Protein glycosylation in bacteria: sweeter than ever

Harald Nothaft and Christine M. Szymanski

Abstract | Investigations into bacterial protein glycosylation continue to progress rapidly. It is now established that bacteria possess both *N*-linked and *O*-linked glycosylation pathways that display many commonalities with their eukaryotic and archaeal counterparts as well as some unexpected variations. In bacteria, protein glycosylation is not restricted to pathogens but also exists in commensal organisms such as certain *Bacteroides* species, and both the *N*-linked and *O*-linked glycosylation pathways can modify multiple proteins. Improving our understanding of the intricacies of bacterial protein glycosylation systems should lead to new opportunities to manipulate these pathways in order to engineer glycoproteins with potential value as novel vaccines.

S-layer
Two-dimensional array of protein or glycoprotein subunits, each with molecular masses ranging from 40 to 200 kDa, that are common constituents of bacterial cell walls.

Department of Biological Sciences and the Alberta Ingenuity Centre for Carbohydrate Science, University of Alberta, Edmonton, Alberta T6G 2E9, Canada.
e-mails: nothaft@ualberta.ca; cszymans@ualberta.ca
doi:10.1038/nrmicro2383

Glycosylation is the most abundant polypeptide chain modification in nature. Glycans can be covalently attached to the amide nitrogen of Asn residues (*N*-glycosylation), to the hydroxyl oxygen of, typically, Ser or Thr residues (*O*-glycosylation), and, in rare cases, to the indole C2 carbon of Trp through a C–C linkage (*C*-mannosylation[1], which will not be discussed further). Protein glycosylation was first demonstrated in the late 1930s[2] and was long thought to exist only in eukaryotes. As more than two-thirds of eukaryotic proteins are predicted to be glycosylated[3] and these modifications are essential for a multitude of cellular functions[4], it is not surprising that countless publications have been dedicated to this topic. It took 40 years until the first bacterial and archaeal glycoproteins were discovered on the surface layers (S-layers) of the archaeon *Halobacterium salinarum*[5] and on the S-layers of two hyperthermophilic *Clostridium* species[6,7]. Glycosylation of bacterial S-layers will not be described further here, but it involves the addition of *O*-linked long-chain glycan repeats to S-layer proteins in a process that is similar to lipopolysaccharide (LPS) biosynthesis (see REF. 8 for a review). A distinguishing feature of S-layer glycosylation is that the *O*-linkages can be formed with Ser, Thr or Tyr residues[9].

More than 20 years later, the first bacterial *N*-linked protein glycosylation (Pgl) pathway was described in the epsilonproteobacterium *Campylobacter jejuni*[10–13] (FIG. 1a). Since then, a wealth of information has been gathered about this system through studies in the native host, heterologous expression in *Escherichia coli* and *in vitro* analyses. The first part of this Review focuses on studies of *N*-glycosylation that have been published in the past 5 years (for a review of earlier work, see REF. 14). Our understanding of bacterial flagellin- and pilin-specific *O*-glycosylation systems has also been growing, and general *O*-glycosylation pathways have been identified recently as a result of the development of new and more sensitive detection methods (BOX 1). These *O*-glycosylation systems are described in the second part of this Review. We also summarize some unconventional pathways for the biosynthesis of both *O*-linked and *N*-linked glycoproteins.

Protein *N*-glycosylation in bacteria

The* Campylobacter jejuni *general glycosylation pathway. Currently, *C. jejuni* is the only bacterium with a well-characterized *N*-glycosylation pathway, and it was the first bacterium for which the glycosylation machinery was fully reconstituted in *E. coli*[13,15]. In *C. jejuni*, a heptasaccharide is built on the cytoplasmic side of the inner membrane on a lipid-linked precursor, undecaprenyl phosphate (Und-P)[15–17] (FIG. 1a); the enzymatic steps are discussed in more detail below. The resulting lipid-linked oligosaccharide (LLO) is then translocated across the inner membrane into the periplasmic space by an ATP-dependent flippase, PglK[18,19], and transferred to Asn residues in target proteins by a bacterial oligosaccharyltransferase (OTase), PglB[13]. In eukaryotes and archaea, the Asn residues are located in an Asn-X-Ser/Thr motif (where X represents any amino acid except Pro), whereas the bacterial system requires an extended *N*-glycosylation consensus sequence that contains an Asp or Glu at the –2 position (Asp/Glu-X_1-Asn-X_2-Ser/Thr, where X_1 and X_2 represent any amino acid except Pro)[20].

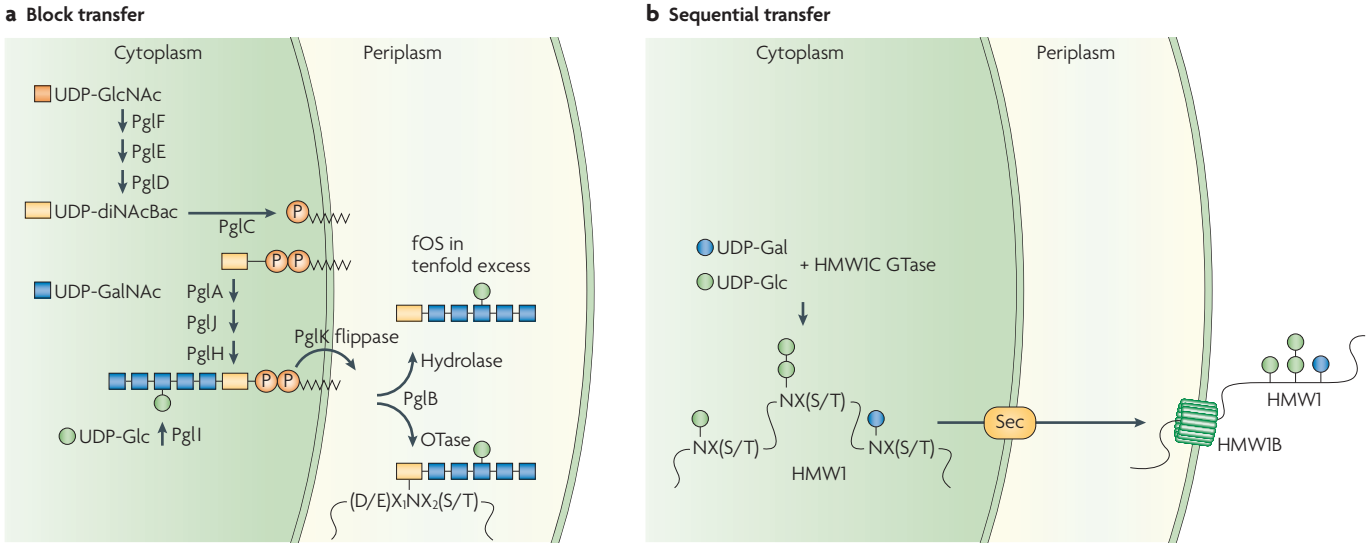


Figure 1 | **Overview of bacterial *N*-linked pathways for protein glycosylation. a** | The mechanism of block transfer for *Campylobacter jejuni*, which is the prototype for the bacterial *N*-linked protein glycosylation system. The undecaprenyl pyrophosphate-linked heptasaccharide is assembled in the cytosol by the addition of the indicated sugars from nucleotide-activated donors (see main text for details). The complete heptasaccharide is translocated across the inner membrane into the periplasm by the protein glycosylation K (PglK) protein, an ATP-binding cassette (ABC)-type transporter. The oligosaccharide is transferred to the amino group of Asn in the protein consensus sequence (Asp/Glu-X_1-Asn-X_2-Ser/Thr, in which X_1 and X_2 are any amino acid except Pro), or released into the periplasm as free oligosaccharides (fOS) by the oligosaccharyltransferase (OTase) PglB. In *C. jejuni*, the fOS/*N*-glycan ratio is approximately 10/1 under standard growth conditions. **b** | The sequential transfer of sugars to proteins in *Haemophilus influenzae*. Sugars (galactose (Gal) and glucose (Glc)) from nucleotide-activated donors are transferred to the eukaryotic-like Asn-X-Ser/Thr sequon of high-molecular-weight adhesin 1 (HMW1) by the glycosyltransferase (GTase) HMW1C in the cytoplasm and are elongated by the same enzyme. The glycoprotein then proceeds through the Sec translocation apparatus and through its cognate outer-membrane channel-forming β-barrel translocator protein, HMW1B, to which it remains tethered. UDP-diNAcBac, UDP-2,4-diacetamido bacillosamine (UDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucose); GalNAc, *N*-acetylgalactosamine; GlcNAc, *N*-acetylglucosamine; P, phosphate; UDP, uridine diphosphate.

Furthermore, in contrast to the eukaryotic OTase, the bacterial OTase can transfer sugars post-translationally to locally flexible structures in folded proteins[21].

More than 65 *C. jejuni* proteins of varying function have been shown to be *N*-glycosylated[22,23], and it is predicted that up to 150 proteins can be modified[22]. Moreover, the increasing number of complete bacterial genome sequences has revealed that genes common to the *N*-glycan pathway are present in all *Campylobacter* species (W. G. Miller, personal communication). Orthologues of Pgl proteins, including the central OTase, have also been found in other related deltaproteobacteria and epsilonproteobacteria (see below and FIG. 2), implying that this protein modification system is more widespread than was originally thought.

Biosynthesis of the* C. jejuni *heptasaccharide. In the *C. jejuni* cytoplasm, *N*-glycan biosynthesis starts with uridine diphosphate (UDP)-activated *N*-acetylglucosamine (UDP-GlcNAc) (FIG. 1a). First PglF, a C6 dehydratase, generates UDP-2-acetamido-2,6-dideoxy-D-xylo-4-hexulose (a UDP-4-keto-sugar) in an NADH-dependent hydride transfer from C4 of UDP-GlcNAc to C6, in conjunction with the elimination of water across the glycosyl C5 and C6 bonds[24]. Then PglE, an aminotransferase, catalyses the pyridoxal-dependent transfer of an amino group from L-glutamate to C4 of the UDP-4-keto-sugar to form UDP-2-acetamido-4-amino-2,4,6-trideoxy-α-D-glucose (a UDP-4-amino-sugar)[24]. PglD then acetylates the C4 group on the UDP-4-amino-sugar to form UDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucose (also known as UDP-2,4-diacetamido bacillosamine; UDP-diNAcBac), using acetyl C oA as an acetyl donor[25–27].

PglC, the first glycosyltransferase (GTase), adds UDP-diNAcBac to Und-P to form diNAcBac-α1-PP-Und[15,28]. *In vitro* biochemical analyses showed that purified PglC accepts synthesized UDP-diNAcBac and UDP-6-hydroxybacillosamine but not UDP-*N*-acetylgalactosamine (UDP-GalNAc) or UDP-GlcNAc[28]. PglA transfers UDP-GalNAc to form a lipid-linked disaccharide, GalNAc-α1,3-diNAcBac-α1-PP-Und. Interestingly, PglA has a relaxed specificity and accepts diNAcBac-PP-Und, GlcNAc-PP-Und and 6-hydroxybacillosamine-PP-Und *in vitro*[29], and diNAcBac-PP-Und and GlcNAc-PP-Und *in vivo* when transferred into the heterologous *E. coli* system[15]. PglJ adds a single α1,4-GalNAc residue to GalNAc-α1,3-diNAcBac-PP-Und to create a trisaccharide. PglH then acts as a polymerase, adding three α1,4-linked GalNAc residues to extend the glycan chain[30]. The transfer of UDP-GalNAc by PglJ in the previous step is required for this PglH-mediated polymerization step, as the combination of PglA and PglH in *in vitro*

Box 1 | **Analysis and identification of *N*-linked glycoproteins**

Advances in bacterial glycoprotein detection and identification and in glycan structure determination have been made recently owing to new developments and increased sensitivity in complementary mass spectrometry (MS) and NMR techniques. For example, high-resolution magic-angle spinning NMR, which allows complex glycan analysis using as little as 40 µl of intact bacterial cells, can detect the *Campylobacter jejuni* heptasaccharide *in vivo*[137] and helped to demonstrate that the *N*-glycan is assembled intracellularly as a block[19]. Another new analytical method uses ion-pairing reagents to separate *N*-glycosylated peptides from non-glycosylated peptides[49] and allowed *N*-linked sugars from a *C. jejuni* protein glycosylation D gene (*pglD*) mutant to be detected, whereas they could not be detected previously[19]. Surprisingly, the technique also showed that wild-type glycoproteins expressed minor amounts of monoacetylated 2,4-diacetamido-2,4,6-trideoxyglucose (NAcBac), leading to the question of whether differential acetylation of *C. jejuni* glycoproteins has any biological consequences.

