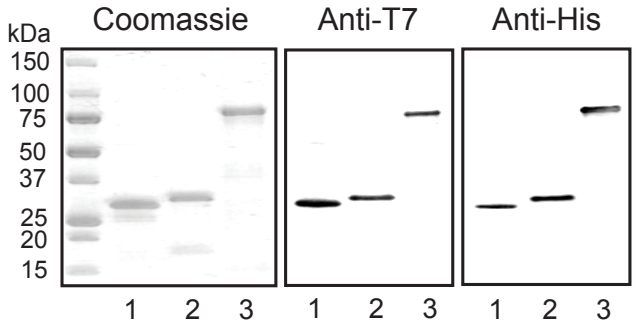
Supplementary Information for the manuscript entitled:

# Biochemical evidence for an alternate pathway in N-linked glycoprotein biosynthesis

A. Larkin, M.M. Chang, G.E. Whitworth, B. Imperiali

## Supplementary Results




**Supplementary Figure 1:** Coomassie-stained SDS-PAGE and Western blot analysis of purified proteins using antibodies specific to an N-terminal T7 or C-terminal His tag. (1) AglK (2) AglC (3) AglB.

**A**

```
AglK (1)   ------------------------------------------------------------MA
Alg5 (1)   MRALRFLIENRNTVFFTLLVALVLSLYLLVYLFSHTPRPPYPEELKYIADEKGHEVSRALPNLNEHQDDE

AglK (3)   DKLIYLIIPAYNEE----KMIKNVVNNLQNHN--YDNIIIVDDGSKDNTYKIMKELEEESKQNNNNNNN
Alg5 (71)  EIFLSVVIPSYNETGRILLMLTDAISFLKEKYGSRWEIVIVDDGSTDNTTQYCLKICKEQFKLN------

AglK (67)  NNNKVIAIKHEQNKGVGGATITGLKKAYELGADIAVTFDADGQHAPDDIAKVIQPIINDSKE--------
Alg5 (135) -YEQFRIIKFSQNRGKGGAVRQGF---LHIRGKYGLFADADGASKFSDVEKLIDAISKIETSSTDLKTTK

AglK (129) --YVVGSR--IKNPKEFKNMPLTKKVGNLGLSFITFLLGGYYVTDSQSGLRAFSKSALKVLSEQLRAKRYE
Alg5 (201) PAVAIGSRAHMVNTEAVIKRSMIRNCLMYGFHTLVFIFGIRSIKDTQCGFKLFNRAAILKIFPYLHTEGWI

AglK (196) TCSEALIIAKKNKLNIGEVPIKTIYTEYSMARGTNVMIGFKIFYRLLMLKMGKVLD--------
Alg5 (272) FDVEILILAIRKRIQIEEIPIS--WHEVDGSKMALAIDSIKMAKDLVIIRMAYLLGIYRDNKKC
```

