Supplementary Information

In Vitro Bacterial Polysaccharide Biosynthesis: Defining the Functions of Wzy and Wzz

Robert Woodward[†‡], Wen Yi[†‡,1], Lei Li[†#‡], Guohui Zhao[†§‡], Hironobu Eguchi[†‡,2], Ramu Sridhar Perali[†,3], Hongjie Guo[†,4], Jing Katherine Song[†,5], Edwin Motari[†], Li Cai[†], Patrick Kelleher[†], Xianwei Liu[†#], Weiqing Han[†], Wenpeng Zhang[†], Yan Ding[#], Mei Li[†,6]and Peng George Wang[†*]

† Departments of Biochemistry and Chemistry, The Ohio State University, Columbus, OH 43210, USA.

[#] National Glycoengineering Research Center and The State Key Laboratory of Microbial Technology, School of Life Sciences, Shandong University, Shandong 250100, China.

§ College of Pharmacy, Nankai University, Tianjin 300071, China.

‡ These authors contributed equally to this work.

1. Present Address: Division of Chemistry and Chemical Engineering, Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125

2. Present Address: Department of Biochemistry, Hyogo College of Medicine, Nishinomiya, Hyogo 663-8501, Japan.

3**.** Present Address: School of Chemistry, University of Hyderabad, Hyderabad 500046, India

4. Present Address: Molecular Microbiology in School of Medicine, Washington University, St.Louis, MO 63110.

5. Present Address: Center for Microbial Pathogenesis, The Research Institute at Nationwide Children's Hospital, Columbus OH 43205.

6. Present Address: Department of Microbiology, University of Alabama, Birmingham, AL 35294.

* To whom correspondence should be addressed: Peng George Wang, Phone: (614) 292-9884. Fax: (614) 688-3106. E-mail: wang.892@osu.edu.

Supplementary Methods:

Construct of *E. coli* O86 (Δ*wec*A, Δ*waa*L) strain

waaL gene knockout primers:
Forward: 5'-
GCAGTTTTGGAAAAGTTATCATCATTATAAAGGTAAAACAT*GTGTAGGCTGGA GCTGCTTCG-3'*
Reverse :5'-
AGTGAGTTTTAACTCACTTCTTAAACTTGTTTATTCTTAA*CATATGAATATCCTC CTTAG-3'*

waaL gene knockout verification primers:
Forward: 5'- TGAAACCTTACACTCTGAAATCATC -3'
Reverse : 5'- CGATTATTAATGACGAGTAAGAGGA -3'

wecA gene knockout primers:
Forward: 5-
ACAGGCCGCTGTTGGCATTGTTATGATGGTGTTCGGCAAGGTGTAGGCTGGA GCTGCTTC-3
Reverse: 5-
TGCAATATCCATAGAGGAAGAATGCTAGCAAAAAGAGCACCATATGAATATC CTCCTTAG-3

*wec*A gene knockout verification primers:
Forward: 5'-GTGAATTTACTGACAGTGAGTACTG-3'
Reverse: 5'-TTATTTGGTT AAATTGGGGC TGCCA-3'

The two genes were knocked out according to a previously established procedure[1]. Briefly, pKD46 and pCP20 plasmids were purified. pKD46 was transferred by electroporation into *E.coli* O86:B7 competent cells. The Cat gene was amplified from pKD3 (or the Kan gene from pKD4) using the corresponding primer above depending on which gene was desired to be knocked out. DpnI digestion was performed for 1 h, after which the amplified gene was purified using a gel-extration kit. 10 μL of pure amplified gene was transferred by electroporation into *E.coli* O86:B7 strain containing pKD46. After the knockout was verified by PCR using the corresponding verification primers, pKD46 was removed by incubation at 42 ^{o}C overnight. Furthermore, for removing the Chloramphenicol resistance, the pCP20 plasmid was transformed into *E.coli* O86:B7 mutant and spread onto an LB plate with Chloramphenicol (34 μg/mL) and Ampicillin (100 μg/mL). The resultant single colonies were then incubated at 42 ^{o}C overnight and screened on LB plates with and without Chloramphenicol. Knockout of the other gene followed the same procedure.

Optimized wzy gene Sequence Length:1367

ATGGTGATTAGTCGTAGCAACTATCGCAATATCTGCTCTTACACCTTTTTCAT
GGTGAACCTGTTCATCCTGATCCTGAGTGTTATCAACGAAGGTTTCTGTGAAA
TCGCGTATGTGATTATCAGCGTGAGCTCTGTTCTGTTTTGCGTTATTATCATTT
GTCTGGAACGTCAGGGCGGTTTCCTGAACCCGATGACCTTCTGCATCATCAGT
GTGTTTTTCTTTATCCTGATCCGCCCGGTTTTCTTTAGCCAGAATATTACCGAA
AACCTGAATGAAGTGATCACGGCCGGCCTGGAAATCGATGAAATCTACGTTT
TCTACTCTCTGGCCGTGGTTAACATTAGTCTGGCATTTACCGTGCTGCTGTAT
AGCGTTCAGAAGGGTACGGTGTCTAAACTGGTTGGCCAGCTGCCGGATCTGT
TCTTTTACAATAAACAGCTGAGCATGATTCTGCTGTGGGGCGGTGGCCTTTCA
GTGCGATCTTTCTGATCAAAAGCTACAAAAAATTCATCATTCTGGGTCAGGTG
TCTGTTTTTGAAGCGGATGCCTACGGCCTGCATGATGAACTGTTCTGTTTACC
CTGAGCAAATATTGTTACATTCTGAGCCTGCTGTTCTCTAAAAACAAAAACTT
CATCCTGTACTCTCTCTGATCTTTATTACCAGTATCGGTTACATTCTGGTGGGT
CTGCGTGGCTATACGATTGCATACGGCTTTCTGCTGCTGTTCTTTCTGGATATT
CGTTATCGCCTGAAAATCAAATGGCTGCTGCTGGTTGCGATTCTGGTTACCAC
GATCAGTAGCCTGATTCTGAACTATCGCATCGGCATTGAAGTGAATAGCGGT
CTGCTGGGCATCATTTTCAACCCGCTGCTGCAGCAGGGTGCGTCTTTTGAAAC
CGTTTATGGCGCCCTGAAATACAACGAAAAAATCCTGTCTTGCATCAGTTACT
ACGATTATTTCTTTACGAACAAAGATATCGGTAGCTGTATCGATATCGCCCGT
GGCGTGTACTTTAAAGAAGGTGGCAGCTTCGCATCTAGTTTTTACTCTGAACT
GATCTACTTTGGTTGGATCATTGGCTCTGTGGCCCTGCTGCTGTTCGCATTTAG
TCTGGCGTTTGTTCAGAGCTGCTACGAAAAAATCATCAAAAACAGCATGAAC
AACAAACTGCATATACCTACCGTCTGATCATTTTCCTGGCACTGCCGAATCTG
ATTTATTTTGCGCGCAGCTCTCTGTTCGATTTTATCACGAAAGTGCTGTTCATC
GCGCTGTTTATTGGTGGCCTGAGCATCGTTCGCCACATTGCCCTGAATATCAA
AAAATGTCATCACCACCATCATCACCATCACCATCACTAA

Identification of Wzy through in-gel digestion and mass spectrometry

Bands were trimmed as close as possible to minimize background polyacrylamide material. Bands were then washed in 50% methanol/5% acetic acid for one hour. The wash step was repeated once more before the bands were dehydrated in acetonitrile. Rehydration of bands followed by incubation with a DTT solution (5mg/ml in 100 mM ammonium bicarbonate) for 30 min was carried out prior to the addition of 15mg/ml Iodoacetamide in 100 mM ammonium bicarbonate solution. Iodoacetamide was incubated with the bands in the dark for 30 min, after which it was removed. Bands were washed with acetonitrile and ammonium bicarbonate (100mM) in a cyclical fashion for 5 min in each solution. After drying in a speed vac, the protease is driven into the bands by rehydrating them in 50 μL of sequencing grade modified trypsin or chymotrypsin at 20 μg/mL in 50 mM ammonium bicarbonate solution for 10 min. 20 μL of 50 mM ammonium bicarbonate was then added and the mixture was incubated overnight at room temperature. The peptides were extracted from the polyacrylamide using 50% acetonitrile and 5% formic acid. After several extractions, the extract was pooled together and concentrated in a speed vac to ~30 μL.

The above gel digestion was performed with sequencing grade trypsin from Promega (Madison, WI) or sequencing grade chymotrypsin from Roche (Indianapolis, IN) using the Multiscreen Solvinert Filter Plates from Millipore (Bedford, MA).

Capillary-liquid chromatography-nanospray tandem mass spectrometry (Nano-LC/MS/MS) was then performed on a Thermo Finnigan LTQ mass spectrometer equipped with a nanospray source operated in positive ion mode. The LC system was an UltiMate™ 3000 system from Dionex (Sunnyvale, CA). Solvent A was water containing 50 mM acetic acid, while solvent B was acetonitrile. 5 μL of each sample was first injected on to the μ-Precolumn Cartridge (Dionex, Sunnyvale, CA), and washed with 50 mM acetic acid. The injector port was switched to inject and the peptides were eluted off of the trap onto the column. A 5 cm 75 μm ID ProteoPep II C18 column (New Objective, Inc. Woburn, MA) packed directly in the nanospray tip was used for chromatographic separations. Peptides were eluted directly off the column into the LTQ system using a gradient of 2-80% solvent B over 45 minutes with a flow rate of 300 nL/min. The total run time was 65 minutes. The MS/MS was acquired according to standard conditions established in the lab. Briefly, a nanospray source operated with a spray voltage of 3 kV and a capillary temperature of 200 ^{o}C is used. The scan sequence of the mass spectrometer was based on the TopTen™ method. The analysis was programmed for a full scan recorded between 350 – 2000 Da and a MS/MS scan. This generated product ion spectra for determination of the amino acid sequence of the ten most abundant peaks in the spectrum over consecutive instrument scans. The CID fragmentation energy was set to 35%. Dynamic exclusion was enabled with a repeat count of 30 s, exclusion duration of 350 s, a low mass width of 0.5 Da and high mass width of 1.50 Da. Sequence information from the MS/MS data was processed by converting the raw dta files into a merged file (.mgf) using an in-house program, RAW2MZXML_n_MGF_batch (merge.pl, a Perl script). The resulting mgf files were searched using Mascot Daemon by Matrix Science version 2.2.1 (Boston, MA), and the database was searched against the full NCBI database (9873339 sequences; 3367482728 residues). The mass accuracy of the precursor ions was set to 2.0 Da given that the data was acquired on an ion trap mass analyzer where the fragment mass accuracy was set to 0.5 Da. Considered modifications (variable) were methionine oxidation and carbamidomethyl cysteine. Two missed cleavages for the enzyme were permitted. Peptides with a score less than 20 were filtered and proteins identified required bold red peptides. Protein identifications were checked manually, and proteins with a Mascot score of 50 or higher with a minimum of two unique peptides from one protein having a -b or -y ion sequence tag of five residues or better were accepted.

Cloning, expression and purification of Wzz from E. coli O86:B7 and O86:H2

The cloning, expression and purification of Wzz from *E. coli* O86:H2 (Wzz_{H2}) has been previously described[2]. The *E. coli* O86:B7 Wzz gene (Wzz_{B7}) was amplified by PCR from the *E. coli* O86:B7 genome. The primers with restriction sites underlined for amplifying the gene were as follows: Forward: 5' CATGCCATGGGGACTTGGAAATTTCC *(NcoI)*, Reverse: 5' CTGAAGCTTCTTCGCGTTGTAATT (*HindIII*). The PCR product was digested with *Nco*I and *Hind*III and subsequently cloned into a pBAD/*Myc*-His A vector digested with the corresponding restriction enzymes. The constructs obtained were checked by

sequencing and subsequently transformed into the *wzz* mutant *E. coli* O86:H7 strain with ampicillin and chloramphenicol selection. The expression and purification of Wzz_{B7} followed published procedures[2].