Furthermore, an MS-based glycomics strategy combining nonspecific proteolytic digestion and permethylation showed for the first time that *C. jejuni* produces the heptasaccharides as free oligosaccharides (fOS) that are present in tenfold excess when compared with the *N*-linked form[138] (see main text, FIG. 1a). Subsequent studies found that fOS release is dependent on both the bacterial growth phase and the osmotic environment, which are conditions that have minimal effects on protein glycosylation[139]. It has therefore been suggested that periplasmic fOS have a similar role to the periplasmic glucans that are common to many proteobacteria (see REF. 140 for a review of periplasmic glucans). It was also shown that *C. jejuni* PglB is responsible for protein *N*-glycosylation and fOS release, and that its conserved Trp-Trp-Asp-Tyr-Gly motif is indispensable for both the *N*-glycosylation oligosaccharyltransferase (*N*-OTase) and hydrolase functions[139]. Eukaryotes also synthesize fOS[141,142], but studies in yeast demonstrated that these fOS are derived solely from misfolded glycoproteins during the degradation pathway[143], and so it is unclear whether fOS play a part in cellular processes in higher eukaryotes. By contrast, the regulated production of fOS in *C. jejuni* may provide another reason why bacteria maintain a general *N*-glycosylation pathway and raises the question of whether the primary purpose of this pathway was to *N*-glycosylate proteins or to produce fOS.

assays failed to yield anything larger than a disaccharide[31]. Interestingly, PglH uses a single active site for multiple GalNAc transfers to the Und-PP-oligosaccharide and does not seem to use a block transfer mechanism, as intermediates with one, two or three GalNAc residues added to the trisaccharide can be detected. It has been suggested that the binding affinity of PglH for the product increases with glycan size and thus serves as a molecular ruler to stop catalysis after the formation of a hexasaccharide[30]. PglI is the GTase that adds the β1,3-linked glucose branch. It has been suggested that PglI releases PglH from the hexasaccharide chain by competing with the substrate[30] to complete the LLO structure[31].

Remarkably, reconstitution of these sequential steps in a single reaction containing all five GTases (PglC, PglA, PglJ, PglH and PglI) resulted in the efficient formation of $GalNAc_2$[Glc]$GalNAc_3$-diNAcBac-PP-Und from Und-P and UDP-sugar donors, demonstrating the complete biosynthesis of an LLO *in vitro* for the first time[31]. As no intermediates were observed, it was concluded that these enzymes must interact in a highly coordinated manner when they are combined.

The lipid carrier. In Eukarya, Archaea and Bacteria, oligosaccharides typically destined for the *N*-glycosylation pathways are assembled onto a polyisoprenoid lipid carrier (for a review, see REF. 32). Although dolichyl pyrophosphate (Dol-PP) is the carrier in the assembly of yeast LLOs[33], both dolichyl phosphate (Dol-P) and Dol-PP have been found as carriers in archaea[34,35]. In bacteria, the LLO is assembled on one polyprenyl pyrophosphate, Und-PP, that serves as the native substrate for PglB[16,17]. It has been shown that the activities of both the GTase PglJ and the OTase PglB were not affected by the presence of a slightly truncated form of undecaprenol, whereas dihydroprenol-11-linked, dolichol-linked, solanesol-linked or geranyl geraniol-linked substrates resulted in a notable decrease in or even a total loss of *in vitro* activities[36,37].

Analyses of LLO formation in *C. jejuni* and in *E. coli* expressing the *C. jejuni N*-glycosylation pathway using lectin-based affinity capture followed by tandem mass spectrometry confirmed heptasaccharide assembly on Und-PP *in vivo*[16]. A more general strategy using porous graphite carbon liquid chromatography in combination with mass spectrometry[17] eliminated the use of lectins and made it possible to analyse low levels of LLOs (from approximately 10^6 bacterial cells), distinguishing LLO species that differ by only one monosaccharide or polyisoprene unit and demonstrating that *C. jejuni* heptasaccharides are assembled on polyisoprenes consisting of 9–12 isoprene units[17].

The acceptor peptide. Using synthetic glycan donors and a short acceptor peptide (KDFNVSKA; the *N*-glycosylation sequon is DFNVS), it was shown that PglB not only accepts saccharides of various lengths (2–7 sugars) *in vitro* but also accepts different sugars at the reducing end (including unnatural 6-hydroxybacillosamine and GlcNAc)[38]. This feature of PglB was also shown in the heterologous *E. coli* system *in vivo*[15,39,40] and, to a lesser extent, when examining *pgl* mutants in *C. jejuni*[41]. The possibility of PglB-dependent *in vitro* glycosylation of a short peptide would argue against the necessity of a particular tertiary structure for protein *N*-glycosylation *in vivo*. As glycosylation in the bacterial system occurs on folded proteins, it was suggested that a functional glycosylation acceptor site must be located in a flexible and surface-exposed region of a folded protein[21].

The X-ray crystal structure of the *C. jejuni N*-glycoprotein Peb3 (encoded by *peb3* at the locus *cj0289c*) showed that the glycosylation sequon at Asn90 is indeed located in a surface-exposed flexible loop that would be accessible to PglB[42]. Examination of the first two visible residues of the *N*-glycan showed that the sugars did not affect the Peb3 structure around the glycosylation sequon[43]. However, thermal denaturation experiments found that the glycoprotein was more stable than the non-glycosylated protein. Although the glycan had no hydrogen bonds with the peptide backbone, it was suggested that it might be involved in stabilizing the loop around Asn90 (REF. 43).

The NMR structures of glycosylated and non-glycosylated versions of truncated AcrA, a *C. jejuni* glycoprotein, confirmed that glycosylation occurs within flexible loops[44]. Only slight changes were observed in the peptide structures of the glycosylated and non-glycosylated proteins, but the attached glycan reduced the side chain flexibility of the required Asn in comparison with the

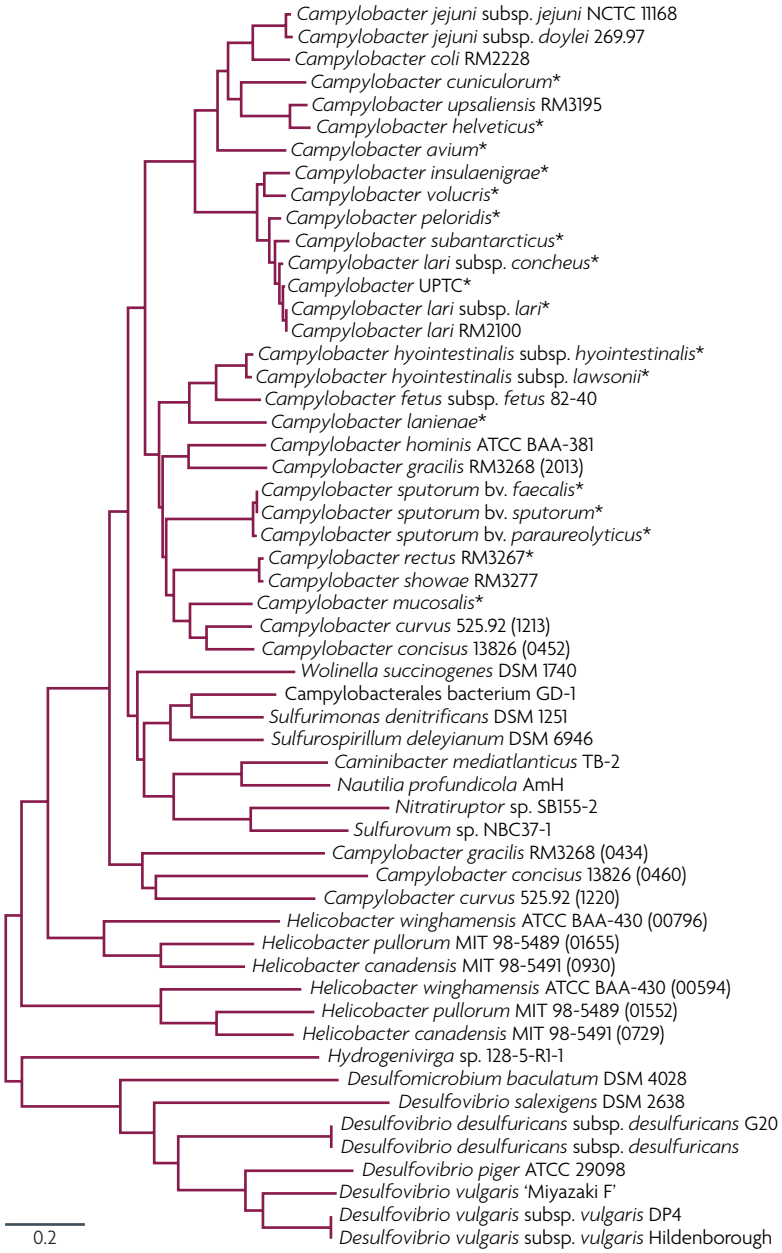


Figure 2 | **Phylogenetic tree comparing protein glycosylation B (PglB) sequences of all sequenced deltaproteobacteria and epsilonproteobacteria.** A phylogenetic tree containing all available full-length bacterial PglB protein sequences was constructed from a multiple sequence alignment with the neighbour-joining method using MEGA 4 software[150]. For PglB protein sequences from genomes that encode two potential PglB orthologues, the number of the corresponding ORF in the genome is indicated in brackets following the strain name. PglB sequences provided by W. G. Miller are indicated with an asterisk. In contrast to the data from the available shotgun genome sequence (Entrez Genome Project 31017), *Campylobacter rectus* RM3267 encodes a full-length PglB orthologue (W. G. Miller, personal communication).

flexibility of the surrounding amino acid residues. This study also demonstrated, for the first time, that the *C. jejuni N*-glycan forms a rod-like structure that is kinked owing to the β1,3-linked diNAcBac and tends to fold back over the protein. As was seen for glycosylated Peb3, minimal carbohydrate–peptide interactions were observed for glycosylated AcrA[44].

Competence
The ability of a bacterium to take up extracellular DNA.

A general function for* C. jejuni N-*glycans? Many diverse proteins are modified by the *N*-glycosylation pathway in *C. jejuni*, so it is not surprising that disruption of this pathway has pleiotropic effects for the bacterium, including reduced protein immunoreactivity with both human and animal sera[10,14], a reduced ability to adhere to and invade human epithelial cells *in vitro* and a decrease in mouse and chicken colonization *in vivo*[45–48]. However, inactivation of *pglG* (a gene that is part of the *C. jejuni pgl* operon and is conserved in certain species (FIG. 3) but currently has an undetermined function) and *pglI* had no effect on chicken colonization, which was corroborated by the observation that mutation of these genes did not affect *N*-glycosylation levels[19]. Interestingly, although disruption of *pglD* results in the loss of chicken colonization, low levels of *N*-glycosylation could still be detected[19,49].

Despite the variety of known and predicted *N*-glycoproteins[22,23], only a few studies have investigated the potential importance of this protein modification. Loss of *N*-glycosylation of the VirB10 component of the type IV secretion system in *C. jejuni* subsp. *jejuni* 81-176 resulted in reduced competence levels and potential changes in the cellular location of VirB10 (REF. 50). By contrast, introducing point mutations at the *N*-glycosylation sites of the adhesin Cj1496c had no effect on chicken colonization when compared with wild-type bacteria, whereas the *cj1496c* knockout mutant is strongly impaired in colonization[51]. More recent studies investigated the function of the glycan on the *C. jejuni* periplasmic zinc uptake A (ZnuA) homologue, Cj0143c[52]. Although the authors demonstrated a role for the *cj0143c* gene product in chicken colonization and zinc acquisition, complementation of a *cj0143c* deletion mutation with a glycosylation-deficient *cj0143c* allele resulted in a wild-type phenotype. Similarly, the *N*-glycan on the *C. jejuni* surface-associated lipoprotein JlpA, an adhesin that is required for chicken colonization[53], was not required for antigenicity[54]. Thus, it seems that *C. jejuni* protein *N*-glycosylation might be important for the function of specific proteins in this organism, maintaining a selective pressure for the pathway, but that the *N*-glycan does not serve the same function for all proteins, at least under the examined conditions.