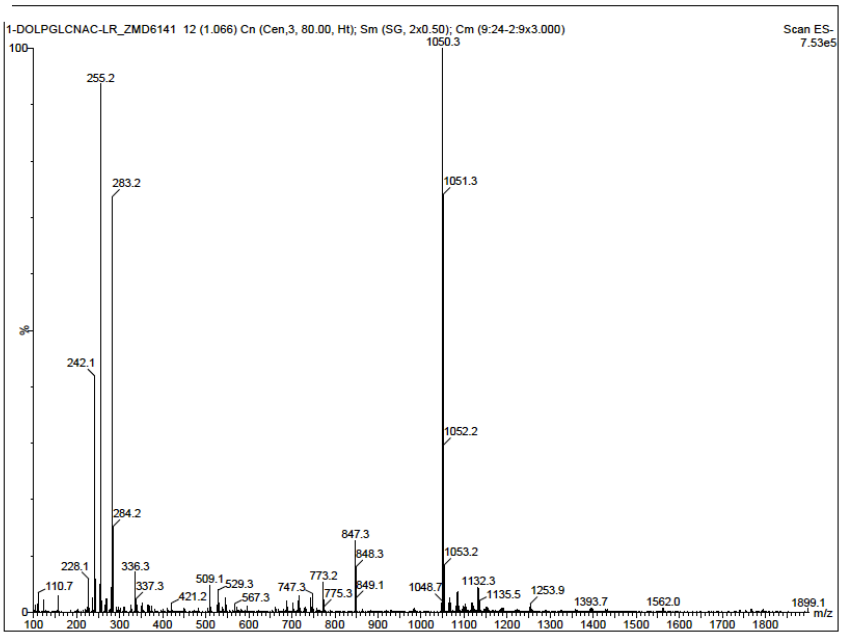
**B**

```
AglK (1)   -MADKLIYLIIPAYNEEKMIKNVVNNLQNHNYDNIIIVDDGSKDNTYKIMKELEEESKQNNNNNNN
AglJ (1)   MPTPDAVCILTPTYNEAETIADVISDYRDEGFANVLVIDGGSTDGTRELAEDAGAHVVVQSGSGKG

AglK (66)  NNNNKVIAIKHEQ-----NKGVGGATITGLKKAYELGAD-----IAVTFDADGQHAPDDIAKVIQP
AglJ (67)  QAVREAVEDHIQAPYVLMLDGDGTYEATDATKMLDPLTEGYDHVIGDRFADMRPGAMTRLNRVGNR

AglK (122) IIN---------DSKEYVVGSRIKNPKEFKNMPLTKK-----------VGNLGLSFITFLLGGYYV
AglJ (133) IINRAFAFIHGQDFRDILSGYRAFTRESFLDMTLTSDGFGIETEMAVECAKRGIKTTVVPTTYYPR

AglK (168) TD-SQSGLRAFSKSALKVLSEQLRAKR---------------YETCSEALIIAKKNKLNIGEVPIK
AglJ (199) PDGSDTNLDPIRDGGIIFLELYRRAKTNNPLFYFGSVGFASTATGLGLALYVAYEWVVRSISHEVI

AglK (218) TIYTEYSMARGTNVMIG---FKIFYRLLMLKMGKVLD--
AglJ (265) AVVSMAGILFGVQLLMFGVLSDLILSLHREQMKRIEELE
```

