.d.	
IX	Xyl	*	4.473	3.29	3.55	3.78	4.10/3.37		

Supplementary Figure 10. Partial assignment of the NMR signals belonging to $Man_3GlcNAc_2$ from protein-released glycans. The chemical shifts obtained from the 2-D spectra were consistent with the main component of the sample being $Man_3GlcNAc_2$. This sample was available in sufficient quantity to allow acquisition of a 2-D HSQC spectrum, which confirmed the proton peak assignments. Xylose was determined to be a contaminant as it was not detected by MS analysis or by NMR of glycans released from lipids. Xylose is a common contaminant of bacterial samples.





Supplementary Figure 11. NMR analysis of *N*-linked glycans released from purified scFv13-R4$^{4x\text{-}DQNAT}$. **(a)** Anomeric region of the resolution-enhanced 1-D proton spectrum acquired at 21°C. **(b)** Partial 2-D gCOSY spectrum acquired at 21°C.



Supplementary Figure 12. (a) Western blot analysis of Fc^{DQNAT} affinity purified from *E. coli* MC4100 *gmd*::kan Δ*waaL* cells carrying pYCG-PglB$_{Cj}$ or pYCG-PglB$_{Cj}$mut as indicated. Fc^{DQNAT} was detected using anti-Human antibodies. **(b)** Western blot analysis of RNaseA S32D and MBP-hHGv^{DQNAT} affinity purified from *E. coli* MC4100 *gmd*::kan Δ*waaL* cells carrying pYCG-PglB$_{Cj}$. RNaseA S32D and MBP-hHGv^{DQNAT} were each detected using anti-His antibodies. **(c)** Western blot analysis of scFv13-R4$^{4x\text{-}DQNAT}$ affinity purified from *E. coli* MC4100 *gmd*::kan Δ*waaL* cells carrying pYCG-PglB$_{Cj}$ for expressing *C. jejuni* PglB or pYCG-PglB$_{Cl}$ for expressing *C. lari* PglB. scFv13-R4$^{4x\text{-}DQNAT}$ was detected using anti-His antibodies. Mannose glycans on all proteins were detected using ConA lectin. All proteins were loaded directly or further treated with PNGase F for removal of *N*-linked glycans.