Glycobiology vol. 21 no. 1 pp. 45–54, 2011
doi:10.1093/glycob/cwq130
Advance Access publication on September 16, 2010

Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo

Flavio Schwarz, Christian Lizak, Yao-Yun Fan, Susanna Fleurkens, Michael Kowarik[2], and Markus Aebi[1]

Department of Biology, Institute of Microbiology, ETH Zürich, Wolfgang-Pauli-Str. 10, CH-8093 Zürich, Switzerland

Received on June 24, 2010; revised on August 20, 2010; accepted on August 25, 2010

A number of proteobacteria carry the genetic information to perform N-linked glycosylation, but only the protein glycosylation (*pgl*) pathway of *Campylobacter jejuni* has been studied to date. Here, we report that the *pgl* gene cluster of *Campylobacter lari* encodes for a functional glycosylation machinery that can be reconstituted in *Escherichia coli*. We determined that the *N*-glycan produced in this system consisted of a linear hexasaccharide. We found that the oligosaccharyltransferase (OST) of *C. lari* conserved a predominant specificity for the primary sequence D/E–X_{-1}–N–X_{+1}–S/T (where X_{-1} and X_{+1} can be any amino acid but proline). At the same time, we observed that this enzyme exhibited a relaxed specificity toward the acceptor site and modified asparagine residues of a protein at sequences DANSG and NNNST. Moreover, *C. lari pgl* glycosylated a native *E. coli* protein. Bacterial N-glycosylation appears as a useful tool to establish a molecular description of how single-subunit OSTs perform selection of glycosyl acceptor sites.

Keywords: Campylobacter lari, N-glycosylation, oligosaccharyltransferase

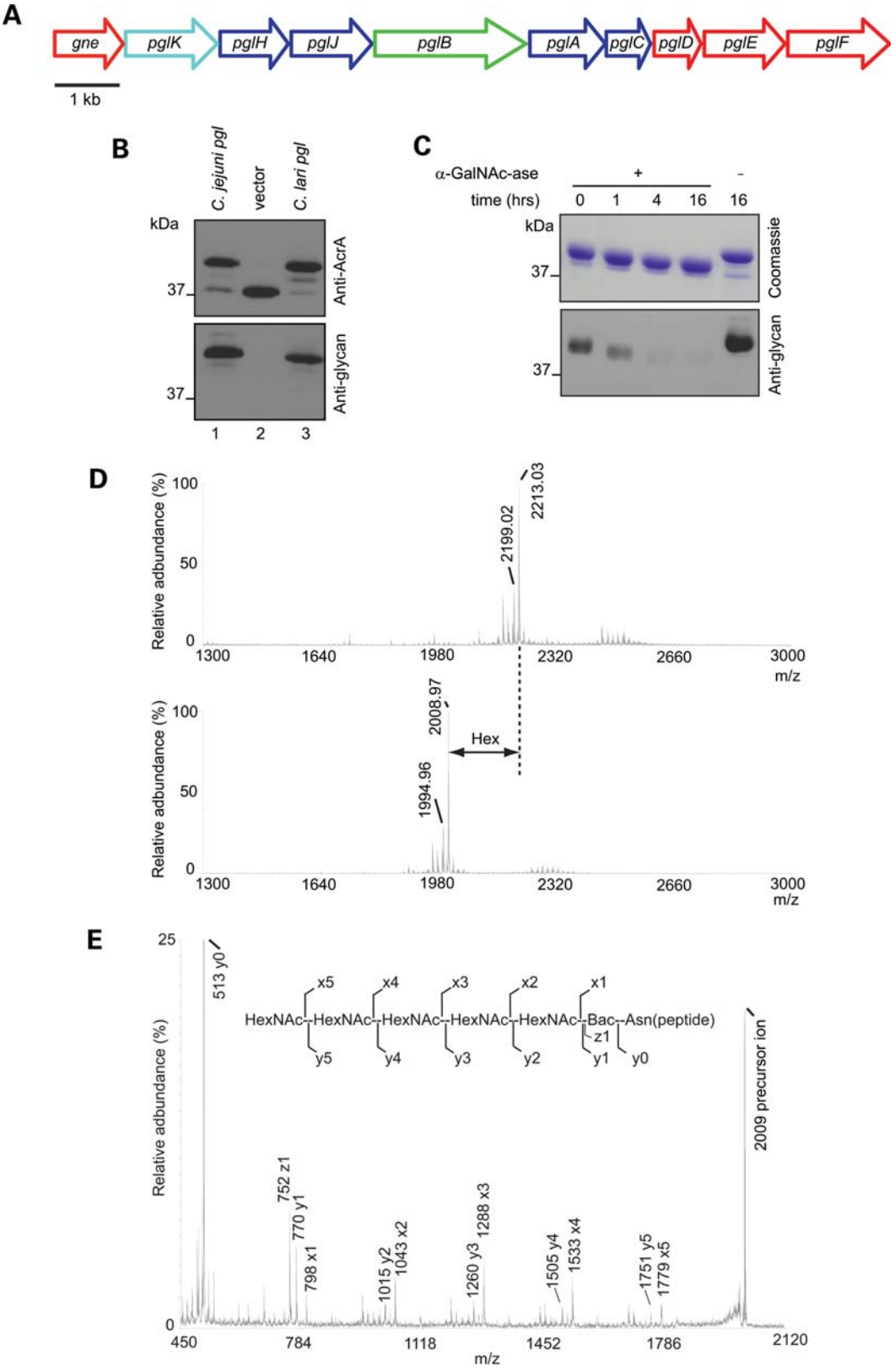
Introduction

N-linked glycosylation is a ubiquitous protein modification that involves the decoration of selected asparagine residues within the sequences N–X_{+1}–S/T (X_{+1}≠P) with an oligosaccharide. This process is organized in a topologically equivalent way in eukaryotes and prokaryotes (Yurist-Doutsch et al. 2008) and involves (I) the assembly of the glycan donor on a lipid anchor at the cytoplasmic side of the ER or of the plasma membrane, (II) reorientation to the opposite face of the membrane and (III) subsequent attachment to the acceptor protein (Burda and Aebi 1999). The defining event of N-glycosylation is the formation of the N-glycosidic bond between the oligosaccharide and the asparagine side chain of the acceptor protein. This reaction is catalyzed by the oligosaccharyltransferase (OST), a single-subunit enzyme in bacteria, archea and protozoa, but a multimeric protein complex in other eukaryotes (Kelleher and Gilmore 2006). The Stt3 protein represents the central component of the OST complex, and it is believed to be the catalytic subunit (Feldman et al. 2005; Igura et al. 2008). However, limited structure/function information on the Stt3 protein limits our understanding of how the poorly reactive amido group of the asparagine side chain is activated to form the N-glycosidic bond to the oligosaccharide (Igura et al. 2008; Maita et al. 2009).