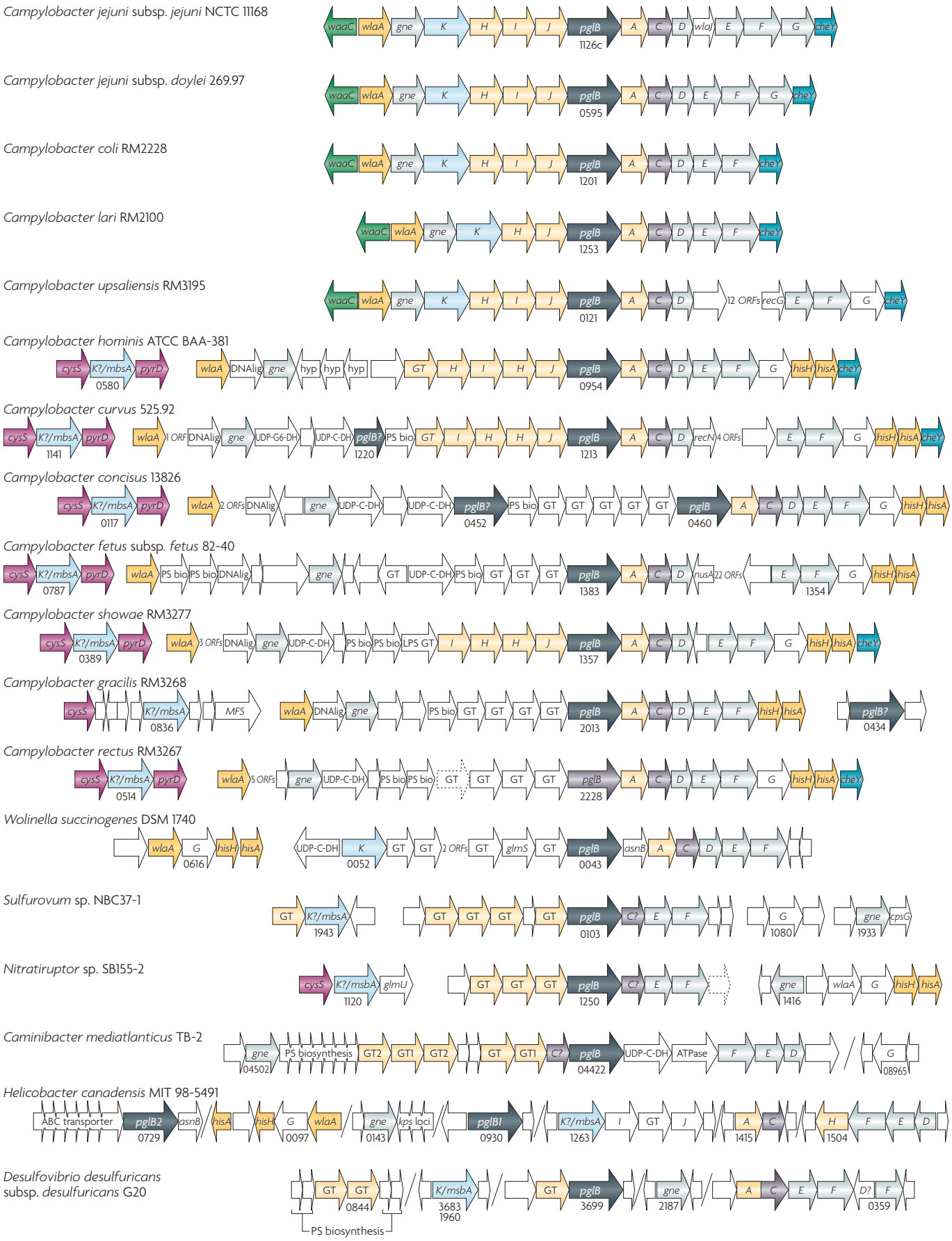
Another role for the *C. jejuni N*-glycan was provided by the observation that the heptasaccharide is recognized by human macrophage galactose-type lectin (MGL)[55]. This specific dendritic cell receptor functions in cell–cell communication, capturing glycosylated antigens for processing and presentation and providing signals that modulate immune cell function. MGL recognizes *C. jejuni* glycoproteins (both derived from *C. jejuni* and expressed by *E. coli*) as well as *C. jejuni* lipooligosaccharides with terminal GalNAc residues. Interestingly, a *C. jejuni pglA* mutant lacking the MGL ligand (that is, devoid of GalNAc residues) resulted in increased production of the cytokine interleukin-6 (IL-6) from human dendritic cells. It was therefore suggested that *C. jejuni* might *N*-glycosylate its proteins to modulate the host immune response in order to limit cellular cytokine production.

Protein* N*-glycosylation resembling the* C. jejuni *system in other bacteria. The increasing number of complete bacterial genome sequences has shown that the organisms encoding orthologues of the *N*-glycan pathway constituents are much more widespread than originally predicted (FIG. 2). Although the GTase content and *pgl* gene distributions vary (FIG. 3), the characteristic enzyme of the pathway, PglB, is present in at least 49 species, with five species possessing two copies of *pglB* and all homologues containing the conserved Trp-Trp-Asp-Tyr/Trp-Gly motif that has been shown to be indispensable for PglB function in *C. jejuni*[13]. Examination of the *pgl* gene clusters and their flanking regions identified putative insertional sites in the respective chromosomes (FIG. 3). The apparent diversity in *pgl* gene distribution contrasts with the partial conservation of the neighbouring chromosomal DNA sequences and suggests that the *pgl* cluster has evolved under selective pressure in its host environment, a typical feature of pathogenicity-related gene clusters[56].

The crystal structures have been determined for the carboxy-terminal, globular domains of *C. jejuni* PglB[57] and the orthologous STT3-like protein archaeal glycosylation B (AglB) from the archeaon *Pyrococcus furiosus*[58]. The structure of *P. furiosus* AglB revealed another motif, Asp-X-X-Lys, that is necessary for protein function and corresponds to the Met-X-X-Ile sequence (residues 568–571) in *C. jejuni* PglB. However, point mutations in the respective amino acids showed that only the Ile571Ala substitution resulted in reduced PglB activity *in vitro*[57]. In addition, all OTases contain a conserved X-X-Asp motif (Ser-Asn-Asp in *C. jejuni* PglB; residues 52–54) that resides in the first lumenal loop of the transmembrane domain and has been implicated as a third member of the catalytic triad. Indeed, substitution of Ser52 and Asp54 in *C. jejuni* PglB led to a loss of PglB activity *in vitro*[57]. Examination of the PglB sequences in FIG. 2 supports the observation that the bacterial *N*-OTases all possess the Trp-Trp-Asp-Tyr/Trp-Gly, Met-X-X-Ile and X-X-Asp motifs.

The genomes of two epsilonproteobacteria that reside in deep-sea vents, *Sulfurovum* sp. NBC37-1 and *Nitratiruptor* sp. SB155-2, possess *pgl* gene clusters[59]. Despite the presence of a well-conserved PglB, both species lack obvious PglA, PglD and PglK orthologues (FIG. 3). The authors speculate that *N*-glycosylation (although not proved to occur in these organisms) might have arisen to establish a relationship between these deep-sea bacteria and hydrothermal-vent invertebrates. *Wolinella succinogenes* DSM1740, a non-pathogenic, bovine-host-adapted epsilonproteobacterium[60], also possesses an *N*-glycosylation gene cluster (FIG. 3). Although the functionality of the pathway in this organism is not yet proved, many protein targets for *N*-glycosylation in *C. jejuni* have orthologues in *W. succinogenes*[60] that contain the extended glycosylation sequon. In the genus *Helicobacter*, the genome of *Helicobacter canadensis* MIT 98-5491, an emerging human pathogen, contains distinct but separated gene loci for *pglDEFH*, *pglAC*, and *pglI* and *pglJ*, as well as two conserved *pglB* loci that encode full-length proteins and are not flanked by other *pgl* orthologues[61] (FIG. 3). The genomes of *Helicobacter winghamensis* ATCC BAA-430 (project ID 32491 in Entrez Genome Project), *Helicobacter pullorum* MIT 98-5489 (project ID 30075) and *H. pullorum* NCTC 12824 (REF. 62) each contain two *pglB* orthologues that can also be found in *Campylobacter curvus* 525.92 (project ID 17161), *Campylobacter concisus* 13826 (RefSeq accession NZ_AAQZ00000000) and *Campylobacter gracilis* RM3268 (project ID 31021). Recently it has been shown that one of the *H. pullorum* NCTC 12842 PglB orthologues (PglB1) indeed possesses OTase activity[62], but the function of the second PglB orthologue (PglB2) remains to be elucidated, as attempts to knock out the corresponding gene were not successful[62]. Interestingly, *Helicobacter pylori* is devoid of any *pgl* family orthologues[14] except for a flippase that is actually used for LPS biosynthesis[63]. This raises the question of why some *Helicobacter* spp. kept (or acquired) these genes, whereas others did not. Perhaps this is a consequence of the niche that each organism inhabits, as *pgl*-negative species belong to the gastric group of *Helicobacter* spp., whereas *pgl*-positive species belong to the enteropathogenic group[64,65]. In the deltaproteobacterium *Desulfovibrio desulfuricans* subsp. *desulfuricans* G20, *pgl* genes are also scattered across the genome (FIG. 3), although other genes clustered around the PglB orthologue may be involved in the synthesis of the *N*-glycan, as the structure differs from the *C. jejuni* heptasaccharide. Direct evidence that *N*-glycosylation occurs in this genus was provided by the crystal structure and mass spectrometry analyses of high-molecular-weight cytochrome *c* (HmcA) isolated from the periplasm of *Desulfovibrio gigas*[66]. Interestingly, the glycan is composed of three *N*-acetylhexosamine (HexNAc) residues that are attached at Asn261 in the eukaryotic-like Asn-X-Ser/Thr consensus sequence[66] (TABLE 1).

Atypical protein* N*-glycosylation in other bacteria. Another *N*-glycosylation system was recently described in the Gram-negative gammaproteobacterium *Haemophilus influenzae*; this system is distinct from the *C. jejuni* pathway and involves sequential transfer of sugars rather than block transfer[67] (FIG. 1b). One of the two *H. influenzae* high-molecular-weight adhesins (HMW1), which interacts with sialylated *N*-glycoproteins on the host cell surface[68], undergoes glycosylation at 31 sites. This glycosylation is necessary for the protein to resist degradation and maintain bacterial cell surface tethering with its cognate outer-membrane channel-forming β-barrel translocator protein, HMW1B[69]. HMW1 and its GTase, HMW1C, seem to associate in the cytoplasm[69], where monohexoses and dihexoses are added to the conventional eukaryotic consensus sequence for *N*-linked glycosylation, Asn-X-Ser/Thr, in all but one case[67]. The suggestion that *H. influenzae* possesses a new type of bacterial GTase was confirmed with purified components *in vitro*, demonstrating that HMW1C can indeed transfer glucose and galactose residues derived from common sugar precursors that are also used for the lipooligosacccharide (LOS) pathway. Another surprising finding is that HMW1C exhibits both OTase activity, adding *N*-linked sugars to HMW1, and GTase



◀ Figure 3 | **Organization of selected *N*-linked protein glycosylation gene clusters.** *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 *N*-linked protein glycosylation (*pgl*) gene orthologues, upstream and downstream ORFs and genes inserted between *pgl* genes (if present) are indicated by arrows. The arrow sizes are not to scale. For other species, genes encoding the essential oligosaccharyltransferase PglB and genes encoding the biosynthetic Pgl enzymes (uridine diphosphate (UDP)-*N*-acetylglucosamine:glucosamine 4-epimerase (*gne*), PglE (*E*), PglF (*F*), PglD (*D*)), glycosyltransferases (PglA (*A*), PglJ (*J*), PglH (*H*), PglI (*I*), PglC (*C*)) and the flanking gene products PglG (*G*) and WlaA are designated according to their orthologues in *C. jejuni* subsp. *jejuni* NCTC 11168 or as 'GT' if no definite homology to any *C. jejuni* subsp. *jejuni* NCTC 11168 Pgl glycosyltransferase was found. Other glycosyltransferases at different locations on the respective genomes are not included in the figure. Orthologues with low similarities but with an identical location in the *pgl* locus are indicated by a question mark. Orthologues to the putative ATP-binding cassette (ABC)-type transporter PglK (*K*) that show higher homology to MsbA than to PglK are labelled '*K*?/*msbA*'. Gaps between *pgl* genes are marked by either the number of ORFs or by a solidus, indicating that these orthologues were found elsewhere in the chromosome. Unlabelled genes do not show particular homology to other genes in that *pgl* locus or to the genes upstream or downstream of the *pgl* loci. Dotted outlines indicate pseudogenes. Locus numbers are given beneath at least one locus per cluster, for identification purposes. The Entrez Genome Project numbers (or RefSeq accession numbers where there is no Genome Project entry) for the species shown are given in brackets: *C. jejuni subsp. doylei* 269.97 (17163), *Campylobacter coli* RM2228 (12516), *Campylobacter lari* RM2100 (RefSeq accession NZ_AAFK00000000), *Campylobacter upsaliensis* RM3195 (12518), *Campylobacter hominis* ATCC BAA-381 (20083), *Campylobacter curvus* 525.92 (17161), *Campylobacter concisus* 13826 (RefSeq accession NZ_AAQZ00000000), *Campylobacter fetus* subsp. *fetus* 82-40 (16293), *Campylobacter showae* RM3277 (31019), *Campylobacter gracilis* RM3268 (31021), *Campylobacter rectus* RM3267 (31017; re-sequencing of two adjacent ORFs, annotated as pseudogenes, revealed that *C. rectus* RM3268 encodes a full-length PglB[24]), *Wolinella succinogenes* DSM 1740 (445), *Sulfurovum* sp. NBC37-1 (18965), *Nitratiruptor* sp. SB155-2 (18963), *Caminibacter mediatlanticus* TB-2 (19293), *Helicobacter canadensis* MIT 98-5491 (30071) and *Desulfovibrio desulfuricans* subsp. *desulfuricans* G20 (329). DNA lig, DNA ligase gene; hyp, hypothetical ORF; LPS, lipopolysaccharide biosynthesis; MFS, major facilitator superfamily; PS bio, polysaccharide biosynthesis; UDP-C-DH, UDP-carbohydrate dehydrogenase; UDP-G6-DH, UDP-glucose-6-dehydrogenase.

activity, capable of generating hexose–hexose bonds[70] (FIG. 1b). It is noteworthy that other two-partner secretion systems (with up to 97% identity to the HMW1–HMW1C pair) exist in *H. influenzae*, *Haemophilus ducreyi*, *Burkholderia xenovorans*, *E. coli*, *Mannheimia succiniciproducens*, *Limnobacter* sp. MED105, *Xanthomonas campestris* and *Yersinia* spp. (such as *Yersinia enterocolitica*, *Yersinia pestis* KIM10 and *Yersinia pseudotuberculosis*)[70]. Similarly, autotransporter proteins (in which the exoprotein and the channel-forming protein are fused into a single multidomain protein with predicted autoproteolytic activity) have been shown to be glycosylated at multiple sites with sugars that are also synthesized for the LPS pathway. However, the *E. coli* autotransporters identified to date — AIDA-I, TibA and antigen 43 — are modified with *O*-linked heptose residues[71–73], and this occurs in the cytoplasm independently of export[74].

