

Rationally Designed Short Polyisoprenol-Linked PglB Substrates for Engineered Polypeptide and Protein N-Glycosylation

Feng Liu,^{†,§} Balakumar Vijaykrishnan,^{†,§} Amirreza Faridmoayer,^{‡,§} Thomas A. Taylor,[†] Thomas B. Parsons,[†] Gonçalo J.L. Bernardes,[†] Michael Kowarik,[‡] and Benjamin G. Davis^{*,†}

[†]Department of Chemistry, Chemistry Research Laboratory, Oxford University, Oxford OX1 3TA United Kingdom

[‡]GlycoVaxyn AG, 8952 Schlieren, Switzerland

Supporting Information

ABSTRACT: The lipid carrier specificity of the protein N-glycosylation enzyme *C. jejuni* PglB was tested using a logical, synthetic array of natural and unnatural C10, C20, C30, and C40 polyisoprenol sugar pyrophosphates, including those bearing repeating *cis*-prenyl units. Unusual, short, synthetically accessible C20 prenols (nerylnerol **1d** and geranylnerol **1e**) were shown to be effective lipid carriers for PglB sugar substrates. Kinetic analyses for PglB revealed clear K_M -only modulation with lipid chain length, thereby implicating successful *in vitro* application at appropriate concentrations. This was confirmed by optimized, efficient *in vitro* synthesis allowing >90% of Asn-linked β -N-GlcNAc-ylated peptide and proteins. This reveals a simple, flexible biocatalytic method for glycoconjugate synthesis using PglB N-glycosylation machinery and varied chemically synthesized glycosylation donor precursors.

Protein glycosylation is a vital co- or post-translational modification that links glycans to proteins typically through N- or O- linkages. Such modifications exist widely in eukaryotic and archaeal organisms and greatly expand the diversity of the proteome.^{1–3} While N-linked glycosylation had been believed to be absent in prokaryotic systems, the discovery of the protein N-glycosyltransferase PglB in Gram-negative bacterium *Campylobacter jejuni* highlighted greater diversity; PglB is responsible for glycosylating over 50 different proteins.^{4–7} Since then PglB orthologues have been found in *Campylobacter lari*,⁸ the *Helicobacter* genus (*H. pullorum*, *H. canadensis*, *H. winghamensis*)⁹ as well as *Desulfovibrio desulfuricans*¹⁰ and, very recently, *Methanococcus voltae*.¹¹ *In vivo* PglB uses a C55 undecaprenylpyrophosphate-linked oligosaccharide as its substrate and glycosylates the primary amide nitrogen of the asparagine side chain in a D/E-X-N-X-T/S consensus sequence (where X can be any amino acid except proline, Figure 1). In contrast to its eukaryotic counterpart Stt3p, PglB does not require other proteins/subunits and can apparently catalyze this N-glycosylation *in vivo* alone.¹² Moreover, *in vivo* biosynthetic studies that allowed exposure of PglB to other lipid-linked oligosaccharides have suggested a possibly relaxed specificity toward the nature of glycan substrate, as compared to the relatively highly specific eukaryotic N-glycosylation enzymes.¹³ Recent elegant approaches flexibly using PglB *in vivo* have been developed to prepare glycoproteins.^{14,15} These distinct characteristics of PglB, as well