An OST identifies acceptor sequences in polypeptides and modifies only some of the potential sites. Amino acids around the consensus sequence impact the probability of glycosylation of the particular site (Petrescu et al. 2004). According to the current model, different strategies evolved to extend the substrate range of N-linked glycosylation. *Campylobacter jejuni* OST (called PglB) transfers a heptasaccharide to the acceptor protein, stringently within the sequon D/E–X_{-1}–N–X_{+1}–S/T (X_{-1}, X_{+1} ≠P) (Kowarik, Young et al. 2006). In protozoa, substrate diversity was extended through duplication of the *STT3* loci with subsequent diversification of OST specificity. In fact, *Leishmania major* and *Trypanosoma brucei* bear different OST isoforms with distinct donor and acceptor specificities (Nasab et al. 2008; Izquierdo et al. 2009). That is, protozoa are able to assemble different glycans on distinct sequons, thus increasing the complexity of their glycoproteome. In other eukaryotes, the core Stt3 is supplemented with a variable number of additional functions that assist and refine the glycosylation process (Kelleher and Gilmore 2006). Along this line, ribophorin I and Ost3/6p are connected to the glycosylation status of membrane proteins and the site occupancy of specific sequons, respectively (Wilson and High 2007; Schulz et al. 2009). These auxiliary OST subunits are believed to preserve distinct protein substrates in an accessible, unfolded state, leading to an increased number of sites accessible to the Stt3 protein.

Here, we attempted to mine the diversity of natural bacterial OSTs in order to extend our understanding of bacterial N-glycosylation. In particular, we investigated the role of the acidic amino acid invariantly present in the consensus sequence of *C. jejuni* glycoproteins and analyzed whether this

[1]To whom correspondence should be addressed: Tel: +41-44-632-64-13; Fax: +41-44-632-13-75; e-mail: aebi@micro.biol.ethz.ch
[2]Present address: GlycoVaxyn AG, Grabenstrasse 3, 8952 Schlieren, Switzerland


A
gne
pglK
pglH
pglJ
pglB
pglA
pglC
pglD
pglE
pglF
1 kb
B
C. jejuni pgl
vector
C. lari pgl
kDa
37
Anti-AcrA
Anti-glycan
1 2 3
C
α-GalNAc-ase
+
-
time (hrs)
0 1 4 16 16
kDa
37
Coomassie
Anti-glycan
D
Relative adbundance (%)
2213.03
2199.02
2008.97
Hex
1994.96
1300 1640 1980 2320 2660 3000
m/z
E
513 y0
HexNAc-HexNAc-HexNAc-HexNAc-HexNAc-Bac-Asn(peptide)
x5 x4 x3 x2 x1
y5 y4 y3 y2 y1 y0
z1
2009 precursor ion
752 z1
770 y1
798 x1
1015 y2
1043 x2
1260 y3
1288 x3
1505 y4
1533 x4
1751 y5
1779 x5
450 784 1118 1452 1786 2120
m/z

requirement was conserved by N-glycosylation machineries of other bacterial species. We found that the protein glycosylation (*pgl*) locus of *Campylobacter lari* encodes for a functional glycosylation system, able to synthesize a linear hexasaccharide and to transfer it to proteins. We observed that *Cl*PglB modified variants of the *C. jejuni* glycoprotein AcrA that lack the canonical consensus sequence D/E–X_{-1}–N–X_{+1}–S/T. We proved that *Cl*PglB transfers the glycan to sites that lack an acidic residue at the −2 position of the AcrA protein, albeit with low efficiency. These data suggested that the negatively charged amino acid of the glycosylation site was not absolutely required for catalysis by bacterial OST, and it raised the question whether the amino acid at the −2 position was involved in the recognition of the acceptor site. Furthermore, we observed that the nonconsensus sequence DANSGT was glycosylated by *Cl*PglB.

Results

C. lari pgl glycosylates proteins with a linear hexasaccharide

The *C. jejuni pgl* cluster is the first and the only bacterial glycosylation system characterized to date. However, in the last few years, genome sequencing unveiled a broad distribution of *pgl* loci among ε- and δ-proteobacteria (Szymanski and Wren 2005). Some of the *pgl* loci share a genetic organization analogous to the one of *C. jejuni*, whereas others present insertions of a variable number of open reading frames (ORFs) between the *pgl* genes. Interestingly, *Helicobacter canadensis* and *Helicobacter pullorum* carry two *pglB* paralogs in distinct locations of the chromosome (Langdon et al. 2009; Jervis et al. 2010). While *pglB* homologs exhibit a variable degree of similarity in their sequences and there is not yet conclusive evidence that they encode for functional OSTs, they all display the conserved motif WWDXG and are predicted to arrange in a comparable topology. Moreover, analysis of the *pgl* loci revealed the presence of functions involved in the bacillosamine biosynthesis, suggesting that this monosaccharide might constitute a common trait of *Campylobacter* N-glycosylation.

We studied the *C. lari* glycosylation system since we identified significant elements of difference among an overall similarity to the *C. jejuni* pathway. The *pgl* cluster extends for about 12 kb on the *C. lari* chromosome and contains 10 ORFs (Figure 1A). The gene (*pglI*) encoding for the glucosyltransferase that attaches the branching glucose to the *C. jejuni* glycan was not found, whereas the other functions of the *C. jejuni pgl* locus were identified based on sequence similarity of the ORFs. *Cl*PglB presents the WWDXG motif and shares 56% identity to *Cj*PglB. We isolated the *C. lari pgl* locus and expressed it in combination with the *C. jejuni* AcrA protein in *Escherichia coli* cells. Immunoblot analysis indicated glycosylation of AcrA (Figure 1B). Purified glycoproteins were detected using serum specific for the *C. jejuni N*-glycan (hR6) (Figure 1C). We have previously showed that this serum recognizes a linear stretch of GalNAc residues (Schwarz et al. 2010). Treatment of glycosylated AcrA with an exo-α-*N*-acetyl-galactosaminidase (α-GalNAc-ase) resulted in a mobility shift after sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis (PAGE) and in the gradual reduction of reactivity to hR6 over time. This suggested that the *C. lari* glycan contained a linear chain of GalNAc residues, linked in α conformation. To determine the structure of the *N*-glycan of *C. lari*, AcrA was purified from *E. coli* cells carrying either the *C. jejuni* or the *C. lari pgl* cluster, incubated with proteinase K, permethylated and subjected to the mass spectrometric (MS) analysis (Figure 1D). Comparison of the two spectra revealed a difference in mass of 204 Da, matching with a permethylated hexose. The MS/MS spectrum of the ion at $m/z = 2008.97$ determined fragmentation behavior compatible with a linear $Bac(HexNAc)_5$ (Figure 1E). Altogether, these data proved that the *C. lari N*-glycan consisted of GalNAc-α-GalNAc-α-GalNAc-α-GalNAc-α-GalNAc-1,3-Bac.