Two-partner secretion system
Form of type V secretion system comprising two distinct proteins: a transporter and a secreted effector.

O-glycosylation in bacteria

O-linked protein glycosylation occurs in all three domains of life, and the eukaryotic and bacterial pathways are well characterized. In this section, we focus on the *O*-glycan pathways that modify bacterial flagella and pili. Excellent reviews on this topic were published in 2006 (REFS 75,76), and thus only recent advances in this field are summarized. The exciting new discovery that some bacteria possess general *O*-linked glycosylation systems that involve glycoprotein synthesis through an amalgamation of conventional *N*-linked and *O*-linked glycosylation mechanisms is also discussed.

Campylobacter *spp. flagella glycosylation*. In addition to the general *N*-linked protein glycosylation pathway discussed above, campylobacters specifically modify their flagellar proteins with *O*-linked glycans (FIG. 4a) that can constitute up to 10% of the protein mass[77]. These modifications are necessary for flagellum assembly[78] and thus affect secretion of virulence-modulating proteins, bacterial colonization of the gastrointestinal tract, auto-agglutination and biofilm formation (for reviews, see REFS 79,80). The predominant *O*-glycans attached to the campylobacter flagellum are derivatives of pseudaminic acid (Pse) or legionaminic acid (Leg), which are C_9 sugars that are related to sialic acids (TABLE 1).

C. jejuni isolates fall into two distinct groups, a livestock-associated clade and a non-livestock-associated clade, the genomic contents of which differ by the presence or absence of the *cj1321–cj1325* or *cj1321–cj1326* cluster, among others[81]. In *C. jejuni* NTCT 11168 subsp. *jejuni* 11168H (from the livestock clade), the protein encoded by *cj1324* is involved in the biosynthesis of the acetamidino (Am) and *N*-methylacetimidoyl (AmNMe) Leg derivatives Leg5Am7Ac and Leg5AmNMe7Ac (where Ac represents an acetamido group), although other Pse and Leg derivatives are also produced[82]. Mutation of the *cj1324* locus resulted in reductions in auto-agglutination and chicken colonization but did not affect cell motility[82]. By contrast, flagella from *C. jejuni* subsp. *jejuni* 81–176 (from the non-livestock clade) are glycosylated with only Pse derivatives and the Pse acetamidino variant substituted with acetyl, *N*-acetylglutamine (GlnAc) or propionyl (Pr) structures (Pse5Ac7Ac, Pse5Am7Ac, Pse5Ac7Ac8OAc, Pse5Am7Ac8GlnAc and Pse5Pr7Pr; where OAc represents an acetoxy (or *O*-acetyl) group); these glycans are added to up to 19 Ser or Thr residues per flagellin A (FlaA) subunit[77,83,84]. Interestingly, specific loss of Pse5Am due to mutation of the Pse biosynthesis A gene (*pseA*; which is at the *cj1316c* locus) resulted in loss of auto-agglutination and reduced adherence to and invasion of intestinal epithelial cells *in vitro*, and reduced virulence in the ferret model[85].

Functional characterization of all 19 *O*-glycosylation sites in the FlaA subunit of *C. jejuni* subsp. *jejuni* 81-176 identified five sites that are necessary for auto-agglutination[86], suggesting that these residues are surface exposed and interact with filaments from other *C. jejuni* cells or with specific ligands on eukaryotic cells. Although no changes in motility were observed, the filaments seemed to be more fragile than those composed of fully glycosylated flagellins[86]. Three other mutants displayed reductions in motility and produced truncated filaments, suggesting that glycans at these positions are important for subunit interactions[86]. Further analyses indicated that glycosylation occurs before flagellin export and independently of the flagellar regulon[86].

Recently, the *C. jejuni* *O*-linked-glycosylation pathway was reconstituted *in vitro*[87]. They purified and biochemically characterized 11 candidate enzymes from *C. jejuni*,

Table 1 | ***N*-linked and *O*-linked bacterial protein glycosylation pathways**

Organism	OTase/ GTase	Linkage	Sequon	Lipid carrier	OTase transfer abilities	Transfer type	Protein targets	Amino acids modified	Sugars transferred
Campylobacter jejuni	PglB*‡	N	(D/E)X_1NX$_2$ (S/T)§	Yes	C2-NAc‖; no β1,4¶; multiple *O*-repeats	Block	>65	1–6	$GalNAc_3$(Glc) $GalNAc_2$diNAcBac
C. jejuni and *Campylobacter coli*	ND	O	S/T	No	ND	Sequential	FlaA	Up to 19	Pse, Leg and derivatives
Clostridium botulinum and *Clostridium difficile*	ND	O	S/T	No	ND	Sequential	Flagellin	7	αLeg5GluNMe7Ac and di-*N*-acetylhexuronic acid (*C. botulinum*); HexNAc-P-Asp-Me, HexNAc, dHexMe, dHex and Hep (*C. difficile*)
Pseudomonas aeruginosa and *Pseudomonas syringae*	ND	O	S/T	No	ND	Sequential	Flagellin	Up to 6	Rha or Rha-linked oligosaccharides (*P. aeruginosa* a-type strains); dHex with an additional mass of 209 Da (*P. aeruginosa* b-type strains); 2Rha with modified 4-amino-4,6-dideoxyGlc attached (plant-pathogenic *P. syringae* strains)
	PilO‡# and TfpW	O	S/T	Yes	1–2 *O*-linked repeats	Block	PilA	1 (*P. aeruginosa* 1244); up to 5 (*P. aeruginosa* 5196)	*N*-hydroxylbutyryl-*N*-formyl-Pse-Xyl-FucNAc (*P. aeruginosa* 1244); mono-Araf, di-Araf and oligo-Araf (*P. aeruginosa* 5196)
Neisseria spp.	PglL*‡	O	S/T plus a region with low complexity	Yes	Promiscuous; multiple *O*-linked repeats	Block	>12, including pilin	1–2	(OAc)Gal-Gal-DATDH and (OAc)Gal-Gal-GATDH
Bacteroides spp.	ND	O	D(S/T)(A/I/L/V/M/T)(S/T)	ND	ND	ND	>8	1–3	Contains fucose
Flavobacterium columnare	ND	O	D(S/T)	ND	ND	ND	Several, including EndoH	1–3	(2-OMe)Man-GlcNAcU-GlcU-Glc(2-OMe)-GlcU-[(2-OMe) Rha]Man
Haemophilus influenzae	HMW1C*‡ and HMW2C	N	NX(S/T)	No	ND	Sequential	HMW1 and HMW2	31	Hex and diHex (where Hex can be Glc or Gal)
Streptomyces coelicolor	Pmt	O	S/T	Yes	ND	Block	At least 1 (PstS)	2–3	TriHex
Desulfovibrio gigas	PglB**	N	NX(S/T)	ND	ND	ND	At least 1 (HmcA)	1	TriHexNAc(where Hex can be GlcNAc or GlcNAc epimers)

αLeg5GluNMe7Ac, 7-acetamido-5-(*N*-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-D-glycero-α-D-galacto-nonulosonic acid; Araf, arabinose (furanose form); DATDH, 2,4-diacetamido-2,4,6-trideoxyhexose; dHex, deoxyhexose; diNAcBac, 2,4-diacetamido bacillosamine (2,4-diacetamido-2,4,6-trideoxy-α-D-glucose); EndoH, endo-β-*N*-acetylglucosaminidase; FlaA, flagellin A; FucNAc, *N*-acetylfucosamine; Gal, galactose; GalNAc, *N*-acetylgalactosamine; GATDH, 2-glyceramido 4-acetamido 2,4,6-trideoxyhexose; Glc, glucose; GlcNAc, *N*-acetylglucosamine; GlcNAcU, 2-acetamido-2-deoxyglucuronic acid; GlcU, glucuronic acid; GTase, glycosyltransferase; Hep, heptose; Hex, hexose; HexNAc, *N*-acetylhexosamine; HmcA, high-molecular-weight cytochrome c; HMW, high-molecular-weight adhesin; Leg, legionaminic acid; Man, mannose; Me, methyl group; ND, not determined; OAc, acetoxy (*O*-acetyl) group; OMe, methoxy (*O*-methyl) group; OTase, oligosaccharyltransferase; P, phosphate; PglB, protein glycosylation B; PglL, pilin glycolsylation L; PilA, pilin;Pse, pseudaminic acid; Rha, rhamnose; Xyl, xylose. *Characterized *in vitro*. ‡Also characterized *in vivo* in the heterologous *E. coli* system. §X_1 and X_2 are any amino acid except Pro. ‖Acetylation at C2 of the reducing end sugar is required. ¶A β1,4 linkage between the reducing end and second sugar residue is not recognized. #Note that only PilO from *P. aeruginosa* 1244 has been characterized. **Based on homology to PglB of *C. jejuni* subsp. *jejuni* NCTC 11168.

fully synthesizing Leg and its CMP-activated form starting from fructose-6-phosphate. Analysis of the pathway identified unique GDP-linked intermediates, which the authors proposed could provide a cellular mechanism for differentiating between similar UDP-linked pathways such as the diNAcBac biosynthesis involved in *N*-linked protein glycosylation (described above).

Interestingly, the related human gastric pathogen *H. pylori* also *O*-glycosylates its flagella with Pse, similarly to *C. jejuni*, and modification is required for bacterial motility and flagellar assembly[88,89]. The *O*-glycosylation gene content of *H. pylori* is far simpler than that of the campylobacters (which display substantial genetic diversity among different species, subspecies

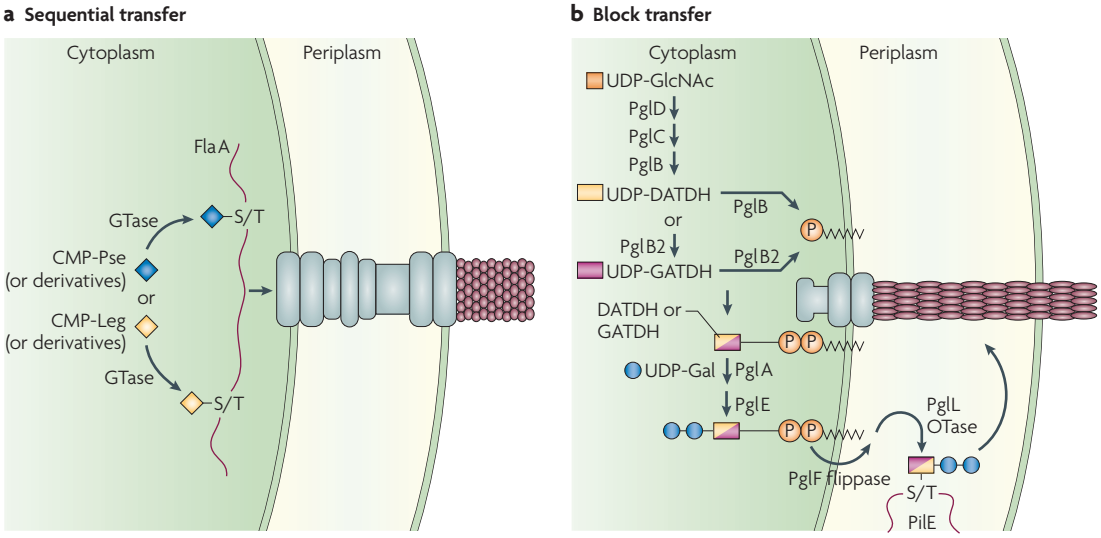


Figure 4 | **Overview of bacterial *O*-linked pathways for protein glycosylation. a** | The sequential transfer of nucleotide-activated sugars to proteins (as occurs in *Campylobacter jejuni*, for example). As the flagellar apparatus of *C. jejuni* spans both the inner and outer membranes, *O*-linked glycosylation of flagellin monomers is thought to occur at the cytoplasm–inner membrane interface. CMP-activated sugars (CMP-pseudaminic acid (CMP-Pse), CMP-legionaminic acid (CMP-Leg) and their derivatives) are individually added to surface-exposed Ser or Thr residues in the flagellin subunit, FlaA, by a glycosyltransferase (GTase). **b** | Block transfer of oligosaccharides from the lipid anchor to the protein target, as occurs in *Neisseria meningitidis*, for example, follows a mechanism that is similar to protein *N*-glycosylation in *C. jejuni*. The saccharide is assembled on the cytoplasmic side of the inner membrane on a lipid anchor. The first sugar, uridine diphospahte (UDP)-2,4-diacetamido-2,4,6-trideoxyhexose (UDP-DATDH) or UDP-2-glyceramido 4-acetamido 2,4,6-trideoxyhexose (UDP-GATDH), is synthesized starting from UDP-*N*-acetylglucosamine (UDP-GlcNAc) by the sequential actions of pilin glycosylation D (PglD), PglC, and PglB (for DATDH) or PglB2 (for GATDH). PglB and PglB2 are bifunctional enzymes involved in the biosynthesis of DATDH and GATDH, respectively, and in their transfer onto the lipid carrier. Galactose (Gal) residues are added to the sugar chain from UDP-Gal donors by the phase-variable GTases PglA and PglE. PglF then 'flips' the lipid-linked saccharide to the periplasm, where the oligosaccharyltransferase (OTase) PglL transfers the glycan to Ser63 of the pilin subunit PilE or to Ser/Thr residues in other periplasmic and membrane proteins. P, phosphate.