Analysis of LPS

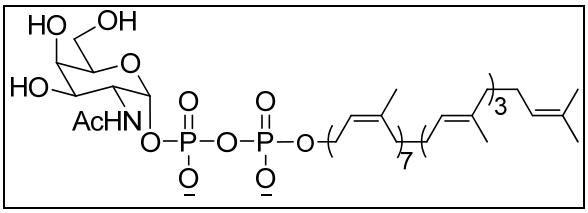
E.coli O86:B7, *E.coli* O86:B7 *Δwzz* and *E.coli* O86:B7 *Δwzz* harboring plasmid pTR-102[3] strains were grown for 16 hrs at 37 °C in LB. 0.5 mL samples were prepared from whole-cell lysate by the proteinase K method as described by Hitchcock and Brown[4]. After electrophoresis on a 12% SDS-PAGE gel, LPS was detected by a previously described silver staining method[5].

General procedure for the preparation of GalNAc-PP-Lipid substrates

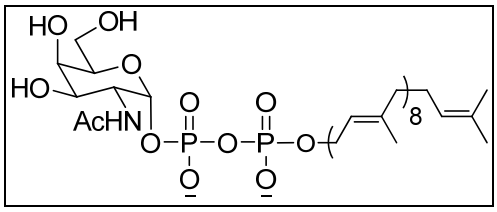
All solvents were dried with a solvent-purification system from Innovative Technology, Inc. All reagents were obtained from commercial sources and used without further purification. The 200-400 mesh silica gel 60 RP-18 (from EMD^{TM}) was utilized for purification. ^{1}H, ^{13}C and ^{31}P NMR spectra were recorded at the indicated field strengths. Mass spectral data was collected in our laboratory using a Shimadzu LCMS-2010A Liquid Chromatograph Mass Spectrometer as well as at The Ohio State University mass spectrometry facility on a Bruker micrOTOF Instrument provided by a grant from the Ohio BioProducts Innovation Center.

N,N'-Carbonyldiimidazole (CDI) (23 mg, 0.14 mmol) was added to a reaction vessel containing the diisopropylethylammonium salt of GalNAc-1-PO_4^{2-} (23 mg, 0.034 mmol)[6-11]. The vessel was flushed with argon, after which anhydrous THF was added (3 mL). The reaction was stirred at room temperature for 2 h. Addition of dry CH_3OH (42 μL) followed by stirring for an additional 1 h at room temperature served to remove excess CDI. The reaction mixture was concentrated and then anhydrous THF (3 mL) was once again added. To a separate reaction vessel, lipid phosphate[12] (0.026 mmol) was added, and the vessel was then flushed with argon. The activated GalNAc-1- PO_4^{2-} was transferred into this vessel via syringe. The reaction was stirred at room temperature for 3 days, after which the reaction mixture was concentrated. Dissolution of the residue in water, followed by filtration through a short bed of C-18 reverse phase silica gel (water then 1:1 Water/Isopropanol) afforded the crude product.

The crude product was treated with a 0.1% $NaOCH_3$ in CH_3OH solution (3 mL). The reaction was stirred at room temperature for 40 min. The reaction mixture was concentrated, after which the residue was dissolved in water. Purification via C-18 reverse phase column chromatography (Solvent A: Isopropanol (Lipids: Und, Solan) or CH_3CN (Lipids: Hept, Pent, cis-Pent, MS-Pent); Solvent B = 1.7% NH_4HCO_3; A/B = 0%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45% and 50% (10 mL each)) provided GalNAc-PP-Lipid as a white solid following lyophilization.



GalNAc-PP-Undecaprenyl: GalNAc-PP-Undecaprenyl was prepared as described above. However, the two purification steps were reversed as the final product proved too insoluble to obtain suitable spectral data. The crude product was thus purified via C-18 reverse phase column chromatography to afford Peracetylated GalNAc-PP-Undecaprenyl as a white solid (20 mg from 23 mg lipid phosphate, 59%). ^{1}H NMR (500 MHz, CD_3OD): δ 5.60 (dd, *J* = 3.2 Hz, *J* = 7.3 Hz, 1H), 5.42 (t, *J* = 6.1 Hz, 2H), 5.21 (dd, *J* = 3.2 Hz, *J* = 11.3 Hz, 1H), 5.14-5.05 (m, 10H), 4.58 (t, *J* = 6.9 Hz, 1H), 4.54-4.45 (m, 3H), 4.21 (dd, *J* = 8.1 Hz, *J* = 10.9 Hz, 1H), 4.04 (dd, *J* = 5.8 Hz, *J* = 10.8 Hz, 1H), 2.10 (s, 3H), 2.08-2.00 (m, 34H), 1.98 (s, 3H), 1.97 (s, 3H), 1.96-1.92 (m, 6H), 1.89 (s, 3H), 1.71 (s, 3H), 1.66-1.62 (m, 21H), 1.58 (s, 3H), 1.57 (s, 9H); ^{13}C NMR (125 MHz, CD_3OD): δ 174.4, 172.3, 172.2, 171.9, 140.6, 136.5, 136.4, 136.3, 136.3, 136.1, 135.9, 132.1, 126.3, 126.3, 126.3, 126.0, 125.6, 125.6, 125.6, 123.6, 123.6, 96.5, 70.4, 68.8, 68.7, 63.8, 63.8, 62.3, 41.0, 41.0, 40.9, 34.4, 33.4, 33.4, 33.4, 33.2, 33.0, 30.8, 30.8, 30.7, 30.6, 30.5, 30.5, 30.4, 28.0, 27.8, 27.8, 27.8, 27.7, 27.7, 27.7, 26.1, 24.0, 23.9, 23.8, 23.0, 20.8, 20.8, 20.7, 17.9, 16.3, 14.5, 14.0; ^{31}P NMR (162 MHz, CD_3OD): δ -9.7, -12.6; LRMS (*m/z*): [M-H]- calcd. for $C_{69}H_{110}NO_{15}P_2$, 1254.7; found, 1254.7.
Final deprotection with 0.1% $NaOCH_3$ in CH_3OH and filtration though a short bed of C-18 reverse phase silica gel afforded GalNAc-PP-Undecaprenyl (18 mg, 99%). LRMS (*m/z*): [M-H]- calcd. for $C_{63}H_{104}NO_{12}P_2$,1128.7; found, 1128.7.

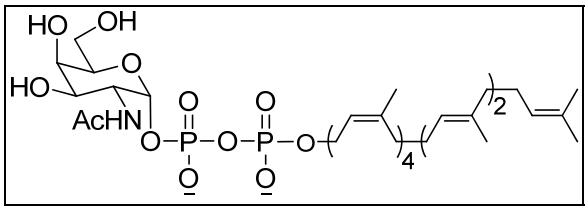


GalNAc-PP-Solanesyl: GalNAc-PP-Solanesyl was prepared as described above. However, the two purification steps were reversed as the final product proved too insoluble to obtain suitable spectral data. The crude product was thus purified via C-18 reverse phase column chromatography to afford Peracetylated GalNAc-PP-Solanesyl as a white solid (35 mg from 45 mg lipid phosphate, 50%). ^{1}H NMR (500 MHz, CD_3OD): δ 5.61 (dd, *J* = 3.3 Hz, *J* = 7.3 Hz, 1H), 5.43 (t, *J* = 6.6 Hz, 2H), 5.22 (dd, *J* = 3.1 Hz, *J* = 11.3 Hz, 1H), 5.14-5.04 (m, 8H), 4.59 (t, *J* = 7.0 Hz, 1H), 4.53 (t, *J* = 6.3 Hz, 3H), 4.21 (dd, *J* = 8.2 Hz, *J* = 10.8 Hz, 1H), 4.04 (dd, *J* = 5.8 Hz, *J* = 10.8 Hz, 1H), 2.11 (s, 3H), 2.10-2.02 (m, 18H), 1.99 (s, 3H), 1.99-1.94 (m, 17H), 1.90 (s, 3H), 1.69 (s, 3H), 1.65 (s, 3H), 1.59 (s, 3H), 1.58 (s, 21H); ^{13}C NMR (125 MHz, CD_3OD): δ 174.5, 172.4, 172.2, 171.9, 140.8, 136.3, 135.9, 135.9, 135.9, 135.8, 132.1, 125.7, 125.7, 125.6, 125.4, 122.7,

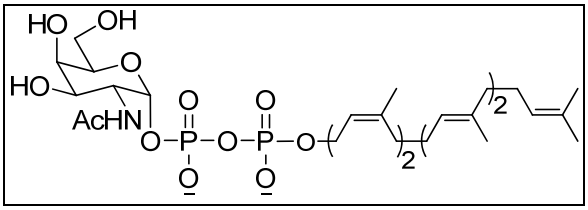
122.6, 96.5, 70.6, 68.7, 64.0, 64.0, 62.3, 41.0, 41.0, 41.0, 40.9, 40.8, 34.5, 33.2, 30.9, 30.8, 30.7, 30.6, 30.5, 28.0, 27.9, 27.8, 27.7, 27.7, 27.7, 26.0, 24.0, 23.8, 23.0, 20.8, 20.7, 20.7, 17.9, 16.8, 16.2, 15.7, 14.5; ^{31}P NMR (162 MHz, CD_3OD): δ -9.6, -12.5; LRMS (*m/z*): [M-H]- calcd. for $C_{59}H_{94}NO_{15}P_2$, 1118.6; found, 1118.5.

Final deprotection with 0.1% $NaOCH_3$ in CH_3OH and filtration though a short bed of C-18 reverse phase silica gel afforded GalNAc-PP-Solanesyl (31 mg, 99%). LRMS (ESI) calcd for $C_{53}H_{88}NO_{12}P_2$ $[M-H]^-$ 992.6, found 992.6.

.

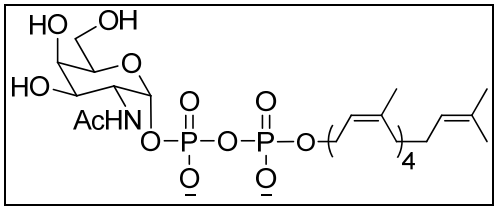


GalNAc-PP-Heptaprenyl: GalNAc-PP-Heptaprenyl was prepared as described above. Following the deprotection step, the crude product was purified via C-18 reverse phase column chromatography to afford GalNAc-PP-Heptaprenyl as a white solid (22 mg from 42 mg lipid phosphate, 36% over 2 steps). ^{1}H NMR (500 MHz, D_2O): δ 5.61 (dd, *J* = 2.7 Hz, *J* = 5.9 Hz, 1H), 5.50 (t, *J* = 5.5 Hz, 1H), 5.24-5.08 (m, 6H), 4.51 (t, *J* = 5.7 Hz, 2H), 4.29 (d, *J* = 10.8 Hz, 1H), 4.23 (t, *J* = 5.5 Hz, 1H), 4.07 (d, *J* = 1.4 Hz, 1H), 4.00 (dd, *J* = 2.2 Hz, *J* = 10.7 Hz, 1H), 3.88-3.75 (m, 2H), 2.20-1.96 (m, 27H), 1.79 (s, 3H), 1.72 (s, 3H), 1.70 (s, 3H), 1.68 (s, 3H), 1.62 (s, 3H), 1.60 (s, 9H); ^{13}C NMR (125 MHz, D_2O): δ 174.5, 172.4, 172.2, 171.9, 140.8, 136.3, 135.9, 135.9, 135.9, 135.8, 132.1, 125.7, 125.7, 125.6, 125.4, 122.7, 122.6, 96.5, 70.6, 68.7, 64.0, 64.0, 62.3, 41.0, 41.0, 41.0, 40.9, 40.8, 34.5, 33.2, 30.9, 30.8, 30.7, 30.6, 30.5, 28.0, 27.9, 27.8, 27.7, 27.7, 27.7, 26.0, 24.0, 23.8, 23.0, 20.8, 20.7, 20.7, 17.9, 16.8, 16.2, 15.7, 14.5; 174.7, 165.3, 134.8, 134.6, 134.3, 130.4, 125.2, 125.0, 124.8, 124.5, 124.3, 124.3, 121.2, 94.6, 72.1, 68.7, 67.7, 62.6, 61.3, 49.9, 39.7, 39.7, 32.0, 31.9, 31.9, 29.9, 29.8, 29.7, 29.5, 26.7, 26.6, 26.2, 25.4, 23.3, 23.2, 23.0, 23.0, 22.3, 17.3, 15.7; ^{31}P NMR (162 MHz, D_2O): δ -9.7, -11.4; LRMS (*m/z*): [M-H]- calcd. for $C_{43}H_{72}NO_{12}P_2$, 856.5; found, 856.7.