C. jejuni and C. lari pgl exhibit different acceptor specificities

We aimed to dissect the *C. lari* PglB specificity toward the acceptor protein and attempted a direct comparison to its *C. jejuni* counterpart. AcrA presents five N-*X*-S/T sequences (Figure 2A); two of them exhibit an acidic amino acid at the −2 position to the asparagine and are glycosylated by *Cj*PglB (Kowarik, Young et al. 2006). Expression of AcrA in *E. coli* bearing the *C. jejuni* or the *C. lari pgl* cluster led to a comparable pattern, as judged by immunoblot, corresponding to AcrA glycosylated at two sites (Figure 2B, lanes 5 and 6). The synthesis of glycosylation-incompetent AcrA $N^{2,4}L$ in the presence of the *C. jejuni pgl* resulted in an unmodified acceptor (Figure 2B, lane 3), but expression of AcrA $N^{2,4}L$ in combination with the *C. lari pgl* gave a pattern indicative of glycosylation (Figure 2B, lane 4). The slowest migrating protein, barely visible in the anti-AcrA immunoblot, produced a signal with hR6, implying double glycosylation. This suggested that the activity of *Cl*PglB did not strictly depend on the 5-mer consensus sequence as defined in *C. jejuni*, but it exhibited a different or broader substrate specificity. Moreover, the hR6 serum specifically reacted with an

Fig. 1. Characterization of the *C. lari N*-glycan. (**A**) Genetic organization of the *pgl* cluster of *C. lari*. ORFs are represented by arrows. The names of genes were assigned by homology to the *C. jejuni pgl* locus. Colors indicate putative functions involved in the biosynthesis of activated monosaccharides (red), glycosyltransferases (blue), flippase (cyan) and OST (green). (**B**) Periplasmic extracts from *E. coli* bearing the *C. jejuni pgl* cluster (lane 1), the plasmid pACYC184 (lane 2) or the *C. lari pgl* cluster (lane 3) and co-expressing AcrA were analyzed by SDS–PAGE and immunoblot and probed with anti-AcrA serum (top) or hR6 (bottom). (**C**) Exo-α-GalNAc-ase digestion of the *C. lari N*-glycan. Purified AcrA from *E. coli* cell bearing the *C. lari pgl* locus was incubated with the glycosidase. Aliquots taken at different time points were analyzed by SDS–PAGE, followed by Coomassie staining (top) or immunoblot with hR6 (bottom). (**D**) Glycosylated AcrA was digested with proteinase K, permethylated and subjected to matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) analysis. Top panel: AcrA carrying the *C. jejuni* glycan; bottom panel: AcrA carrying the *C. lari* glycan. The two main peaks display a difference in mass of 204 Da, matching with a permethylated hexose. Minor peaks correspond to under-methylated glycans. (**E**) MALDI-MS/MS spectrum of the precursor ion at $m/z = 2009$, from AcrA carrying the *C. lari* glycan. The fragment ions originating from the sequential loss of monosaccharide residues are indicated in the spectrum.

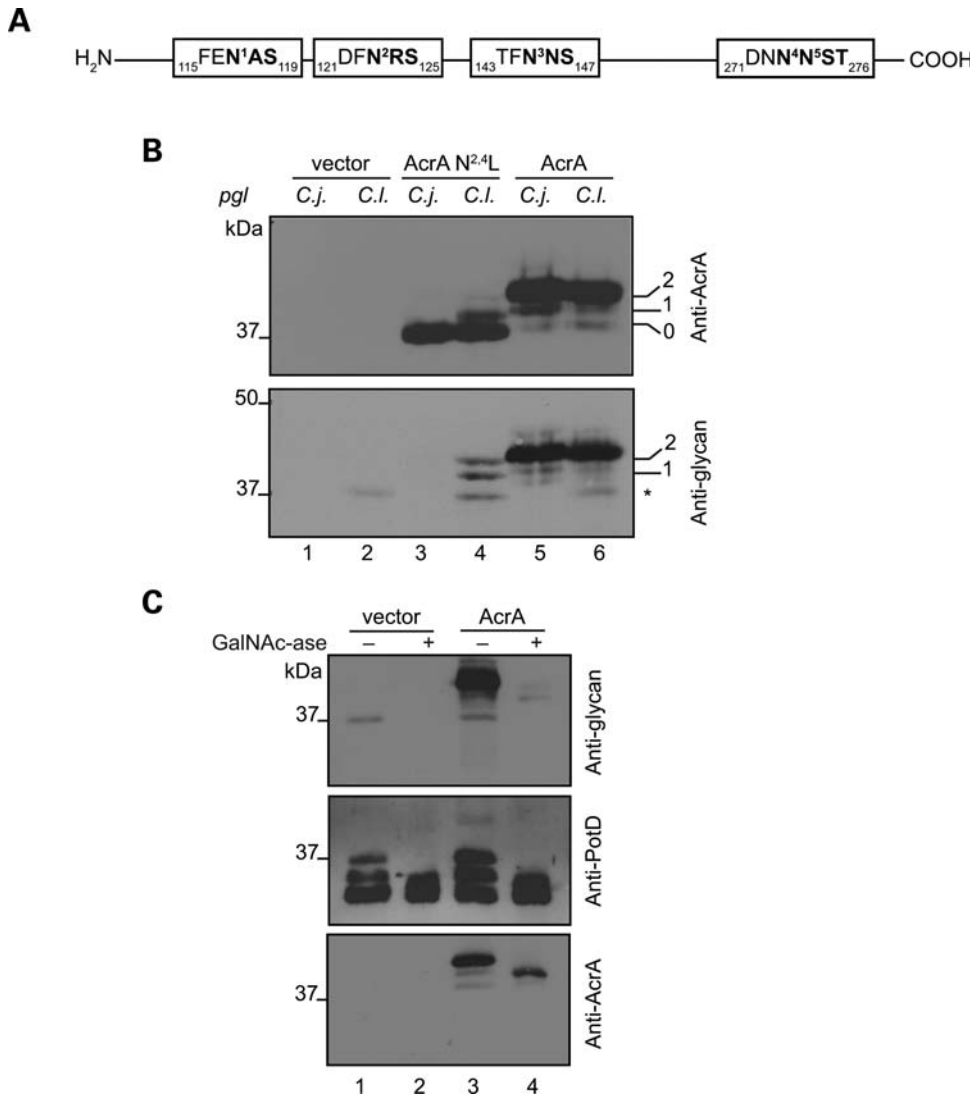


Fig. 2. *C. jejuni* and *C. lari pgl* present different acceptor specificities. (**A**) Schematic representation of the five glycosylation sites of AcrA. (**B**) Periplasmic extracts from *E. coli* bearing the *C. jejuni* or the *C. lari pgl* cluster and expressing either wild-type AcrA or the variant N123L/N273L ($N^{2,4}L$) were analyzed by SDS–PAGE and immunoblot and probed with anti-AcrA serum (top panel) or anti-glycan hR6 (lower panel). Star indicates the *C. lari pgl*-specific protein which reacts to the hR6 serum. Numbers on the left of the gel frame show the electrophoretic mobility of the molecular weight marker. The numbers of *N*-glycans on AcrA are indicated at the right side of the gel frame. (**C**) Periplasmic extracts from *E. coli* bearing the *C. lari pgl* cluster in combination with the pBR322 vector (lanes 1 and 2) or a plasmid encoding for AcrA (lanes 3 and 4) were analyzed by immunoblot and probed with anti-glycan (top), anti-PotD (middle) or anti-AcrA serum (bottom). Glycosylated proteins were incubated with (lanes 2 and 4) or without (lanes 1 and 3) α-GalNAc-ase.

additional periplasmic protein of *E. coli* cells bearing the *C. lari pgl* cluster (Figure 2B, indicated with a star). In order to identify this protein, we separated glycosylated proteins from periplasmic extracts by soy bean agglutinin (SBA) lectin affinity chromatography (Supplementary data, Figure S1A) and analyzed the resulting specimens by MS. We identified the spermidine–putrescine-binding protein PotD of *E. coli*, previously reported not to be a substrate of the *C. jejuni pgl* machinery (Kowarik, Young et al. 2006). Immunoblot analysis confirmed that endogenous PotD was modified by the *C. lari* machinery (Figure 2C). Lastly, we confirmed that the PotD protein was modified at the sites $DD\underline{N}_{26}NT$ and $ES\underline{N}_{62}ET$ by expression of histidine-tagged PotD variants presenting point mutations in these sites (Supplementary data, Figure S1B). Altogether, these data suggested that *C. lari* and *C. jejuni pgl* possess different yet overlapping acceptor specificities.