and individual isolates). Reconstitution and biochemical characterization of the Pse biosynthesis pathway showed that *H. pylori* uses a similar set of enzymes: the proteins encoded by the loci HP_0840, HP_0366, HP_0327, HP_0326B, HP_0178 and HP_0326A correspond to PseB (encoded by the locus *cj1293*), PseC (*cj1294*), PseH (*cj1313*), PseG (*cj1312*), PseI (*cj1317*) and PseF (*cj1311*) in *C. jejuni*[90]. However, the extensive diversity that is seen in the glycan structures of *C. jejuni* is not observed in *H. pylori*, presumably because the sheath surrounding the flagellar filament negates this need[88].

Clostridium *spp. flagella glycosylation.* An *O*-linked glycosylation system was also identified in *Clostridium botulinum*, in which the flagella are modified with the Leg derivative 7-acetamido-5-(*N*-methyl-glutam-4-yl)-amino-3,5,7,9-tetradeoxy-D-glycero-α-D-galacto-nonulosonic acid (αLeg5GluNMe7Ac) and with di-*N*-acetylhexuronic acid, which are attached to seven Ser or Thr sites in each protein subunit[91]. By contrast, the flagellin of *Clostridium difficile* 630 is *O*-glycosylated at up to seven sites, with HexNAc residues modified with methylated Asp through a phosphate linkage. A mutant lacking the saccharide showed a defect in flagellum assembly, leading to a loss of motility[92]. Interestingly, a highly conserved flagellar-glycan biosynthesis locus in *C. difficile* isolates from more recent outbreaks showed considerable genetic diversity and increased locus size compared with the *C. difficile* 630 locus[92]. In these isolates, the flagella were found to be modified with a heterogeneous glycan containing up to five monosaccharide residues consisting of HexNAc, deoxyhexose, methylated deoxyhexose and heptose[92] (TABLE 1).

Pseudomonas *spp. flagella glycosylation.* In *Pseudomonas aeruginosa*, type-specific antisera and molecular mass classifies flagellin into two major types: a-type and b-type[93,94]. Flagellins of both types are known to be glycosylated, and the genes involved in carbohydrate synthesis cluster together in genomic islands[95,96]. Two a-type strains (*P. aeruginosa* PAK and *P. aeruginosa* JJ692) were shown to be *O*-glycosylated at two sites on their flagellin subunits. Although the glycosyl moiety is linked through a rhamnose residue in both strains, the *P. aeruginosa* PAK glycan contained up to 11 additional monosaccharides, whereas the *P. aeruginosa* JJ692 glycan is glycosylated with only a single rhamnose[97] (TABLE 1). This heterogeneity among a-type strains is the result of variations in gene content in the glycosylation islands[98]. By contrast, the genes encoding enzymes for *O*-glycan biosynthesis in the b-type strains are much more conserved[98]. In the b-type strain

P. aeruginosa PAO1, flagellin contains a single deoxyhexose residue attached to two nearby Ser residues, and each glycan can be linked to a unique modification of mass 209 Da containing a phosphate moiety[99]. In the case of the plant pathogens *Pseudomonas syringae* pv. *tabaci* and *P. syringae* pv. *glycinea*, the flagella were found to be glycosylated at six Ser residues. Here, the *O*-linked glycan was shown to be composed of a unique trisaccharide containing two rhamnosyl residues and one modified 4-amino-4,6-dideoxyglucosyl residue[100].

Flagellins from most bacteria, including *P. aeruginosa* PAK flagellin, are ligands for Toll-like receptor 5 (TLR5), which results in the stimulation of IL-8 production[101,102]. Interestingly, flagellin isolated from *P. aeruginosa* PAK glycosylation mutants stimulated 50% less IL-8 production from A549 human alveolar epithelial cells than wild-type flagellin[102]. By contrast, *C. jejuni* flagella do not induce TLR5, and removal of the *O*-glycans does not provide an enhanced response[103].

Pseudomonas *pilin glycosylation*. Instead of glycosylating their flagella, some *P. aeruginosa* strains modify their pili; modification of both appendages in a single strain has not been demonstrated. *P. aeruginosa* 1244 modifies its pilin with *N*-hydroxylbutyryl-*N*-formyl-Pse-xylose-*N*-acetylfucosamine linked to Ser148, the carboxy-terminal amino acid of pilin[104,105] (TABLE 1). It has been shown that the charge of the pilin disulphide loop is important and that the terminal Ser is the major pilin glycosylation recognition feature that cannot be substituted[106]. The structural pilus gene, *pilA*, and the *O*-OTase gene, *pilO*, are co-transcribed, and mutation of *pilO* results in loss of glycosylation[107]. Interestingly, the *O*-glycan has antigenic similarity to LPS[104] and was later shown to be derived from the LPS biosynthesis pathway[108]. By contrast, the pilin of the group IV strain *P. aeruginosa* PA5196 is still modified in an LPS mutant with a homo-oligomer of α1,5-linked D-arabinofuranose, a mycobacterium-like α1,5-linked oligosaccharide[109]. *P. aeruginosa* PA5196 pilin is glycosylated at multiple sites by TfpW, a GTase with a unique, strain-specific glycosylation activity[109] (TABLE 1).

The pili of group IV *Pseudomonas* spp. are also key to the ability of these organisms to adhere to a range of surfaces. It was suggested that pilin glycosylation is involved in virulence, as *pilO* mutation in *P. aeruginosa* 1244 resulted in reduced twitching motility, increased sensitivity to pilus-specific bacteriophages, and lack of colonization in a mouse respiratory model in competition experiments between the *pilO* mutant and wild type[110]. By contrast, pilus morphology seemed normal in the mutant, and the bacteria formed typical biofilms[110].

Neisseria *spp. pilin glycosylation*. Several pilin glycosylation (*pgl*) genes have been identified in both *Neisseria meningitidis* and *Neisseria gonorrhoeae*, which encode GTases and sugar-modifying enzymes required for the biosynthesis of the oligosaccharides[111,112] (FIG. 4b). In *N. meningitidis*, the glycan moiety comprises a trisaccharide with the structure Gal-β1,4-Gal-α1,3-DATDH (where DATDH represents 2,4-diacetamido-2,4,6-trideoxyhexose and is similar to the reducing-end sugar in the *C. jejuni N*-glycan) or Gal-β1,4-Gal-α1,3-GATDH (where GATDH represents 2-glyceramido-4-acetamido-2,4,6-trideoxyhexose)[113] (TABLE 1). However, there are variations in the length of the pilin glycan, and this is influenced by the on–off states of the biosynthesis genes, which contain homopolymeric tracts and simple repeats prone to slipped-strand mispairing[111,114,115]. In *N. gonorrhoeae*, the structure of the glycan on the intact pilin was confirmed to be (OAc)Hex-Hex-DATDH with non-stoichiometric amounts of acetylation[112]. *N. gonorrhoeae* PglO is the transferase for (OAc)Hex-Hex-DATDH and is homologous to PglL in *N. meningitidis*[111]. Both contain the domain Wzy_C (Pfam accession PF04932), a signature that is common to ligases involved in O antigen biosynthesis and that is also present in PilO in *P. aeruginosa* 1244 (REF. 107).

General *O*-linked protein glycosylation pathways

A new development in the bacterial glycosylation field was the discovery that several bacteria possess general *O*-linked glycosylation systems capable of modifying various proteins through the assembly of sugars onto lipid carriers in a process similar to *N*-linked glycosylation.

Neisseria *spp. possess a general* O-*OTase*. Two studies in 2009 demonstrated not only that *Neisseria* spp. *O*-glycosylate their pilin but also that their *O*-OTases are central enzymes in a general *O*-linked glycosylation system[116,117] (FIG. 4b). Immunoblotting of *N. gonorrhoeae* whole-cell lysates with serum raised against (OAc)Hex-DATDH-modified PilE showed that the serum reacted with multiple proteins that showed no reactivity in a *pglO* mutant[117]. Eleven membrane-associated proteins were identified, and each was *O*-glycosylated with the pilus glycan. Although no common amino acid motif around the Ser or Thr residues could be identified, all *O*-linked substrates were found to share domains bearing signatures of low complexity that are rich in Ser, Ala and Pro, which is similar to the *O*-linked-glycosylation recognition domains of mucin-type glycoproteins in eukaryotes[118]. In parallel, it was shown that the general *O*-glycosylation pathway in *N. meningitidis* modifies the outer-membrane, surface-exposed *O*-glycoprotein AniA, a nitrite reductase that carries either one or two pilin glycans in a non-structured carboxy-terminal repeat region[116] (the AniA orthologue in *N. gonorrhoeae* is also *O*-glycosylated[117]). Interestingly, AniA glycosylation increases in a pilin mutant, implying increased occupancy of partially used glycosylation sites or occupancy of additional sites and suggesting that AniA and pilin are competing for substrates from the same pathway[116].

Although it is well established that type IV pili are the major virulence determinant in *Neisseria* spp.[119], the function of the general *O*-glycosylation system still remains obscure. However, based on the structural model of AniA, which has been shown to provide protection against killing by human sera[120], it was proposed that the

Slipped-strand mispairing
Mispairing of tandem direct repeat DNA that occurs owing to slippage between the template and newly synthesized DNA strands during replication. Such mispairing can change the number of repeats in the newly synthesised strand relative to the template DNA.

Box 2 | **Exploitations for glyco-engineering**

An important achievement by Wacker and colleagues[13] was the functional transfer of the *Campylobacter jejuni N*-glycan pathway into *Escherichia coli* in 2002 (and later into *Salmonella enterica*[40]). This started a new era in glycoprotein engineering that allowed the *in vivo* production of glycosylated, homologous proteins in an easy-to-manipulate and fast-growing bacterial host. Since then, the design of specific *E. coli* expression systems (for examples, see REFS 18,39) has allowed larger-scale and more efficient production of these proteins, and advances in NMR and mass spectrometry techniques provide quick and cost-effective methods for their analyses. Moreover, novel tools such as the recently described glycophage system[144], a genetic platform to investigate the bacterial protein glycosylation pathway, allow the screening, optimization and engineering of novel or specifically designed glycosyltransferase (GTase) or oligosaccharyltransferase (OTase) activities and acceptor proteins. Thus, it is now possible to 'mix and match' and even to create novel sugar biosynthesis pathways to produce oligosaccharides linked to undecaprenyl pyrophosphate (Und-PP). Taking advantage of the relaxed substrate specificities of the *N*-OTase protein glycosylation B (PglB) or the *O*-OTases PglL and PilO, these systems have been shown to have massive potential for the synthesis of bacterial glycoconjugates by linking, for example, O antigen, peptidoglycan and capsule structures to a protein acceptor for the development of novel vaccines[40,145] (reviewed in REF. 146). However, *in vitro* and *in vivo* approaches might be limited by the availability, expression or function of the required biosynthetic enzymes and by the availability of specific precursors to reconstitute the underlying pathways. A recently described system for the production of eukaryotic *N*-glycoproteins combined *in vivo* biosynthesis with *in vitro* chemoenzymatic manipulations[147]. Using a modified glycosylation locus of *C. jejuni* in the heterologous *E. coli* system, an *N*-linked glycan with the common eukaryotic initiating sugar *N*-acetylglucosamine (GlcNAc) could be transferred to the *C. jejuni* glycoprotein AcrA *in vivo*. Following purification and subsequent *in vitro* enzymatic trimming of the glycoconjugate to GlcNAc-Asn, complex-type glycans were synthesized and enzymatically transferred not only to the GlcNAc-Asn of AcrA but also, albeit with lower efficiency, to a human antibody fragment and a single-chain antibody that contained the extended bacterial glycosylation sequon in place of the Asn-X-Ser/Thr sequon of human glycoproteins. This combination of *in vivo* and *in vitro* methods will aid in the overall goal of producing homogenous eukaryotic glycoproteins.

