**Functional analysis of *N*-linking oligosaccharyl transferase enzymes encoded by deep-sea vent proteobacteria**

**Dominic C Mills**1,3**, Adrian J Jervis**2**, Sherif Abouelhadid**1**, Laura E Yates**1,3**, Jon Cuccui**1**, Dennis Linton**2**, \*Brendan W Wren**1

1London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT. UK

2Faculty of Life Sciences, Michael Smith Building, University of Manchester, Manchester M13 9PT, UK

3 Present address: School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, 14853, USA

\*Corresponding author. Mailing address: Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, University of London, Keppel Street, London WC1E 7HT. Tel: +44 (0) 207 927 2288, e-mail: brendan.wren@lshtm.ac.uk.

Running head: Novel bacterial N-linking oligosaccharyl transferases

Supplementary material:

Three figures, 1 phylogenetic tree (S1), 1 amino acid alignment (S2) and one additional western blot (S3)

**Abstract**

Bacterial *N-*linking oligosaccharyl transferases (OTase enzymes) transfer lipid-linked glycans to selected proteins in the periplasm and were first described in the intestinal pathogen *Campylobacter jejuni,* a member of the ε-proteobacteria-subdivision of bacteria. More recently, orthologues from other ε-proteobacterial *Campylobacter* and *Helicobacter* species and a δ-proteobacterium, *Desulfovibrio desulfuricans,* have been described, suggesting that these two subdivisions of bacteria may be a source of further *N*-linked protein glycosylation systems. Whole-genome sequencing of both ε- and δ-proteobacteria from deep-sea vent habitats, a rich source of species from these subdivisions, revealed putative ORFs encoding OTase enzymes and associated adjacent glycosyltransferases similar to the *C. jejuni* *N*-linked glycosylation locus. We expressed putative OTase ORFs from the deep-sea vent species *Nitratiruptor tergarcus, Sulfurovum lithotrophicum and Deferribacter desulfuricans* in *Escherichia coli* and showed they were able to functionally complement the *C. jejuni* OTase, CjPglB . The enzymes were shown to possess relaxed glycan specificity, transferring diverse glycan structures and demonstrated different glycosylation sequon specificities.Additionally a permissive *D. desulfuricans* acceptor protein was identified, and we provide evidence that the *N-*linked glycan synthesised by *N. tergarcus* and *S. lithotrophicum* contains an acetylated sugar at the reducing end. This work demonstrates that deep-sea vent bacteria encode functional *N-*glycosylation machineries and are a potential source of biotechnologically important OTase enzymes.

**Introduction**

Asparagine-linked glycosylation is a common post-translational modification of eukaryotic proteins, and is involved in many cellular functions such as quality control, protein folding and secretion (Helenius & Aebi 2004). The eukaryotic *N*-glycosylation machinery is located within the endoplasmic reticulum (ER), where the glycan is assembled on the lipid carrier dolichyl pyrophosphate in the membrane by the action of several glycosyl transferases, flipped across the membrane and finally transferred to an acceptor protein at the consensus sequon N-X-S/T, where X can be any amino acid except proline (Aebi 2013). This transfer is accomplished by the action of the oligosaccharyltransferase (OTase) enzyme complex. Initially believed to be limited to eukaryotes, *N-*linked glycosylation was also identified in Archaea, where a surface layer (S-layer) protein was shown to be *N*-glycosylated in *Halobacterium salinarum* (Mescher & Strominger 1976). Further investigation revealed that this type of protein modification was present in many Archaeal species, displaying a variety of *N-*glycan structures (Eichler 2013). The S-layer protein is the best characterised Archaeal glycoprotein, but other proteins have been identified such as the ABC transporter SSO1273 in *Sulfolobus solfataricus* P2 (Palmieri et al. 2013) and an unusual type IV pilus protein in *Methanococcus maripaludis* (Ng et al. 2011). It was not until 1999 that the first bacterial *N-*linked protein glycosylation system was identified in the intestinal pathogen *Campylobacter jejuni* (Szymanski et al. 1999). Intriguingly, while the transfer of the *N*-linked glycan to acceptor proteins in the eukaryotic model organism *S. cerevisiae* requires the action of the multi-enzyme OTase complex, (Aebi 2013), a single enzyme, termed PglB, with significant levels of amino acid similarity to the STT3 subunit of the yeast OTase complex, was sufficient for glycan transfer (Wacker et al. 2002; Young et al. 2002). The functional transfer of the *C. jejuni N*-linked protein glycosylation machinery encoded by the protein glycosylation (*pgl*) locus into *Escherichia coli* allowed structural determination of the glycan and functional characterisation of individual gene products involved (Wacker et al. 2002; Linton et al. 2005). Subsequent demonstration of relaxed glycan substrate specificity of the *C. jejuni* OTase PglB (CjPglB) paved the way for the development of recombinant glycoengineering, an approach by which a desired glycoconjugate can be generated entirelyin *E. coli* by co-expression of a glycan-coding locus, an acceptor protein and the OTase enzyme (Feldman et al. 2005). The *C. jejuni* *N*-linked glycosylation acceptor sequon was shown to differ from the eukaryotic N-X-S/T with an extended motif containing a negatively charged amino acid at the -2 position D/E-Z-N-X-S/T (where Z, like X, can be any amino acid except proline) (Kowarik, Young, et al. 2006). The most recent survey of *C. jejuni* glycoproteins identified 154 glycopeptides corresponding to 53 glycoproteins, confirming the general nature of the glycosylation machinery (Scott et al. 2011).

Orthologues of the CjPglB enzyme have been found in all *Campylobacter* species genome sequences, a subset of *Helicobacter* species, some members of the genus *Desulfovibrio*, as well as some δ- and ε-proteobacterial species found in deep-sea vent habitats (Nothaft & Szymanski 2010). Three of these orthologues have been functionally characterised; those from *Campylobacter lari* (ClPglB) (Schwarz et al. 2011), *Helicobacter pullorum* (HpPglB)(Jervis et al. 2010)*,* and the δ-proteobacterium, *Desulfovibrio desulfuricans* (DdPglB)(Ielmini & Feldman 2011).

Deep-sea hydrothermal vents are light-independent environments on the sea-floor that host a large community of chemolithoautotrophic proteobacterial symbionts, predominantly ε-proteobacteria (Huber et al. 2007). The genome sequences of several bacterial species isolated from deep-sea vents have been determined (Nakagawa et al. 2007; Nakagawa et al. 2005; Inagaki et al. 2004; Takaki et al. 2010) and several of the species were found to possess orthologues of the *C. jejuni* *pglB* gene. In the ε-proteobacteria *N. tergarcus* and *S. lithotrophicum,* the region adjacent to the putative *pglB* genes contain genes encoding proteins predicted to be involved in the generation and transfer of nucleotide-activated sugars, indicating the presence of at least a partial *pgl* operon similar to the one encoded by *C. jejuni*  (Fig 1A) (Nothaft & Szymanski 2010). In contrast, the *pglB* gene in the δ-proteobacterium *D. desulfuricans* appears to be an orphan gene and not part of a *pgl* operon*.* The presence of these putative *pglB* genes suggests that these uncharacterised organisms may possess functional protein *N­-*glycosylation machineries. Investigation of the putative *N-*glycosylation systems directly in these species is complicated by their relatively complex and unusual growth requirements, such as growth in supplemented synthetic sea water and at higher temperatures (Takai et al. 2003; Nakagawa et al. 2005). We therefore characterised the putative OTase enzymes in *E. coli* by co-expression with an acceptor protein and a lipid-linked glycan substrate, as done previously reported for the OTase enzymes of *D. desulfuricans* and *H. pullorum* (Ielmini & Feldman 2011; Jervis et al. 2010).