C. lari PglB glycosylates AcrA lacking the D/E–X_{-1}–N–X_{+1}–S/T site

To investigate whether bacterial OSTs generally exhibit substrate specificities different from that of *C. jejuni*, we designed two versions of the AcrA acceptor protein. In the *C. jejuni* system, the AcrA N273Q mutant conserves a single glycosylation site, while the double mutation D121A/N273Q prevents glycosylation. We analyzed the glycosylation profile of these AcrA variants in *E. coli* SCM3 bearing a plasmid to direct synthesis of the *C. jejuni* glycan and a plasmid encoding for the PglB from *C. jejuni*, *Campylobacter coli* or *C. lari* (Figure 3).

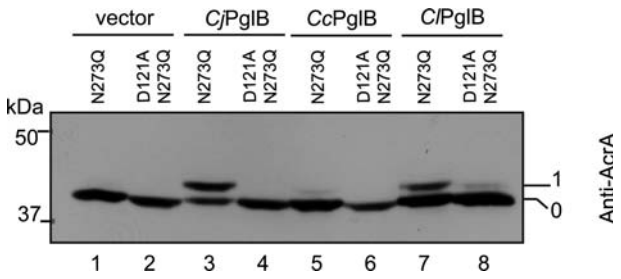


Fig. 3. AcrA glycosylation by bacterial OSTs. Periplasmic extracts of *E. coli* cells transformed with the mutated *C. jejuni pgl* cluster (*pglMut*), expressing either AcrA N273Q (lanes 1, 3, 5 and 7) or AcrA D121A/N273Q (lanes 2, 4, 6 and 8), and bearing a plasmid encoding for *C. jejuni* (lanes 3 and 4), *C. coli* (lanes 5 and 6) or *C. lari* (lanes 7 and 8). Proteins were separated by SDS–PAGE and probed with anti-AcrA serum. The numbers of *N*-glycans on AcrA are indicated at the right side of the gel frame.

Both *C. coli* and *C. lari* PglB transferred the *C. jejuni* glycan to AcrA N273Q, albeit with very low efficiency in the case of *C. coli* (Figure 3, lanes 5 and 7). Strikingly, we could observe glycosylation of the AcrA D121A/N273Q variant only by *C. lari* PglB (Figure 3, lane 8). We speculated that the detected glycosylated product could be either due to glycosylation at a distinct site or due to a mixed population of AcrA glycoforms, each weakly glycosylated at a different site.

Therefore, we undertook a comprehensive analysis of AcrA glycosylation by the *C. lari* system. We first generated mutant forms of AcrA containing N-to-L replacements at each of the N–*X*–S/T sites and studied their glycosylation profile (Figure 4A). Disruption of the sites at N117 (N^1L), N145 (N^3L) and N274 (N^5L) gave a phenotype similar to the wild-type AcrA (Figure 4A, lanes 3, 5 and 7). The N123L (N^2L) mutation clearly affected glycosylation, implying that the site N123 was used by *Cl*PglB (Figure 4A, lane 4). Interestingly, mutation of the N273 (N^4L) reduced the ratio between doubly and mono-glycosylated AcrA (Figure 4A, lane 6). This suggested that the inactivation of the N^4 site led to inefficient modification of an alternative site. At this point, we speculated that the mutation of N^4 directed glycosylation of the adjacent site N^5.

We further constructed a series of AcrA variants presenting multiple point mutations. These proteins were expressed in *E. coli* carrying the *C. lari pgl* locus, purified via nickel affinity chromatography and analyzed by SDS–PAGE. To convincingly demonstrate glycosylation, we incubated the AcrA preparations with α-GalNAc-ase. Purified fractions from cells expressing the triple mutants N117L/N123L/N273L ($N^{1,2,4}$L) and N123L/N145L/N273L ($N^{2,3,4}$L) AcrAs were resolved as a major protein with a mass of about 37 kDa and two larger proteins as visualized by Coomassie staining (Figure 4B, lanes 1 and 3). These proteins were sensitive to glycosidase treatment, suggesting glycosylation (lanes 2 and 4). The hR6 immunoblot supported this observation. Concomitant mutation of the N123, N273 and N274 sites ($N^{2,4,5}$L) resulted in a single hR6-reactive protein (Figure 4B, lane 5). hR6 reactivity disappeared after glycosidase treatment (lane 6). This led us to conclude that N^5 was glycosylated by *Cl*PglB. Surprisingly, simultaneous mutation of four ($N^{1,2,4,5}$L) or all five ($N^{1,2,3,4,5}$L) asparagine residues within the N–*X*–S/T sites gave a phenotype equivalent to the triple mutant $N^{2,4,5}$L (Figure 4B, lanes 7–10). After careful inspection of the AcrA sequence, we suspected that the site $_{257}$DANSGT$_{262}$ might represent an alternative, suboptimal glycosylation site. Indeed, inactivation of this site by mutation of the asparagine residue (N^6) prevented glycosylation by *Cl*PglB (Figure 4B, lane 11).

In order to conclusively prove that the N^5 and N^6 sites were glycosylated, we performed the MS analysis. To overcome the issue of limiting ionization efficiency of glycopeptides, we took advantage of the newly generated data on the *C. lari N*-glycan and developed a procedure to produce useful *N*-glycopeptides. We incubated nickel affinity-purified AcrA variants with trypsin and then, after heat inactivation of the protease, added α-GalNAc-ase. In this way, we produced a mixture of peptides and corresponding bacillosamine-tagged glycopeptides that were subjected to the liquid chromatography-electrospray ionization mass spectrometry (LC-ESI-MS) analysis. The fragmentation spectrum of the triply charged precursor ion at $m/z = 923.8$ matched with the unmodified peptide AVFDNQN^{5}STLLPGAFATITSEGFIQK (Supplementary data, Figure S2A), whereas the fragmentation profile of the triply charged precursor ion at $m/z = 999.8$ was consistent with the presence of bacillosamine at N^5 of the same peptide (Figure 4C). Similarly, the MS/MS profile of the doubly charged precursor ion at $m/z = 871.5$ matched with the peptide LYFIDSVIDAN^{6}SGTVK (Supplementary data, Figure S2B), and ions generated by the fragmentation of the species at $m/z = 985.5$ (doubly charged) were in agreement with bacillosamine-tagged N^6 of the same peptide (Figure 4D). Consequently, glycosylation at the sites DNQNST, DNLNST and DANSGT illustrated that precise positioning of the D/E and S/T residues was not essential on specific sites of AcrA.