Another recent application describes the expression of the S-layer glycoprotein SgsE of *Geobacillus stearothermophilus* in *E. coli* after engineering one of the natural protein *O*-glycosylation sites into a target site for *N*-glycosylation[148]. Co-expression of the *pgl* genes of *C. jejuni* and the O7 biosynthesis genes of *E. coli* resulted in the presentation of the heptasaccharide and the O7 polysaccharide on the surface of *E. coli*. The authors point out that these nano-patterned, self-assembling glycoproteins may open up new strategies for influencing and controlling complex biological systems with potential applications in the areas of biomimetics, drug targeting, vaccine design or diagnostics[148].

Although most groups have been focusing on the general *N*-linked and *O*-linked protein glycosylation systems, another recent study describes a glyco-engineering application using the *Campylobacter* spp. flagellin-specific *O*-glycosylation system, in which flagella are modified with the bioorthogonal chemical tag azido-pseudaminic acid (azido-Pse)[149]. Feeding an azido-labelled Pse precursor to a *C. jejuni* mutant deficient in the synthesis of the precursor resulted in uptake and conversion of the precursor *in vivo*, with subsequent incorporation into the flagella. As the azido-Pse was amenable to further chemical modification, it was suggested that these tags could be used for cell labelling, *in vivo* animal models of infection, biophysical studies of bacterial motility and studies probing flagellum assembly or the surface accessibility of glycan modifications on flagellins.

O-glycosylated carboxyl terminus shields the substrate recognition domain of AniA, thereby protecting it from immune recognition[116].

General* O-*glycosylation in* Bacteroides *spp. *Bacteroides* is the numerically dominant genus of the human intestinal microbiota, and members of this genus produce a vast number of glycan structures, including at least eight different phase-variable capsular polysaccharides, four of which may contain a terminal fucose residue[121–124]. Investigations into fucose utilization by *Bacteroides* spp. found that exogenous L-fucose from the host is incorporated not only into multiple bacterial capsular polysaccharides but also into oligosaccharides that are added onto glycoproteins[123,125,126]. Mutational analyses of one of the eight identified glycoproteins, BF2494, revealed that three sites are modified, but the glycan structure could not be determined. Subsequent mutational analyses of the glycosylation sites uncovered an unexpected amino acid motif for *O*-glycosylation in *Bacteroides fragilis* — Asp-Ser/Thr-Ala/Ile/Leu/Val/Met/Thr — in which the last amino acid requires a methyl group. The gene locus that was identified as being involved in *O*-glycosylation (termed *lfg* for locus of *B. fragilis* glycosylation) contained, among other genes, a putative flippase and five putative GTases. Similar genetic loci were found in *Bacteroides caccae*, *Bacteroides ovatus*, *Bacteroides thetaiotaomicron*, *Bacteroides uniformis* and *Bacteroides vulgatus*, and all of these species were shown to produce *O*-linked glycoproteins[126]. Interestingly, the Gram-negative bacteria *Flavobacterium meningosepticum* (now called *Elizabethkingia meningoseptica*) and *Flavobacterium columnare* also *O*-glycosylate several secreted proteins (including the first-described endo-β-*N*-acetylglucosaminidase (EndoH), capable of cleaving eukaryotic *N*-linked glycans[127]) with a heptasaccharide containing mannose, rhamnose, glucuronic acid, 2-acetamido-2-deoxyglucuronic acid and its methylated derivatives. The glycan is localized at Asp-Ser and Asp-Thr-Thr consensus sites similar to the sequon identified for the *Bacteroides* spp. described above[128–130].

A *B. fragilis* strain deficient in the ability to fucosylate its surface glycoproteins and/or capsular polysaccharides was out-competed by the wild type in mouse co-infection studies[123]. This suggests that fucosylation gives these bacteria a survival advantage in the mammalian intestinal ecosystem[123,126] and allows them to be immunologically inert through molecular mimicry of their host[131]. However, it has yet to be shown whether protein glycosylation has a specific function or whether these phenotypes are due to changes in bacterial capsular polysaccharides, which have already been shown to play many important roles in interactions with the host.

O-mannosylation in Gram-positive actinomycetes. It has long been known that, among the actinomycetes, *Streptomyces* spp. and *Mycobacterium* spp. glycosylate several bioactive natural products and secreted antigens[132]. For example, two surface lipoproteins of *Mycobacterium tuberculosis* are *O*-glycosylated, and the implicated transferase has structural similarity with eukaryotic protein mannosyltransferases[133,134]. The process of *O*-mannosylation in all three domains of life was recently reviewed[135]. In *Streptomyces coelicolor*, the phosphate-binding protein PstS is *O*-glycosylated with a trihexose by the mannosyltransferase Pmt, a membrane-bound lipoprotein[136]. Thus, *O*-mannosylation seems to be a general pathway found in all actinomycetes. This process also requires a lipid carrier, similarly to what has been described for *Neisseria* spp. and *Pseudomonas* spp.

(FIG. 4b), but the LLO would need to be flipped across the outer membrane of these Gram-positive organisms. *In vivo* glycosylation studies suggest that Pmt does not glycosylate Ser or Thr residues nonspecifically but, rather, has a target site preference[136]. However, despite the fact that all glycosylated residues in the actinomycetes are located close to the amino or carboxyl termini of the target proteins, a consensus sequence has not been identified so far.

Conclusions and perspectives

The study of bacterial glycoprotein pathways has been hindered in the past by the inherent complexity of these systems and the lack of specialized analytical methods for their interpretation. However, a growing number of bacterial *N*-linked protein glycosylation pathways have now been identified that are predicted to proceed through the sequential addition of nucleotide-activated sugars onto membrane-anchored lipids, followed by the translocation of the synthesized oligosaccharide across the inner membrane and subsequent attachment of the sugar chain onto Asn residues of specific protein sequons. The prototype for the study of this pathway has been *C. jejuni*. By contrast, there are several examples of bacterial *O*-linked glycosylation systems in which GTases sequentially add nucleotide-activated sugars directly onto specific proteins at Ser or Thr residues with no obvious sequon requirement. These two mechanisms of bacterial protein glycosylation resemble those that have been extensively studied in eukaryotes. In addition, general *O*-linked pathways have been described that involve a lipid-linked anchor from which a block of sugars are transferred onto multiple proteins and, in some cases, a sequon has been identified. Even more unexpectedly, a process for *N*-glycosylation in the bacterial cytoplasm has been described that utilizes multifunctional GTases capable of transferring different nucleotide-activated sugars onto Asn as well as onto other sugars.

This rapidly increasing repertoire of GTases and OTases provides an enormous array of both sugar-specific and promiscuous enzymes that may eventually allow the synthesis of any saccharide and its transfer onto a protein (BOX 2). The potential of combinatorial glyco-engineering strategies from both evolutionary and applied perspectives, together with the optimization and exploitation of currently available *in vivo* and *in vitro* systems, should enable the large-scale production of recombinant glycoproteins for the generation of glycoconjugates with industrial and medical applications. In turn, this will no doubt revolutionize the methods by which glycoconjugate vaccines are produced in the future.

1. Furmanek, A. & Hofsteenge, J. Protein C-mannosylation: facts and questions. *Acta Biochim. Pol.* **47**, 781–789 (2000).
2. Neuberger, A. Carbohydrates in protein: the carbohydrate component of crystalline egg albumin. *Biochem. J.* **32**, 1435–1451 (1938).
3. Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8 (1999).
4. Moens, S. & Vanderleyden, J. Glycoproteins in prokaryotes. *Arch. Microbiol.* **168**, 169–175 (1997).
5. Mescher, M. F. & Strominger, J. L. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. *J. Biol. Chem.* **251**, 2005–2014 (1976).
6. Sleytr, U. B. Heterologous reattachment of regular arrays of glycoproteins on bacterial surfaces. *Nature* **257**, 400–402 (1975).
7. Sleytr, U. B. & Thorne, K. J. Chemical characterization of the regularly arranged surface layers of *Clostridium thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*. *J. Bacteriol.* **126**, 377–383 (1976).
8. Messner, P., Steiner, K., Zarschler, K. & Schaffer, C. S-layer nanoglycobiology of bacteria. *Carbohydr. Res.* **343**, 1934–1951 (2008).
Review covering the recent advances made in the field of bacterial S-layer nanoglycobiology.
9. Zarschler, K. *et al.* Protein tyrosine *O*-glycosylation – a rather unexplored prokaryotic glycosylation system. *Glycobiology* **20**, 787–798 (2010).
10. Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. & Guerry, P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* **32**, 1022–1030 (1999).
11. Linton, D., Allan, E., Karlyshev, A. V., Cronshaw, A. D. & Wren, B. W. Identification of *N*-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. *Mol. Microbiol.* **43**, 497–508 (2002).
12. Young, N. M. *et al.* Structure of the *N*-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem.* **277**, 42530–42539 (2002).
First definitive proof of *N*-linked protein glycosylation in bacteria and structural determination of the *C. jejuni* glycan.
13. Wacker, M. *et al.* N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**, 1790–1793 (2002).
First study to demonstrate the use of bacteria for recombinant protein glyco-engineering.
14. Szymanski, C. M. & Wren, B. W. Protein glycosylation in bacterial mucosal pathogens. *Nature Rev. Microbiol.* **3**, 225–237 (2005).
15. Linton, D. *et al.* Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol. Microbiol.* **55**, 1695–1703 (2005).
16. Reid, C. W. *et al.* Affinity-capture tandem mass spectrometric characterization of polyprenyl-linked oligosaccharides: tool to study protein *N*-glycosylation pathways. *Anal. Chem.* **80**, 5468–5475 (2008).
17. Reid, C. W., Stupak, J., Szymanski, C. M. & Li, J. Analysis of bacterial lipid-linked oligosaccharide intermediates using porous graphitic carbon liquid chromatography-electrospray ionization mass spectrometry: heterogeneity in the polyisoprenyl carrier revealed. *Anal. Chem.* **81**, 8472–8478 (2009).
Description of a general technique for the isolation and characterization of lipid-linked oligosaccharides.
18. Alaimo, C. *et al.* Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25**, 967–976 (2006).
19. Kelly, J. *et al.* Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. *J. Bacteriol.* **188**, 2427–2434 (2006).
20. Kowarik, M. *et al.* Definition of the bacterial *N*-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966 (2006).
Demonstration that the *N*-glycosylation consensus sequence requirement is extended for the *C. jejuni* OTase.
21. Kowarik, M. *et al.* N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **314**, 1148–1150 (2006).
Study showing that, in contrast to the glycosylation of unfolded proteins in eukaryotes, *N*-glycosylation in bacteria occurs at locally flexible structures in folded proteins.
22. Nothaft, H., Amber, S., Aebi, M. & Szymanski, C. M. in *Campylobacter* 3rd edn (eds Nachamkin, I., Szymanski, C. M. & Blaser, M. J.) 447–469 (ASM, Washington DC, 2008).
23. Scott, N. E. *et al.* Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, HCD and ETD-MS applied to the *N*-linked glycoproteome of *Campylobacter jejuni*. *Mol. Cell Proteomics* 1 Apr 2010 (doi:10.1074/mcp.M000031-MCP201).
24. Schoenhofen, I. C. *et al.* Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*: enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. *J. Biol. Chem.* **281**, 723–732 (2006).
25. Olivier, N. B., Chen, M. M., Behr, J. R. & Imperiali, B. *In vitro* biosynthesis of UDP-*N*,*N*′-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system. *Biochemistry* **45**, 13659–13669 (2006).
26. Olivier, N. B. & Imperiali, B. Crystal structure and catalytic mechanism of PglD from *Campylobacter jejuni*. *J. Biol. Chem.* **283**, 27937–27946 (2008).
27. Rangarajan, E. S. *et al.* Structure and active site residues of PglD, an *N*-acetyltransferase from the bacillosamine synthetic pathway required for *N*-glycan synthesis in *Campylobacter jejuni*. *Biochemistry* **47**, 1827–1836 (2008).
28. Glover, K. J., Weerapana, E., Chen, M. M. & Imperiali, B. Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry* **45**, 5343–5350 (2006).
29. Weerapana, E., Glover, K. J., Chen, M. M. & Imperiali, B. Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. *J. Am. Chem. Soc.* **127**, 13766–13767 (2005).
30. Troutman, J. M. & Imperiali, B. *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. *Biochemistry* **48**, 2807–2816 (2009).
31. Glover, K. J., Weerapana, E. & Imperiali, B. *In vitro* assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Natl Acad. Sci. USA* **102**, 14255–14259 (2005).
First *in vitro* synthesis of the complete lipid-linked *N*-glycan of *C. jejuni* using purified enzyme components and chemically synthesized Und-PP-diNAcBac.