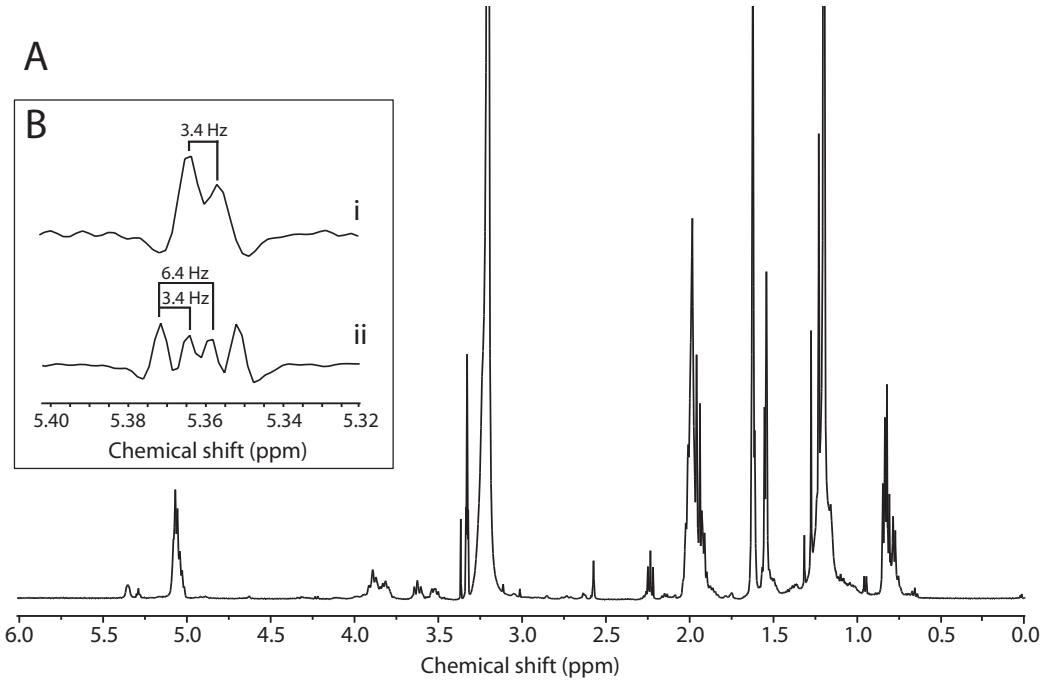
**C**

```
AglK     (1)   MADKLIYLIIPAYNEEKMIKNVVNNLQNHNYDNIIIVDDGSKDNTYKIMKELEEESKQNNNNN
MMP1170  (1)   MEKNDIFVVIPAYNEEKMIKNTLINLKSHGYENIIVVDDGSRDNTSKIAISEE----------

AglK     (64)  NNNNNNKVIAIKHEQNKGVGGATITGLKKAYELGADIAVTFDADGQHAPDDIAKVIQPIINDS
MMP1170  (54)  -------VIVCKHIINRGLGGALKTGLKCAVKYNPKVIVTFDADGQHDPEDIFKVSEPILEDS

AglK     (127) KEYVVGSRIKNPKEFKNMPLTKKVGNLGLSFITFLLGGYYVTDSQSGLRAFSKSALKVLSEQL
MMP1170  (110) FDVVVGSRLIDENELKNMPLIKKIGNWGLNFITYLMGGRMVTDSQGGLRAFSYDAAEIVSKQL

AglK     (190) RAKRYETCSEALIIAKKNKLNIGEVPIKTIYTEYSMARGTNVMIGFKIFYRLLMLKMGKVLD
MMP1170  (173) KSNRYEVSSEFIVLFKKNNLKFKEVPIKTIYTEYSMARGTNVITGFKILFKLLIQKLI----
```

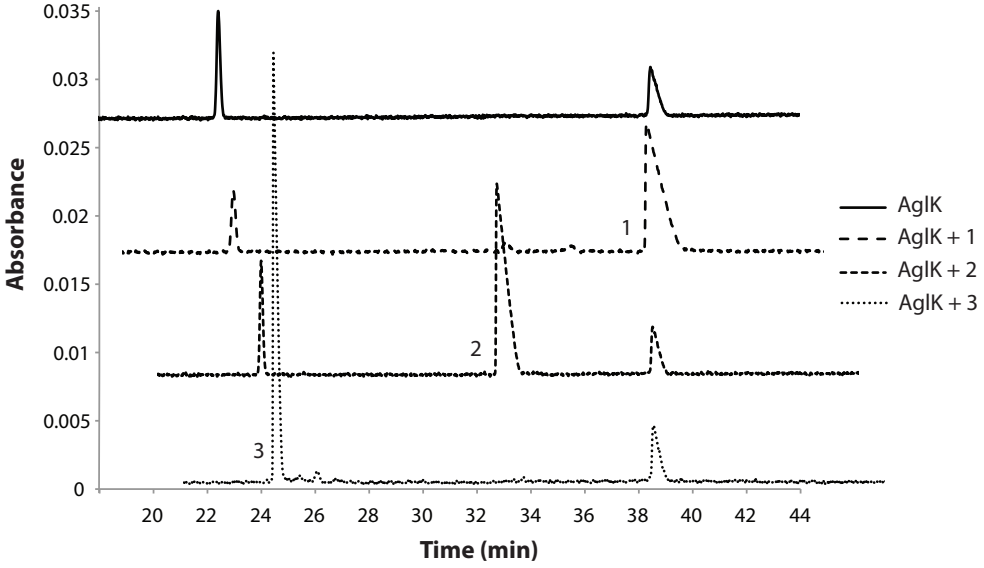
**Supplementary Figure 2:** Alignment of AglK from *M. voltae* with (**A**) the dolichyl-phosphate β-glucosyltransferase (DPG) Alg5 from *S. cerevisiae*, (**B**) AglJ from *H. volcanii*, or (**C**) MMP1170 from *M. maripaludis*. Residues highlighted in black indicate sequence identity (25% for Alg5, 18% for AglJ, and 60% for MMP1170) and those in gray and black denote sequence similarity (53% for Alg5, 43% for AglJ, and 79% for MMP1170). The alignment was performed using ClustalW.