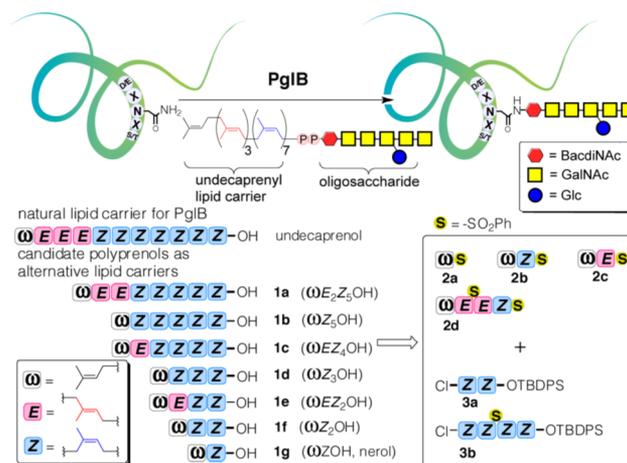


Figure 1. Bacterial N-linked glycosylation and designed unnatural candidate polyisoprenols **1a–g** as alternative short lipid carriers for PglB-catalyzed glycosylation. The polyisoprenols were synthesized from building blocks **2a–d** and **3a** and **3b**.

as the fact that *C. jejuni* PglB can be readily overexpressed in functional form in *Escherichia coli*,¹⁶ highlight PglB's potential as a synthetic biocatalyst. They suggest it as a potentially ideal model from which to generate a ready synthetic system for *in vitro* protein glycosylation. However, the donor substrates used normally by PglB *in vivo* (C55 lipid pyrophosphoryl-linked oligosaccharides containing the rare, bacterial sugar bacillosamine (Bac)) would restrict this system (both in substrate accessibility and product relevance).

In an attempt to optimize the PglB protein N-glycosylation platform for practical, synthetic (and hence *in vitro*) use, we designed an array of chemically generated polyisoprenol variants to find those simpler and shorter than the natural undecaprenol (Figure 1) and that might serve as alternative lipid carriers that could be recognized by PglB. Insightful prior work has elucidated some aspects of the polyisoprenol specificity of PglB.^{17,18} However, estimated conversions for these reactions were $\leq 20\%$. In addition, many of these prior substrates were prepared enzymatically from polyisoprenols isolated from natural sources, enabling only nmol analysis of lipid pyrophosphates containing the natural *Campylobacter* (GalNAc-GalNAc-BacdiNAc) gly-

Received: September 11, 2013

Published: December 16, 2013

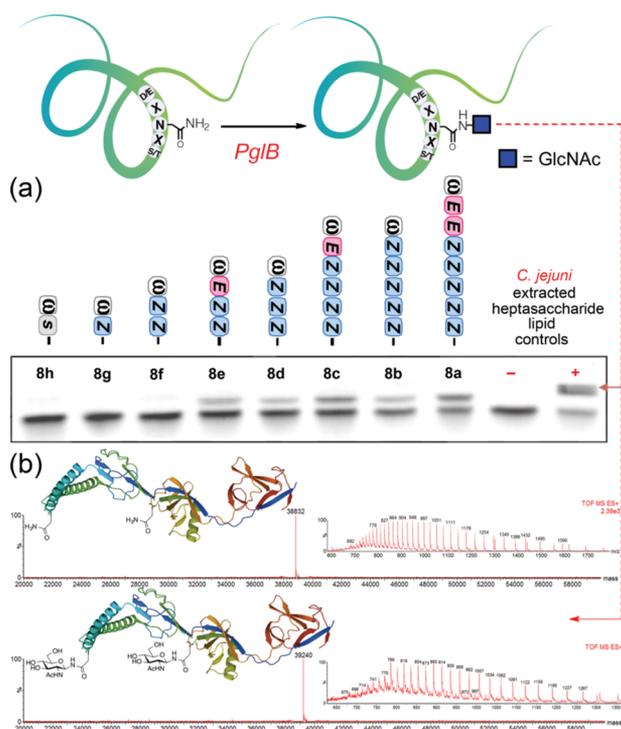


Figure 2. Peptide and protein glycosylation. (a) Fluorescent electrophoretic analysis of peptide glycosylation with lipids (**8a–h**) (– = no lipid; + = lipid extracted and enriched from *E. coli* cells producing *C. jejuni* heptasaccharide-linked undecaprenyl pyrophosphate); [glycolipid] = [peptide] 20 μM ; [PglB] = 0.44 μM . (b) ES-MS of *in vitro* N-linked protein (AcrA) glycosylation with **8c**; >95% diglycosylation.

in the *in vitro* assay. These activities ‘mapped’ tight proximal-site activity and relaxed distal-site activity consistent with model shown in Figure S12.