32. Jones, M. B., Rosenberg, J. N., Betenbaugh, M. J. & Krag, S. S. Structure and synthesis of polyisoprenoids used in N-glycosylation across the three domains of life. *Biochim. Biophys. Acta* **1790**, 485–494 (2009).
33. Burda, P. & Aebi, M. The dolichol pathway of *N*-linked glycosylation. *Biochim. Biophys. Acta* **1426**, 239–257 (1999).
34. Abu-Qarn, M., Eichler, J. & Sharon, N. Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.* **18**, 544–550 (2008).
35. Chaban, B., Voisin, S., Kelly, J., Logan, S. M. & Jarrell, K. F. Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae N*-linked glycans: insight into *N*-linked glycosylation pathways in Archaea. *Mol. Microbiol.* **61**, 259–268 (2006).
36. Chen, M. M. *et al.* Polyisoprenol specificity in the *Campylobacter jejuni* N-linked glycosylation pathway. *Biochemistry* **46**, 14342–14348 (2007).
37. Chen, M. M., Glover, K. J. & Imperiali, B. From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in *Campylobacter jejuni*. *Biochemistry* **46**, 5579–5585 (2007).
38. Glover, K. J., Weerapana, E., Numao, S. & Imperiali, B. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol.* **12**, 1311–1315 (2005).
39. Feldman, M. F. *et al.* Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **102**, 3016–3021 (2005).
40. Wacker, M. *et al.* Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl Acad. Sci. USA* **103**, 7088–7093 (2006).
41. Nothaft, H., Liu, X., McNally, D. J. & Szymanski, C. M. *N*-linked protein glycosylation in a bacterial system. *Methods Mol. Biol.* **600**, 227–243 (2010).
42. Rangarajan, E. S. *et al.* Structural context for protein *N*-glycosylation in bacteria: The structure of PEB3, an adhesin from *Campylobacter jejuni*. *Protein Sci.* **16**, 990–995 (2007).
43. Min, T. *et al.* Specificity of *Campylobacter jejuni* adhesin PEB3 for phosphates and structural differences among its ligand complexes. *Biochemistry* **48**, 3057–3067 (2009).
44. Slynko, V. *et al.* NMR structure determination of a segmentally labeled glycoprotein using *in vitro* glycosylation. *J. Am. Chem. Soc.* **131**, 1274–1281 (2009).
45. Jones, M. A. *et al.* Adaptation of *Campylobacter jejuni* NCTC11168 to high-level colonization of the avian gastrointestinal tract. *Infect. Immun.* **72**, 3769–3776 (2004).
46. Karlyshev, A. V. *et al.* The *Campylobacter jejuni* general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology* **150**, 1957–1964 (2004).
47. Szymanski, C. M., Burr, D. H. & Guerry, P. Campylobacter protein glycosylation affects host cell interactions. *Infect. Immun.* **70**, 2242–2244 (2002).
48. Hendrixson, D. R. & DiRita, V. J. Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract. *Mol. Microbiol.* **52**, 471–484 (2004).
49. Ding, W., Nothaft, H., Szymanski, C. M. & Kelly, J. Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry. *Mol. Cell Proteomics* **8**, 2170–2185 (2009).
50. Larsen, J. C., Szymanski, C. & Guerry, P. N-linked protein glycosylation is required for full competence in *Campylobacter jejuni* 81-176. *J. Bacteriol.* **186**, 6508–6514 (2004).
51. Kakuda, T. & DiRita, V. J. Cj1496c encodes a *Campylobacter jejuni* glycoprotein that influences invasion of human epithelial cells and colonization of the chick gastrointestinal tract. *Infect. Immun.* **74**, 4715–4723 (2006).
52. Davis, L. M., Kakuda, T. & DiRita, V. J. A *Campylobacter jejuni znuA* orthologue is essential for growth in low-zinc environments and chick colonization. *J. Bacteriol.* **191**, 1631–1640 (2009).
53. Flanagan, R. C., Neal-McKinney, J. M., Dhillon, A. S., Miller, W. G. & Konkel, M. E. Examination of *Campylobacter jejuni* putative adhesins leads to the identification of a new protein, designated FlpA, required for chicken colonization. *Infect. Immun.* **77**, 2399–2407 (2009).
54. Scott, N. E. *et al.* Mass spectrometric characterization of the surface-associated 42 kDa lipoprotein JlpA as a glycosylated antigen in strains of *Campylobacter jejuni*. *J. Proteome Res.* **8**, 4654–4664 (2009).
55. van Sorge, N. M. *et al.* *N*-glycosylated proteins and distinct lipooligosaccharide glycoforms of *Campylobacter jejuni* target the human C-type lectin receptor MGL. *Cell. Microbiol.* **11**, 1768–1781 (2009).
Study showing that the *C. jejuni N*-glycan binds to human MGL and may thereby modulate the host immune response.
56. Hacker, J., Blum-Oehler, G., Muhldorfer, I. & Tschape, H. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* **23**, 1089–1097 (1997).
57. Maita, N., Nyirenda, J., Igura, M., Kamishikiryo, J. & Kohda, D. Comparative structural biology of Eubacterial and Archaeal oligosaccharyltransferases. *J. Biol. Chem.* **285**, 4941–4950 (2010).
58. Igura, M. *et al.* Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* **27**, 234–243 (2008).
59. Nakagawa, S. *et al.* Deep-sea vent epsilon-proteobacterial genomes provide insights into emergence of pathogens. *Proc. Natl Acad. Sci. USA* **104**, 12146–12150 (2007).
60. Baar, C. *et al.* Complete genome sequence and analysis of *Wolinella succinogenes*. *Proc. Natl Acad. Sci. USA* **100**, 11690–11695 (2003).
61. Loman, N. J. *et al.* Genome sequence of the emerging pathogen *Helicobacter canadensis*. *J. Bacteriol.* **191**, 5566–5567 (2009).
62. Jervis, A. J. *et al.* Characterisation of *N*-linked protein glycosylation in *Helicobacter pullorum*. *J. Bacteriol.* **192**, 5228–5236.
Description of another bacterial *N*-glycosylation pathway resembling the *C. jejuni* prototype.
63. Hug, I. *et al. Helicobacter pylori* lipopolysaccharide is synthesized via a novel pathway with an evolutionary connection to protein *N*-glycosylation. *PLoS Pathog.* **6**, e1000819 (2010).
64. Fox, J. G. The non-*H. pylori* helicobacters: their expanding role in gastrointestinal and systemic diseases. *Gut* **50**, 273–283 (2002).
65. Mobley, H. L. T., Mendz, G. L. & Hazell, S. L. in *Helicobacter pylori: Physiology and Genetics* (eds Vandamme, P. & Solnick, J. V.) 39–52 (ASM, Washington DC, 2001).
66. Santos-Silva, T. *et al.* Crystal structure of the 16 heme cytochrome from *Desulfovibrio gigas*: a glycosylated protein in a sulphate-reducing bacterium. *J. Mol. Biol.* **370**, 659–673 (2007).
67. Gross, J. *et al.* The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an unusual *N*-linked carbohydrate modification. *J. Biol. Chem.* **283**, 26010–26015 (2008).
68. St. Geme, J. W. 3rd. The HMW1 adhesin of nontypeable *Haemophilus influenzae* recognizes sialylated glycoprotein receptors on cultured human epithelial cells. *Infect. Immun.* **62**, 3881–3889 (1994).
69. Grass, S. *et al.* The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol. Microbiol.* **48**, 737–751 (2003).
70. Grass, S., Lichti, C. F., Townsend, R. R., Gross, J. & St. Geme, J. W. 3rd. The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. *PLoS Pathog.* **6**, e1000919 (2010).
Recent publication describing the unusual *N*-linked protein glycosylation system of *Haemophilus influenzae*.
71. Benz, I. & Schmidt, M. A. Glycosylation with heptose residues mediated by the *aah* gene product is essential for adherence of the AIDA-I adhesin. *Mol. Microbiol.* **40**, 1403–1413 (2001).
72. Lindenthal, C. & Elsinghorst, E. A. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect. Immun.* **67**, 4084–4091 (1999).
73. Sherlock, O., Dobrindt, U., Jensen, J. B., Munk Vejborg, R. & Klemm, P. Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein. *J. Bacteriol.* **188**, 1798–1807 (2006).
74. Charbonneau, M. E. & Mourez, M. The *Escherichia coli* AIDA-I autotransporter undergoes cytoplasmic glycosylation independently of export. *Res. Microbiol.* **159**, 537–544 (2008).
75. Logan, S. M. Flagellar glycosylation — a new component of the motility repertoire? *Microbiology* **152**, 1249–1262 (2006).
Review on bacterial flagellar glycosylation.
76. Ng, S. Y., Chaban, B. & Jarrell, K. F. Archaeal flagella, bacterial flagella and type IV pili: a comparison of genes and posttranslational modifications. *J. Mol. Microbiol. Biotechnol.* **11**, 167–191 (2006).
77. Thibault, P. *et al.* Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *J. Biol. Chem.* **276**, 34862–34870 (2001).
78. Goon, S., Kelly, J. F., Logan, S. M., Ewing, C. P. & Guerry, P. Pseudaminic acid, the major modification on *Campylobacter flagellin*, is synthesized via the Cj1293 gene. *Mol. Microbiol.* **50**, 659–671 (2003).
79. Guerry, P. *Campylobacter* flagella: not just for motility. *Trends Microbiol.* **15**, 456–461 (2007).
Review summarizing the biological functions of *O*-linked flagella glycosylation in *Campylobacter* spp.
80. Guerry, P. & Szymanski, C. M. *Campylobacter* sugars sticking out. *Trends Microbiol.* **16**, 428–435 (2008).
81. Champion, O. L. *et al.* Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source. *Proc. Natl Acad. Sci. USA* **102**, 16043–16048 (2005).
82. Howard, S. L. *et al. Campylobacter jejuni* glycosylation island important in cell charge, legionaminic acid biosynthesis, and colonization of chickens. *Infect. Immun.* **77**, 2544–2556 (2009).
83. McNally, D. J. *et al.* Functional characterization of the flagellar glycosylation locus in *Campylobacter jejuni* 81-176 using a focused metabolomics approach. *J. Biol. Chem.* **281**, 18489–18498 (2006).
84. Schirm, M., Schoenhofen, I. C., Logan, S. M., Waldron, K. C. & Thibault, P. Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins. *Anal. Chem.* **77**, 7774–7782 (2005).
85. Guerry, P. *et al.* Changes in flagellin glycosylation affect *Campylobacter* autoagglutination and virulence. *Mol. Microbiol.* **60**, 299–311 (2006).
86. Ewing, C. P., Andreishcheva, E. & Guerry, P. Functional characterization of flagellin glycosylation in *Campylobacter jejuni* 81-176. *J. Bacteriol.* **191**, 7086–7093 (2009).
87. Schoenhofen, I. C., Vinogradov, E., Whitfield, D. M., Brisson, J. R. & Logan, S. M. The CMP-legionaminic acid pathway in *Campylobacter*: Biosynthesis involving novel GDP-linked precursors. *Glycobiology* **19**, 715–725 (2009).
88. Josenhans, C., Vossebein, L., Friedrich, S. & Suerbaum, S. The *neuA*/*flmD* gene cluster of *Helicobacter pylori* is involved in flagellar biosynthesis and flagellin glycosylation. *FEMS Microbiol. Lett.* **210**, 165–172 (2002).
89. Schirm, M. *et al.* Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* **48**, 1579–1592 (2003).
90. Schoenhofen, I. C., McNally, D. J., Brisson, J. R. & Logan, S. M. Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: synthesis from UDP-*N*-acetylglucosamine by a single enzymatic reaction. *Glycobiology* **16**, 8C–14C (2006).
91. Twine, S. M. *et al.* Flagellar glycosylation in *Clostridium botulinum*. *FEBS J.* **275**, 4428–4444 (2008).
92. Twine, S. M. *et al.* Motility and flagellar glycosylation in *Clostridium difficile*. *J. Bacteriol.* **191**, 7050–7062 (2009).
93. Allison, J. S., Dawson, M., Drake, D. & Montie, T. C. Electrophoretic separation and molecular weight characterization of *Pseudomonas aeruginosa* H-antigen flagellins. *Infect. Immun.* **49**, 770–774 (1985).
94. Lanyi, B. Serological properties of *Pseudomonas aeruginosa*. II. Type-specific thermolabile (flagellar) antigens. *Acta Microbiol. Acad. Sci. Hung.* **17**, 35–48 (1970).
95. Arora, S. K., Bangera, M., Lory, S. & Ramphal, R. A genomic island in *Pseudomonas aeruginosa* carries the determinants of flagellin glycosylation. *Proc. Natl Acad. Sci. USA* **98**, 9342–9347 (2001).
96. Brimer, C. D. & Montie, T. C. Cloning and comparison of *fliC* genes and identification of glycosylation in the flagellin of *Pseudomonas aeruginosa* a-type strains. *J. Bacteriol.* **180**, 3209–3217 (1998).
97. Schirm, M. *et al.* Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *J. Bacteriol.* **186**, 2523–2531 (2004).