A D/E residue at −2 position is not absolutely required for ClPglB glycosylation

Next, we examined the glycosylation profile of AcrA N^2L presenting a D-to-A substitution in the $_{271}$DNN^{4}N^{5}ST$_{276}$ site (Figure 5A). Remarkably, *Cl*PglB transferred a glycan to the ANN^{4}N^{5}ST site, implying that aspartic acid was not essential for catalysis (lanes 1 and 2 and 7 and 8). MS/MS analysis of the precursor ion at the mass corresponding to the glycosylated peptide ($m/z = 2940$) did not allow us to clearly assign the glycan at N^4 or N^5 (Supplementary data, Figure S3).

We extended the study to glycosylation of the DNN^{4}N^{5}AT and DNN^{4}N^{5}SA sites. We observed a gradual increase in the glycosylation efficiency, with ANN^{4}N^{5}ST site being modified the least (Figure 5A, lane 1) and DNN^{4}N^{5}SA the most (Figure 5A, lane 5). This indicated that the aspartic acid had an active role in recruiting the substrate to PglB. Although we did not determine the position of the glycan in the DNN^{4}N^{5}AT site, we reasoned that this was a similar situation to DNQN^{5}ST and DNLN^{5}ST. Accordingly, the S275A mutation provided a suboptimal substrate, whereas T276A resulted in an optimal acceptor site sequence for *Cl*PglB.

ClPglB does not glycosylate a eukaryotic protein at the native glycosylation site

We ruled out the hypothesis that the acidic amino acid in the −2 position was strictly required for catalysis by *Cl*PglB. In

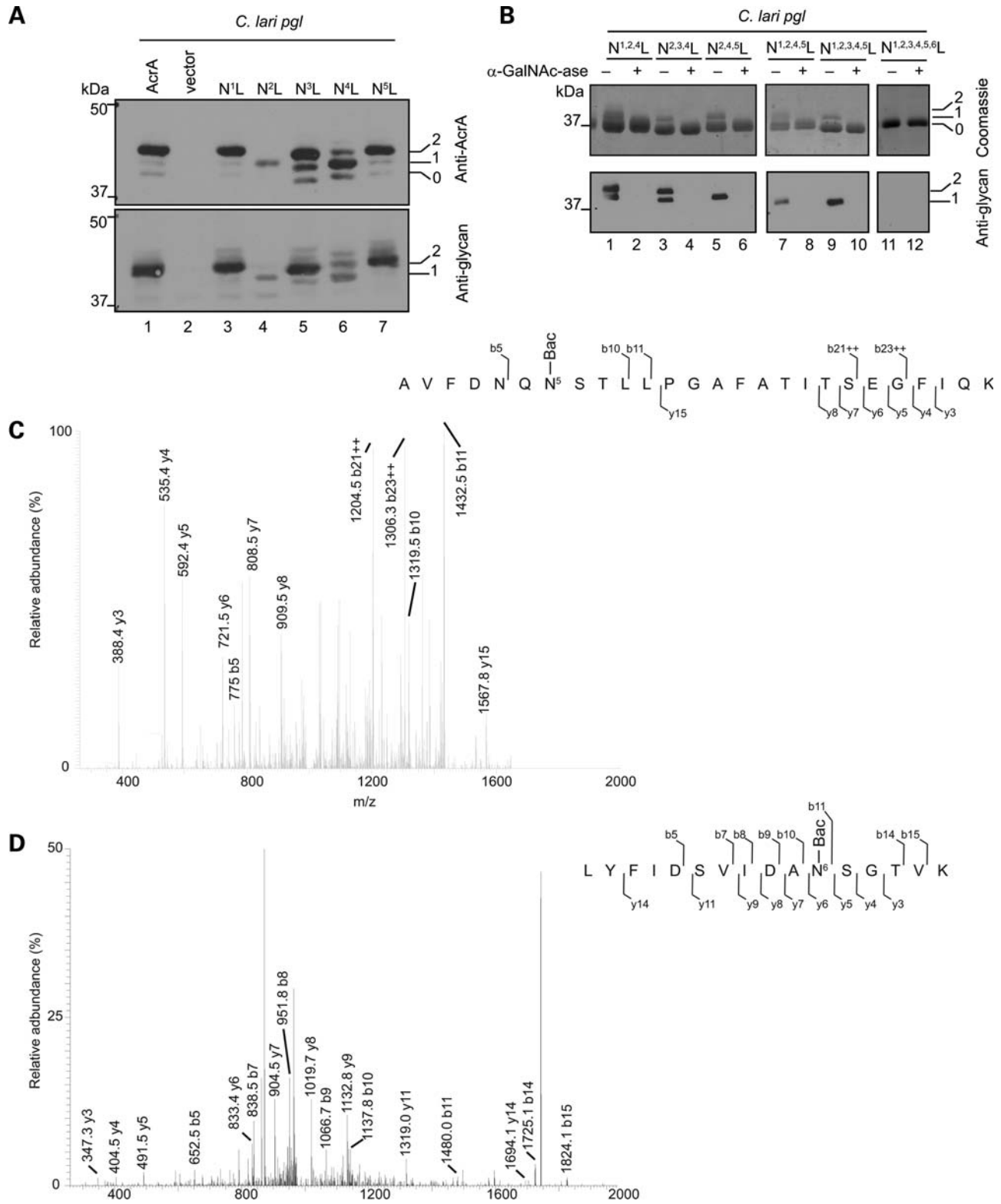


Fig. 4. *Cl*PglB glycosylates AcrA at two noncanonical sites. (**A**) Periplasmic extracts from *E. coli* expressing single-site AcrA variants were analyzed by SDS–PAGE and immunoblot and probed with anti-AcrA serum (top) or hR6 (bottom). Numbers at the right of the gel frame indicate the *N*-glycans on AcrA. (**B**) Nickel affinity-purified AcrA variants were separated by SDS–PAGE, stained by Coomassie (top) or transferred to nitrocellulose membranes and probed with hR6 serum (bottom). AcrAs were incubated with (+) or without (−) α-GalNAc-ase. (**C**) LC-ESI-MS/MS analysis of AcrA D121A/N273Q, purified from *E. coli* cells bearing the *C. lari pgl* cluster. Spectrum from fragmentation of the triply charged precursor ion at $m/z = 999.8$ corresponds to the glycopeptide AVFDNQN(Bac)STLLPGAFATITSEGFIQK. (**D**) LC-ESI-MS/MS analysis of AcrA N123L D271A. Spectrum from fragmentation of the doubly charged ion at $m/z = 985.5$ matches with the glycopeptides LYFIDSVIDAN(Bac)SGTVK.

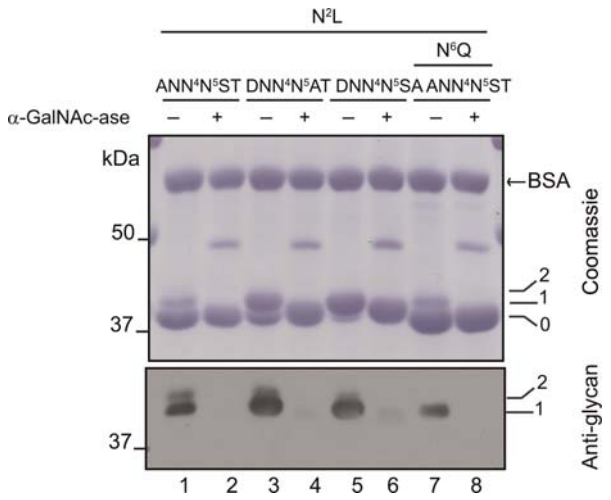


Fig. 5. Acidic amino acid is dispensable for AcrA glycosylation at a distinct site. Nickel affinity-purified AcrA variants were separated by SDS–PAGE, stained by Coomassie (top) or transferred to nitrocellulose membranes and probed with hR6 (bottom) sera. AcrA variants were incubated with (+) or without (−) α-GalNAc-ase.