We present the functional expression and characterisation of three novel OTase enzymes from deep-sea vent bacteria, identify a possible glycoprotein encoded by one of these species and gain an insight to the nature of the native *N­-*linked glycan structures.

**Results**

**Identification of putative N-linking OTase enzymes from deep-sea vent bacteria**

In order to identify putative orthologues of CjPglB, the amino acid sequence was used as the query against all databases of prokaryotic proteins using blastp. Numerous non-*Campylobacter* orthologues were identified, and three encoded by deep-sea vent bacteria were chosen for further analysis (Fig 1B and supplementary Fig 1). These included *Nitratiruptor tergarcus* (54 % amino acid similarity, 34 % identity to the *C. jejuni* PglB enzyme)and *Sulfurovum lithotrophicum* (55 % similarity, 37 % identity)*,* two representatives of the ε-proteobacteria, and *Deferribacter desulfuricans* (40 % similarity, 24 % identity), a species classified within the phylum Deferribacteres from the δ-proteobacteria.

Several amino acid residues and structural features have been identified as important for OTase activity in ClPglB (Gerber et al. 2013; Lizak et al. 2011; Ihssen et al. 2012; Lizak et al. 2014). An amino acid alignment of the three deep-sea vent OTase enzymes with CjPglB indicates that the deep-sea vent OTase enzymes possess all the important residues for the function of ClPglB (highlighted in supplementary Fig. S2 and Table II), aside from residues R331 and I572 that are absent in DfdPglB.

In most *Campylobacter* species the OTase gene is present within the locus encoding for the assembly of the *N*-linked oligosaccharide (Nothaft & Szymanski 2010).The genes flanking the deep-sea vent putative *pglB* genes from *N. tergarcus* (NtPglB) and *S. lithotrophicum* (SlPglB) include genes predicted to encode an initiating undecaprenol-phosphate sugar phosphotransferase (PglC) and a number of glycosyltransferases. However no gene encoding a “flippase” enzyme required for membrane translocation of the lipid-linked oligosaccharide (LLO) into the periplasm was identified. In contrast, there are no orthologues of *C. jejuni* *N*-linked glycosylation pathway genes adjacent to the *D. desulfuricans* (DfdPglB) orthologue (Fig. 1A).

**The three deep-sea vent OTase enzymes are functional in *E. coli***

For functional analysis the putative OTase enzymes were tested for their ability to complement CjPglB in *E. coli*. The predicted ORFs coding for NtPglB, SlPglB and DfdPglBwere codon-optimised, synthesised and cloned into inducible expression vectors of the pEXT family (see Methods) (Dykxhoorn et al. 1996). Initial activity assays tested the ability of the OTase enzymes to transfer the *C. jejuni N-*linked heptasaccharide to the commonly used *C. jejuni* reporter glycoproteins, AcrAand Cj0114The three deep sea vent OTase enzymes (and CjPglB as a positive control) were co-expressed with hexa-his-tagged AcrA and the *C. jejuni pgl* gene locus with an insertionally inactivated *pglB* gene in *E. coli* strain CLM24 (Feldman et al. 2005). Analysis of purified AcrA by Western blotting, using anti-hexa-his and anti-heptasaccharide antibodies demonstrated CjPglB-mediated glycosylation of AcrA at two sites as previously reported (ref), but no modification was observed by the three putative deep-sea vent OTase enzymes (Fig. 2A).

Analysis of purified Cj0114 confirmed CjPglB-dependent glycosylation of Cj0114 at four extended *N*-linked glycosylation sequons as previously reported (Jervis *et al*., 2010) (Fig. 2C-D). Two of the four Cj0114 glycoforms were observed with NtPglB and SlPglB (Fig. 2C) demonstrating OTase activity of these two enzymes. No extra bands were observed for DfdPglB, suggesting that this enzyme was not able to transfer the *C. jejuni* heptasaccharide to the Cj0114 protein. These data confirm protein *N*-linked glycosylation activity of two of three deep-sea vent OTase enzymes.

**The sequon is required but not sufficient for glycosylation by NtPglB**

To further investigate NtPglB and SlPglB-mediated Cj0114 glycosylation, we employed Cj0114 mutants in which each of the asparagine residues within the four glycosylation sequons (N100, N154, N172 and N178) were replaced with glutamine (Jervis et al. 2010). Co-expression of these four variants with pACYCpglΔ*pglB* and CjPglB generated three Cj0114 glycoforms as expected (Fig. 3A, left panel). Co-expression of Cj0114 N100Q and N178Q with pACYCpglΔ*pglB* and NtPglB resulted in two glycoforms as for wild type Cj0114, however, Cj0114 N154Q and N172Q produced only a single glycoform (Fig. 3A right panel). This demonstrates that asparagine residues N154 and N172, but not N100 and N178, are required for modification with the *C. jejuni* heptasaccharide by NtPglB. The sequons surrounding N154 and N172 contain an aspartic acid at the -2 position, while the sequons surrounding N100 and N178 contain glutamic acid, suggesting that NtPglB may display a preference towards the former. However, replacement of aspartic acid at position -2 of N172 with a glutamic acid (D170E) did not disrupt glycosylation with NtPglB (data not shown). To investigate the requirement for a negatively charged residue at the -2 position for activity of NtPglB and SlPglB, the aspartic acid at the -2 position of N172 was replaced with alanine (mutant D170A). Thisresulted in one less Cj0114 glycoform produced by CjPglB, NtPglB and SlPglB (Fig. 3B), demonstrating that as for CjPglB both NtPglB and SlPglB require a negatively charged amino acid at the -2 position.

**The deep-sea vent OTase enzymes can transfer non-campylobacter glycan structures**

CjPglB possesses relaxed glycan specificity (Feldman et al. 2005). In order to assess the ability of the three deep-sea vent OTase enzymes to transfer a variety of glycan moieties, the enzymes were co-expressed with Cj0114 in *E. coli* strain E69 that synthesises the *O*9 *O*-antigen with *N-*acetylglucosamine (GlcNAc) at the reducing end (McCallum et al. 1989), and in *E. coli* strain CLM24 producing the *F. tularensis O*-antigen with QuiNAc (2-acetamido-2,6-dideoxy-O-d-glucose) at the reducing end (Cuccui et al. 2013). All three OTase enzymes were able to transfer both structures to Cj0114 (Fig. 4), demonstrating that similar to CjPglB, the deep-sea vent OTase enzymes possess relaxed glycan specificity. Interestingly, both NtPglB and SlPglB preferentially transferred shorter chains of the *F. tularensis O*-antigen compared to CjPglB, while the transfer efficiency of DfdPglB appeared very low.