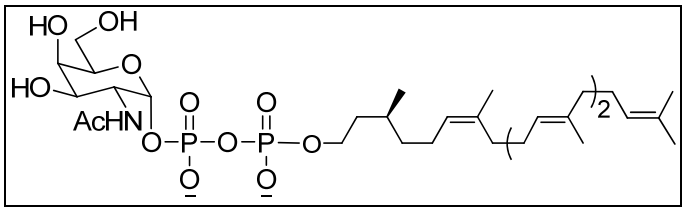


GalNAc-PP-Pentaprenyl: GalNAc-PP-Pentaprenyl was prepared as described above. Following the deprotection step, the crude product was purified via C-18 reverse phase column chromatography to afford GalNAc-PP-Pentaprenyl as a white solid (22 mg from 28 mg lipid phosphate, 49% over 2 steps). ^{1}H NMR (500 MHz, CD_3OD): δ 5.67 (dd, *J* =

3.1 Hz, *J* = 6.8 Hz, 1H), 5.56 (t, *J* = 6.6 Hz, 1H), 5.30-5.16 (m, 4H), 4.58 (t, *J* = 6.5 Hz, 2H), 4.36 (d, *J* = 10.9 Hz, 1H), 4.30 (dd, *J* = 6.0 Hz, 1H), 4.14 (d, *J* = 3.0 Hz, 1H), 4.07 (dd, *J* = 2.9 Hz, *J* = 10.9 Hz, 1H), 3.94-3.84 (m, 2H), 2.30-2.23 (m, 2H), 2.22-2.11 (m, 13H), 2.10-2.02 (m, 4H), 1.85 (s, 3H), 1.80 (s, 3H), 1.75 (s, 3H), 1.71 (s, 3H), 1.68 (s, 6H); ^{13}C NMR (125 MHz, D_2O): δ 174.7, 160.3, 141.2, 135.8, 134.9, 134.4, 130.6, 124.6, 124.5, 124.3, 121.2, 94.7, 72.2, 68.7, 67.7, 62.7, 61.3, 49.9, 48.9, 39.7, 39.7, 32.0, 31.9, 26.7, 26.6, 26.1, 25.4, 23.1, 22.9, 22.3, 17.4, 15.8, 15.7; ^{31}P NMR (162 MHz, D_2O): δ -10.4, -13.1; LRMS (*m/z*): [M-H]- calcd. for $C_{33}H_{56}NO_{12}P_2$, 720.3; found, 720.3.



GalNAc-PP-cis-Pentaprenyl: GalNAc-PP-cis-Pentaprenyl was prepared as described above. Following the deprotection step, the crude product was purified via C-18 reverse phase column chromatography to afford GalNAc-PP-cis-Pentaprenyl as a white solid (15 mg from 48 mg lipid phosphate, 20% over 2 steps). ^{1}H NMR (500 MHz, D_2O): δ 5.62 (dd, *J* = 3.0 Hz, *J* = 6.5 Hz, 1H), 5.51 (t, *J* = 6.0 Hz, 1H), 5.24-5.12 (m, 4H), 4.53 (t, *J* = 6.2 Hz, 2H), 4.30 (d, *J* = 11.1 Hz, 1H), 4.25 (t, *J* = 5.9 Hz, 1H), 4.09 (d, *J* = 2.4 Hz, 1H), 4.01 (dd, *J* = 2.6 Hz, *J* = 10.9 Hz, 1H), 3.88-3.78 (m, 2H), 2.22-2.00 (m, 19H), 1.79 (s, 3H), 1.73 (s, 3H), 1.72-1.68 (m, 9H), 1.63 (s, 3H); ^{13}C NMR (125 MHz, D_2O): δ 174.7, 141.3, 135.4, 135.4, 134.8, 134.7, 125.2, 125.1, 124.9, 124.4, 124.4, 94.8, 72.2, 68.6, 67.7, 62.8, 61.3, 49.8, 32.1, 32.1, 31.8, 26.6, 26.4, 26.2, 25.5, 25.4, 23.2, 23.2, 23.0, 23.0, 22.3, 17.3, 15.7; ^{31}P NMR (162 MHz, D_2O): δ -10.1, -12.3; LRMS (*m/z*): [M-H]- calcd. for $C_{33}H_{56}NO_{12}P_2$, 720.3; found, 720.4.

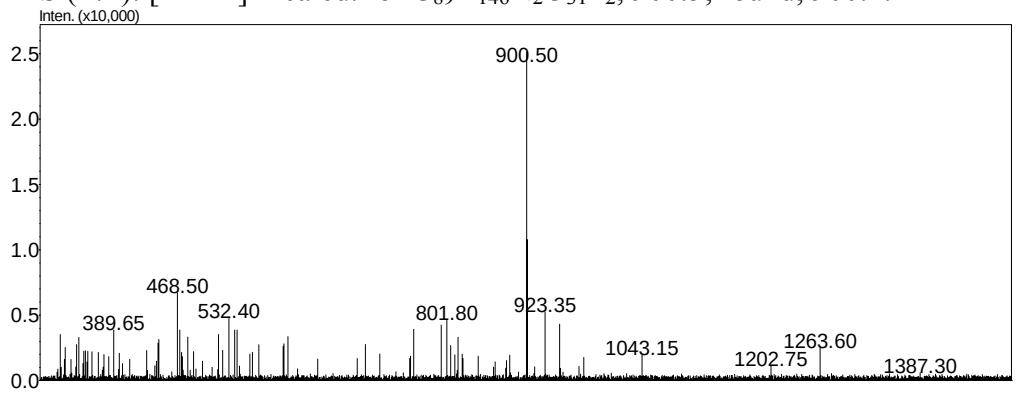


GalNAc-PP-MS-Pentaprenyl: GalNAc-PP-MS-Pentaprenyl was prepared as described above. Following the deprotection step, the crude product was purified via C-18 reverse phase column chromatography to afford GalNAc-PP-MS-Pentaprenyl as a white solid (27 mg from 65 mg lipid phosphate, 26% over 2 steps). ^{1}H NMR (500 MHz, D_2O): δ 5.55 (dd, *J* = 3.3 Hz, *J* = 7.1 Hz, 1H), 5.22-5.05 (m, 4H), 4.26 (dt, *J* = 2.9 Hz, *J* = 11.0 Hz, 1H), 4.19 (dd, *J* = 5.2 Hz, *J* = 7.1 Hz, 1H), 4.06-3.93 (m, 4H), 3.84-3.72 (m, 2H), 2.11-2.02 (m, 11H), 2.01-1.92 (m, 6H), 1.69 (s, 3H), 1.66 (s, 3H) 1.61 (s, 3H), 1.58 (s, 6H), 1.54-1.24

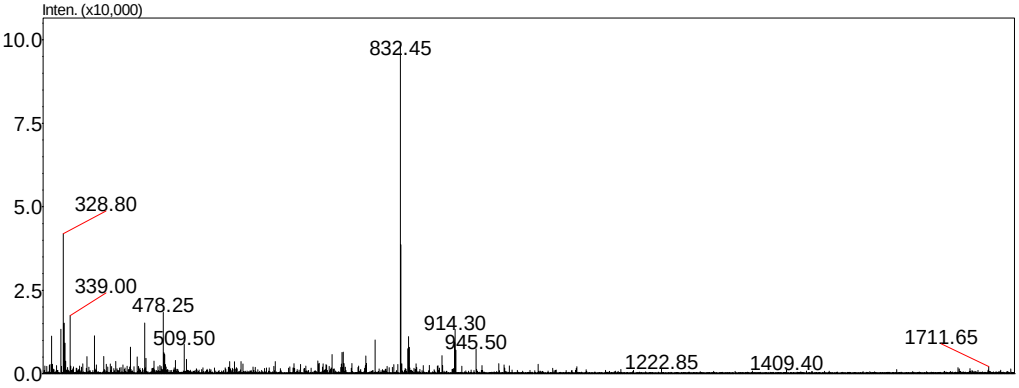
(m, 4H), 1.22-1.10 (m, 1H), 0.92 (d, J = 6.5 Hz, 3H); ^{13}C NMR (125 MHz, D_2O): δ 174.5, 135.0, 134.8, 134.5, 130.8, 125.7, 124.4, 124.2, 94.7, 72.1, 68.6, 67.7, 64.9, 61.3, 49.9, 49.8, 39.6, 37.3, 37.1, 31.8, 29.0, 26.6, 26.6, 26.5, 25.4, 24.9, 23.2, 22.3, 18.8, 18.8, 17.4, 15.7, 15.7; ^{31}P NMR (162 MHz, D_2O): δ -9.9, -11.8; LRMS (*m/z*): [M-H]- calcd. for $C_{33}H_{58}NO_{12}P_2$, 722.3; found, 722.3.

ESI-MS for RU-PP-Lipid analogs

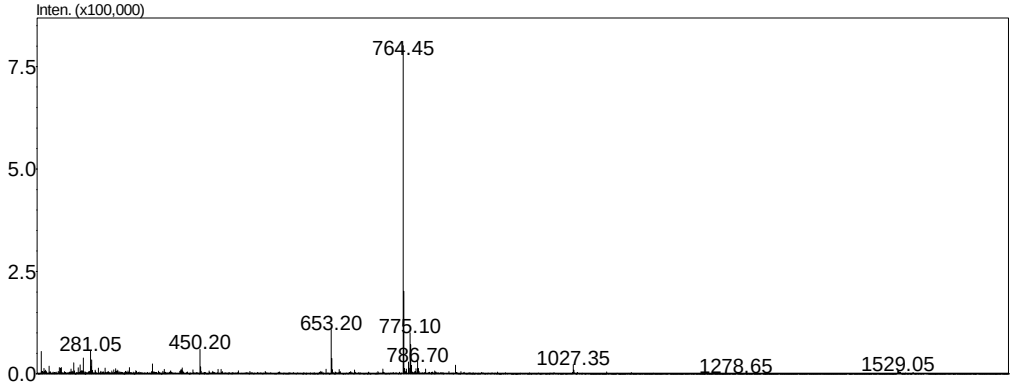
RU-PP-Undecaprenyl:
LRMS (*m/z*): [M-2H]2- calcd. for $C_{89}H_{146}N_2O_{31}P_2$, 900.5; found, 900.4.