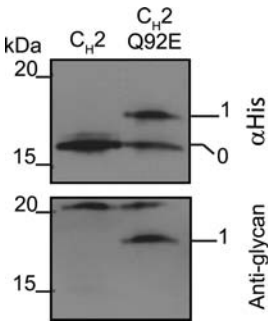


Fig. 6. *Cl*PglB does not glycosylate a eukaryotic protein at the native glycosylation site. SDS–PAGE and immunoblot analysis of periplasmic extracts from *E. coli* cells expressing histidine-tagged C_H2 protein variants. Membranes were probed with anti-histidine tag antibodies (top) or hR6 serum (bottom).

order to test whether *Cl*PglB was active toward the short, eukaryotic glycosylation site N–*X*–S/T, we examined the glycosylation of a eukaryotic protein. We chose immunoglobulin G (IgG) as model acceptor protein because detailed structural information is available for this molecule. IgGs are invariably decorated with a glycan within the QYNST sequon of the C_H2 domain. Analyses of glycosylated and un-glycosylated IgG fragments have revealed that the glycosylation site is located in an extended loop connecting two β-strands (Prabakaran et al. 2008; Girardi et al. 2009). This implies that the acceptor site folds in a flexible structure, a possible prerequisite for *Cj*PglB (Kowarik, Numao et al. 2006). However, expression of wild-type C_H2 domain of a human IgG1 in *E. coli* bearing the *pgl* machinery did not result in glycosylation of the C_H2 protein (Figure 6A). In contrast, the protein variant Q92E was recognized as a substrate and glycosylated.

Discussion

During N-linked *glycosylation*, only a subset of potential acceptor sites is chosen for glycosylation. In yeast and higher eukaryotes, site selection is aided by the concerted activity of the OST subunits, which present specific polypeptides to the Stt3 protein. A similar advantage has been achieved in some protozoa by duplication of the *STT3* gene: OST isoforms featuring various degrees of similarity direct glycosylation of distinct subsets of polypeptides. A limited number of proteobacteria possess a protein N-glycosylation system, probably acquired by the horizontal transfer from archaea. Although duplication of the gene encoding for the OST has been observed in rare cases, most bacteria carry a single copy of *pglB* on the genome. In this respect, N-glycosylating bacteria are likely to mimic an early evolutionary stage. Therefore, bacterial systems appear as ideal candidates to establish an accurate molecular description of how a single-subunit OST performs site selection. This will in turn shed light on the catalytic mechanism of N-glycosylation. Lastly, access to bacterial glycosylation systems represents an important means to explore and exploit the diversity of this glycosylation.

*Cj*PglB modifies asparagine residues within the sequence D/E–X_{-1}–N–X_{+1}–S/T (where X_{-1} and X_{+1} are any amino acids but proline) when located in flexible, exposed structural elements (Kowarik, Numao et al. 2006; Kowarik, Young et al. 2006). The presence of an acidic group at the −2 position to the asparagine seems to be a common prerequisite of *Cj*PglB and *Cl*PglB for efficient glycosylation. However, weak glycosylation of AcrA at the NNNST site by *Cl*PglB illustrates that the acidic amino acid is dispensable for catalysis in vivo and corroborates the hypothesis that the formation of the N-glycosidic bond proceeds via a reaction scheme conserved in bacteria and eukaryotes. Similar conclusions have been drawn from the observation that tripeptides of sequence NLT are modified in vitro by membrane fractions containing *Cj*PglB, in the presence of synthetic lipid-linked oligosaccharides (Glover et al. 2005). Both in vivo and in vitro, the presence of the aspartic acid at the −2 position greatly enhances the efficiency of the reaction. Although it is tempting to invoke involvement of the negatively charged amino acid in the recruitment of acceptor substrate to PglB, the exact role of the −2 position remains elusive. In fact, peptides with the semi-consensus sequence DFNVX (*X* = cysteine, homoserine or diaminopropionic acid) are neither substrates of *Cj*PglB nor behave as inhibitors in vitro (Chen et al. 2007). Similarly, DFQRT peptides do not seem to reduce glycosylation of AcrA in vitro (data not shown). This evidence supports a more complex role for the −2 position. It might be that the aspartic acid, which is a poor α-helix former (Pace and Scholtz 1998), enhances the flexibility required by PglB. Thus, the D121A mutation on AcrA might preclude glycosylation at N^2 by a double effect—reduction of flexibility plus a drop in recognition by PglB. In contrast, the N^4/N^5 sites are predicted to fold in a flexible conformation (Kowarik, Young et al. 2006), and their local structure might be unaffected by the mutation of D271. However, while this hypothesis supports site occupancy of AcrA, it does not explain why IgG1-C_H2 is not glycosylated.

Although we cannot formally exclude the possibility that heterologous expression of *pgl* genes in *E. coli* might affect

their specificities or that Pgl protein levels might differ, selective modification of the PotD protein by *Cl*PglB demonstrates that bacterial glycosylation machineries feature different glycosylation phenotypes. However, the evidence that native *E. coli* proteins are rarely glycosylated suggests that *Campylobacter* acceptor proteins co-evolved with *pgl* systems and are thus tuned to accept *N*-glycans.

Glycosylation of the sequence DANSG is unusual. The vast majority of all *N*-glycans is found within the sequence N-X_{+1}–S/T (where X_{+1} can be any amino acid but proline) (Marshall 1974). Atypical N-glycosylation sites bearing the NXC sequence have been reported for eukaryotic proteins including CD69 (Vance et al. 1997), immunoglobulin μ chain (Kehry et al. 1979), human leptin receptor (Haniu et al. 1998) and von Willebrand factor (Titani et al. 1986). Additionally, a recent global LC-MS/MS analysis of murine glycopeptides has extended the spectrum of nonconsensus recognition motifs to include N–G–X and N–X–V sequences (Zielinska et al. 2010). Structural studies have indicated that the amide of the asparagine networks to the hydroxyl group of the +2 position to form an Asn turn (Imperiali and Hendrickson 1995) and that the side chain of the amino acid at the +1 position are displaced out of this conformation. However, glycosylation of nonconsensus sequences shows that the Asn-turn motif might not be strictly required for protein modification. Recently, it has been described that the asparagine of the noncanonical NSG sequence of the C_H1 domain of human IgG1 is partially glycosylated (Valliere-Douglass et al. 2009). This finding seems similar to what we report on the NSG site of AcrA. Both observations demonstrate that bacterial and eukaryotic OSTs share a similar basis for the recognition of glycosyl acceptor sites. Moreover, occurrence of this particular glycosylation both in bacteria and in eukaryotes implies that this activity is intrinsic in the PglB/Stt3 protein, and it is not due to the absence of potential acceptor substrates. In a follow-up paper, Valliere-Douglass *et al*. (2010) identified additional noncanonical glycosylation sites in IgG1 and proposed that the common feature of these unusual low-efficiency glycosylation sites is the presence of a "reverse-consensus" motif (serine/threonine residue at the −2 position relative to the modified asparagine residue). Therefore, the DANSG site of AcrA might not belong to this class of atypical glycosylation sequences.