**Analysis of putative glycoproteins encoded by the deep-sea vent bacteria**

As Cj0114 was glycosylated by the deep-sea vent OTase enzymes in the recombinant *E. coli* system, a BLAST search was performed to identify potential orthologues encoded by the three species. A Cj0114 orthologue was identified in each species, and designated Nt0114 (34 % identity to Cj0114), Sl0114 (30 % identity to Cj0114) and Dfd0114 (29 % identity to Cj0114). Only Nt0114contained an extended bacterial *N*-glycosylation sequon, whilst Sl0114 and Dfd0114 contained two and three eukaryotic N-X-S/T sequons, respectively (Fig. 5, supplementary figures S3 and S4).

Both Nt0114 and Dfd0114 were tested for their capacity to be glycosylated through expression in *N*-linked glycosylation competent *E. coli* CLM24 producing either the *C. jejuni* *N*-linked heptasaccharide glycan or *F. tularensis* O-antigen. In this system Nt0114 was glycosylated with the *C. jejuni* heptasaccharide and *F. tularensis* O-antigen by CjPglB but not by any of the three deep-sea OTase enzymes (Supplementary figure S3). Interestingly, when Dfd0114 lacking extended bacterial *N*-linked glycosylation sequons was co-expressed in *E. coli* CLM24 with OTase enzymes and the *C. jejuni* heptasaccharide*,* a glycosylated form was detected in the presence of CjPglB, suggesting transfer of glycan to a eukaryotic sequon (Fig. 5B). When Dfd0114 was co-expressed with the *F. tularensis O*-antigen and the OTase enzymes, glycoforms were observed in the presence of CjPglB and DfdPglB, but not NtPglB or SlPglB (Fig. 5C). This suggested glycan transfer by both CjPglB and DfdPglB to a eukaryotic-type sequon. CjPglB was previously believed to strictly require a negatively charged residue at the -2 position of the sequon (Kowarik, Young, et al. 2006). However, recent work has demonstrated modification of asparagine residues not located within an extended sequon by CjPglB, both in the native host as well as in *E. coli* (Scott et al. 2014; Ollis et al. 2014)*.* To identify the Dfd0114 sequon glycosylated by CjPglB, site-directed mutant constructs were generated by individually changing the asparagine residue in each of the three eukaryotic sequons to alanine. The mutant N107A was no longer glycosylated by CjPglB, while mutants N101A and N118A remained glycosylated, suggesting transfer of the glycan to asparagine N101 by CjPglB (Fig 5D).

**Functional characterisation of putative initiating glycosyl transferase enzymes involved in generation of the *N*-linked LLO in *N. tergarcus* and *S. lithotrophicum***

Unlike *C. jejuni*, the genes involved in the synthesis of the *N­-*linked LLO in *N. tergarcus,* *S. lithotrophicum* and *Deferribacter desulfuricans* are not encoded within a single locus (Nothaft & Szymanski 2010) (Fig. 1A). However, a putative undecaprenol-phosphate UDP-glycosyl transferase (*pglC*) is located immediately downstream of *pglB* in both *N. tergarcus* and *S. lithotrophicum* (Fig. 1A). Additionally, orthologues of two of the three enzymes involved in the synthesis of UDP-N,N′-diacetylbacillosamine (diNAcBac) from UDP-N-acetylglucosamine (UDP-GlcNAc), *pglE* and *pglF*, but not *pglD* were identified downstream of *pglB* and *pglC* (Morrison & Imperiali 2014). No orthologues of *pglC, E* or *F* were identified from *D. desulfuricans*, suggesting that the native glycan transferred in this organism is unlikely to contain diNAcBac at the reducing end. Recombinant expression of NtPglC or SlPglC together with Cj0114 and pACYCpgl*pglC*::Km in *E. coli* strain CLM37, which lacks the initiating transferase of the *O*-antigen biosynthesis repeat unit synthesis machinery (WecA), resulted in Cj0114 glycosylation with a glycan recognised by the anti-*C. jejuni* heptasaccharide antiserum HR6 (Fig. 6A). This demonstrates that both deep-sea vent PglC orthologues can functionally complement Cj PglC and are able to transfer a glycan moiety to undecaprenol-phosphate, on which the full *C. jejuni* heptasaccharide can be synthesized by enzymes encoded by the *C. jejuni pgl* locus. It has previously been shown that CjPglB can transfer heptasaccharides with either a diNAcBac or a GlcNAc residue at the reducing end (Linton et al. 2005). However, CjPglC has been shown to only catalyse the transfer of UDP-Bac to undecaprenol-phosphate (Glover et al. 2006). Elucidation of the nature of the reducing end glycan requires mass spectrometric analysis; however, the full length Cj0114 protein was unsuitable for mass spectrometry analysis. A modified C-terminal truncation of Cj0114 termed NGRP, which displays improved performance of the N172 tryptic peptide during MS analysis has previously been generated (Adrian J. Jervis, unpublished). This protein was used in an analogous PglC complementation experiment, and the resultant glycoform subjected to tandem MS analysis to identify the nature of the glycan at the reducing end of the oligosaccharide (Fig 6B). This showed the presence of a 228 Da residue consistent with a diNAcBac at the reducing end of the glycan, demonstrating that this was the substrate for both NtPglC and SlPglC. Two further minor peaks were observed in the NtPglC sample. The peak at 2822 corresponds to a *C. jejuni* heptasaccharide with a HexNAc at the reducing end based on mass difference . The peak at 2832 either corresponds to a *C. jejuni* heptasaccharide with an unknown glycan at the reducing end or is the result of partial fragmentation. Unfortunately, the low intensity of these peaks did not allow for full structural elucidation. This result suggests that NtPglC may also be able to transfer a HexNAc residue to undecaprenol-phosphate with low efficiency. Unfortunately, the intensity of the peaks was too low to sequence the glycan and confirm this hypothesis.