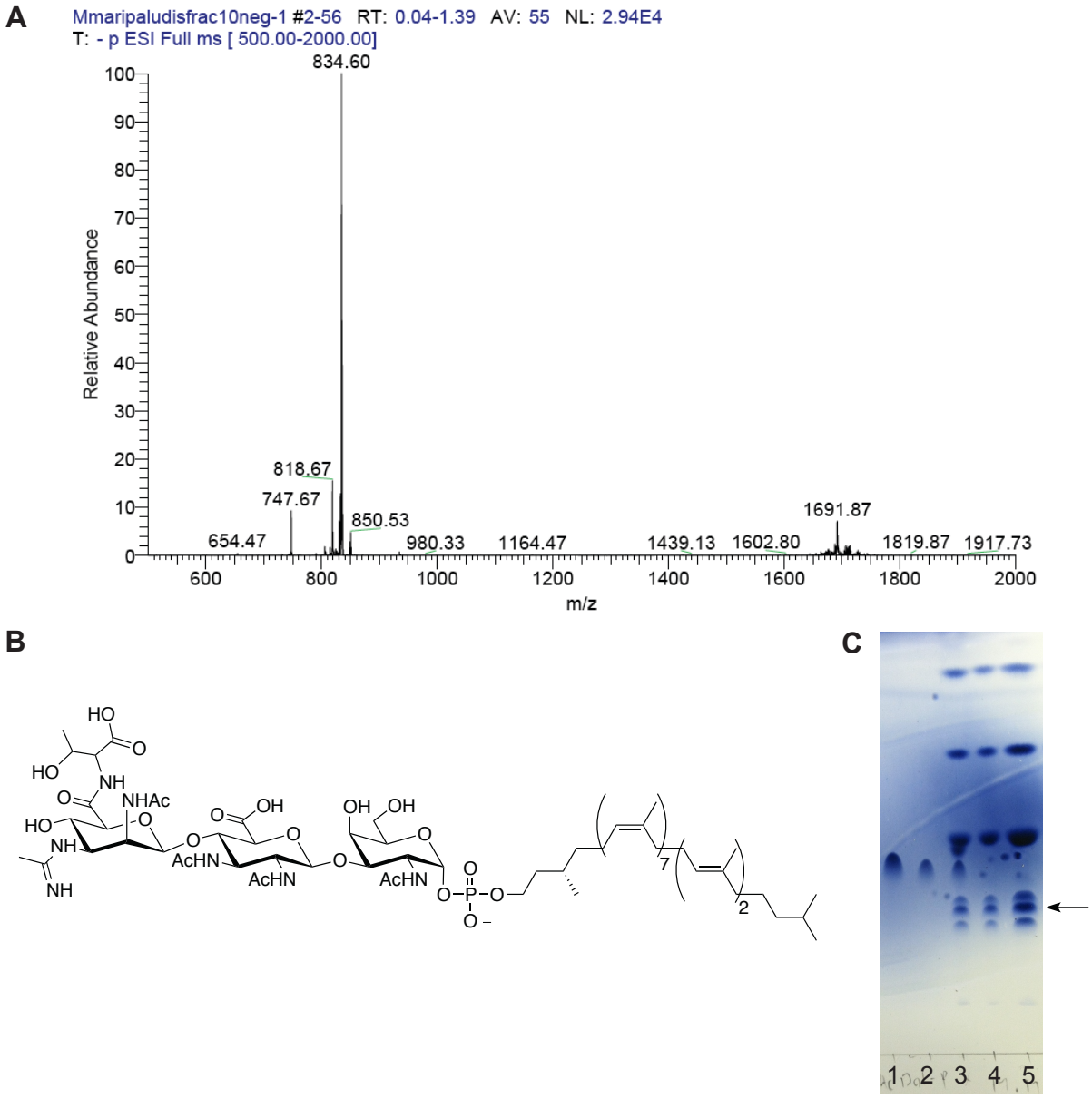

**Supplementary Figure 3**: ESI-MS (negative ion mode) of purified Dol-P-GlcNAc (calcd. $[M-H]^-$ = 1050.8). The (C55) Dol-linked product was enriched during HPLC purification



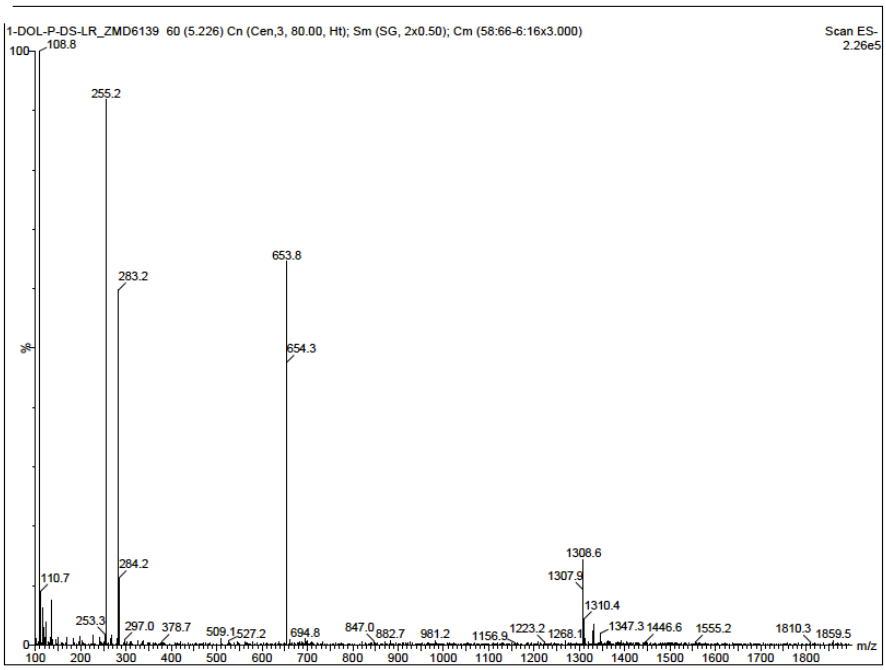

**Supplementary Figure 4**: (**A**) $^{1}$H NMR spectrum of Dol-P-GlcNAc. (**B**) (i) Expansion of the $^{31}$P decoupled $^{1}$H NMR spectrum of the anomeric proton for Dol-P-GlcNAc, a doublet with a coupling constant of 3.4 Hz ($J_{1,2}$). (ii) Expansion of the $^{1}$H NMR spectrum of the anomeric proton for Dol-P-GlcNAc, a doublet of doublets with coupling constants of 3.4 ($J_{1,2}$) and 6.4 ($J_{1,P}$) Hz.