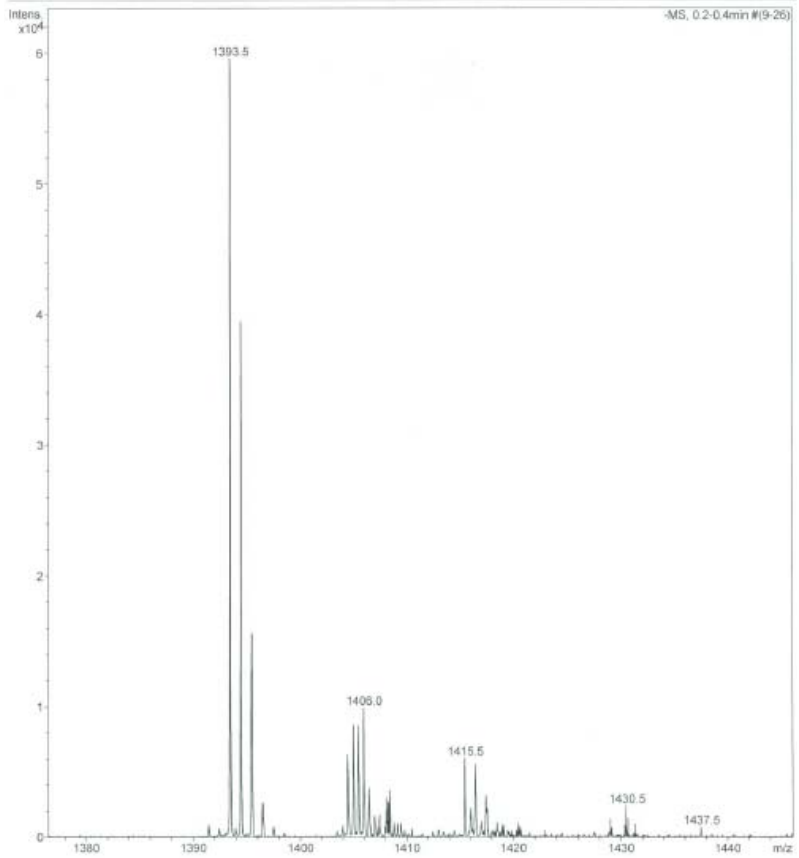
RU-PP-Solanesyl:
LRMS (*m/z*): [M-2H]2- calcd. for $C_{79}H_{130}N_2O_{31}P_2$, 832.4; found, 832.4.



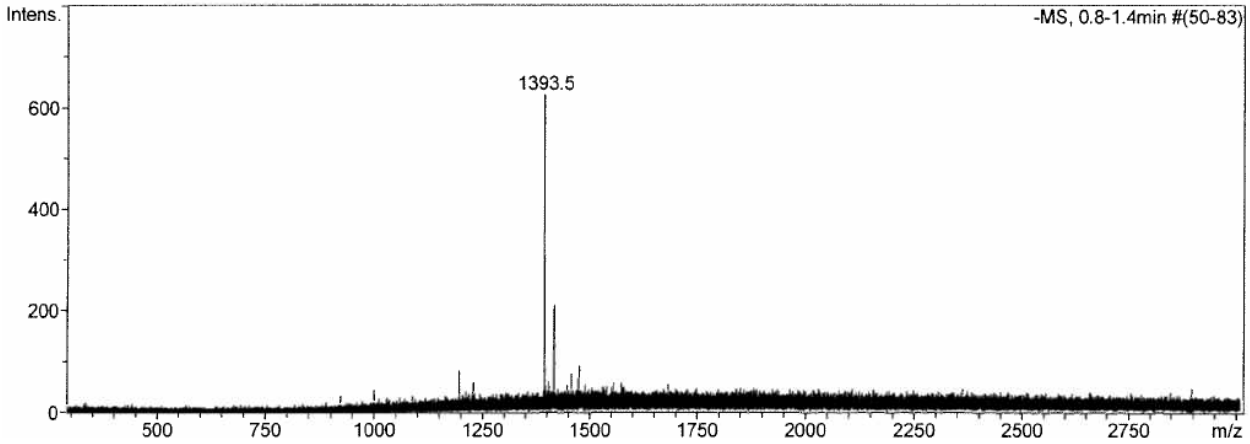
RU-PP-Heptaprenyl:
LRMS (*m/z*): [M-2H]2- calcd. for $C_{69}H_{114}N_2O_{31}P_2$, 764.3; found, 764.4.



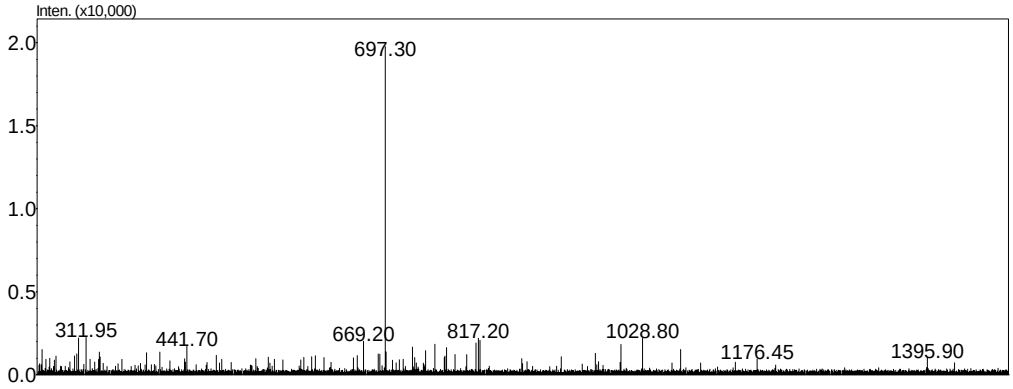
RU-PP-Pentaprenyl:
LRMS (*m/z*): [M-H]- calcd. for $C_{59}H_{99}N_2O_{31}P_2$, 1393.6; found, 1393.5.



RU-PP-cis-Pentaprenyl:
LRMS (*m/z*): [M-H]- calcd. for $C_{59}H_{99}N_2O_{31}P_2$, 1393.6; found, 1393.5.



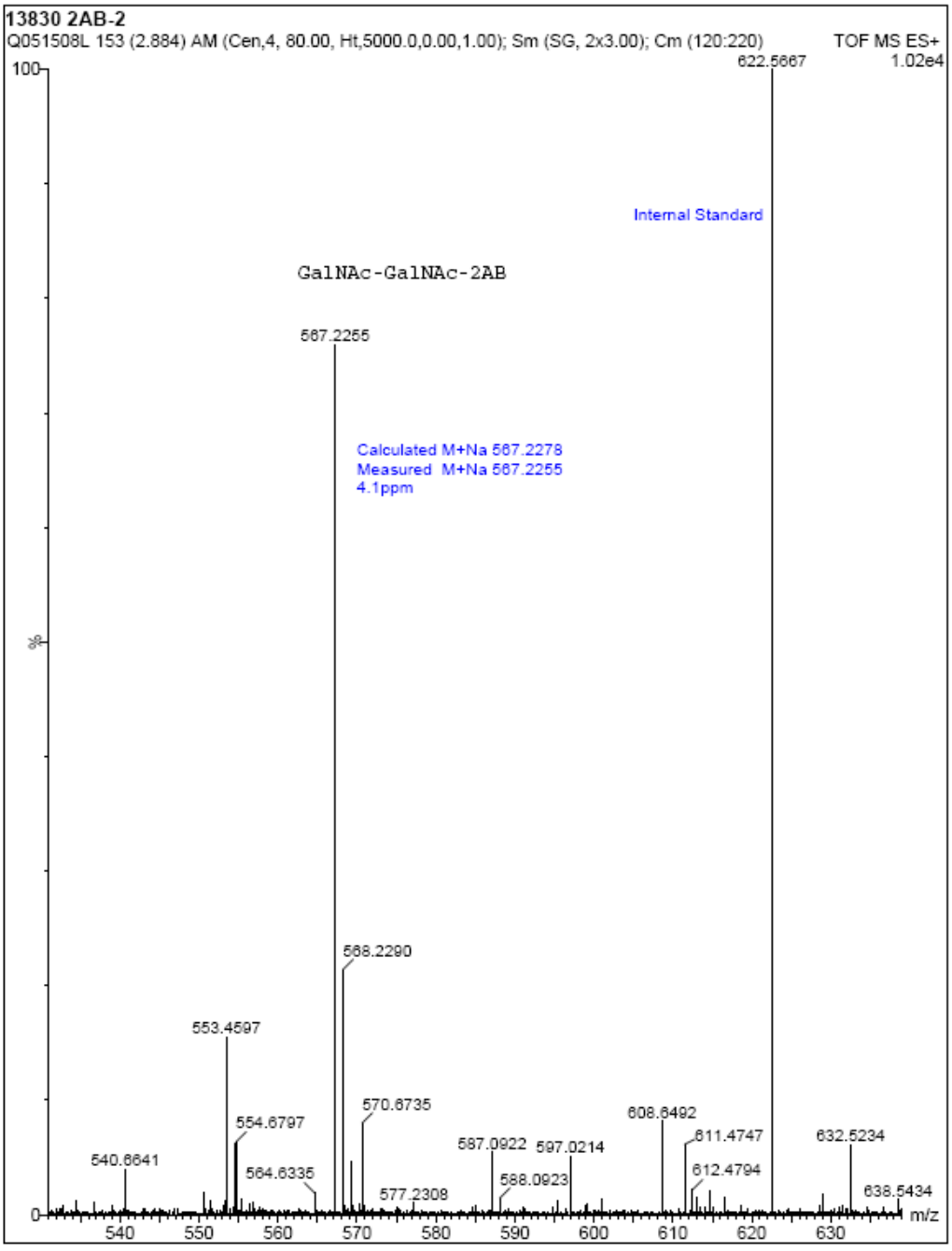
RU-PP-MS-Pentaprenyl:
LRMS (*m/z*): [M-2H]2- calcd. for $C_{59}H_{100}N_2O_{31}P_2$, 697.3; found, 697.3.



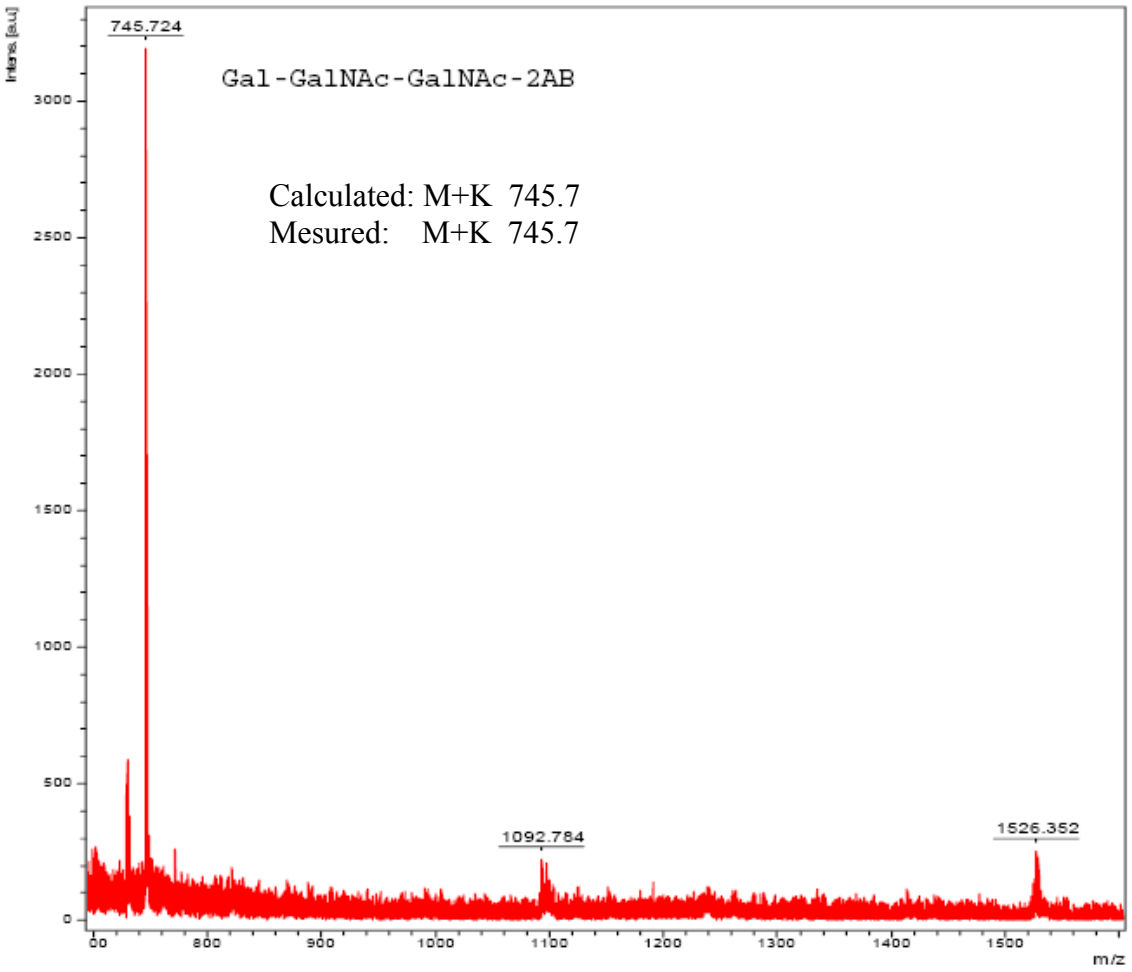
Supplementary Results:

Supplementary Figure 1 MALDI-Mass spectra for the di- (**a**), tri- (**b**), tetra- (**c**) and pentasaccharide (**d**) products which were labeled with 2-AB.