**Discussion**

The discovery and functional characterisation of a bacterial *N*-linked general protein glycosylation system in *C. jejuni* (Szymanski et al. 1999; Wacker et al. 2002) challenged the dogma that this type of post-translational protein modification was limited to the eukaryotic and archaeal kingdoms. Since then, further bacterial *N*-linking OTase enzymes have been functionally characterised (Schwarz et al. 2011; Jervis et al. 2010; Ielmini & Feldman 2011). Intriguingly, genome sequencing of three bacterial species from a deep-sea vent habitat identified genes encoding putative orthologues of the *C. jejuni* OTase PglB (Nothaft & Szymanski 2010; Nakagawa et al. 2007; Takaki et al. 2010). We report the functional characterisation of the OTase orthologues encoded by these three species (Fig. 1) using a recombinant approach in *E. coli.*

We have demonstrated that the *N. tergarcus,* *S. lithotrophicum* and *D. desulfuricans* *pglB* orthologues encode functional OTase enzymes that are able to transfer lipid-linked oligo- and polysaccharides to an acceptor protein. However, while it has been shown that CjPglB is able to glycosylate any *N-*glycosylation sequon as long as the protein is targeted to the periplasm and the glycosylation sequon is present within a flexible, accessible loop (Fisher et al. 2011; Kowarik, Young, et al. 2006; Kowarik, Numao, et al. 2006), our data suggests a more stringent acceptor protein requirement for the three deep-sea vent OTase enzymes, as only one acceptor protein, Cj0114, was glycosylated and at only two of the four possible sites. Such an acceptor protein specificity has not been reported previously for *N-*linking OTase enzymes, but has been demonstrated for two bacterial *O­-*linking OTase enzymes (Horzempa et al. 2006; Harding et al. 2015). It is therefore possible that the deep-sea vent enzymes present a new class of *N*-linking OTase enzymes with more stringent acceptor protein specificity.

Even within the one protein that was successfully glycosylated by the deep-sea vent OTase enzymes, a preference for particular sequons was observed. This preference was not a result of the primary amino acid sequence of the sequon, as altering the non-modified sequons did not result in glycosylation. The sequon preference was also not due to general unavailability of the sequon within the secondary structure of the protein, as CjPglB was able to glycosylate all four sequons. It has previously been shown that the *H. pullorum* PglB1 was only able to glycosylate two of the four sequons within the Cj0114 protein (Jervis et al. 2010), and alternative glycosylation of the *C. jejuni* protein AcrA by the enzymes from *C. lari* and *D. desulfuricans* has also been reported (Ielmini & Feldman 2011; Schwarz et al. 2011), suggesting that sequon usage may be variable among bacterial *N*-OTase enzymes.

As both NtPglB and SlPglB possess the R331 amino acid residue (Table II) which is implicated in the interaction of ClPglB with the negatively charged residue at the -2 position of the extended glycosylation sequon (Lizak et al. 2011) it was unsurprising that glycosylation by both enzymes required this negative charge. R331 is absent in DfdPglB, similar to the orthologue encoded by *D. desulfuricans*, which is able to glycosylate sequons lacking the negatively charged residue at the -2 position (Ielmini & Feldman 2011). However, it was not possible to identify the Cj0114 sequon preference of this OTase due to low levels of activity. Similar to CjPglB (Wacker et al. 2006; Cuccui et al. 2013), all three OTase enzymes displayed relaxed glycan specificity, and were able to transfer both the *E. coli O*9 *O*-antigen and the *F. tularensis O*-antigen.

In order to support the hypothesis of a fully functional *N-*glycosylation machinery encoded by deep-sea vent bacteria, we sought to identify putative native glycoprotein. We identified orthologues of Cj0114 in all three bacterial species. The *S. lithotrophicum* orthologue (Sl0114) does not contain any glycosylation sequons, while the *N. tergarcus* protein (Nt0114) contains one extended sequon and the *D. desulfuricans* orthologue (Dfd0114) contains 3 eukaryotic sequons lacking the -2 negatively charged residues. Recombinant expression of Dfd0114 with the *C. jejuni N*-linked heptasaccharide and CjPglB resulted in transfer of the glycan to Dfd0114 at an asparagine residue within the sequence PNNNIS. The *C. lari* PglB has been shown to glycosylate asparagine residues not located within an extended bacterial sequon both *in vivo* (Schwarz et al. 2011) and *in vitro* (Gerber et al. 2013) and recent evidence suggests that CjPglB has similar activity, both in the native host as well as in a recombinant *E. coli* system (Scott et al. 2014; Ollis et al. 2014). No glycosylation of Dfd0114 with the *C. jejuni* heptasaccharide was observed for DfdPglB. However, low levels of Ddf0114 glycosylation by DfdPglB was observed when co-expressed with the *F. tularensis O*-antigen. This indicates that DfdPglB, similar to the enzyme from *Desulfovibrio desulfuricans* is able to glycosylate short, eukaryotic sequons (Ielmini & Feldman 2011). Further studies in *Deferribacter desulfuricans* are required to confirm the glycosylation status of Dfd0114 in the native organism as well as to identify the native glycoproteome and investigate the presence of eukaryotic glycosylation sequons within these proteins. In contrast, Nt0114 was only glycosylated by CjPglB, and not by any of the deep-sea vent OTase enzymes, including the “native” NtPglB (Supplementary figure S3). Glycosylation of Nt0114 by CjPglB demonstrated that the protein is present in the correct subcellular compartment and the sequon is in principle accessible to the OTase enzymes. This suggests that the lack of glycosylation of Nt0114 by the deep-sea vent OTase enzymes is likely due to the more stringent acceptor protein specificity of these enzymes. This does not, however, rule out the possibility that this protein may be glycosylated in *N. tergarcus*. Further studies in *N. tergarcus* are required to address this and to identify the complete native glycoproteome.

No data are available regarding *N-*linked glycan structure in *N. tergarcus* or *S. lithotrophicum*. To investigate the nature of the sugar residue present at the reducing end of the *N*-glycan produced by these species, a CjPglC complementation experiment using the two putative initiating glycosyl transferase enzymes encoded by *N. tergarcus* and *S. lithotrophicum* was performed. This demonstrated that both NtPglC and SlPglC are able to transfer a diNAcBac residue to undecaprenol-phosphate, on top of which the remaining heptasaccharide was assembled. In the case of NtPglC, a small amount of glycan likely containing a HexNAc at the reducing end as judged by mass difference was also observed, suggesting that NtPglC is able to transfer a HexNAc to undecaprenol-phosphate at a very low rate. Analysis of the structure of the glycans in the native organisms, or attempts to reconstitute the complete *N-*glycan biosynthesis pathway in *E. coli* are required to fully investigate the structure of the glycan synthesised by the two species.

The demonstration of functional *N*-linked glycosylation systems encoded by deep-sea vent bacteria raises interesting questions regarding the evolution of this post-translational modification system and the role of *N*-linked protein glycosylation in the biology of these species. While this type of post-translational modification is ubiquitous amongst higher organisms, and has also been found to be encoded by 166 of 168 archaeal genome sequences obtained to date (Kaminski, Eichler 2013), it has so far only been shown in a small number of bacterial species (Jervis et al. 2010; Ielmini & Feldman 2011; Schwarz et al. 2011; Nothaft & Szymanski 2010). In eukaryotes, the role of the *N-­*glycan is multifunctional, ranging from protein quality control to secretion and interaction between proteins, and the modification is essential for the function of the cell (Aebi 2013). In the archaeal species *H. volcanii* and members of the genus *Methanococcus*, the *N-*glycosylation machinery is not essential for cell viability (Jarrell et al. 2010). However, disruption of *N-*glycosylation in *H. volcanii* led to decreased ability to grow in high salt concentrations (Kaminski, Naparstek, et al. 2013). Additionally, *H. volcanii* cells containing a disrupted OTase gene were unable to produce intact flagella and were non-motile (Tripepi et al. 2012). Glycosylation of the S-layer protein in *H. volcanii* was recently shown to be dependent on salinity levels, suggesting a role in survival in the relatively harsh environment (Kaminski, Guan, et al. 2013). It has been demonstrated that CjPglB can function as a hydrolase in addition to an OTase, resulting in the release of the *N-*glycan as a free oligosaccharide into the periplasm of the bacterium (Nothaft et al. 2009). This release has been shown to be influenced by altering the salt and osmolyte concentration of the environment, suggesting an adaptive function similar to that observed for archaea. It can be speculated that the OTase enzymes encoded by deep-sea species, living predominantly under harsh environmental conditions such as high temperature and high osmolarity, may contribute to the survival in those conditions by generation of free oligosaccharides to counteract the high osmotic levels of sea water. It has also been shown that a glycosylated form of the *C. jejuni* glycoprotein PEB3 is more thermostable than a non-glycosylated protein (Min et al. 2009). Therefore, protein glycosylation in the deep-sea vent bacteria may increase the overall thermostability of the glycosylated subsection of the proteome.