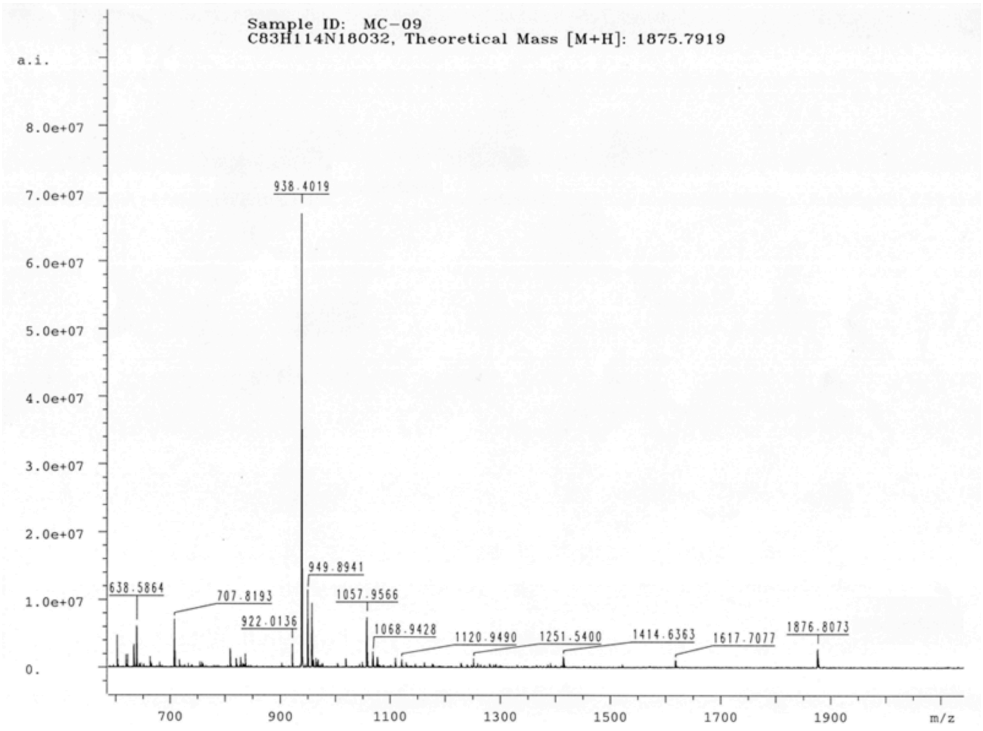

**Supplementary Figure 5**: Capillary electrophoresis of AglK reaction products after aqueous extractions of the reaction mixture in order to identify the nucleotide product as UDP. In each case, (1) UDP, (2) UMP, or (3) UDP-GlcNAc was spiked into the reaction mixture in order to identify the nucleotide peak. CE was performed on a Hewlett-Packard 3D CE system. A bare silica capillary (75 μm x 80 cm) was used with a detector distance of 72 cm. The running buffer was 25 mM sodium tetraborate (pH 9.5). The capillary was conditioned before each run with a 0.4 M NaOH wash for 2 min, $H_2O$ for 2 min, and running buffer for 2 min. Samples were introduced by pressure injection for 16 s at 30 mbar, and the separation was performed at 25 kV and monitored at 254 nm.