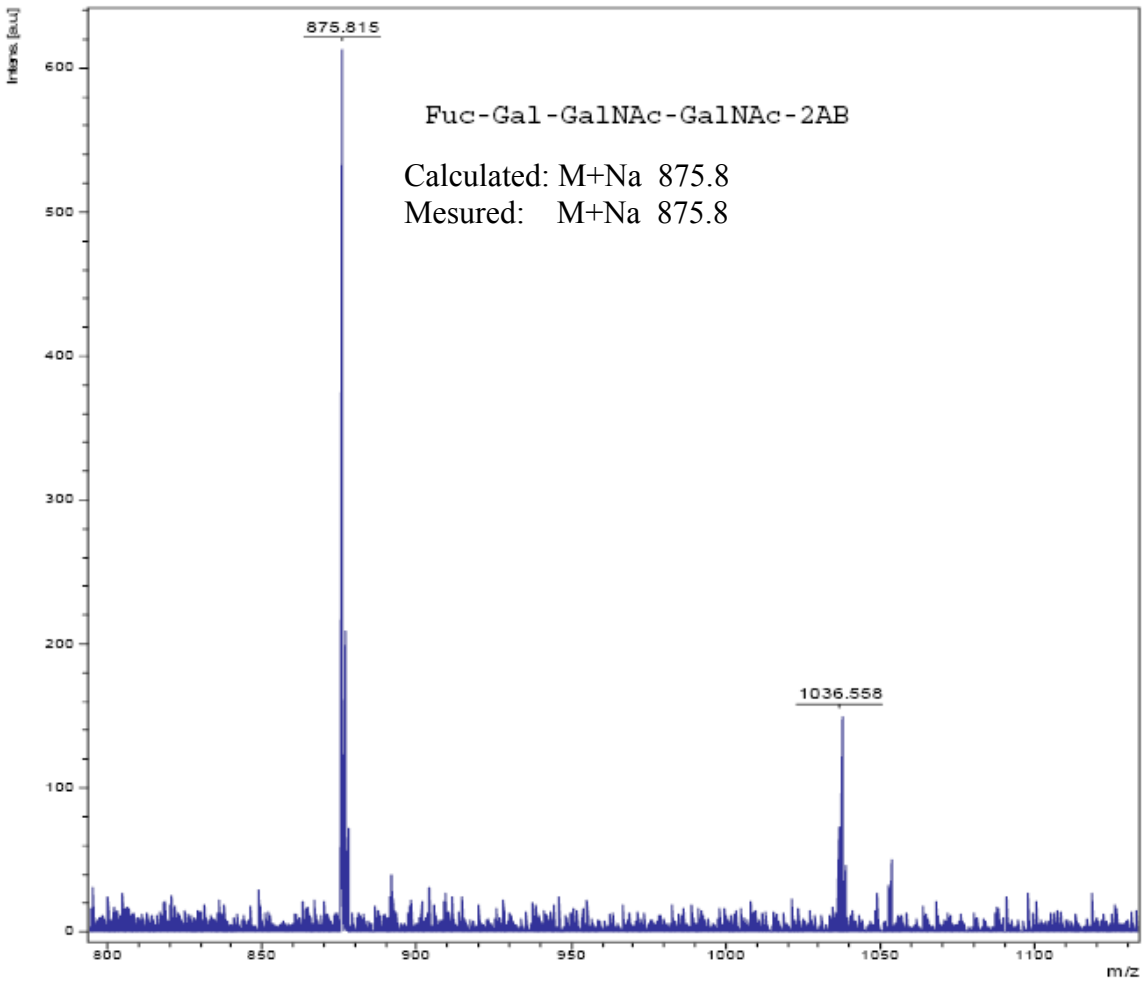
a)



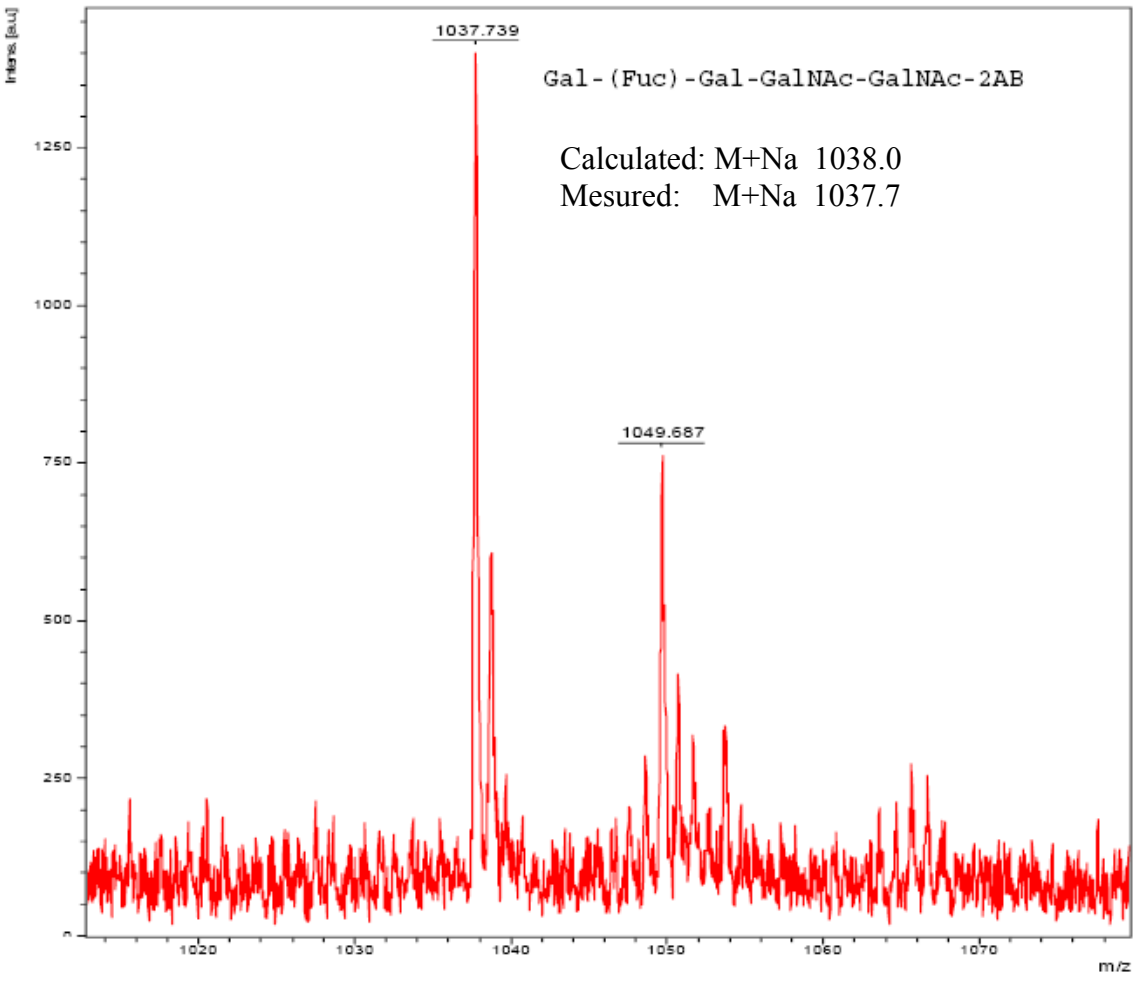
b)


745.724
Gal-GalNAc-GalNAc-2AB
Calculated: M+K 745.7
Mesured: M+K 745.7
1092.784
1526.352
Intens. [a.u.]
3000
2500
2000
1500
1000
500
0
800
900
1000
1100
1200
1300
1400
1500
m/z

c)

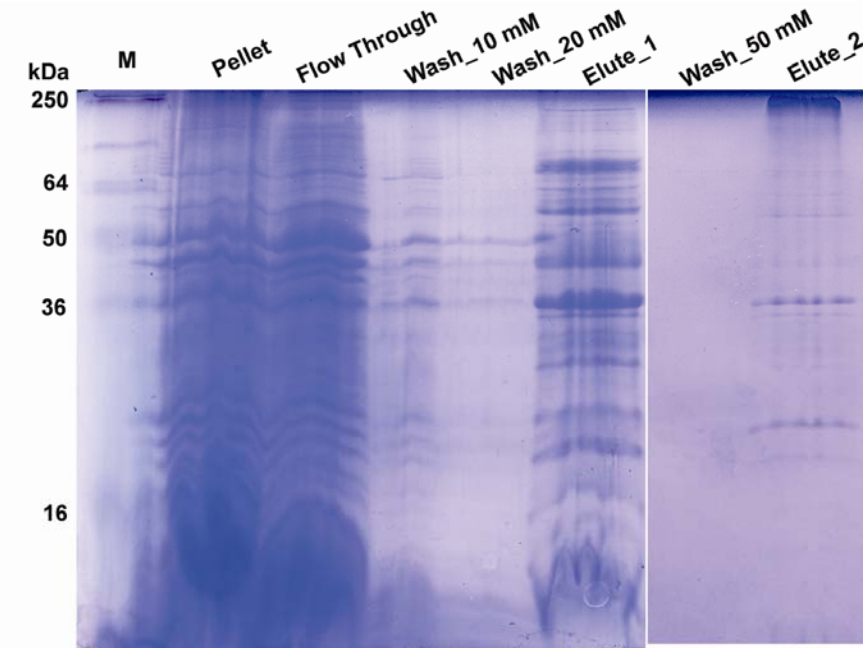

875.815
Fuc-Gal-GalNAc-GalNAc-2AB
Calculated: M+Na 875.8
Mesured: M+Na 875.8
1036.558
Intens [a.u.]
600
500
400
300
200
100
0
800
850
900
950
1000
1050
1100
m/z

d)


1037.739
Gal-(Fuc)-Gal-GalNAc-GalNAc-2AB
Calculated: M+Na 1038.0
Mesured: M+Na 1037.7
1049.687
Intens. [a.u.]
1250
1000
750
500
250
0
1020
1030
1040
1050
1060
1070
m/z

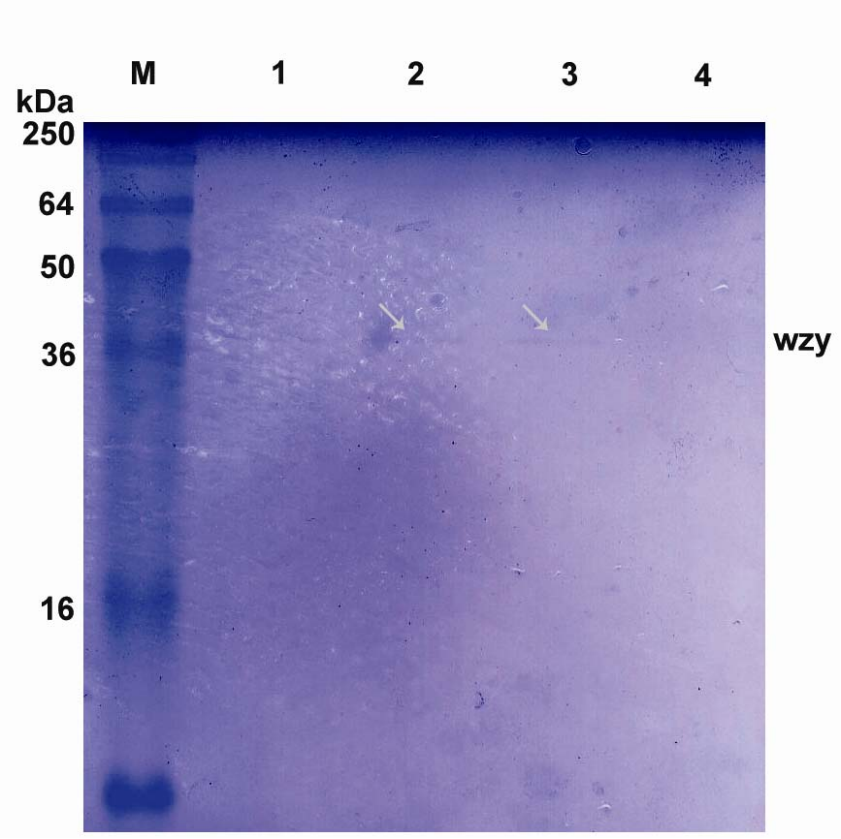
Supplementary Figure 2 Purification of Wzy: (**a**) Ni-NTA column purification, (**b**) Ion-exchange and gel-filtration column purification, and (**c**) Final concentrated sample.