Studies in *C. jejuni* using cells deficient in either the OTase CjPglB or enzymes involved in synthesis of the lipid-linked heptasaccharide have suggested roles for *N*-linked glycosylation in chicken colonisation and adhesion and invasion of epithelial cells(Szymanski et al. 2002; Karlyshev et al. 2004). However, as a total number of 53 *C. jejuni* proteins have been shown to be *N-*glycosylated (Scott et al. 2011), it has not been possible to identify the precise role of glycosylation in this pathogen. It has also been demonstrated that the *C. jejuni N*-glycan is recognised by the human galactose-type lectin MGL, suggesting a potential role for the glycan in modulation of the immune system (van Sorge et al. 2009). As many deep-sea vent bacterial species can exist as free-living biomass as well as symbionts on other deep-sea vent animals such as polychaetes and shrimps (Polz & Cavanaugh 1995) it can be speculated that the bacteria may employ an *N-*linked glycan to interact with and modulate the immune system of their symbiotic partners. However, more studies in the native organisms are needed to address the function and scope of protein glycosylation in these species.

From a ‘glycoengineering’ point of view, it is interesting to note that two of the three organisms that encode the OTase enzymes characterised here require higher growth temperature than those used for the recombinant expression in the *E. coli* host. This has two potential implications. Firstly, these enzymes may possess different biophysical characteristics compared to the best studied CjPglB, such as better thermostability/’shelf life’. Secondly, the activity of the recombinant OTase enzymes in the *in vivo E. coli* expression host may be different from the activity in the native host. Further studies in the native organisms, as well as adaptation of published *in vitro* glycosylation experiments (Kowarik, Numao, et al. 2006; Jervis et al. 2010) for testing of these OTase enzymes under different conditions are required to investigate these possibilities.

In summary, this is the first functional characterisation of bacterial OTase enzymes encoded by three bacterial species from a deep-sea vent habitat and paves the way for further studies of the role of protein *N-*glycosylation in these specialised bacteria and other bacterial species such as *C. jejuni.* Additionally, the study provides a deeper understanding of a biotechnologically important class of enzymes.

**Materials and Methods**

**Bacterial strains, plasmids and growth conditions**

*E. coli* strains were grown in Lysogeny Broth (LB) or on LB agar at 37°C. Where required, the medium was supplemented with antibiotics at the following concentrations: 100 µg ml-1 ampicillin, 34 µg ml-1 chloramphenicol, 50 µg ml-1 kanamycin and 100 µg ml-1 spectinomycin. *E. coli* DH5α library efficiency cells (Invitrogen, Carlsbad, USA) were routinely used as a host for cloning experiments*. E. coli* strain E69 was kindly provided by Chris Whitfield (University of Guelph, Canada). All strains and plasmids are listed in Table I.

**Synthesis and subcloning of genes from deep-sea vent bacteria**

The ORFs encoding NtPglB, SlPglB and DfdPglB, as well as the putative acceptor proteins were codon-optimised for expression in *E. coli,* synthesised by Celtek Genes (Celtek-genes) and delivered in vector pGH flanked by recognition sites for restriction endonucleases SacI and XbaI for the *pglB* genes and EcoRI and XbaI for the acceptor protein and *pglC* genes. The ORFs were subcloned by restriction digestion into vectors pEXT20, pEXT21, pEXT22 or pMLBAD (Table I) and sequence verified.

**Functional analysis of the deep-sea vent PglB proteins**

**Transfer of the *C. jejuni* heptasaccharide in *E. coli***

*E. coli* CLM24 cells were transformed with plasmids pACYCpglΔ*pglB*, along with a plasmid encoding an acceptor protein (See Table I for details) and either a control plasmid or a plasmid encoding a PglB. One colony was grown in LB broth to an OD600 of 0.4 - 0.6 at 37°C, and *pglB* expression induced with 1 mM IPTG. Cultures were grown for an additional 16 hours, the cells harvested by centrifugation, lysed (see below) and acceptor proteins purified by immobilized metal ion affinity chromatography (IMAC).

**Transfer of the *E. coli* O9 *O*-antigen polysaccharide**

Plasmid pBRCj0114 was co-expressed with a plasmid encoding the desired PglB protein in *E. coli* strain E69 which synthesises the O9 ­*O*-antigen, and induction and sample preparation were performed as above.

**Transfer of the *F. tularensis* *O*-antigen polysaccharide**

*E. coli* CLM24 cells were transformed with plasmids pGAB2, pBRCj0114 and either a control plasmid or a plasmid encoding the desired PglB protein. Protein expression and sample preparation were performed as above.

**Purification of acceptor proteins using IMAC**

Briefly, cell pellets obtained after overnight induction of the acceptor proteins were resuspended in lysis solution (500 mM NaCl, 25 mM NaH2PO4, 15 mM imidazole containing 1 mg/ml lysozyme, pH 7.5), and sonicated. Lysates were clarified by centrifugation, and Ni-NTA agarose added. After 1 hour with mixing, the slurry was loaded onto a Pierce Spin cup, washed five times with wash solution (500 mM NaCl, 25 mM NaH2PO4, 25 mM imidazole, pH 7.5), and bound proteins eluted with elution buffer (500 mM NaCl, 25 mM NaH2PO4, 500 mM imidazole, pH 7.5).

**Analysis of glycosylated acceptor protein**

Purified acceptor proteins were separated by SDS-PAGE using NuPAGE™ Novex™ 4-12 % Bis-Tris protein gels (Invitrogen), transferred to nitrocellulose membrane and analysed by two colour immunoblot using anti-hexahistidine, anti-glycan and corresponding fluorescent-labelled secondary antibodies (Table III) using an Odyssey near-infrared imager (LiCOR biosciences).

**Site-directed mutagenesis of acceptor sequons**

Site-directed mutagenesis of acceptor protein sequons was performed using the QuikChange XL site-directed mutagenesis kit (Agilent) according to the manufacturer’s instructions.