These first kinetic parameters (Table 1) for PglB suggested key features. In particular, the variation of activity with lipid length in the substrate is strikingly only dependent on K_M ; k_{cat} remains essentially unaltered. This suggests that the lipid may not play a primary role in catalytic turnover but is a key regulator of substrate uptake. This suggested too that *in vitro* reactions conducted at sufficiently high concentrations $>K_M$ would allow transfer efficiencies equal to those found for full length lipid substrates. This was valuably confirmed in synthetic reactions that allowed the synthesis of GlcNAc-ylated glycopeptide in yields >90% using 0.1 mM glycosyl donor substrate **8c** with 20 μM of acceptor peptide. These >90% reactions usefully extend^{17,18} the synthetic utility of PglB.

Having elucidated valuable plasticity toward unnatural lipid-variant substrates, we next examined glycan breadth beyond the atypical monosaccharide GlcNAc already demonstrated. Conjugates (**13** and **14**) that contain both unnatural sugar and lipid carrier (Scheme 2) were prepared by coupling nerylnerylphos-

phate (**5d**) with the 6-azido-GlcNAc (**9**) and 2-azidoGlcNAc (**10**, GlcNAz). These compounds would allow subsequent flexible postexpressional modifications on proteins that contain D/E-X-N-X-S/T tag via reaction of the introduced azide with a number of compatible methods. Both azido analogues (**13** and **14**) failed to undergo glycosylation with the peptide in the presence of PglB. The failure of these glycolipids to act as substrates may be explained by the use of nonpreferred moieties in both halves of these unnatural sugar-unnatural lipid conjugates. It may also suggest a particular lack of plasticity by PglB toward alterations at sugar sites C-2 or C-6; glycolipids **13** and **14** did not inhibit PglB glycosylations using **8c** (see SI).

GlcNAc-ylated peptide could then be extended (Scheme 2) with the use of endoglycosidases^{35,36} (EndoS,³⁷ EndoA,^{15,38,39} Scheme 2) and glycosyltransferases (β -1,4-GalT).⁴⁰ This allowed ready access to differently elaborated glycopeptides bearing, e.g., LacNAc (94%), or the eukaryotic N-glycan core-pentasaccharide (64%). Finally, we then tested the ability of this system to glycosylate proteins. Using short GlcNAc-lipids **8c** and **8e**, we could effect *in vitro* glycosylation >95% at the two consensus Asn sites in the *C. jejuni* AcrA¹⁴ protein.

In conclusion, a wide variety of *cis*-polyisoprenol variants were chemically synthesized and studied for their binding specificity against PglB. For the first time, LPPS's with only a single sugar and lipid chain lengths as short as C20 and C30 have been shown to be effective substrates for PglB in glycosylating specific peptide motifs. This reveals unexpected breadth for PglB beyond the minimal C40 lipid-trisaccharide substrate determined previously.¹⁷ Our experimental catalytic data are consistent with previous crystallographic and modeling analyses. A closer examination of the lipid-binding pocket also reveals a relationship between the PglB structure and the chain length requirements that we have discovered here. The narrow pocket that precedes the hydrophobic groove is surrounded by polar residues Ser198, Ser201, Arg375. Longer lipid chains beyond the third isoprenyl unit may be required for increased affinity, by favorable interaction with the hydrophobic groove. This explains the observation that **8e** was an active substrate, but **8f** was not. Similar studies on the lipid carrier specificity of MurG have however shown a quite different trend, in which nerol (**1g**) and nerylnerol (**1d**) conjugates were much better substrates than those of longer lipids bearing repeating *cis* units.^{24,41,42} Investigation of the crystal structures of MurG and/or PglB with lipid carrier bound, once available, would shed light on these clear differences in lipid carrier specificities.