Further investigation will be needed to unravel the biological significance of this relaxed glycosylation, whether this has to be regarded as a limitation in the fidelity of OST or as a strategy to extend the number of N-glycosylations.

Materials and methods

Construction of plasmids

All the strains and plasmids used in this study are listed in Supplementary data, Table SI. Oligonucleotides were purchased from Microsynth (Balgach, Switzerland). Unless otherwise stated, *E. coli* DH5α was chosen as host for cloning. Restriction enzymes were purchased from NEB or Fermentas. T4 DNA ligase was from NEB. Plasmid pFLA29 was constructed by annealing of oligos 5′-TGTAGTTAATTAACGATGATAAAAATCTAGCGCTGAGGCAGTGCTGAGC-3′ and 5′-GCTCAGCACTGCCTCAGCGCTAGATTTTTATCATCGTTAATTAAC-3′ and subsequent ligation into pACYC184, previously cut with *Bfm*I and Bpu1102I. Cloning of the *C. lari pgl* cluster was performed by three overlapping polymerase chain reaction (PCR) fragments and homologous recombination in yeast (Oldenburg et al. 1997). Integration of the PCR fragments into the *Sma*I-linearized YEp352 vector was achieved by homologous recombination in SS328 yeast strain during transformation. Transformants were selected on synthetic minimal medium lacking uracil, and recombined plasmids were isolated from yeast cells (Sherman 2002). After re-transformation in *E. coli*, plasmids were confirmed by restriction digestion. A 12 kb fragment was released from YEp352(pgl4) by digestion with *Pac*I and Bpu1102I and ligated into the corresponding sites of pFLA29, thus producing pACYC(pgl4). Plasmid pFLA26 was constructed by subcloning the PsiI/SpeIAcrA fragment from pMIK43 into pFLA19. The *pglB* ORFs were amplified by PCR using *C. coli* gDNA and *C. lari* gDNA. Fragments containing *CcpglB* were cut with *Xba*I/*Hin*dIII and ligated into pMLBAD, previously cut with the same enzymes. Fragments containing *ClpglB* were cut with *Eco*RI and *Pst*I and ligated into pMLBAD, previously cut with the same enzymes. All *pglB* ORFs are in frame with an HA tag at the C-terminus. The *potD* gene of *E. coli* MC1067 was amplified and cloned into pEC415 previously cut with *Nde*I and *Eco*RI. All point mutations were inserted by site-directed mutagenesis with the QuickChange Kit from Stratagene. All constructs were confirmed by restriction analysis and sequencing of relevant fragments (Synergene AG, Switzerland).

Protein expression, purification and analysis

E. coli SCM3 or SCM6 harboring pACYC(pgl) or pACYC (pgl4) and a plasmid coding for an acceptor protein were grown overnight from single colonies at 37°C in volumes of 10 mL of Luria-Bertani (LB) medium. Ampicillin (100 mg/L), trimethoprim (100 mg/L) and chloramphenicol (25 mg/L) were added to the medium, as needed. Preparation of periplasmic extracts was carried out by lysozyme treatment, consisting of an incubation in 30 mM Tris–HCl pH 8.5, 20% (w/v) sucrose, 1 mM EDTA, 1 mg/mL lysozyme (Sigma), for 1 h on ice. Spheroblasts were removed by centrifugation. For protein purification, cultures were scaled up to 0.5 L. For AcrA purification, periplasmic extracts were supplemented with imidazole to reach a final concentration of 10 mM, sterile-filtered (0.22 μm) and purified by nickel affinity chromatography using Ni-NTA agarose (Qiagen). For the separation of glycosylated proteins, periplasmic extracts were dialyzed twice against phosphate-buffered saline (PBS) pH 7.4 and loaded onto an SBA-agarose (Vector laboratories Inc., CA) column. After washing with PBS, bound proteins were eluted with PBS containing 0.5 M galactose. SDS–PAGE was performed according to Lämmli; immunodetection was performed with polyclonal anti-AcrA serum (Wacker et al. 2002), polyclonal hR6 serum (S. Amber and M. Aebi, in preparation), polyclonal anti-PotD serum (Furuchi et al. 1991) and monoclonal anti-His antibody (Qiagen). For the removal of the GalNAc residues from *N*-glycan, 3 μg of glycoproteins was incubated for 3 h (unless differently specified) with 40 U of exo-α-*N*-acetyl-galacosaminidase (NEB) in the presence of 4 μg of bovine serum albumin in 50 mM Tris buffer pH 7.5.

MS analysis

For the characterization of the *N*-glycan of *C. lari*, 50 μg of glycoproteins was incubated with proteinase K (Sigma), overnight at 37°C, at a 1:50 (w/w) enzyme: substrate ratio. Digestion products were passed through a carbon column (ENVI-Carb SPE, Supelco), washed with 0.1% formic acid (FA) and eluted with 25% acetonitrile (ACN), 0.1% FA. Eluent was permethylated and spotted on the plate with 2,5-dihydroxybenzoic acid (DHB) matrix (10 mg/mL in 50% ACN and 0.1% FA). MALDI-MS and MS/MS were acquired in the positive ion reflectron mode by MALDI-TOF/TOFMS with 4800 MALDI TOF/TOF Analyzer instrument (Applied Biosystems, CA). For glycopeptide identification, 0.1 mg of glycoproteins was incubated with trypsin (Promega), overnight at 37°C. After heat inactivation, tryptic digestion products were further treated with exo-α-*N*-acetyl-galacosaminidase from 8 h to overnight. Sep-Pak C18 was used to desalt and remove enzymes. Peptides were eluted with 70% ACN and 0.1% FA and analyzed with an LTQ-FT-ICR-MS instrument (Thermo Scientific, MA). Samples were injected into an Eksigent-nano-HPLC system (Eksigent Technologies, CA) by an autosampler and separated on a self-made reverse-phase tip column (75 μm × 80 mm) packed with C18 material (AQ, 3 μm, 200 A, Bischoff GmbH, Germany). The column was equilibrated with solvent A (A: 3% ACN and 0.2% FA, in water). Peptides were eluted using the following gradient: 0–50 min, 0–60% B; 50–53 min, 60–97% B; 53–60 min, 97% B (B: 80% ACN and 0.2% FA, in water) at a flow rate of 0.2 μL/min. High accuracy mass spectra were acquired at an LTQ-ICR-FT in the mass range of 300–2000 *m*/*z* and a target value of 5×10^5 ions. Up to four data-dependent MS/MS spectra of the most intense ions with charge state 2+ or higher were recorded in parallel at the ion trap using collision-induced dissociation.

Supplementary data

Supplementary data for this article is available online at http://glycob.oxfordjournals.org/.

Funding

This work was supported by the Swiss National Science Foundation (grant 3100AQ-105541 to M.A.) and the ETH Zurich.

Acknowledgments

We thank the members of the Aebi laboratory for fruitful discussions. We are grateful to Manuela Mally and Karin Ilg for critically reading the manuscript. We thank Miguel Valvano and Cristina Marolda for the gift of *E. coli* SCM6 and Brendan Wren and Herbert Hächler for the genomic DNA of *C. lari* and *C. coli*. F.S., C.L. and Y.-Y.F. are students of the Life Sciences Zurich Graduate School.