**MALDI-MS analysis of glycopeptides**

Coomassie-stained bands in SDS-PAGE gels were excised and subjected to in-gel trypsin digestion followed by cleanup on a C18 Zip-Tip (Millipore). MALDI-TOF MS and MALDI-LIFT-TOF/TOF MS spectra were acquired by laser-induced dissociation (LID) using a Bruker Ultraflex II mass spectrometer in the positive-ion reflection mode with 2,5-dihydroxybenzoic acid (DHB) (20 mg ml-1 in 0.1 % formic acid, 30 % acetonitrile) as the matrix. Data were analysed with FlexAnalysis 3.0 software (Bruker Daltonics).

**Funding**

This work was supported by the Wellcome Trust and the Biotechnology and Biological Sciences Research Council.

**Acknowledgements**

Plasmid pGVXN114 encoding CjPglB was obtained from Glycovaxyn AG, Switzerland under MTA.

We would also like to thank David Knight and the staff of the Biomolecular Analysis Core Research Facility, Faculty of Life Sciences, University of Manchester, for assistance with MS.

**References**

Aebi, M., 2013. N-linked protein glycosylation in the ER. *Biochimica et biophysica acta*.

Bolivar, F. et al., 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene*, 2(2), pp.95–113.

Cuccui, J. et al., 2013. Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis. *Open biology*, 3(5), p.130002.

Cuthbertson, L., Powers, J. & Whitfield, C., 2005. The C-terminal domain of the nucleotide-binding domain protein Wzt determines substrate specificity in the ATP-binding cassette transporter for the lipopolysaccharide O-antigens in Escherichia coli serotypes O8 and O9a. *The Journal of biological chemistry*, 280(34), pp.30310–9.

Dykxhoorn, D.M., St Pierre, R. & Linn, T., 1996. A set of compatible tac promoter expression vectors. *Gene*, 177(1-2), pp.133–6.

Feldman, M.F. et al., 2005. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 102(8), pp.3016–21.

Gerber, S. et al., 2013. Mechanism of bacterial oligosaccharyltransferase: in vitro quantification of sequon binding and catalysis. *The Journal of biological chemistry*, 288(13), pp.8849–61.

Harding, C.M. et al., 2015. Acinetobacter strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple proteins. *Molecular microbiology*.

Homonylo, M.K. et al., 1988. Monoclonal antibodies against the capsular K antigen of Escherichia coli (O9:K30(A):H12): characterisation and use in analysis of K antigen organisation on the cell surface. *Canadian journal of microbiology*, 34(10), pp.1159–65.

Horzempa, J. et al., 2006. Glycosylation substrate specificity of Pseudomonas aeruginosa 1244 pilin. *The Journal of biological chemistry*, 281(2), pp.1128–36.

Huber, J.A. et al., 2007. Microbial population structures in the deep marine biosphere. *Science (New York, N.Y.)*, 318(5847), pp.97–100.

Ielmini, M. V & Feldman, M.F., 2011. Desulfovibrio desulfuricans PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology*, 21(6), pp.734–42.

Ihssen, J. et al., 2010. Production of glycoprotein vaccines in Escherichia coli. *Microbial cell factories*, 9, p.61.

Ihssen, J. et al., 2012. Structural insights from random mutagenesis of Campylobacter jejuni oligosaccharyltransferase PglB. *BMC biotechnology*, 12, p.67.

Jarrell, K.F. et al., 2010. S-layer glycoproteins and flagellins: reporters of archaeal posttranslational modifications. *Archaea (Vancouver, B.C.)*, 2010.

Jervis, A.J. et al., 2010. Characterization of N-linked protein glycosylation in Helicobacter pullorum. *Journal of bacteriology*, 192(19), pp.5228–36.

Kaminski, L., Naparstek, S., et al., 2013. Add salt, add sugar: N-glycosylation in Haloferax volcanii. *Biochemical Society transactions*, 41(1), pp.432–5.

Kaminski, L., Guan, Z., et al., 2013. Two distinct N-glycosylation pathways process the Haloferax volcanii S-layer glycoprotein upon changes in environmental salinity. *mBio*, 4(6), pp.e00716–13.

Karlyshev, A. V et al., 2004. The Campylobacter jejuni general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology (Reading, England)*, 150(Pt 6), pp.1957–64.

Kowarik, M., Young, N.M., et al., 2006. Definition of the bacterial N-glycosylation site consensus sequence. *The EMBO journal*, 25(9), pp.1957–66.

Kowarik, M., Numao, S., et al., 2006. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science (New York, N.Y.)*, 314(5802), pp.1148–50.

Larkin, M.A. et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), pp.2947–8.

Lefebre, M.D. & Valvano, M.A., 2002. Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in Burkholderia cepacia complex isolates. *Applied and environmental microbiology*, 68(12), pp.5956–64.

Linton, D. et al., 2005. Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. *Molecular microbiology*, 55(6), pp.1695–703.

Lizak, C. et al., 2014. A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. *The Journal of biological chemistry*, 289(2), pp.735–46.

Lizak, C. et al., 2011. X-ray structure of a bacterial oligosaccharyltransferase. *Nature*, 474(7351), pp.350–5.

McCallum, K.L., Laakso, D.H. & Whitfield, C., 1989. Use of a bacteriophage-encoded glycanase enzyme in the generation of lipopolysaccharide O side chain deficient mutants of Escherichia coli O9:K30 and Klebsiella O1:K20: role of O and K antigens in resistance to complement-mediated serum killing. *Canadian journal of microbiology*, 35(11), pp.994–9.

Min, T. et al., 2009. Specificity of Campylobacter jejuni adhesin PEB3 for phosphates and structural differences among its ligand complexes. *Biochemistry*, 48(14), pp.3057–67.

Nakagawa, S. et al., 2007. Deep-sea vent epsilon-proteobacterial genomes provide insights into emergence of pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 104(29), pp.12146–50.

Nakagawa, S. et al., 2005. Nitratiruptor tergarcus gen. nov., sp. nov. and Nitratifractor salsuginis gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the epsilon-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. *International journal of systematic and evolutionary microbiology*, 55(Pt 2), pp.925–33.

Nothaft, H. et al., 2009. Study of free oligosaccharides derived from the bacterial N-glycosylation pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 106(35), pp.15019–24.

Nothaft, H. & Szymanski, C.M., 2010. Protein glycosylation in bacteria: sweeter than ever. *Nature reviews. Microbiology*, 8(11), pp.765–78.

Ollis, A.A. et al., 2014. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nature chemical biology*.

Polz, M.F. & Cavanaugh, C.M., 1995. Dominance of one bacterial phylotype at a Mid-Atlantic Ridge hydrothermal vent site. *Proceedings of the National Academy of Sciences of the United States of America*, 92(16), pp.7232–6.

Schwarz, F. et al., 2011. Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology*, 21(1), pp.45–54.

Scott, N.E. et al., 2014. Comparative Proteomics and Glycoproteomics Reveal Increased N-Linked Glycosylation and Relaxed Sequon Specificity in Campylobacter jejuni NCTC11168 O. *Journal of proteome research*.