The discovery of the breadth of PglB and these accessible lipid carriers now effectively enables the synthesis of lipid-pyrophosphate-linked substrates suitable for the *in vitro* generation of tailor-made glycoproteins. Importantly, to our knowledge, this currently represents the only *in vitro* biocatalytic system for the formation of the vital GlcNAc- β -1-N-Asn linkage (and importantly can be driven to >95% on proteins, here for AcrA); the recently discovered *Methanococcus* AglB, e.g.,¹¹ does

Table 1. Kinetic Parameters for PglB with GlcNAc Lipids^a

glycolipid substrate	k_{cat} [min^{-1}]	K_M [mM]	k_{cat}/K_M [$\text{min}^{-1}\text{mM}^{-1}$]
8d GlcNAc-PP- ω Z ₃	0.0234 \pm 0.0021	0.077 \pm 0.016	0.30
8c GlcNAc-PP- ω EZ ₄	0.0231 \pm 0.0016	0.055 \pm 0.01	0.42
8a GlcNAc-PP-prenol ₃	0.0225 \pm 0.0021	0.034 \pm 0.01	0.66

^aUPLC using TAMRA-fluorescence intensity; substrate concentrations: [glycolipid] = 1, 5, 10, 50, 100, 200 μM , [peptide] = 20 μM ; enzyme concentration [PglB] = 0.44 μM ; reaction time <2h, 30 $^{\circ}\text{C}$; all conducted in duplicate.

not transfer GlcNAc and requires an unusual disaccharide. This discovery complements prior GalNAc transfer and confirms a predicted activity.¹⁸ Notably, WecA,⁴³ the enzyme that would generate PP-linked glycolipid substrates for PglB, is membrane-associated and cannot be readily exploited *in vitro*. *In vitro* biocatalytic installation of GlcNAc, shown here, also creates a useful precursor sugar site for carbohydrate-processing enzyme-mediated extension, as shown here (Scheme 2). The >90% *in vitro* efficiencies shown here therefore make PglB a highly viable synthetic biocatalyst for varied glycopolypeptides, coupled with substrate accessibility and potential for further enzymatic transformation.

■ ASSOCIATED CONTENT

Supporting Information

Full procedures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

Ben.Davis@chem.ox.ac.uk

Author Contributions

[§]These authors contributed equally.

Notes

The authors declare the following competing financial interest(s): M.K. and A.F. are employees of Glycovaxyn. A patent has been filed and will afford inventors royalties, if licensed, in line with university guidelines.

■ ACKNOWLEDGMENTS

We thank the BBSRC (T.A.T.), Glycovaxyn (F.L., B.V., B.G.D.) for funding, Drs P. Carranza, C. Bich for method development and glycopeptide analysis, and Drs L. Gong, J. Wickens for HRMS analysis. B.G.D. is a Royal Society Wolfson Merit Research Award Recipient.