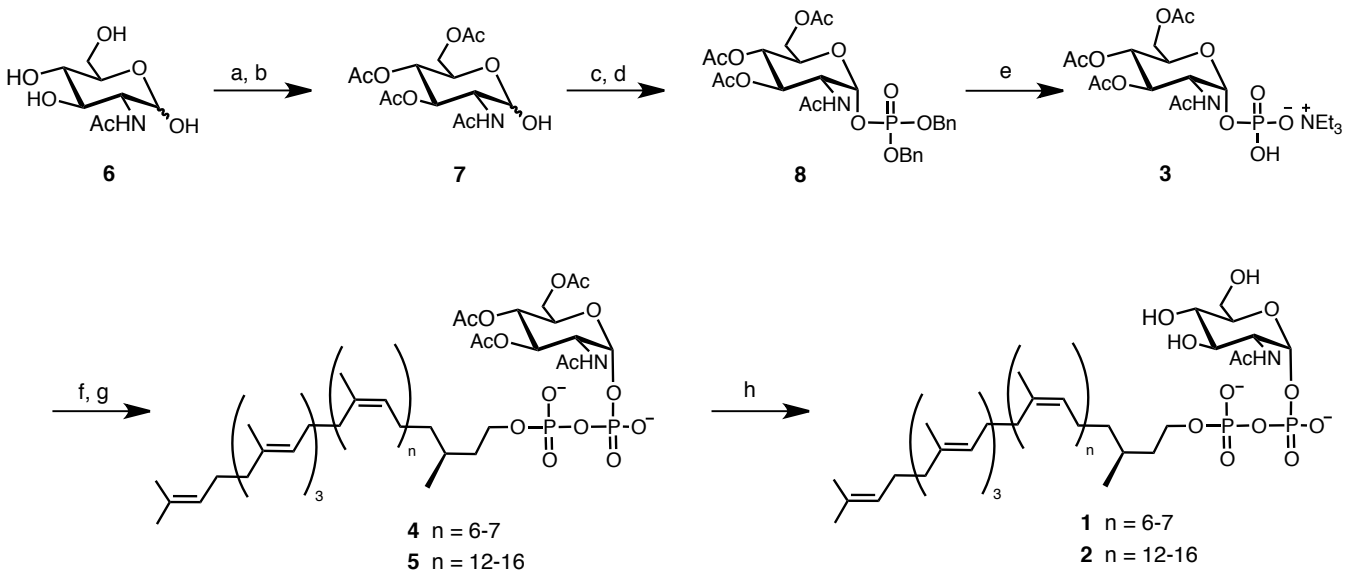

**Supplementary Figure 6**: (**A**) ESI-MS (negative ion mode) of *M. maripaludis* lipid extraction after HPLC purification, where calcd. $[M\text{-}2H+Na]^-$ = 1692.0 and $[M\text{-}2H]^{2-}$ = 834.5. (**B**) Structure of the α- and ω-saturated Dol-P-trisaccharide, ManNAc3NAm(6Thr)A-β1,4-Glc-2,3-diNAcA-β1,3-GalNAc-P-Dol. (**C**) TLC of (C55) Dol-P (1-2), *M. maripaludis* lipid extract (4-5), and co-spot (3). The Dol-P-trisaccharide purified by HPLC is indicated by an arrow.




**Supplementary Figure 7**: ESI-MS (negative ion mode) of Dol-P-GlcNAc-Glc-2,3-diNAcA (calcd. $[M-H]^-$ = 1308.8).



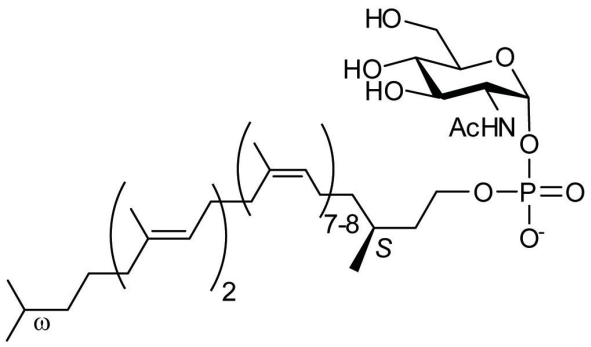

**Supplementary Figure 8**: ESI-MS (positive ion mode) of the glycopeptide produced by AglB (calcd. $[M+H]^+$ = 1875.8).



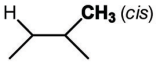
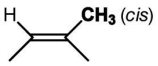
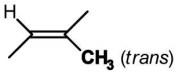
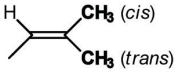