Scott, N.E. et al., 2011. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobact. *Molecular & cellular proteomics : MCP*, 10(2), pp.M000031–MCP201.

van Sorge, N.M. et al., 2009. N-glycosylated proteins and distinct lipooligosaccharide glycoforms of Campylobacter jejuni target the human C-type lectin receptor MGL. *Cellular microbiology*, 11(12), pp.1768–81.

Szymanski, C.M. et al., 1999. Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Molecular microbiology*, 32(5), pp.1022–30.

Szymanski, C.M., Burr, D.H. & Guerry, P., 2002. Campylobacter protein glycosylation affects host cell interactions. *Infection and immunity*, 70(4), pp.2242–4.

Takai, K. et al., 2003. Deferribacter desulfuricans sp. nov., a novel sulfur-, nitrate- and arsenate-reducing thermophile isolated from a deep-sea hydrothermal vent. *International journal of systematic and evolutionary microbiology*, 53(Pt 3), pp.839–46.

Takaki, Y. et al., 2010. Bacterial lifestyle in a deep-sea hydrothermal vent chimney revealed by the genome sequence of the thermophilic bacterium Deferribacter desulfuricans SSM1. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 17(3), pp.123–37.

Tamura, K. et al., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, 30(12), pp.2725–9.

Tripepi, M. et al., 2012. N-glycosylation of Haloferax volcanii flagellins requires known Agl proteins and is essential for biosynthesis of stable flagella. *Journal of bacteriology*, 194(18), pp.4876–87.

Wacker, M. et al., 2002. N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science (New York, N.Y.)*, 298(5599), pp.1790–3.

Wacker, M. et al., 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proceedings of the National Academy of Sciences of the United States of America*, 103(18), pp.7088–93.

Whitfield, C., Schoenhals, G. & Graham, L., 1989. Mutants of Escherichia coli O9:K30 with altered synthesis and expression of the capsular K30 antigen. *Journal of general microbiology*, 135(10), pp.2589–99.

Young, N.M. et al., 2002. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. *The Journal of biological chemistry*, 277(45), pp.42530–9.

**Figure 1. Analysis of the *pgl* loci of the deep-sea vent bacteria and phylogenetic tree of PglB/AglB/STT3 OTase enzymes** A:Depiction of the genomic localisation of the PglB-encoding genes of *C. jejuni N. tergarcus, S. lithotrophicum* and *Deferribacter desulfuricans***.** GT: Glycosyl transferase. White arrows: genes with no involvement in sugar biosynthesis. ORFs annotated as C? encode putative initiating glycosyl transferase enzymes with low homology to *C. jejuni* PglC. B: Subset view of the phylogenetic tree of the oligosaccharyl transferase protein sequence (PglB, AglB and STT3 subunit) from the three domains of life (Bacteria,   
Archaea and Eukarya, respectively) (full tree shown in S1). The tree was a CLUSTAL W (Larkin et al. 2007) generated protein sequence alignment using the neighbour-joining (NJ) method with 1000 bootstrap replicates in phylogenetic package MEGA6 (Tamura et al. 2013). The PglB orthologues investigated in this study are boxed.

**Figure 2. Functional expression of deep-sea vent OTase enzymes**

A: Amino acid sequence of the *C. jejuni* glycoprotein AcrA. Bold sequence: Two extended bacterial N-linked glycosylation sequons. Underlined amino acids: acceptor asparagines B: NtPglB, SlPglB and DfdPglB OTase enzymes were co-expressed in *E. coli* CLM24 with plasmids coding for the synthesis of the *C. jejuni N*-linked glycan and the affinity-tagged *C. jejuni* glycoprotein AcrA. Empty vector (no OTase) and CjPglB expressing cultures were included as controls. Following purification, the acceptor proteins were separated by SDS-PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green). C: Amino acid sequence of Cj0114. Italicised sequence: Engineered *pelB* leader sequence for efficient targeting of the protein to the periplasm. Bold sequence: Four extended bacterial glycosylation sequons. Underlined amino acids: acceptor asparagines D:NtPglB, SlPglB and DfdPglB OTase enzymes were co-expressed in *E. coli* CLM24 with plasmids coding for the synthesis of the *C. jejuni N*-linked glycan and the affinity-tagged *C. jejuni* glycoprotein Cj0114. Empty vector (no OTase) and CjPglB expressing cultures were included as controls. After purification, the acceptor proteins were separated by SDS-PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green).

**Figure 3. Analysis of the sequon usage by NtPglB and requirement for a negative charge at the -2 position for NtPglB and SlPglB.**

A: CjPglB and NtPglB were co-expressed in *E. coli* CLM24 with plasmids coding for the synthesis of the *C. jejuni N*-linked glycan and four site-directed mutant versions of affinity-tagged *C. jejuni* glycoprotein Cj0114, where the acceptor asparagine residue of each site had been replaced with a glutamine. After purification, the acceptor proteins were separated by SDS-PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green). B: CjPglB, NtPglB and SlPglB were co-expressed in *E. coli* CLM24 with a plasmid coding for the synthesis of the *C. jejuni N*-linked glycan and either the wild-type Cj0114 protein or a site-directed mutant version of Cj0114 where the aspartic acid residue at the -2 position of sequon 3 had been replaced with an alanine (D170A).

**Figure 4. NtPglB, SlPglB and DfdPglB possess relaxed glycan specificity**

NtPglB, SlPglB and DfdPglB OTase enzymes were co-expressed with affinity-tagged *C. jejuni* glycoprotein Cj0114 in A: *E. coli* E69 which synthesises O9 *O*-antigen and B: *E. coli* CLM24 containing a plasmid coding for the synthesis of the *F. tularensis O­-*antigen. After purification, the acceptor proteins were separated by SDS-PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green).

**Figure 5.** **A putative deep-sea vent glycoprotein encoded by *D. desulfuricans* is glycosylated by CjPglB and DfdPglB at a non-canonical asparagine.**

A: Amino acid sequence of Dfd0114. B: Affinity tagged Dfd0114 was co-expressed with the OTase enzymes in *E. coli* CLM24 with a plasmid coding for the synthesis of the *C. jejuni N*-linked glycan C: Affinity tagged Dfd0114 was co-expressed with the OTase enzymes in *E. coli* CLM24 containing a plasmid coding for the synthesis of the *F. tularensis O­-*antigen. D: CjPglB was co-expressed in *E. coli* CLM24 with a plasmid coding for the synthesis of the *C. jejuni N*-linked glycan and either the wild type or three site-directed mutant versions of affinity-tagged *C. jejuni* glycoprotein Dfd0114, where the putative acceptor asparagine residue had been replaced with an alanine.