■ REFERENCES

- (1) Spiro, R. G. *Glycobiology* **2002**, *12*, 43R.
- (2) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. *Angew. Chem., Int. Ed.* **2005**, *44*, 7342.
- (3) Ohtsubo, K.; Marth, J. D. *Cell* **2006**, *126*, 855.
- (4) Szymanski, C. M.; Yao, R.; Ewing, C. P.; Trust, T. J.; Guerry, P. *Mol. Microbiol.* **1999**, *32*, 1022.
- (5) Young, N. M.; Brisson, J. R.; Kelly, J.; Watson, D. C.; Tessier, L.; Lanthier, P. H.; Jarrell, H. C.; Cadotte, N.; St Michael, F.; Aberg, E.; Szymanski, C. M. *J. Biol. Chem.* **2002**, *277*, 42530.
- (6) Linton, D.; Allan, E.; Karlyshev, A. V.; Cronshaw, A. D.; Wren, B. *W. Mol. Microbiol.* **2002**, *43*, 497.
- (7) Scott, N. E.; Parker, B. L.; Connolly, A. M.; Paulech, J.; Edwards, A. V. G.; Crossett, B.; Falconer, L.; Kolarich, D.; Djordjevic, S. P.; Hojrup, P.; Packer, N. H.; Larsen, M. R.; Cordwell, S. J. *Mol. Cell. Proteomics* **2011**, *10*, M000031.
- (8) Schwarz, F.; Lizak, C.; Fan, Y.-Y.; Fleurkens, S.; Kowarik, M.; Aebi, M. *Glycobiology* **2011**, *21*, 45.
- (9) Jervis, A. J.; Langdon, R.; Hitchen, P.; Lawson, A. J.; Wood, A.; Fothergill, J. L.; Morris, H. R.; Dell, A.; Wren, B.; Linton, D. *J. Bacteriol.* **2010**, *192*, 5228.
- (10) Ielmini, M. V.; Feldman, M. F. *Glycobiology* **2011**, *21*, 734.
- (11) Larkin, A.; Chang, M. M.; Whitworth, G. E.; Imperiali, B. *Nat. Chem. Biol.* **2013**, *9*, 367.
- (12) Glover, K. J.; Weerapana, E.; Numao, S.; Imperiali, B. *Chem. Biol.* **2005**, *12*, 1311.
- (13) Wacker, M.; Feldman, M. F.; Callewaert, N.; Kowarik, M.; Clarke, B. R.; Pohl, N. L.; Hernandez, M.; Vines, E. D.; Valvano, M. A.; Whitfield, C.; Aebi, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7088.
- (14) Feldman, M. F.; Wacker, M.; Hernandez, M.; Hitchen, P. G.; Marolda, C. L.; Kowarik, M.; Morris, H. R.; Dell, A.; Valvano, M. A.; Aebi, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3016.
- (15) Schwarz, F.; Huang, W.; Li, C.; Schulz, B. L.; Lizak, C.; Palumbo, A.; Numao, S.; Neri, D.; Aebi, M.; Wang, L.-X. *Nat. Chem. Biol.* **2010**, *6*, 264.
- (16) Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-Lazar, M.; Haslam, S. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M. *Science* **2002**, *298*, 1790.
- (17) Chen, M. M.; Weerapana, E.; Ciepichal, E.; Stupak, J.; Reid, C. W.; Swiezewska, E.; Imperiali, B. *Biochemistry* **2007**, *46*, 14342.
- (18) Li, L.; Woodward, R.; Ding, Y.; Liu, X.-w.; Yi, W.; Bhatt, V. S.; Chen, M.; Zhang, L.-w.; Wang, P. G. *Biochem. Biophys. Res. Commun.* **2010**, *394*, 1069.
- (19) Sato, K.; Inoue, S.; Onishi, A.; Uchida, N.; Minowa, N. *J. Chem. Soc., Perkin Trans. 1* **1981**, 761.
- (20) Sato, K.; Miyamoto, O.; Inoue, S.; Furusawa, F.; Matsushashi, Y. *Chem. Lett.* **1983**, *12*, 725.