Conflict of interest statement

None declared.

Abbreviations

ACN, acetonitrile; Bac, bacillosamine; *Cj*PglB, *Campylobacter jejuni* PglB; *Cl*PglB, *Campylobacter lari* PglB; DHB, 2,5-dihydroxybenzoic acid; FA, formic acid; GalNAc, *N*-acetyl-galactosamine; IgG, immunoglobulin G; LB, Luria-Bertani; LC-ESI, liquid chromatography-electrospray ionization mass spectrometry; MALDI-MS, matrix-assisted laser desorption/ionization mass spectrometry; MS, mass spectrometry; ORF, open reading frame; OST, oligosaccharyltransferase; PBS, phosphate-buffered saline; PCR, polymerase chain reaction; pgl, protein glycosylation; SBA, soy bean agglutinin; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

References

Burda P, Aebi M. 1999. The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta*. 1426:239–257.

Chen MM, Glover KJ, Imperiali B. 2007. From peptide to protein: Comparative analysis of the substrate specificity of N-linked glycosylation in *C. jejuni*. *Biochemistry*. 46:5579–5585.

Feldman MF, Wacker M, Hernandez M, Hitchen PG, Marolda CL, Kowarik M, Morris HR, Dell A, Valvano MA, Aebi M. 2005. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc Natl Acad Sci USA*. 102:3016–3021.

Furuchi T, Kashiwagi K, Kobayashi H, Igarashi K. 1991. Characteristics of the gene for a spermidine and putrescine transport system that maps at 15 min on the *Escherichia coli* chromosome. *J Biol Chem*. 266:20928–20933.

Girardi E, Holdom MD, Davies AM, Sutton BJ, Beavil AJ. 2009. The crystal structure of rabbit IgG-Fc. *Biochem J*. 417:77–83.

Glover KJ, Weerapana E, Numao S, Imperiali B. 2005. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem Biol*. 12:1311–1315.

Haniu M, Arakawa T, Bures EJ, Young Y, Hui JO, Rohde MF, Welcher AA, Horan T. 1998. Human leptin receptor. Determination of disulfide structure and N-glycosylation sites of the extracellular domain. *J Biol Chem*. 273:28691–28699.

Igura M, Maita N, Kamishikiryo J, Yamada M, Obita T, Maenaka K, Kohda D. 2008. Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J*. 27:234–243.

Imperiali B, Hendrickson TL. 1995. Asparagine-linked glycosylation: Specificity and function of oligosaccharyl transferase. *Bioorg Med Chem*. 3:1565–1578.

Izquierdo L, Schulz BL, Rodrigues JA, Guther ML, Procter JB, Barton GJ, Aebi M, Ferguson MA. 2009. Distinct donor and acceptor specificities of *Trypanosoma brucei* oligosaccharyltransferases. *EMBO J*. 28:2650–2661.

Jervis AJ, Langdon R, Hitchen P, Lawson AJ, Wood A, Fothergill JL, Morris HR, Dell A, Wren B, Linton D. 2010. Characterisation of N-linked protein glycosylation in *Helicobacter pullorum*. *J Bacteriol*. in press.

Kehry M, Sibley C, Fuhrman J, Schilling J, Hood LE. 1979. Amino acid sequence of a mouse immunoglobulin mu chain. *Proc Natl Acad Sci USA*. 76:2932–2936.

Kelleher DJ, Gilmore R. 2006. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology*. 16:47–62.

Kowarik M, Numao S, Feldman MF, Schulz BL, Callewaert N, Kiermaier E, Catrein I, Aebi M. 2006a. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science*. 314:1148–1150.

Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, *et al* 2006b. Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J*. 25: 1957–1966.

Langdon RH, Cuccui J, Wren BW. 2009. N-linked glycosylation in bacteria: An unexpected application. *Future Microbiol*. 4:401–412.

Maita N, Nyirenda J, Igura M, Kamishikiryo J, Kohda D. 2009. Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J Biol Chem*. 285:4941–4950.

Marshall RD. 1974. The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp*. 40:17–26.

Nasab FP, Schulz BL, Gamarro F, Parodi AJ, Aebi M. 2008. All in one: *Leishmania major* STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. *Mol Biol Cell*. 19:3758–3768.

Oldenburg KR, Vo KT, Michaelis S, Paddon C. 1997. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res*. 25:451–452.

Pace CN, Scholtz JM. 1998. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys J*. 75:422–427.

Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR. 2004. Statistical analysis of the protein environment of N-glycosylation sites: Implications for occupancy, structure, and folding. *Glycobiology*. 14:103–114.

Prabakaran P, Vu BK, Gan J, Feng Y, Dimitrov DS, Ji X. 2008. Structure of an isolated unglycosylated antibody C(H)2 domain. *Acta Crystallogr D Biol Crystallogr*. 64:1062–1067.

Schulz BL, Stirnimann CU, Grimshaw JP, Brozzo MS, Fritsch F, Mohorko E, Capitani G, Glockshuber R, Grutter MG, Aebi M. 2009. Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc Natl Acad Sci USA*. 106:11061–11066.

Schwarz F, Huang W, Li C, Schulz BL, Lizak C, Palumbo A, Numao S, Neri D, Aebi M, Wang LX. 2010. A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat Chem Biol*. 6:264–266.

Sherman F. 2002. Getting started with yeast. *Methods Enzymol*. 350:3–41.

Szymanski CM, Wren BW. 2005. Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol*. 3:225–237.

Titani K, Kumar S, Takio K, Ericsson LH, Wade RD, Ashida K, Walsh KA, Chopek MW, Sadler JE, Fujikawa K. 1986. Amino acid sequence of human von Willebrand factor. *Biochemistry*. 25:3171–3184.

Valliere-Douglass JF, Kodama P, Mujacic M, Brady LJ, Wang W, Wallace A, Yan B, Reddy P, Treuheit MJ, Balland A. 2009. Asparagine-linked oligosaccharides present on a non-consensus amino acid sequence in the CH1 domain of human antibodies. *J Biol Chem*. 284:32493–32506.

Valliere-Douglass JF, Eakin CM, Wallace A, Ketchem RR, Wang W, Treuheit MJ, Balland A. 2010. Glutamine-linked and non-consensus asparagine-linked oligosaccharides present in human recombinant antibodies define novel protein glycosylation motifs. *J Biol Chem*. 285: 16012–16022.

Vance BA, Wu W, Ribaudo RK, Segal DM, Kearse KP. 1997. Multiple dimeric forms of human CD69 result from differential addition of *N*-glycans to typical (Asn-X-Ser/Thr) and atypical (Asn-X-cys) glycosylation motifs. *J Biol Chem*. 272:23117–23122.

Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, *et al* 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*. 298:1790–1793.

Wilson CM, High S. 2007. Ribophorin I acts as a substrate-specific facilitator of N-glycosylation. *J Cell Sci*. 120:648–657.

Yurist-Doutsch S, Chaban B, VanDyke DJ, Jarrell KF, Eichler J. 2008. Sweet to the extreme: Protein glycosylation in Archaea. *Mol Microbiol*. 68:1079–1084.

Zielinska DF, Gnad F, Wisniewski JR, Mann M. 2010. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell*. 141:897–907.