**Supplementary Figure 9**: Synthesis of (C55-60) Dol-PP-GlcNAc (1) and (C85-105) Dol-PP-GlcNAc. Reagents and conditions: (a) $Ac_2O$, pyr; (b) $HN(CH_3)_2$, $CH_3CN$; (c) dibenzyl *N,N*-phosphoramidite, 1,2,4-triazole, $CH_2Cl_2$; (d) 1% $H_2O_2$, -78 °C; (e) $Pd(OH)_2/C$, $H_2$, $NEt_3$; (f) CDI, DMF; (g) MeOH, then Dol-P (C55-60 or C85-105), $CH_2Cl_2$; (h) 1% NaOMe/MeOH.

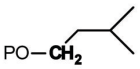
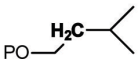
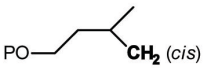
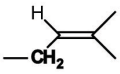
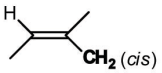
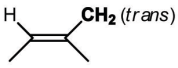
**Supplementary Table 1**: (**A**) Dol-P-GlcNAc (**B**) $^{1}H$ and $^{13}C$ NMR assignments (ppm) and coupling constants (Hz) for the dolichyl portion of Dol-P-GlcNAc. (**C**) $^{1}H$ and $^{13}C$ NMR assignments (ppm) and coupling constants (Hz) for the GlcNAc portion of Dol-P-GlcNAc.

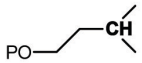
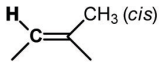
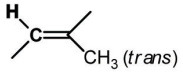
**A**

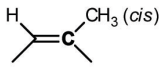
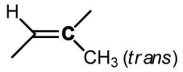
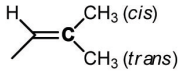



**B** **Dolichyl-Phosphate**

| Isoprene unit | Methyl groups | δC | δH |
|---|---|---|---|
| α |  | 20.0 | 0.79 |
| β to ω-3 |  | 23.2 | 1.55 |
| ω-1, ω-2 |  | 16.0, 17.6 | 1.55 |
| ω |  | 25.8 (*cis*), 15.5 (*trans*) | 1.55 |

| Isoprene unit | Methylene group | δC | δH |
|---|---|---|---|
| α |  | 64.9 | 3.86 |
| α |  | 37.5 | 1.29 |
| α |  | 37.9 | 1.39, 1.62 |
| β to ω |  | 26.6 | 2.01 |
| β to ω-3 |  | 32.1 | 1.98 |
| ω-1, ω-2 |  | 39.7 | 1.94 |

| Isoprene unit | Methine groups | δC | δH |
|---|---|---|---|
| α |  | 29.1 | 1.28 |
| β to ω-3 |  | 124.4 | 5.06 |
| ω to ω-2 |  | 129.9 | 5.09 |

| Isoprene unit | Quaternary carbons | δC |
|---|---|---|
| α |  | 135.5 |
| β to ω-3 |  | 134.9 |
| ω to ω-2 |  | 131.3 |

**C** **GlcNAc**

| Proton | δH (Hz) | Carbon | δC |
|---|---|---|---|
| H-1 ($J_{1,2}$) | 5.36 (3.4) | C-1 | 94.2 |
| H-1 ($J_{1,P}$) | 5.36 (6.4) | | |
| H-2 ($J_{2,3}$) | 3.89 (9.4) | C-2 | 54.0 |
| H-3 ($J_{3,4}$) | 3.63 (10.0) | C-3 | 71.5 |
| H-4 | 3.26 | C-4 | 67.7 |
| H-5 | 3.82 | C-5 | 73.7 |
| H-6 ($J_{5,6}$) | 3.53 (7.6) | C-6 | 62.3 |
| H-6 ($J_{6,6'}$) | 3.53 (9.9) | | |
| H-6' | 3.91 | | |
| | | C=O | 171.3 |
| $CH_3$ | 1.96 | | 22.6 |