- (21) Jaenicke, L.; Siegmund, H. *Biol. Chem. Hoppe-Seyler* **1986**, *367*, 787.
- (22) Inoue, S.; Kaneko, T.; Takahashi, Y.; Miyamoto, O.; Sato, K. *J. Chem. Soc., Chem. Commun.* **1987**, 1036.
- (23) Grassi, D.; Lippuner, V.; Aebi, M.; Brunner, J.; Vasella, A. *J. Am. Chem. Soc.* **1997**, *119*, 10992.
- (24) Chen, L.; Men, H.; Ha, S.; Ye, X.-Y.; Brunner, L.; Hu, Y.; Walker, S. *Biochemistry* **2002**, *41*, 6824.
- (25) Chang, Y.-F.; Liu, C.-Y.; Guo, C.-W.; Wang, Y.-C.; Fang, J.-M.; Cheng, W.-C. *J. Org. Chem.* **2008**, *73*, 7197.
- (26) Meng, F.-C.; Chen, K.-T.; Huang, L.-Y.; Shih, H.-W.; Chang, H.-H.; Nien, F.-Y.; Liang, P.-H.; Cheng, T.-J. R.; Wong, C.-H.; Cheng, W.-C. *Org. Lett.* **2011**, *13*, 5306.
- (27) Heseck, D.; Lee, M.; Zajčček, J.; Fisher, J. F.; Mobashery, S. *J. Am. Chem. Soc.* **2012**, *134*, 13881.
- (28) Shih, H.-W.; Chang, Y.-F.; Li, W.-J.; Meng, F.-C.; Huang, C.-Y.; Ma, C.; Cheng, T.-J. R.; Wong, C.-H.; Cheng, W.-C. *Angew. Chem., Int. Ed.* **2012**, *51*, 10123.
- (29) Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P. *Nature* **2011**, *474*, 350.
- (30) Jaenicke, L.; Siegmund, H.-U. *Chem. Phys. Lipids* **1989**, *51*, 159.
- (31) Danilov, L. L.; Druzhinina, T. N.; Kalinchuk, N. A.; Maltsev, S. D.; Shibaev, V. N. *Chem. Phys. Lipids* **1989**, *51*, 191.
- (32) Bernardes, G. J. L.; Davis, B. G. unpublished results, University of Oxford, 2007.
- (33) Bernardes, G. J. L.; Kikkeri, R.; Magliano, M.; Laurino, P.; Collot, M.; Hong, S. Y.; Lepenies, B.; Seeberger, P. H. *Org. Biomol. Chem.* **2010**, *8*, 4987.
- (34) Schagger, H. *Nat. Protoc.* **2006**, *1*, 16.
- (35) Wang, L.-X. *Carbohydr. Res.* **2008**, *343*, 1509.
- (36) Rich, J. R.; Withers, S. G. *Nat. Chem. Biol.* **2009**, *5*, 206.
- (37) Goodfellow, J. J.; Baruah, K.; Yamamoto, K.; Bonomelli, C.; Krishna, B.; Harvey, D. J.; Crispin, M.; Scanlan, C. N.; Davis, B. G. *J. Am. Chem. Soc.* **2012**, *134*, 8030.
- (38) Fernández-González, M.; Boutureira, O.; Bernardes, G. J.; Chalker, J. M.; Young, M.; Errey, J.; Davis, B. G. *Chem. Sci.* **2010**, *1*, 709.
- (39) Lomino, J. V.; Naegeli, A.; Orwenyo, J.; Amin, M. N.; Aebi, M.; Wang, L.-X. *Bioorg. Med. Chem.* **2013**, *21*, 2262.
- (40) Gamblin, D. P.; Garnier, P.; van Kasteren, S.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. *Angew. Chem., Int. Ed.* **2004**, *43*, 828.
- (41) Ye, X.-Y.; Lo, M.-C.; Brunner, L.; Walker, D.; Kahne, D.; Walker, S. *J. Am. Chem. Soc.* **2001**, *123*, 3155.
- (42) Liu, C.-Y.; Guo, C.-W.; Chang, Y.-F.; Wang, J.-T.; Shih, H.-W.; Hsu, Y.-F.; Chen, C.-W.; Chen, S.-K.; Wang, Y.-C.; Cheng, T.-J.; Ma, C.; Wong, C.-H.; Fang, J.-M.; Cheng, W.-C. *Org. Lett.* **2010**, *12*, 1608.
- (43) Lehrer, J.; Vigeant, K. A.; Tatar, L. D.; Valvano, M. A. *J. Bacteriol.* **2007**, *189*, 2618.