a)



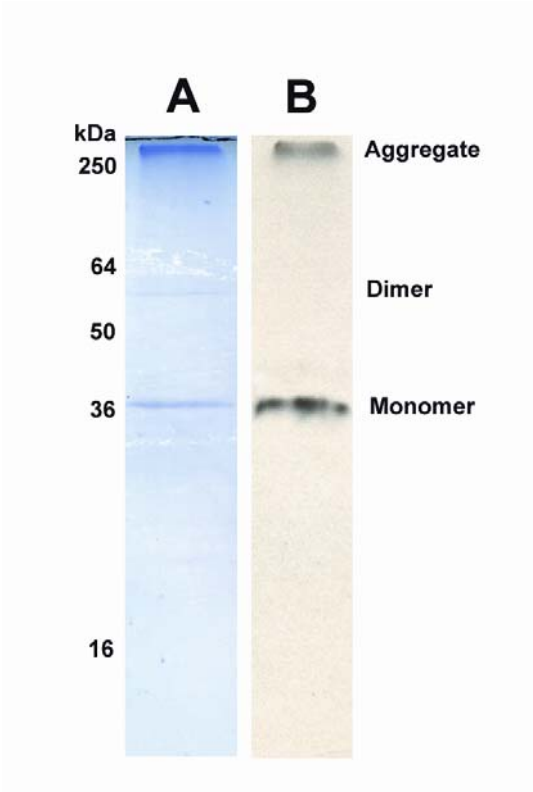
After the first washing step (20 mM imidazole), the eluent containing Wzy proved to be highly contaminated. A second washing step was therefore performed using 50 mM imidazole. The purity of the eluted wzy was significiantly improved.

b)



Purification using an SP Sepharose Fast Flow column (Pharmacia) and superdex 200 column (Pharmacia) yielded fractions containing Wzy with much higher apparent purify (Fractions **2**-**3**). These samples were pooled and concentrated for further assessment of purity.

c)



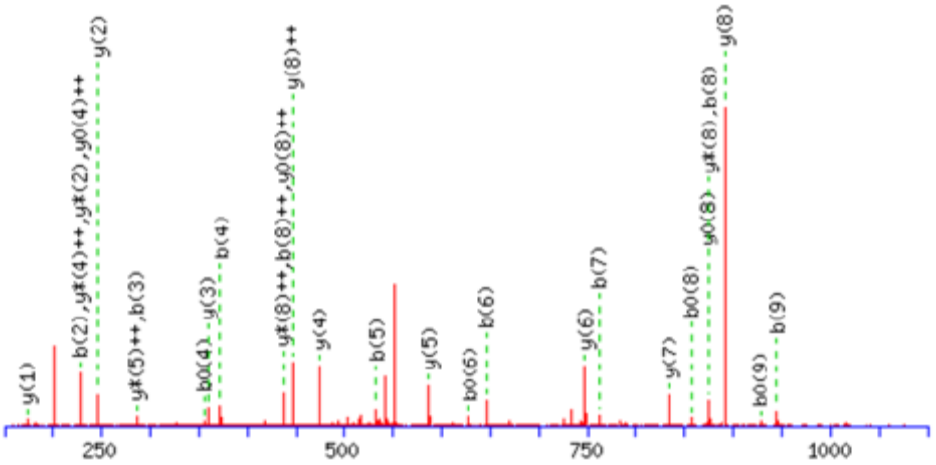
(**A**) Coomassie blue stained gel showing Wzy monomer (36 kDa), dimer (58 kDa) and aggregate (>250 kDa). (**B**) Western blot of Wzy showing only monomer and aggregate. Sample was probed by incubation with an anti-His antibody at 1:2000 dilution, followed by incubation of a secondary antibody HRP-linked anti-mouse IgG at a dilution of 1:2000.

It should be noted that the apparent molecular weight of Wzy observed from the gel is less than the theoretical value (~50 kDa) calculated from the amino acid sequence. It is not unexpected since membrane proteins have been shown in a number of studies[13-15] to migrate abnormally in SDS-PAGE, likely due to the hydrophobic properties of such proteins. In addition, high molecular weight aggregation has been a common observation in membrane protein expression[16]. In our study, co-expression of the GroES/GroEL chaperone system greatly reduced the aggregation of Wzy, as solely protein aggregates were observed in the absence of the chaperone system (**Supplementary Fig. 5** online). Thus, it was speculated that the GroES/GroEL chaperone facilitated the expression and correct folding of Wzy. It should also be noted that in preliminary studies, we found that a pBAD-*wzy-myc-his*$_6$ plasmid construct (non-optimized wzy gene) with tandem C-terminal myc and His_6 epitopes can fully complement the LPS phenotype in the *wzy*-depleted strain (**Supplementary Fig. 6** online). Such a result suggests that incorporation of epitopes at the C-terminus of Wzy does not affect enzymatic activity *in vivo*.

Supplementary Figure 3 Identification of Wzy peptides by LC/MS/MS. Possible sequences and match scores are listed below for the (**a**) Aggregate, (**b**) Dimer and (**c**) Monomer.

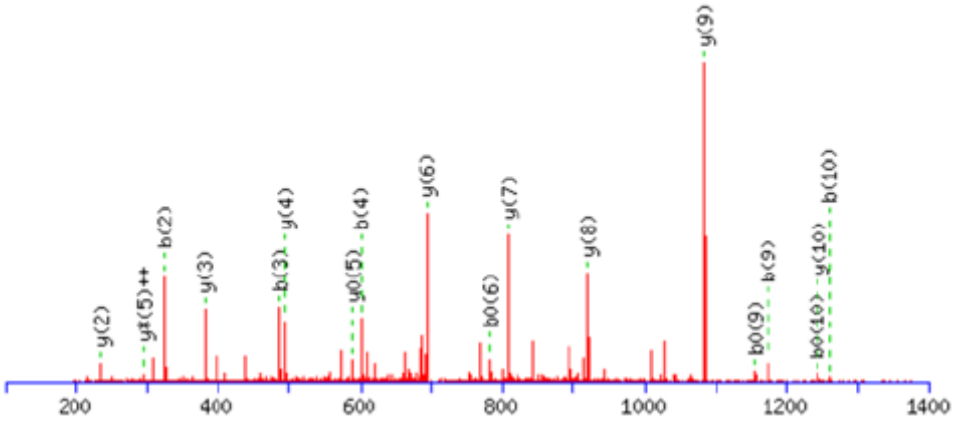
a) Aggregate:

DIGSCIDIAR (53.9)
Monoisotopic mass of neutral peptide Mr(calc): 1118.5390
Fixed modifications: Carbamidomethyl (C)
Ions Score: 54 Expect: 0.00086



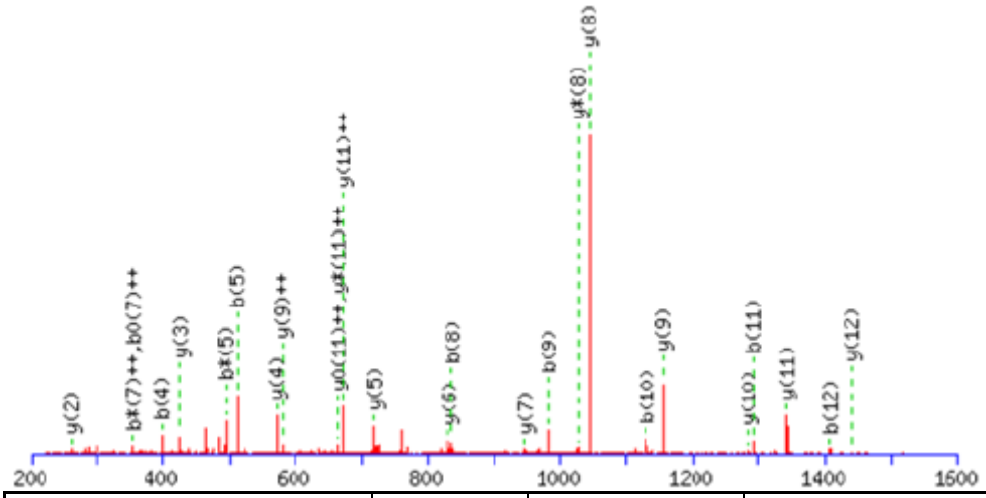
Score (> 36 indicates identity)	Expect	Protein	Peptide
53.9	0.00086	gi 56384981	DIGSCIDIAR
20.9	1.7	N/A	MVLNLDRSR
12.7	11	N/A	GTSGMYFLAR
12.3	12	N/A	TVYIGIDIAR
11.8	14	N/A	VEGMSLEEAR
11.8	14	N/A	VEGXSLEEAR
11.2	16	N/A	LNGSVADMRR
9.9	21	N/A	SDSDQLDVIK
9.9	22	N/A	DMEEKKLAR
9.4	24	N/A	NIRMMTLLK

YCYILSLLFSK (60.7)
Monoisotopic mass of neutral peptide Mr(calc): 1405.7315
Fixed modifications: Carbamidomethyl (C)
Ions Score: 61 Expect: 0.00022



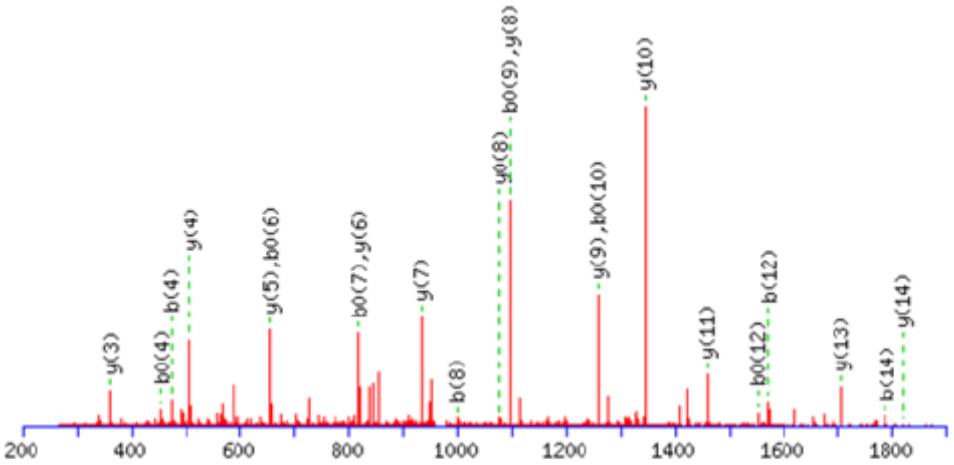
Score (> 37 indicates identity)	Expect	Protein	Peptide
60.7	0.00022	gi 56384981	YCYILSLLFSK
20.8	2.2	N/A	ALGRLSIAMGDSSK
6.7	56	N/A	MSAPYKDLLVKK
4	1.00e+002	N/A	FAHMAIKLNSMK
3.9	1.10e+002	N/A	VENGLSVGEYIVK
3.3	1.20e+002	N/A	VEGVSKVDVAFEK
3.2	1.30e+002	N/A	DGHAIEVLIRQR
3.1	1.30e+002	N/A	RSLEIFSGAGGLAK
2.3	1.60e+002	N/A	MNNIIAVVLYNK
2	1.70e+002	N/A	VVDQTXSVTTESK

LVGQLPDLFFYNK (82.6)
Monoisotopic mass of neutral peptide Mr(calc): 1552.8290
Fixed modifications: Carbamidomethyl (C)
Ions Score: 83 Expect: 1.4e-006



Score (> 30 indicates homology; > 37 indicates identity)	Expect	Protein	Peptide
82.6	1.40e-006	gi 56384981	LVGQLPDLFFYNK
12.6	14	N/A	NGRMLAVIGMLHDK
12	16	N/A	EYQILLPGEIHIK
7.2	48	N/A	TLALLERLRDEAR
4	1.00e+002	N/A	FVIGGPMGDCGLTGR
4	1.00e+002	N/A	FVIGGPXGDCGLTGR
1.6	1.80e+002	N/A	MSVIKMTDLDLAGK
0.2	2.40e+002	N/A	KSALLDNKEHSADK
0.1	2.50e+002	N/A	VALNIHLSVTSCIK

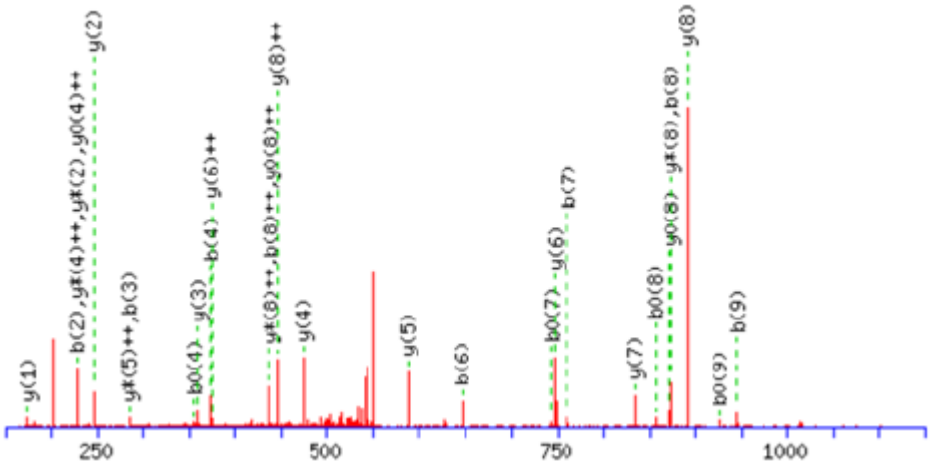
ILSCISYYDYFFTNK (91)
Monoisotopic mass of neutral peptide Mr(calc): 1932.8968
Fixed modifications: Carbamidomethyl (C)
Ions Score: 91 Expect: 2.4e-007



Score (> 30 indicates homology; > 37 indicates identity)	Expect	Protein	Peptide
90.6	2.40e-007	gi 56384981	ILSCISYYDYFFTNK
12.0	18	N/A	YNPVAELQSPPAIPPVDK
10.2	27	N/A	NEGVVASSAWPYQTNALK
4.8	93	N/A	CTQKLIAAIPHIPEWR
4.7	94	N/A	TLMYEFNHSHPSEVDK
3.7	1.2e+002	N/A	HIPGALHLPHKQMTAER
3.5	1.3e+002	N/A	AEIAFLSSIIGKLKTTIK
3.2	1.3e+002	N/A	TLGSGVIXDQRGYIITNK
2.9	1.4e+002	N/A	SKKIAVIGECMIELSEK
2.5	1.6e+002	N/A	YATDITDPQSGEFMTIK

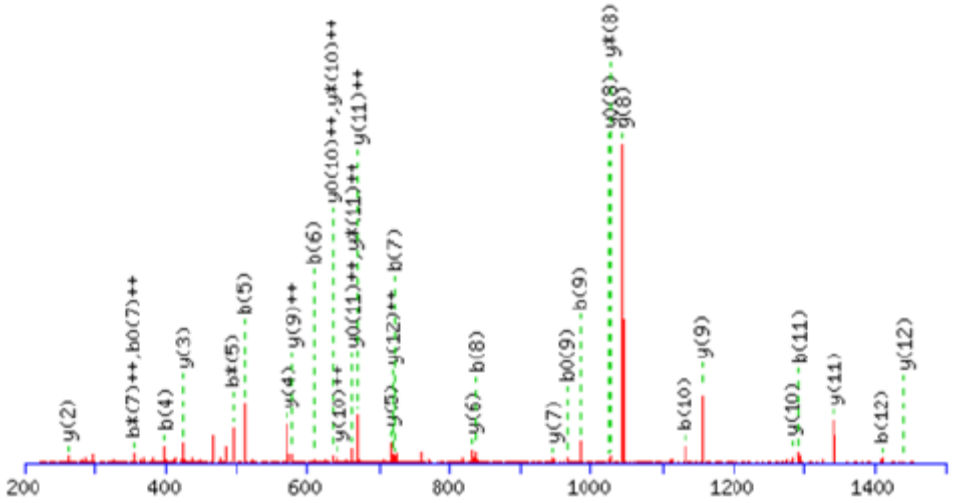
b) Dimer

DIGSCIDIAR
Monoisotopic mass of neutral peptide Mr(calc): 1118.5390
Fixed modifications: Carbamidomethyl (C)
Ions Score: 57 Expect: 0.00045



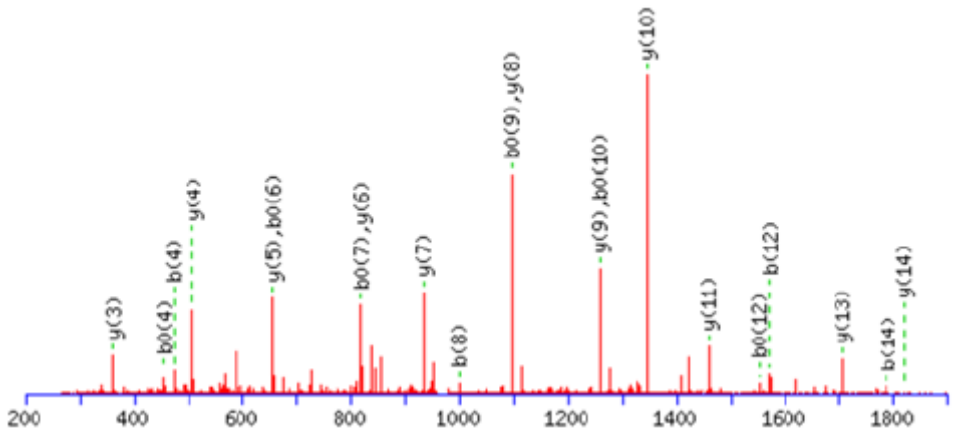
Score (> 33 indicates homology; > 36 indicates identity)	Expect	Protein	Peptide
56.7	0.00045	gi 56384981	DIGSCIDIAR
17.4	3.8	N/A	TVYIGIDIAR
14.1	8.2	N/A	MVLNLDRSR
13.8	8.7	N/A	MDSARALIAR
12.3	12	N/A	LNGSVADMRR
11.3	15	N/A	RFTIEDIAR
11.1	16	N/A	AGDMVIDGSVR
10.8	18	N/A	DMEEKKLAR
10.2	20	N/A	EAEQNSSEAR
9.9	21	N/A	LDPTRYQAR

LVGQLPDLFFYNK
Monoisotopic mass of neutral peptide Mr(calc): 1552.8290
Fixed modifications: Carbamidomethyl (C)
Ions Score: 75 Expect: 8e-006



Score (> 33 indicates homology; > 36 indicates identity)	Expect	Protein	Peptide
75.0	8e-006	gi 56384981	LVGQLPDLFFYNK
11.8	17	N/A	EYQILLPGEIHIK
10.0	26	N/A	NGRMLAVIGMLHDK
7.1	50	N/A	TLALLERLRDEAR
6.1	63	N/A	MSVIKMTDLDLAGK
4.0	1e+002	N/A	FVIGGPMGDCGLTGR
4.0	1e+002	N/A	FVIGGPXGDCGLTGR
2.0	1.6e+002	N/A	VALNIHLSVTSCIK
1.9	1.7e+002	N/A	INLGIGFYKDETGK
0.2	2.4e+002	N/A	KSALLDNKEHSADK

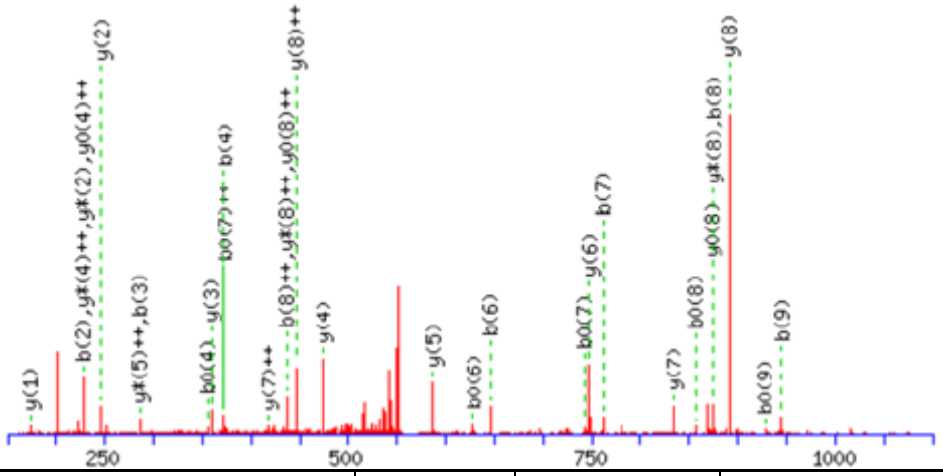
ILSCISYYDYFFTNK (96.0)
Monoisotopic mass of neutral peptide Mr(calc): 1932.8968
Fixed modifications: Carbamidomethyl (C)
Ions Score: 96 Expect: 5.8e-008



Score (> 33 indicates homology; > 36 indicates identity)	Expect	Protein	Peptide
96	5.80e-008	gi 56384981	ILSCISYYDYFFTNK
14.6	8	N/A	TLGSGVIXDQRGYIITNK
11.9	15	N/A	NEGVVASSAWPYQTNALK
9.9	24	N/A	CTQKLIAAIPHIPEWR
7.2	44	N/A	LQGSVDKDVFTRLLEGR
6.1	56	N/A	KMIMAMVQDIRVILIK
6	58	N/A	AEIAFLSSIIGKLKTTIK
5.6	63	N/A	KMIMAMVQDIRVILIK
5.5	64	N/A	VPLYHIMDVAKQMENN
5.4	66	N/A	TLMYEFNHSHPSEVDK

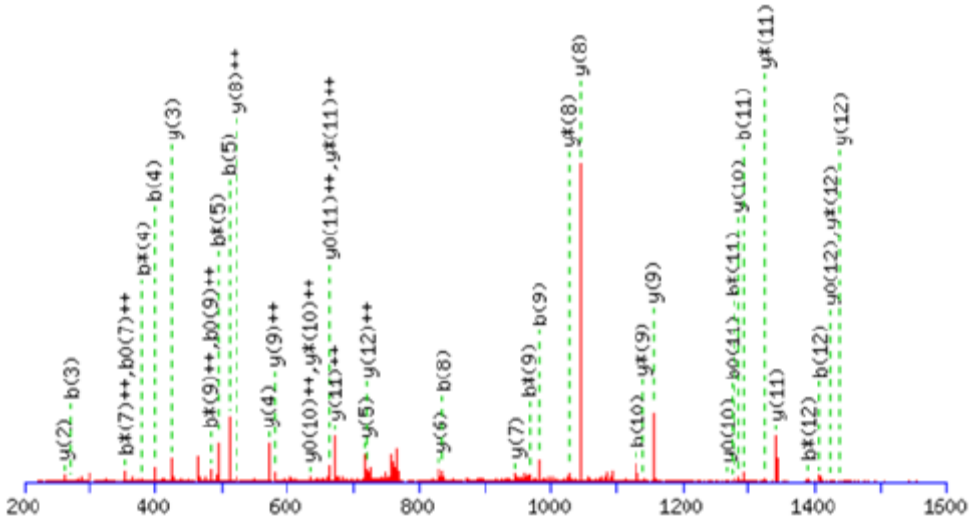
c) Monomer

DIGSCIDIAR (46)
Monoisotopic mass of neutral peptide Mr(calc): 1118.5390
Fixed modifications: Carbamidomethyl (C)
Ions Score: 46 Expect: 0.0067



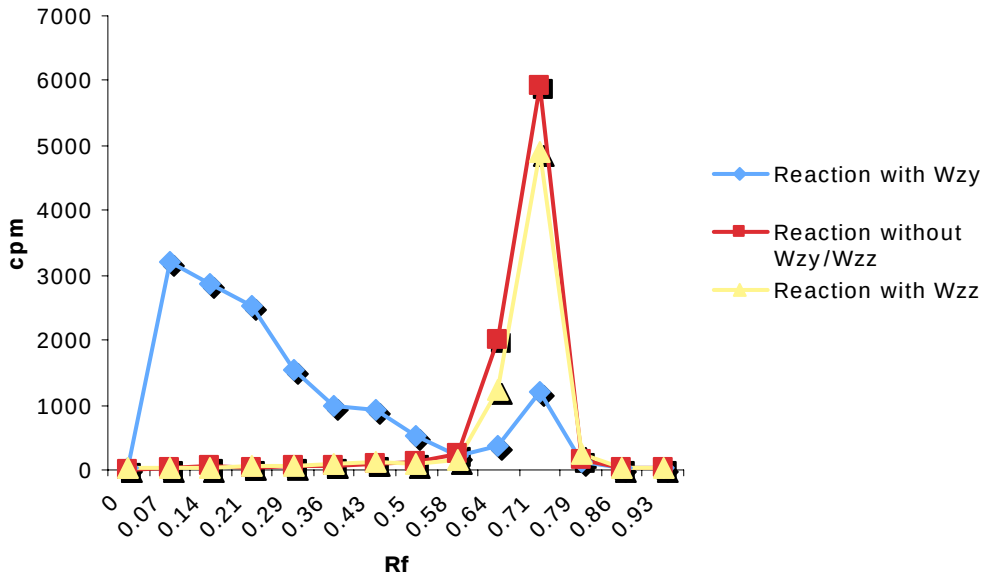
Score (> 27 indicates homology; > 37 indicates identity)	Expect	Protein	Peptide
46.1	0.0067	gi 56384981	DIGSCIDIAR
11.4	20	N/A	MVLNLDRSR
10.4	25	N/A	LNGSVADMRR
9.4	32	N/A	VWADDSAPEK
8.8	36	N/A	AESLDLNMVK
8.5	39	N/A	MARSKTAQPK
8.3	41	N/A	MDSARALIAR
8.2	41	N/A	NIRMMTLLK
7.9	44	N/A	LLMNDGLVSR
7.5	49	N/A	TVYIGIDIAR

LVGQLPDLFFYNK (54)
Monoisotopic mass of neutral peptide Mr(calc): 1552.8290
Fixed modifications: Carbamidomethyl (C)
Ions Score: 54 Expect: 0.00091



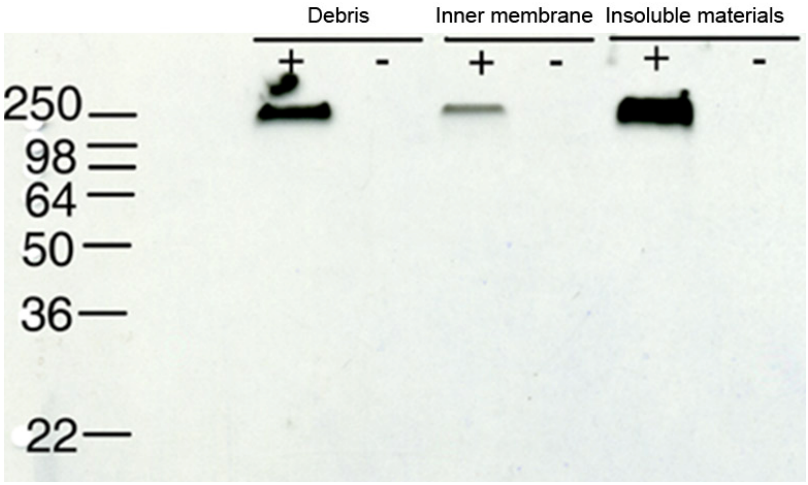
Score (> 22 indicates homology; > 37 indicates identity)	Expect	Protein	Peptide
54.5	0.00091	gi 56384981	LVGQLPDLFFYNK
5.6	70	N/A	NGRMLAVIGMLHDK
2.8	1.3e+002	N/A	FVIGGPMGDCGLTGR
2.8	1.3e+002	N/A	FVIGGPXGDCGLTGR
2.6	1.4e+002	N/A	KSALLDNKEHSADK
1.7	1.7e+002	N/A	EYQILLPGEIHIK
1.6	1.8e+002	N/A	ERDLLERLITLGK
1.3	1.9e+002	N/A	TLALLERLRDEAR
0.9	2.1e+002	N/A	MSVIKMTDLDLAGK
0.4	2.3e+002	N/A	VLKAKLTIDIPTIK

Supplementary Figure 4 Wzy activity assay using RU-PP-Und as a substrate.



The figure was generated using a previously described method[11]. Briefly, the reaction mixtures and control (Ni-NTA elution from cells containing an empty pBAD vector) were spotted on Whatman 3 MM chromatography paper (1x14 cm). The strips were developed in isobutyric acid/1 N NH_4OH (5:3 V/V). After the solvent front migrated to the top, each strip was cut into 1 cm squares and counted for radioactivity. Results were plotted as cpm versus R_f (the relative mobility compared to the solvent front).

Supplementary Figure 5 Western Blot of Wzy expressed without chaperone expression vector (+ = plasmid with Wzy, - = plasmid without Wzy)



Cells were disrupted by sonication, followed by centrifugation (10,000 g, 10 min, 4°C) to remove debris. The membrane fractions were isolated by ultracentrifugation (150,000 g, 75 min, 4 °C) and solubilized in 5 mL buffer A (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 2% n-Decyl-β-D-maltopyranoside (DDM)) overnight at 4 °C. The unsolubilized material was separated by ultracentrifugation (150,000 g, 75 min, 4 °C). Debris and unsolubilized material from the total membrane fraction were further denatured using 8 M Urea and 2% SDS before applicaton to SDS-PAGE and Western Blotting. Western Blotting was performed as described in the manuscript. The results suggest that Wzy expressed without chaperone expression vector forms aggregates with most of the Wzy protein being insoluble.

Supplementary Figure 6 In vivo disruption and complementation of wzy gene. 1, LPS from *E. coli* O86:B7 wild type strain; 2, LPS from *E. coli* O86:B7 *Δwzy*; 3, LPS from *E. coli* O86:B7 *Δwzy* + wzy

The wzy gene from *E. coli* O86 was replaced by the chloramphenicol acetyltransferase (CAT) gene using the RED recombination system of phage lambda. The CAT gene was PCR-amplified from plasmid pKK232 using primers binding to the 5' and 3' ends of the gene. Each primer also carries 40 bp gene segments flanking the target genes. The PCR product was electroporated into *E. coli* O86 strain carrying the pKD20 vector. The chloramphenicol resistant transformants were selected after induction of the RED gene according to the protocol described previously[1,17]. To complement the mutant strain, the pBAD-wzy recombinant vector was transformed into the mutant strain, which was subsequently grown in LB media supplemented with 0.1% arabinose. LPS from wild-type, wzy-mutant and complemented strains were extracted according to a previously established protocol[18] and visualized on SDS-PAGE by silver staining. Lane 1, the LPS from wild-type O86 displays a typical ladder-like pattern, a smooth LPS phenotype. Lane 2, the LPS from the mutant strain in which the *wzy* gene was deleted shows a semi-rough phenotype, in which only one repeating unit is attached to the Lipid A-core. Lane 3, the LPS from *wzy*-complemented strain shows a smooth phenotype, resembling that of the wild-type strain. This *in vivo* study confirmed that Wzy is involved in polymerization and that the recombinant *wzy* gene (with addition of C-terminal Myc and His tags) is fully functional.

References:

1. Datsenko, K.A., Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97,** 6640-6645 (2000).
2. Guo, H. *et al.* Overexpression and characterization of Wzz of Escherichia coli O86:H2. *Protein. Expr. Purif.* **48**, 49-55 (2006).
3. Yi, W. *et al.* Escherichia coli O86 O-antigen biosynthetic gene cluster and stepwise enzymatic synthesis of human blood group B antigen tetrasaccharide. *J. Am. Chem. Soc.* **127**, 2040-2041 (2005).
4. Hitchcock, P.J. & Brown, T.M. Morphological heterogeneity among Salmonella lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. *J. Bacteriol.* **154**, 269-77 (1983).
5. Wang, L. & Reeves, P.R. Involvement of the galactosyl-1-phosphate transferase encoded by the Salmonella enterica rfbP gene in O-antigen subunit processing. *J. Bacteriol.* **176**, 4348-56 (1994).
6. Lee, J. & Coward, J.K. Enzyme-catalyzed glycosylation of peptides using a synthetic lipid disaccharide substrate. *J. Org. Chem.* **57,** 4126-4135 (1992).
7. Sim, M.M., Kondo, H. & Wong, C.H. Synthesis of dibenzyl glycosyl phosphites using dibenzyl N,N-diethylphosphoramidite as phosphitylating reagent: an effective route to glycosyl phosphates, nucleotides, and glycosides. *J. Am. Chem. Soc.* **115**, 2260-7 (1993).
8. Ye, X.Y. *et al.* Better Substrates for Bacterial Transglycosylases. *J. Am. Chem. Soc.* **123**, 3155-3156 (2001).
9. Weerapana, E., Glover, K.J., Chen, M.C. & Imperiali, B. *J. Am. Chem. Soc.* **127**, 13766-13767 (2005).
10. Montoya-Peleaz, P.J. *et al.* Identification of a UDP-Gal:GlcNAc-R galactosyltransferase activity in Escherichia coli VW187. *Bioorgan. Med. Chem. Lett.* **15**, 1205-1211 (2005).
11. Ginsberg, C., Zhang, Y.H., Yuan, Y. & Walker, S. In vitro reconstitution of two essential steps in wall teichoic acid biosynthesis. *ACS Chem. Biol.* **1**, 25-28 (2006).
12. Branch, C.L., Burton, G. & Moss, S.F. An expedient synthesis of allylic polyprenyl phosphates. *Synth. Commun.* **29**, 2639-2644 (1999).
13. Abeyrathne, P.D. & Lam, J.S. WaaL of Pseudomonas aeruginosa utilizes ATP in in vitro ligation of O antigen onto lipid A-core. *Mol. Microbiol.* **65**, 1345-59 (2007).
14. Al-Dabbagh, B., Mengin-Lecreulx, D. & Bouhss, A. Purification and characterization of the bacterial UDP-GlcNAc:undecaprenyl-phosphate GlcNAc-1-phosphate transferase WecA. *J. Bacteriol.* **190**, 7141-6 (2008).
15. Zhang, Y. *et al.* Synthesis of heptaprenyl-lipid IV to analyze peptidoglycan glycosyltransferases. *J Am Chem Soc* **129**, 3080-1 (2007).
16. 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).
17. Yu, D. *et al.* An efficient recombination system for chromosome engineering in

Escherichia coli. *Proc. Natl. Acad. Sci. USA* **91.** 5978-5983 (2000).
18. Guo, H. et al. Molecular analysis of the O-antigen gene cluster of Escherichia coli O86:B7 and characterization of the chain length determinant gene (wzz). *Appl. Environ. Microbiol.* **71**, 7995-8001 (2005).