**Figure 6. Complementation of *C. jejuni* PglC by two deep-sea orthologues.**

A: NtPglC and SlPglC were co-expressed in *E. coli* CLM37 (Δ*wecA*) with plasmid pACYCpgl*pglC*::Km and affinity-tagged *C. jejuni* glycoprotein Cj0114. After purification, the acceptor proteins were separated by SDS-PAGE, western blotted and probed with antisera against the affinity tag (red) and glycan (green). B: MALDI-MS analysis of the reporter protein NGRP glycosylated by the *C. jejuni* Pgl system in *E. coli* and an isogenic *pglC* mutant complemented by PglC orthologues from deep-sea species *N. tergarcus* and *S. lithotrophicum* in *E. coli* CLM37*.* (B) Full mass spectra of tryptic digests in the *m/z* range containing the glycosylated peptide from reporter protein NGRP. The peak with an *m/z* of 2847.2 corresponds to the peptide modified with *C. jejuni* heptasaccharide with the reducing end sugar diNAcBac as confirmed by MALDI-LIFT-TOF/TOF MS (C) and represented in (D). The low intensity peak with an *m/z* of 2822.2 corresponds to the predicted mass of peptide glycosylated with a heptasaccharide containing a HexNAc at the reducing end.

**Table I. Bacterial strains and plasmids.**

|  |  |  |
| --- | --- | --- |
| **Bacterial strains** | Genotype | Source |
| *E. coli* DH5α | F- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rk-, mk+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ- | Invitrogen |
| *E. coli* E69 | Prototroph; O9a:K30:H12 | (Homonylo et al. 1988) |
| *E. coli* CLM24 | W3110 derivative, Δ*waaL* | (Feldman et al. 2005) |
| *E. coli* CLM37 | W3110 derivative, Δ*wecA* | (Linton et al. 2005) |
| **Plasmids** |  |  |
| pBR322 | Expression vector, constitutive expression, AmpR | (Bolivar et al. 1977) |
| pMLBAD | Expression vector, arabinose inducible, TmpR | (Lefebre & Valvano 2002) |
| pEXT20 | Expression vector, IPTG inducible, AmpR | (Dykxhoorn et al. 1996) |
| pEXT21 | Expression vector, IPTG inducible, SpecR | (Dykxhoorn et al. 1996) |
| pEXT22 | Expression vector, IPTG inducible, KanR | (Dykxhoorn et al. 1996) |
| pNtPglB21 | Codon-optimised *N. tergarcus pglB* cloned into pEXT21 | This study |
| pNtPglB22 | Codon-optimised *S. lithotropicum pglB* cloned into pEXT21 | This study |
| pSlPglB21 | Codon-optimised *N. tergarcus pglB* cloned into pEXT22 | This study |
| pSlPglB22 | Codon-optimised *S. lithotropicum pglB* cloned into pEXT22 | This study |
| pDfdPglB21 | Codon-optimised *D. desulfuricans pglB* cloned into pEXT21 | This study |
| pDfdPglB22 | Codon-optimised *D. desulfuricans pglB* cloned into pEXT22 | This study |
| pGVXN114 | HA-tagged *C. jejuni* PglB cloned into pEXT21 | (Ihssen et al. 2010) |
| pMAF10 | HA-tagged PglB from *C. jejuni* cloned in pMLBAD, TmpR | (Feldman et al. 2005) |
| pACYCpgl*pglB*Kan | *pgl* locus from *C. jejuni* cloned into pACYC184, KanR transposon in *pglB* | (Linton et al. 2005) |
| pGAB2 | Recombinant expression of the *F. tularensis O*-antigen | (Cuccui et al. 2013) |
| pWA2 | His-tagged soluble AcrA expressed from pBR322 | (Feldman et al. 2005) |
| pBRCj0114 | His-tagged CJ0114 constitutively expressed from pBR322 | (Jervis et al. 2010) |
| pBRCj0114N100Q | Derivative of pBRCj0114 with mutated glycosylation site: N100Q | This study |
| pBRCj0114N154Q | Derivative of pBRCj0114 with mutated glycosylation site: N154Q | This study |
| pBRCj0114N172Q | Derivative of pBRCj0114 with mutated glycosylation site: N172Q | This study |
| pBRCj0114N178Q | Derivative of pBRCj0114 with mutated glycosylation site: N178Q | This study |
| pBRCj0114D172E | Derivative of pBRCj0114 with mutated glycosylation site: D172E | This study |
| pBRCj0114D172A | Derivative of pBRCj0114 with mutated glycosylation site: D172A | This study |
| pBRCj0114N172QN100Q | Derivative of pBRCj0114 with mutated glycosylation sites: N172QN100Q | This study |
| pBRCj0114N172QN178Q | Derivative of pBRCj0114 with mutated glycosylation sites: N172QN178Q | This study |
| pBRDfd0114 | His-tagged Dfd0114 constitutively expressed from pBR322 | This study |
| pBRDfd0114N101Q | Derivative of pBRDfd0114 with mutated glycosylation site: N101Q | This study |
| pBRDfd0114N107Q | Derivative of pBRDfd0114 with mutated glycosylation site: N107Q | This study |
| pBRNt0114 | His-tagged Nt0114 constitutively expressed from pBR322 | This study |
| pBRDfd0114N118Q | Derivative of pBRDfd0114 with mutated glycosylation site: N118Q | This study |
| pMLBADNtPglC | Codon-optimised *N. tergarcus pglC* cloned into pMLBAD | This study |
| pMLBADSlPglC | Codon-optimised *S. lithotrophicum pglC* cloned into pMLBAD | This study |
| pNGRP | His-tagged, truncated version (amino acids 1 to 180) of Cj0114 in pEXT20 | This study |

**Table II.** Presence of critical amino acid residues in PglB orthologues from *N. tergarcus, S. lithotrophicum* and *Deferribacter desulfuricans.*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **D56** | **R147** | **D154/156** | **E319** | **R331** | **R375** | **WWDYG** | **I572** |
| **NtPglB** | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** |
| **SlPglB** | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** | **✓** |
| **DfdPglB** | **✓** | **✓** | **✓** | **✓** | **🗶** | **✓** | **✓** | **🗶** |

**Table III.** Antibodies used in this study

|  |  |  |
| --- | --- | --- |
| **Antibody** | **Dilution** | **Source/Reference** |
| **Mouse anti-hexahistidine monoclonal** | **1:1000** | **Abcam** |
| **Rabbit anti-hexahistidine polyclonal** | **1:2000** | **Abcam** |
| **Mouse anti-*F. tularensis* LPS monoclonal** | **1:1000** | **Abcam** |
| **Rabbit anti-*C. jejuni* heptasaccharide polyclonal** | **1:500** | **(Schwarz et al. 2011)** |
| **Rabbit anti *E. coli* O9 LPS polyclonal** | **1:500** | **(Cuthbertson et al. 2005)** |
| **IRDye® 680RD Goat-anti mouse** | **1:20000** | **LI-COR bioscience** |
| **IRDye® 680RD Goat-anti rabbit** | **1:20000** | **LI-COR bioscience** |
| **IRDye® 800 CW Goat-anti mouse** | **1:20000** | **LI-COR bioscience** |
| **IRDye® 800 CW Goat-anti rabbit** | **1:20000** | **LI-COR bioscience** |