98. Arora, S. K., Wolfgang, M. C., Lory, S. & Ramphal, R. Sequence polymorphism in the glycosylation island and flagellins of *Pseudomonas aeruginosa*. *J. Bacteriol.* **186**, 2115–2122 (2004).
99. Verma, A. *et al.* Glycosylation of b-type flagellin of *Pseudomonas aeruginosa*: structural and genetic basis. *J. Bacteriol.* **188**, 4395–4403 (2006).
100. Takeuchi, K. *et al.* Flagellin glycans from two pathovars of *Pseudomonas syringae* contain rhamnose in D and L configurations in different ratios and modified 4-amino-4, 6-dideoxyglucose. *J. Bacteriol.* **189**, 6945–6956 (2007).
101. Hayashi, F. *et al.* The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099–1103 (2001).
102. Verma, A., Arora, S. K., Kuravi, S. K. & Ramphal, R. Roles of specific amino acids in the N-terminus of *Pseudomonas aeruginosa* flagellin and of flagellin glycosylation in the innate immune response. *Infect. Immun.* **73**, 8237–8246 (2005).
103. de Zoete, M. R., Keestra, A. M., Wagenaar, J. A. & van Putten, J. P. Reconstitution of a functional Toll-like receptor 5 binding site in *Campylobacter jejuni* flagellin. *J. Biol. Chem.* **285**, 12149–12158 (2010).
104. Castric, P., Cassels, F. J. & Carlson, R. W. Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. *J. Biol. Chem.* **276**, 26479–26485 (2001).
105. Comer, J. E., Marshall, M. A., Blanch, V. J., Deal, C. D. & Castric, P. Identification of the *Pseudomonas aeruginosa* 1244 pilin glycosylation site. *Infect. Immun.* **70**, 2837–2845 (2002).
106. Horzempa, J., Comer, J. E., Davis, S. A. & Castric, P. Glycosylation substrate specificity of *Pseudomonas aeruginosa* 1244 pilin. *J. Biol. Chem.* **281**, 1128–1136 (2006).
107. Castric, P. *pilO*, a gene required for glycosylation of *Pseudomonas aeruginosa* 1244 pilin. *Microbiology* **141**, 1247–1254 (1995).
108. DiGiandomenico, A. *et al.* Glycosylation of *Pseudomonas aeruginosa* 1244 pilin: glycan substrate specificity. *Mol. Microbiol.* **46**, 519–530 (2002).
109. Kus, J. V. *et al.* Modification of *Pseudomonas aeruginosa* Pa5196 type IV pilins at multiple sites with D-Ara*f* by a novel GT-C family arabinosyltransferase, TfpW. *J. Bacteriol.* **190**, 7464–7478 (2008).
110. Smedley, J. G. 3rd *et al.* Influence of pilin glycosylation on *Pseudomonas aeruginosa* 1244 pilus function. *Infect. Immun.* **73**, 7922–7931 (2005).
111. Power, P. M. *et al.* Genetic characterization of pilin glycosylation and phase variation in *Neisseria meningitidis*. *Mol. Microbiol.* **49**, 833–847 (2003).
112. Aas, F. E., Vik, A., Vedde, J., Koomey, M. & Egge-Jacobsen, W. *Neisseria gonorrhoeae O*-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. *Mol. Microbiol.* **65**, 607–624 (2007).
113. Chamot-Rooke, J. *et al.* Alternative *Neisseria* spp. type IV pilin glycosylation with a glyceramido acetamido trideoxyhexose residue. *Proc. Natl Acad. Sci. USA* **104**, 14783–14788 (2007).
114. Jennings, M. P. *et al.* Identification of a novel gene involved in pilin glycosylation in *Neisseria meningitidis*. *Mol. Microbiol.* **29**, 975–984 (1998).
115. Banerjee, A. *et al.* Implications of phase variation of a gene (*pgtA*) encoding a pilin galactosyl transferase in gonococcal pathogenesis. *J. Exp. Med.* **196**, 147–162 (2002).
116. Ku, S. C., Schulz, B. L., Power, P. M. & Jennings, M. P. The pilin *O*-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase. *Biochem. Biophys. Res. Commun.* **378**, 84–89 (2009).
117. Vik, A. *et al.* Broad spectrum *O*-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proc. Natl Acad. Sci. USA* **106**, 4447–4452 (2009).
Demonstration of general protein *O*-glycosylation in *Neisseria* spp., along with reference 116.
118. Gerken, T. A., Owens, C. L. & Pasumarthy, M. Determination of the site-specific *O*-glycosylation pattern of the porcine submaxillary mucin tandem repeat glycopeptide. Model proposed for the polypeptide:GalNAc transferase peptide binding site. *J. Biol. Chem.* **272**, 9709–9719 (1997).
119. Mattick, J. S. Type IV pili and twitching motility. *Annu. Rev. Microbiol.* **56**, 289–314 (2002).
120. Cardinale, J. A. & Clark, V. L. Expression of AniA, the major anaerobically induced outer membrane protein of *Neisseria gonorrhoeae*, provides protection against killing by normal human sera. *Infect. Immun.* **68**, 4368–4369 (2000).
121. Coyne, M. J., Weinacht, K. G., Krinos, C. M. & Comstock, L. E. Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc. Natl Acad. Sci. USA* **100**, 10446–10451 (2003).
122. Krinos, C. M. *et al.* Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414**, 555–558 (2001).
123. Coyne, M. J., Reinap, B., Lee, M. M. & Comstock, L. E. Human symbionts use a host-like pathway for surface fucosylation. *Science* **307**, 1778–1781 (2005).
124. Kalka-Moll, W. M. *et al.* Immunochemical and biological characterization of three capsular polysaccharides from a single *Bacteroides fragilis* strain. *Infect. Immun.* **69**, 2339–2344 (2001).
125. Fletcher, C. M., Coyne, M. J., Bentley, D. L., Villa, O. F. & Comstock, L. E. Phase-variable expression of a family of glycoproteins imparts a dynamic surface to a symbiont in its human intestinal ecosystem. *Proc. Natl Acad. Sci. USA* **104**, 2413–2418 (2007).
126. Fletcher, C. M., Coyne, M. J., Villa, O. F., Chatzidaki-Livanis, M. & Comstock, L. E. A general *O*-glycosylation system important to the physiology of a major human intestinal symbiont. *Cell* **137**, 321–331 (2009).
Recent literature on general *B. fragilis* protein *O*-glycosylation, which is important for colonization of the mammalian intestine.
127. Elder, J. H. & Alexander, S. *endo*-β-*N*-Acetylglucosaminidase F: endoglycosidase from *Flavobacterium meningosepticum* that cleaves both high-mannose and complex glycoproteins. *Proc. Natl Acad. Sci. USA* **79**, 4540–4544 (1982).
128. Reinhold, B. B., Hauer, C. R., Plummer, T. H. & Reinhold, V. N. Detailed structural analysis of a novel, specific *O*-linked glycan from the prokaryote *Flavobacterium meningosepticum*. *J. Biol. Chem.* **270**, 13197–13203 (1995).
129. Vinogradov, E., Perry, M. B. & Kay, W. W. The structure of the glycopeptides from the fish pathogen *Flavobacterium columnare*. *Carbohydr. Res.* **338**, 2653–2658 (2003).
130. Plummer, T. H. Jr, Tarentino, A. L. & Hauer, C. R. Novel, specific *O*-glycosylation of secreted *Flavobacterium meningosepticum* proteins. Asp-Ser and Asp-Thr-Thr consensus sites. *J. Biol. Chem.* **270**, 13192–13196 (1995).
131. Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. A model of host-microbial interactions in an open mammalian ecosystem. *Science* **273**, 1380–1383 (1996).
132. Schmidt, M. A., Riley, L. W. & Benz, I. Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol.* **11**, 554–561 (2003).
133. Dobos, K. M., Khoo, K. H., Swiderek, K. M., Brennan, P. J. & Belisle, J. T. Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. *J. Bacteriol.* **178**, 2498–2506 (1996).
134. VanderVen, B. C., Harder, J. D., Crick, D. C. & Belisle, J. T. Export-mediated assembly of mycobacterial glycoproteins parallels eukaryotic pathways. *Science* **309**, 941–943 (2005).
135. Lommel, M. & Strahl, S. Protein *O*-mannosylation: conserved from bacteria to humans. *Glycobiology* **19**, 816–828 (2009).
Review summarizing the main features of the biosynthesis pathway of *O*-mannosyl glycans and addressing their biological relevance in bacteria and eukaryotes.
136. Wehmeier, S. *et al.* Glycosylation of the phosphate binding protein, PstS, in *Streptomyces coelicolor* by a pathway that resembles protein *O*-mannosylation in eukaryotes. *Mol. Microbiol.* **71**, 421–433 (2009).
137. Szymanski, C. M. *et al.* Detection of conserved *N*-linked glycans and phase-variable lipooligosaccharides and capsules from campylobacter cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy. *J. Biol. Chem.* **278**, 24509–24520 (2003).
138. Liu, X. *et al.* Mass spectrometry-based glycomics strategy for exploring N-linked glycosylation in eukaryotes and bacteria. *Anal. Chem.* **78**, 6081–6087 (2006).
First description of the existence of fOS derived from the bacterial *N*-glycosylation pathway.
139. Nothaft, H., Liu, X., McNally, D. J., Li, J. & Szymanski, C. M. Study of free oligosaccharides derived from the bacterial *N*-glycosylation pathway. *Proc. Natl Acad. Sci. USA* **106**, 15019–15024 (2009).
Demonstration that PglB of *C. jejuni* possesses both OTase and hydrolase activity and that release of fOS is sensitive to osmolyte concentration.
140. Bohin, J. P. Osmoregulated periplasmic glucans in Proteobacteria. *FEMS Microbiol. Lett.* **186**, 11–19 (2000).
141. Suzuki, T. & Funakoshi, Y. Free *N*-linked oligosaccharide chains: formation and degradation. *Glycoconj. J.* **23**, 291–302 (2006).
142. Chantret, I. & Moore, S. E. Free oligosaccharide regulation during mammalian protein *N*-glycosylation. *Glycobiology* **18**, 210–224 (2008).
143. Hirayama, H., Seino, J., Kitajima, T., Jigami, Y. & Suzuki, T. Free oligosaccharides to monitor glycoprotein endoplasmic reticulum-associated degradation in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **285**, 12390–12404 (2010).
144. Durr, C., Nothaft, H., Lizak, C., Glockshuber, R. & Aebi, M. The *Escherichia coli* glycophage display system. *Glycobiology* 25 Jun 2010 (doi:10.1093/glycob/cwq102).
145. Faridmoayer, A. *et al.* Extreme substrate promiscuity of the *Neisseria* oligosaccharyl transferase involved in protein *O*-glycosylation. *J. Biol. Chem.* **283**, 34596–34604 (2008).
146. Feldman, M. F. in *Microbial Glycobiology: Structures, Relevance and Applications* (eds Moran, A. P. *et al.*) 903–914 (Elsevier, Burlington, Massachusetts, 2009).
Recent review on the emerging area of bacterial glycoprotein engineering, with a focus on the bacterial OTases.
147. Schwarz, F. *et al.* A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nature Chem. Biol.* **6**, 264–266 (2010).
Important proof-of-concept study using a combined *in vivo* and *in vitro* chemoenzymatic approach for the biosynthesis of eukaryotic glycoproteins.
148. Steiner, K. *et al.* Recombinant glycans on an S-layer self-assembly protein: a new dimension for nanopatterned biomaterials. *Small* **4**, 1728–1740 (2008).
149. Liu, F., Aubry, A. J., Schoenhofen, I. C., Logan, S. M. & Tanner, M. E. The engineering of bacteria bearing azido-pseudaminic acid-modified flagella. *Chembiochem* **10**, 1317–1320 (2009).
150. Kumar, S., Nei, M., Dudley, J. & Tamura, K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* **9**, 299–306 (2008).

Acknowledgements

We thank W. G. Miller for providing sequence information on the *pgl* loci, and M. Klotz and J. St.Geme for stimulating discussions. This work has been supported by the Alberta Ingenuity Centre for Carbohydrate Science and the Canadian National Research Council Genomics and Health Initiative. C.M.S. is an Alberta Ingenuity Scholar.

Competing interests statement

The authors declare no competing financial interests.

FURTHER INFORMATION

Harald Nothaft and Christine M. Szymanski's homepage: http://www.biology.ualberta.ca/faculty/christine_szymanski
Entrez Genome Project: http://www.ncbi.nlm.nih.gov/genomeprj
Pfam: http://pfam.janelia.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF