PROTEIN GLYCOSYLATION IN BACTERIAL MUCOSAL PATHOGENS

Christine M. Szymanski and Brendan W. Wren‡*

Abstract | In eukaryotes, glycosylated proteins are ubiquitous components of extracellular matrices and cellular surfaces. Their oligosaccharide moieties are implicated in a wide range of cell–cell and cell–matrix recognition events that are required for biological processes ranging from immune recognition to cancer development. Glycosylation was previously considered to be restricted to eukaryotes; however, through advances in analytical methods and genome sequencing, there have been increasing reports of both *O*-linked and *N*-linked protein glycosylation pathways in bacteria, particularly amongst mucosal-associated pathogens. Studying glycosylation in relatively less-complicated bacterial systems provides the opportunity to elucidate and exploit glycoprotein biosynthetic pathways. We will review the genetic organization, glycan structures and function of glycosylation systems in mucosal bacterial pathogens, and speculate on how this knowledge may help us to understand glycosylation processes in more complex eukaryotic systems and how it can be used for glycoengineering.

**Institute for Biological Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada.*
‡Department of Infectious and Tropical Disease, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1 7HT, UK.
e-mails: Christine.Szymanski@nrc-cnrc.gc.ca; brendan.wren@lshtm.ac.uk
doi:10.1038/nrmicro1100
Published online
10 February 2005

It is now evident that glycoproteins are a common feature in all domains of life. However, there is limited information available about the structures of the linked glycans, the detailed process of protein glycosylation, the importance of these modifications and how these systems can be exploited. Glycan structures are usually attached to proteins at either an Asn-X-Ser/Thr consensus sequence (*N*-linked, where X represents any amino acid except proline) or to Ser/Thr residues (*O*-linked). Recent publications by the authors described the first confirmed report of a bacterial *N*-linked glycosylation pathway in *Campylobacter jejuni*[1,2]. In addition, an *O*-linked glycosylation pathway has also been identified and characterized in *C. jejuni*[3–5]. Orthologues of the genes in both pathways are found in other bacteria, particularly mucosal-associated pathogens, and both pathways have similarity to the respective *N*- and *O*-linked glycosylation processes in eukaryotes. Although bacterial glycans are more complex than their eukaryotic counterparts, there are many similarities in the structures and biosynthetic pathways in mucosal pathogens, indicating that these modifications might have similar roles in multiple organisms. In this review, we relate genetic and structural data for the biosynthesis of glycosylation pathways in several mucosal-associated pathogens and describe what is known and/or predicted for the biological roles of these glycans. The review highlights the latest developments and excitement in the study of bacterial glycosylation pathways. Rather than providing an exhaustive literature review, which has been provided elsewhere[6–9], we focus on the best-studied example, that of *Campylobacter*. We also describe how studying glycosylation in bacterial systems may provide the opportunity for elucidating glycosylation processes in eukaryotic systems and how these systems could be exploited for glycoengineering.

C. jejuni: *N*- and *O*-linked glycosylation systems

C. jejuni is a human gut mucosal pathogen that is one of the main causes of bacterial gastroenteritis worldwide. Consequently, there has been much interest in studying the pathogenic mechanisms of this problematic organism. One of the most striking findings from these endeavours has been the identification and characterization of two glycosylation loci, which

Table 1a | **Summary of flagellar glycosylation locus in *Campylobacter jejuni* NCTC 11168**

Genes	Annotation	Additional information	Biological effect	References
Sugar biosynthesis				
Cj1293	*pseB*, sugar nucleotide epimerase/dehydratase	Orthologue of *H. pylori flaA1*	Loss of flagella assembly/motility in Cj 81-176; loss of Pse with replacement by PseAm/altered solubility in Cc VC167	5,31, 100,101
Cj1294	Aminotransferase			
Cj1311	*neuA2*, *N*-acetylneuraminic acid cytidylyltransferase		Loss of motility/Pse synthesis in Cj 81-176	3,31
Cj1316c	*pseA*		Loss of PseAm in Cj 81-176	3
Cj1317c	*neuB3*, *N*-acetylneuraminic acid synthetase		Loss of flagella assembly in Cj 81-176, 81116 and 11168; loss of Pse synthesis in Cj 81-176	3,31, 102,103
Cj1327	*neuB2*, *N*-acetylneuraminic acid synthetase, *ptmC*		Loss of flagella modification in Cj G1 but not 11168; still motile but involved in PseAm synthesis in Cc VC167	4,102
Cj1328	*neuC2*, putative *N*-acetyl-glucosamine-6-phosphate 2-epimerase/*N*-acetylglucosamine-6-phosphate, *ptmD*		Still motile but involved in PseAm synthesis in Cc VC167	4
Cj1329	Putative sugar–phosphate nucleotide transferase, *ptmE*		Still motile but involved in PseAm synthesis in Cc VC167	4
Cj1330	Unknown, *ptmF*		Still motile but involved in PseAm synthesis in Cc VC167	4
Cj1331	*ptmB*, *neuA3*, acylneuraminate cytidylyltransferase		Still motile but involved in PseAm synthesis in Cc VC167	4,13
Cj1332	*ptmA*, putative oxidoreductase		Still motile but involved in PseAm synthesis and protective immunity in Cc VC167	4,13
1318 family				
Cj1318	*maf1*	A polymorphic G tract	Loss of motility in Cj 11168	16
Cj1333, Cj1334	*maf2*, *maf3*		Loss of motility in Cj 11168	16
Cj1335–Cj1336	*maf4*	A polymorphic G tract that could allow translation into Cj1336	Loss of motility in Cj 11168	16
Cj1337	*maf5*		Loss of motility in Cj 11168	16
Cj1340c, Cj1341c	*maf6, maf7*			16

encode both *O*-linked and *N*-linked protein glycosylation pathways. The presence of two characterized glycosylation pathways makes campylobacters a useful model system for understanding bacterial protein glycosylation and for comparative studies with respective eukaryotic systems (see REF. 10 for a comprehensive review on this topic). This has been the driving force for the detailed genetic and structural characterization of these pathways in *Campylobacter*.

The* O*-linked flagellin glycosylation system. The *O*-linked FLAGELLIN glycosylation system was originally proposed based on the results of studies demonstrating that *Campylobacter* flagellins are sensitive to periodate oxidation and that they bind to a sialic-acid-specific lectin[11]. Subsequent studies, including determination of the genome sequence of *C. jejuni* strain NCTC 11168, revealed a flagellar glycosylation cluster of approximately 50 genes, including genes encoding the flagellin structural proteins FlaA and FlaB (TABLE 1). Annotation of the glycosylation locus suggested that approximately half of the genes could be involved in glycan biosynthesis. This, together with mutagenesis studies, has provided evidence for a role for many of these genes in either motility or flagellar glycosylation (TABLE 1). Multiple orthologues of the sialic-acid biosynthesis genes were identified and several of these — Cj1311 (*neuA2*), Cj1317 (*neuB3*), Cj1327 (*neuB2*), Cj1328 (*neuC2*) and Cj1331 (*neuA3*) — are located in the flagellar glycosylation locus[12]. Initially, these genes were implicated in sialic-acid biosynthesis[11,13,14]; however, recent structural analysis of the flagellins from *C. jejuni*[3] and the related species *Campylobacter coli*[4] has shown that flagellins are not modified with sialic acid, but with several monosaccharide analogues of the related sugar pseudaminic acid (Pse), which is a nine-carbon sugar and a member of the 5,7-diamino-3,5,7,9-tetradeoxynon-2-ulosonic acids (FIG. 1). In *C. jejuni* strain 81-176, glycosylation is confined to 19 serine and threonine residues along the central domain of the flagellin protein, the region known to be surface-exposed in the flagellar filament[3]. Thus, the *Campylobacter* flagella are extensively modified, with approximately 10% of the total FlaA glycoprotein

FLAGELLIN
The structural protein of which the bacterial flagellum is constructed.

Table 1b | **Summary of flagellar glycosylation locus in *Campylobacter jejuni* NCTC 11168**

Genes	Annotation	Additional information	Biological effect	References
617 family				
Cj1305c, Cj1306c	Unknown	Homopolymeric tracts		
Cj1310c	Unknown	Homopolymeric tracts		
Cj1342c	Unknown	Homopolymeric tracts		
Structural proteins				
Cj1338c	*flaB*, flagellin B		Less motile with normal flagella	104
Cj1339c	*flaA*, flagellin A		Non-motile with stubby flagella	104
Unknown				
Cj1295	Unknown			
Cj1296–Cj1297	Unknown	G(9) tract that could vary to allow translation into Cj1297		
Cj1298	Unknown			
Cj1299	*acpP2*, probable ACP			
Cj1300–Cj1302	Unknown			
Cj1303	*fabH2*, probable 3-oxoacyl-ACP synthase			
Cj1304	*acpP3*, probable ACP			
Cj1307	Putative amino-acid-activating enzyme			
Cj1308	*acpP4*, possible ACP			
Cj1309c	Unknown			
Cj1312, Cj1313	Possible flagellar protein			
Cj1314c	Probable cyclase			
Cj1315c	Probable amidotransferase			
Cj1319	Putative nucleotide sugar dehydratase			
Cj1320	Putative aminotransferase			
Cj1321	Putative transferase			
Cj1322–Cj1324	Unknown			
Cj1325–Cj1326	Unconfirmed	G(10) tract that could vary to allow translation into Cj1326		

ACP, acyl carrier protein; Cc, *Campylobacter coli*; Cj, *Campylobacter jejuni*; Pse, pseudaminic acid; PseAm, 5-acetamidino analogue of Pse.

mass consisting of *O*-linked carbohydrate and each filament containing approximately 20,000 protein subunits[15].

Many genes in the *O*-linked glycosylation locus of *C. jejuni* strain NCTC 11168 are hypothetical and are predicted to encode proteins that belong to two paralogous gene families (TABLE 1). One of these families, known as the 1318 motility accessory factor (1318/*maf*) family of flagellin-associated proteins, has seven members (Cj1318, Cj1333, Cj1334, Cj1335/Cj1336, Cj1337, Cj1340 and Cj1341). Several *maf* members are involved in motility variation by a slipped-strand mispairing of single nucleotide repeat sequences[16]. The second paralogous gene family (the 617 family, which is represented by Cj0617 and Cj0618) includes four genes in the glycosylation locus (Cj1305c, Cj1306c, Cj1310c and Cj1342c) that also contain intragenic single nucleotide repeats, indicating phase-variable gene expression. No homologues of this family are found in other bacterial species. The functions of the proteins encoded by the 617 and 1318/*maf* gene families are currently unknown. Perhaps most surprising is the lack of a characterized glycosyltransferase in the flagellar glycosylation locus, which contrasts with the *N*-linked glycosylation locus described below.

LIPOPOLYSACCHARIDE (LPS). An important amphiphilic molecule integrated in and extending outward from the outer membrane of the Gram-negative bacterial cell wall; structurally composed of hydrophobic Lipid A (responsible for endotoxin activity), core polysaccharide and hydrophilic O-antigen polysaccharide side chains.

The N-linked protein glycosylation system. The *N*-linked general glycosylation system was originally thought to be involved in LIPOPOLYSACCHARIDE (LPS) biosynthesis[17,18]. It was subsequently shown to have a role in the glycosylation of multiple *C. jejuni* proteins[19]. The genes from this 16-kb glycosylation locus were named *pgl* (Cj*pgl*) for protein glycosylation and are summarized in online supplementary information S1 (Table). We have speculated on the potential roles of the genes in the pathway on the basis of sequence similarities to genes that encode glycosyltransferases and enzymes that are required for sugar modification, as well as information obtained from the glycan structure. Mass spectrometry and NMR spectroscopy determined the glycan to be the heptasaccharide GalNAc-α1,4-GalNAc-α1,4-(Glcβ1,3)-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-β1,*N*-Asn, where Bac is bacillosamine (2,4-diacetamido-2,4,6-trideoxyglucose)[2] (FIG. 1).

PglB has significant amino acid similarity to STT3, which is found exclusively in eukaryotes and in archaea[1,19]. Both genetic and biochemical studies in yeast have shown that STT3 is an essential component of the nine-member oligosaccharyltransferase complex and is crucial for the proper functioning of all eukaryotic

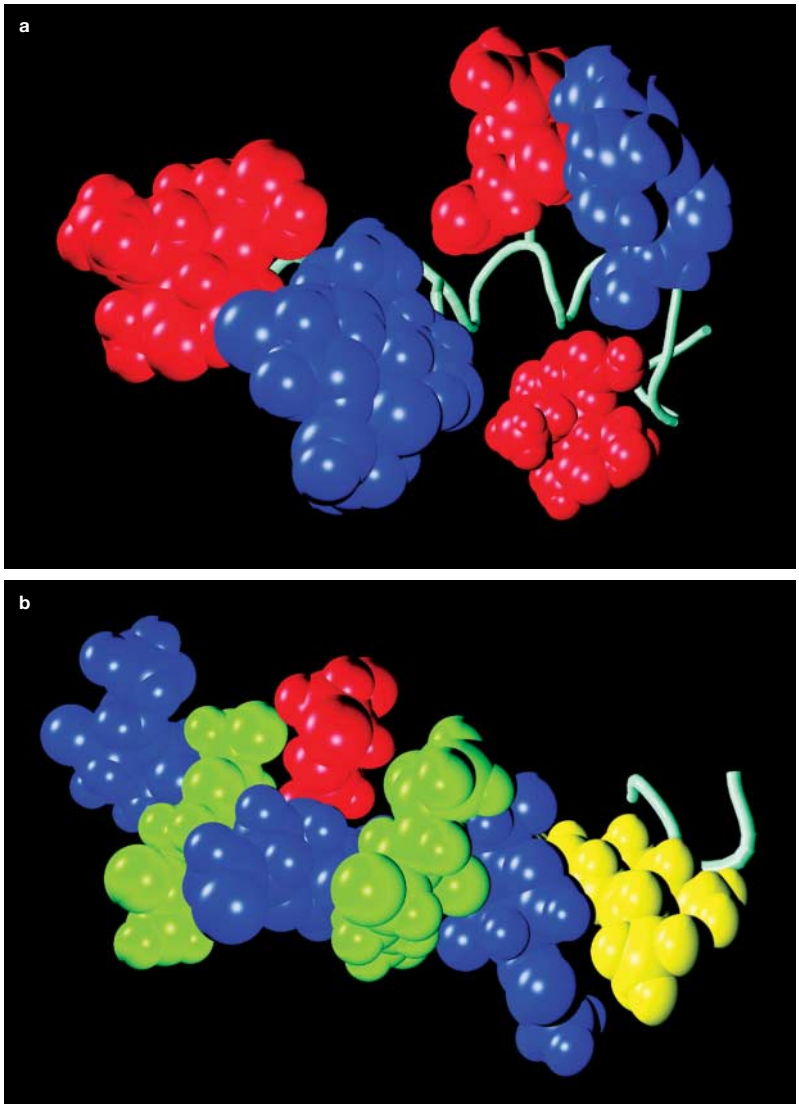

Figure 1 | **Space-filling models of the *C. jejuni* glycans from the *O*- and *N*-linked protein glycosylation systems. a** | Space-filling model of pseudaminic acid, Pse_5Ac_7Ac (red), and the dihydroxyproprionyl derivative, Pse_5Pr_7Pr (blue), decorating the FlaA peptide $_{392}$FTQNVSSISAFMSAQGSGF$_{410}$ at the underlined residues[3]. Note that the assignment of specific sugars to the modified residues is arbitrary. **b** | An *N*-linked heptasaccharide attached to a peptide with bacillosamine (yellow), five GalNAc residues (shown in alternating blue and green for clarity) and the branching glucose (red).

cells[20–23]. In eukaryotes, STT3 has been shown to have an important role in the oligosaccharyltransferase complex and is evolutionarily conserved[24]. STT3 orthologues, including PglB, contain a conserved carboxy-terminal catalytic motif (WWDYG), of which the second and third amino acids are essential[1,21,24]. *Neisseria* species, in which the pilin glycan is *O*-linked rather than *N*-linked, lack the STT3/PglB orthologue. On the basis of these data, we propose that *C. jejuni* PglB has a role in coupling the glycan to the asparagine residue to form the *N*-linked glycoprotein, which is a similar role to that of STT3 in the oligosaccharyltransferase complex (FIG. 2). However, some components of the *C. jejuni* Pgl pathway have amino acid similarity to components of the *O*-linked glycosylation pathways of *Neisseria meningitidis* pilin and other organisms that synthesize bacillosamine derivatives (see below). In our proposed model, UDP-HexNAc is converted to bacillosamine by sequential modification by PglF (dehydratase), PglE (aminotransferase) and PglD (acetyltransferase) in the cytoplasm. PglC then attaches the bacillosamine residue to a lipid carrier. In *N. meningitidis*, the PglA orthologue has been shown to attach the α-1,3-linked galactose moiety to diacetamido-trideoxy hexose (DATDH, a sugar that is similar to bacillosamine but for which the hexose identity has not been determined[25,26]). Therefore, we speculate that in *C. jejuni* PglA adds the α-1,3-linked *N*-acetylglucosamine (GalNAc) moiety to bacillosamine. It has recently been demonstrated that PglH is an α-1,4-GalNAc transferase[27]. PglH and PglJ have similar amino acid sequences, indicating that both might be involved in transferring the next four α-1,4-linked GalNAc moieties (FIG. 2). We propose that the final glycosyltransferase, PglI, adds the branching glucose moiety. Once assembled, it is proposed that the entire heptasaccharide is 'flipped' across the inner membrane into the periplasm by the ATP-binding cassette (ABC) transporter orthologue WlaB. This is analogous to the eukaryotic pathway in which the assembled sugars are flipped from the cytoplasm into the lumen of the endoplasmic reticulum (ER) (FIG. 2).

Mucosal-associated pathogens: common trends

Many of the pathogens that thrive within the mucosal linings of the human body have relatively small genomes and yet devote significant proportions of their genome content to glycoconjugate biosynthesis. Occasionally, the enzymes that are required for glycoprotein biosynthesis are shared between other polysaccharide biosynthetic pathways, such as the LIPOOLIGOSACCHARIDE (LOS) and capsule biosynthetic pathways, but often these pathways are separate without any apparent compartmentalization. Although bacteria are able to synthesize a greater variety of sugars than their eukaryotic counterparts (discussed below), many organisms express similar sugars with common functions. A summary of the mucosal pathogens described in this review is shown in TABLE 2.

Commonality of sugars. Bacteria are capable of synthesizing a variety of amino and deoxy sugars (reviewed in REF. 28). Bacillosamine (FIG. 1) derivatives are synthesized by *Bacillus*, *Neisseria*, *Streptococcus*, *Pseudomonas*, *Vibrio*, *Fusobacterium*, *Thiobacillus*, *Alteromonas*, *Pseudoalteromonas* and *Clostridium* species, as well as by *Campylobacter* and related organisms (see below). Comparison of the *Campylobacter* and *Neisseria pgl* loci indicates that most of the genes are involved in synthesis of bacillosamine or its derivative DATDH, which results from the conversion of an acetamido sugar by sequential dehydration, transamination and transacetylation reactions. Thus, the predicted orthologues of *C. jejuni* (PglDEF) and *Neisseria* (PglBCD) are common to many organisms that are capable of DATDH sugar biosynthesis. Pseudaminic acid derivatives are also common to *Campylobacter*, *Helicobacter* and *Pseudomonas* species.

LIPOOLIGOSACCHARIDE (LOS). Similar to LPS but lacking the O-antigen polysaccharide side chain repeats.

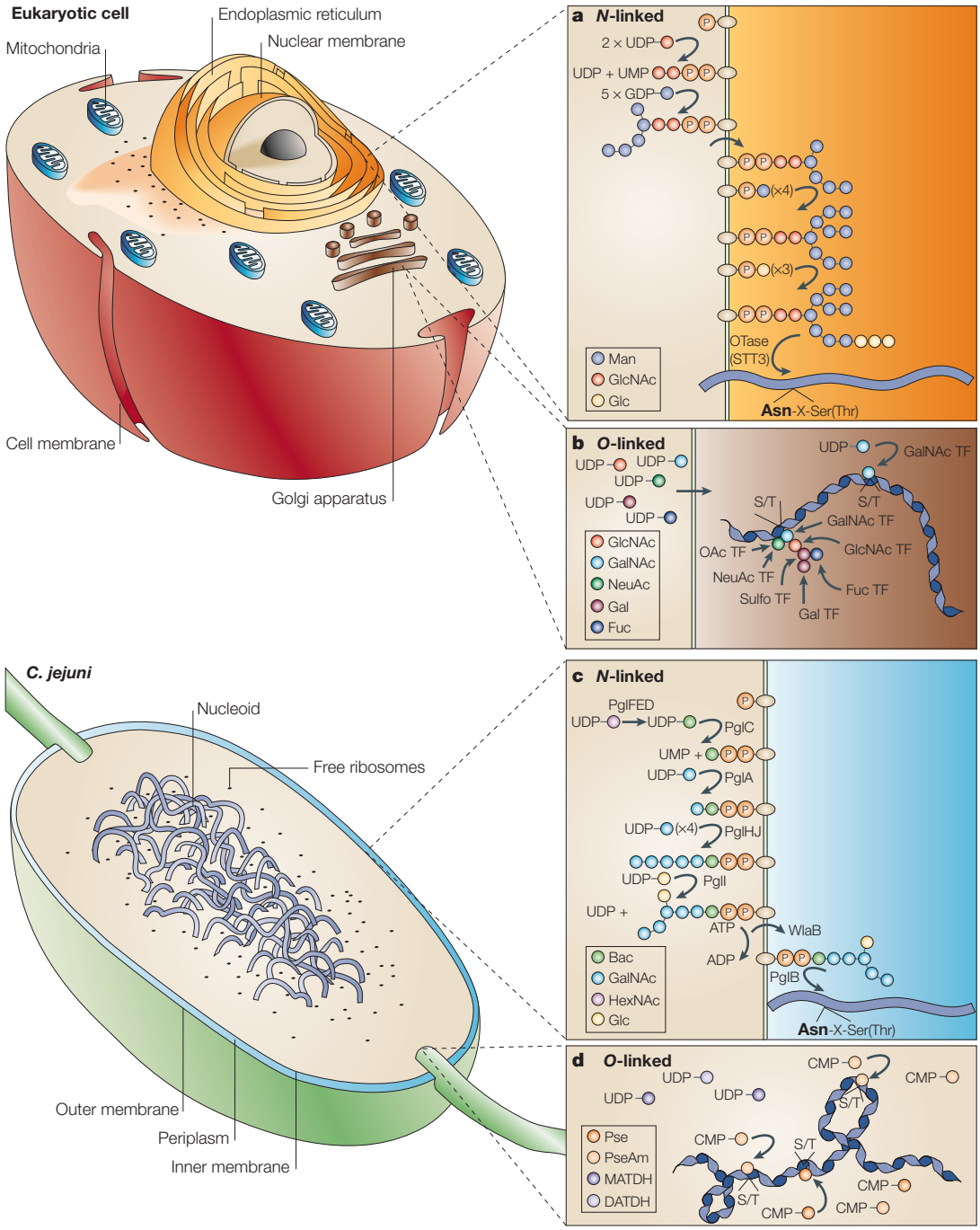

Figure 2 | **Model for the biosynthesis of *N*- and *O*-linked glycoproteins in eukaryotes (a,b) in comparison to bacteria (c,d).** In *C. jejuni*, *N*-linked glycosylation is proposed to proceed through the sequential addition of nucleotide-activated sugars onto a lipid carrier, resulting in the formation of a branched heptasaccharide (**c**). This glycan is then 'flipped' across the inner membrane into the periplasm by a putative ATP-binding cassette (ABC) transporter. These proposed steps are identical to the pathway that has been identified in eukaryotes (**a**), with the bacterial periplasm being functionally equivalent to the endoplasmic reticulum (ER). However, in eukaryotes the sugars are further processed to a 14-mer that is then transferred to the growing polypeptide by a complex of proteins collectively known as the oligosaccharyltransferase (OTase). One of the key members of this complex is STT3, the orthologue of the bacterial PglB. In *C. jejuni*, PglB is the only component that is necessary for transfer of the heptasaccharide to the asparagine residue (**c**). In eukaryotes, *N*-linked glycans are variable due to the trimming and processing reactions that further occur in the ER and Golgi (not shown), whereas the *N*-linked heptasaccharide in *C. jejuni* remains unmodified. The process of *O*-linked glycosylation in eukaryotes is less conserved, but generally proceeds through the step-wise transfer of nucleotide-activated monosaccharides to serine or threonine residues on proteins in the Golgi (**b**) that can then be further modified by processes such as sulphation or *O*-acetylation. However, reports have demonstrated that *O*-linked glycosylation can also occur in the ER and cytoplasm of eukaryotes[96–99] (not shown) and, in yeast, is initiated in the ER with the addition of a single mannose (Man) from a lipid-linked carrier rather than nucleotide[98]. As the bipolar flagella of *C. jejuni* span both the inner and outer membranes, *O*-linked glycosylation of flagellin monomers is proposed to occur in the cytoplasm/inner membrane where nucleotide-activated sugars are individually added to serine or threonine residues that are surface exposed (**d**). Bac, 2,4-diacetamido-2,4,6-trideoxyglucose; DATDH, diacetamido-trideoxyhexose; Fuc, fucose; Gal, galactose; GalNAc, *N*-acetylgalactosamine; Glc, glucose; GlcNAc, *N*-acetylglucosamine; HexNAc, *N*-acetylhexosamine; MATDH, monoacetamido-trideoxyhexose; NeuAc, *N*-acetylneuraminic acid; Pse, 5,7-diamino-3,5,7,9-tetradeoxynon-2-ulosonic acid (Pse_5Ac_7Ac); PseAm, 5-acetamidino analogue of Pse (Pse_5Am_7Ac); TF, transferase.

Table 2 | **Summary of discussed bacterial protein glycosylation systems**

Organism	Protein	Glycan structure	Var	Link	Sites	Sugars per site	Transfer of sugars	Biological effects
Campylobacter jejuni	Multiple, secreted	$GalNAc_2(Glc)GalNAc_3Bac$	No	*N**	1–3	7	Block	T4SS assembly/competence, animal colonization, adherence/invasion, protein antigenicity
Campylobacter coli VC167	FlaA flagellin	Pseudaminic acid, acetamidino, deoxypentose + Pse/PseAm	Yes	*O*	16	1–2	Sequential	Flagella assembly, loss of motility, protein antigenicity
Campylobacter jejuni 81-176	FlaA flagellin	Pseudaminic acid, acetamidino, *O*-acetylacetamidino, dihydroxyproprionyl	Yes	*O*	19	1	Sequential	Flagella assembly, loss of motility, protein antigenicity
Helicobacter pylori	FlaA, FlaB	Pseudaminic acid	No	*O*	7,10	1	Sequential	Flagella assembly, loss of motility
Neisseria meningitidis	Pilin	$Gal_{1\text{-}2}DATDH$	Yes	*O*	1	2–3	Sequential	Complement-mediated killing
Neisseria gonorrhoeae	Pilin	$Gal_{1\text{-}2}DATDH$	Yes	*O*	1	2–3	Sequential	Correlation with disease
Pseudomonas aeruginosa PAK	Flagellin	Up to 11 sugars linked through rhamnose including 4-amino 4,6-dideoxyhexose viosamine	Yes	*O*	2	Various	Sequential	Unknown
Pseudomonas aeruginosa JJ692	Flagellin	Rhamnose	–	*O*	2	1	Sequential	unknown
Pseudomonas aeruginosa 1244	Pilin	Hydroxybutyryl-formyl-pseudaminic acid with xylose and *N*-acetylfucosamine	–	*O*	1	3	Block	Unknown
Escherichia coli H10407	TibA adhesin	Heptoses	–	–	–	–	–	Adherence/invasion
Escherichia coli 2787	AIDA-I adhesin	Heptoses	–	*O*	19	1	Sequential	Adherence
Haemophilus influenzae	HMW1 adhesin	Gal, Glc, Man	–	–	–	–	–	Protein stability/adherence

Link, type of glycan linkage to protein; Sites, number of sites per protein modified; Sugars per site, number of sugars per modified site; Var, variability in glycan structure. *Although this is the only *N*-linked glycosylation system that has been characterized, several homologous Pgl systems have recently been identified in *C. coli*, *C. upsaliensis*, *C. lari*, *W. succinogenes* and *D. desulfuricans* (see FIG. 3). AIDA-I, adhesin involved in diffuse adherence; Bac, bacillosamine; DADTH, diacetamido-trideoxy hexose; Gal, galactose; Glc, glucose; HMW1, high-molecular-weight glycoprotein adhesin; Man, mannose; PseAm, 5-acetamidino analogue of pseudaminic acid; T4SS, type IV secretion system.

In *Pseudomonas aeruginosa* strain 1244, the pilin glycan contains a serine-linked trisaccharide — hydroxybutyryl-formyl-pseudaminic acid linked to xylose and *N*-acetyl-fucosamine[29]. Interestingly, this structure is also found in *P. aeruginosa* LPS[29]. In *Helicobacter pylori*, both FlaA and FlaB flagellin subunits are modified with pseudaminic acid; however, unlike *C. jejuni*, none of the other pseudaminic acid derivatives are observed[30]. Novel methods such as CAPILLARY ELECTROPHORESIS COUPLED TO ELECTROSPRAY MASS SPECTROMETRY (CE-ESMS) with precursor ion scanning have been instrumental in probing bacterial cell lysates for sugar nucleotide precursors[31]. This technique has enabled the identification of pseudaminic acid nucleotide metabolites in *H. pylori* — for example, CMP-Pse, UDP-monodiacetamido-trideoxyhexose and UDP-diacetamido-trideoxyhexose[31]. CE-ESMS studies can facilitate the elucidation of carbohydrate biosynthetic processes and assist in the functional characterization of unknown gene products such as those encoded by the *Campylobacter* flagellar gene locus (TABLE 1).

Sharing enzymes from different biosynthetic pathways. To maintain compact genomes and avoid redundancy, many organisms have evolved to use the same biosynthetic machinery for different carbohydrate pathways. Kneidinger *et al.* recently identified a pathway for GDP-heptose biosynthesis that is used to glycosylate the *Aneurinibacillus thermoaerophilus* S-LAYER PROTEIN[32]. *C. jejuni* expresses both ADP- and GDP-heptose pathways for modification of LOS and CAPSULAR POLYSACCHARIDE (CPS); however, both pathways share a common phosphatase[33]. Mutation of the *C. jejuni* phosphatase results in loss of heptose from both LOS and CPS. Even more resourceful is the example of GalE, which is a bifunctional epimerase that can convert glucose (Glc) to galactose (Gal) and *N*-acetylglucosamine (GlcNAc) to GalNAc[34]. As expected, inactivation of *C. jejuni galE* results in truncation of LOS[34,35]. However, the concomitant inhibition of GalNAc conversion also results in loss of both the capsular polysaccharide and *N*-linked glycoprotein modifications[34]. Thus, one enzyme affects three important pathways for glycoconjugate biosynthesis. In *N. meningitidis*, *galE* inactivation has also been shown to cause truncation of both the pilin trisaccharide Gal(β1-4)-Gal(α1-3)-2,4-diacetamido-2,4,6-trideoxyhexose and the LOS outer core[36]. Fischer and Haas have recently suggested that RecA from *H. pylori* is glycosylated and have shown that inactivation of *galE* and *pmi* (which encodes a bifunctional mannose-6-phosphate isomerase/GDP-mannose pyrophosphorylase), both of which are involved in LPS biosynthesis, affects

CAPILLARY ELECTROPHORESIS COUPLED TO ELECTROSPRAY MASS SPECTROMETRY (CE-ESMS). A coupled system where complex mixtures are first separated by capillary electrophoresis before being introduced into the mass spectrometer using electrospray ionization.

S-LAYER PROTEIN Surface (S) layers are composed of a crystalline array of high-molecular-weight protein or glycoprotein subunits and form a matrix surrounding some bacterial cells.

CAPSULAR POLYSACCHARIDE (CPS). Also sometimes referred to as K-antigen, capsules are composed of polysaccharide repeats that surround some bacterial cells and are anchored in the membrane by a terminal lipid moiety.

migration of RecA by gel electrophoresis[37]. Other examples of overlapping pathways include the *Haemophilus influenzae* phosphoglucomutase, which affects both LOS and the high-molecular-weight glycoprotein adhesin (HMW1)[38], and the *Escherichia coli* bifunctional ADP-heptose kinase/nucleotidyltransferase (HldE), which provides heptose for LPS, and, in some enterotoxigenic *E. coli* (ETEC), the adhesin involved in diffuse adherence (AIDA-I)[39]. In *P. aeruginosa* strain 1244, structural similarity between the pilin glycan and the O-ANTIGEN also indicates that these two pathways share common metabolites and biosynthetic enzymes[40] (see above). Indeed, mutation of two genes that are essential for the first steps of O-antigen biosynthesis (*wbpM* and *wbpL*) resulted in loss of both LPS O-antigen and pilin glycan in this organism.

Genetic variation in* O*-linked systems. Pathogens often avoid the mucosal immune response by two common mechanisms of genetic variation: the presence of nucleotide repeats in selected genes, which leads to FRAMESHIFT MUTATIONS, or gene polymorphisms in the glycosylation clusters, which results in changes in the composition of biosynthetic pathways. The neisserial type IV pilin glycosylation pathway takes advantage of both mechanisms for variation. The gene encoding the galactosyltransferase PglA, which is responsible for the addition of α1-3Gal to the *N. meningitidis* pilin, contains a homopolymeric G tract. This is in contrast to the *C. jejuni pglA* gene from the *N*-linked pathway, which lacks the homopolymeric tract and is invariant. In addition, the neisserial *pglG* and *pglH* genes contain polyC tracts, the *pglB* gene contains a polyA tract and the *pglE* gene contains a heptanucleotide repeat. Variation in the *pglE* gene allows the terminal β1-4Gal to undergo PHASE VARIATION, resulting in conversion from a trisaccharide to a disaccharide structure[6]. Kahler *et al.* have also described polymorphisms in the *pgl* locus of *N. meningitidis* and *Neisseria gonnorhoea*, identifying gene fusions, deletions and insertions in the respective loci[41]. Again, this is unlike the *N*-linked pathways of *Campylobacter* species, in which the corresponding *pgl* genes are conserved in both sequence and co-linear chromosomal arrangement and show few differences across species (see below). However, similar to *Neisseria* spp., *P. aeruginosa* also has both sequence and gene polymorphisms within the glycosylation islands[42]. The authors speculate that although this organism has few different flagellin types, it might compensate by acquiring the ability to further diversify this antigenic surface protein by glycosylation. Microarray comparison did not show an absolute correlation between the site of infection and the composition of the glycosylation island; however, general trends were observed, indicating that different glycans can be expressed in different hosts or environments, which could provide them with specific survival advantages. For example, in *Pseudomonas syringae*, modification of the glycans on the flagella determines plant host specificity[43].

Extensive polymorphism has also been demonstrated in the *Campylobacter O*-linked glycosylation clusters, indicating that there is a selective pressure for bacteria to alter their surface structures[10]. This is further emphasized by comparison of the carbohydrate structures on the flagella of *C. jejuni* 81-176 and *C. coli* VC167. In addition to decorating their flagella with pseudaminic acid, these strains produce modifications that are unique to each strain. Furthermore, although both strains express an acetamidino variant of pseudaminic acid (PseAm), the PseAm structures that are produced differ, as shown by mass spectrometry and antibody reactivity[4]. In addition, the locus involved in PseAm biosynthesis also differs between these two species[4]. These reports for *O*-linked glycan variation on surface-exposed structures are in contrast to the presence of pseudaminic acid moieties on FlaA and FlaB monomers on the *H. pylori* sheathed flagella[30] and in gene clusters and structures observed in bacteria expressing *N*-linked glycoproteins (which are predominantly located in the periplasm, see below).

Common features of* N*-linked glycosylation pathways in bacteria. Sequencing, microarray comparisons and HIGH-RESOLUTION MAGIC-ANGLE SPINNING NMR (HR-MAS NMR) studies have demonstrated that both the *pgl* genes and the *N*-linked heptasaccharide structure are conserved in all *C. jejuni* and *C. coli* strains that have been examined[19,44–48] (FIG. 3). Recently, genome data for more distantly related *Campylobacter* species have become available[49]. Examination of genome sequence data from *C. jejuni* RM1221, *C. coli* RM2228, *Campylobacter lari* RM2100 and *Campylobacter upsaliensis* RM3195 demonstrated that *pgl* gene homologues are highly conserved, co-linear and clustered (FIG. 3). In addition, all clusters encode the essential oligosaccharyltransferase PglB (57–99% amino acid sequence identity; online supplementary information S1 (Table)), all of which contain the conserved catalytic motif WWDYG (REF. 49; W. Miller, personal communication). All these genomes lack the *wlaJ* insertion (probable integral membrane protein) found in *C. jejuni* NCTC 11168, which was not required in *E. coli* to generate recombinant *N*-linked glycoproteins[1]. In addition, the *C. coli* sequence lacks a *pglG* homologue. *C. lari* showed similar results to *C. coli*, but *C. lari* also lacks the putative glucosyltransferase gene *pglI*. *C. upsaliensis* contains all *pgl* homologues in the same order as *C. jejuni*; however, the *pglEFG* homologues have been separated from the other members of the pathway by an insertion of 14 genes.

Recently, Baar *et al.* identified another *pgl* gene cluster in the related ε-proteobacteria *Wolinella succinogenes*[50] (FIG. 3). Remarkably, the locus is similar to those found in the campylobacters. However, in contrast to *C. jejuni*, the *W. succinogenes pgl* cluster contains two separate gene insertions, has inverted the order of *pglH* and *pglI*, and has retained *galE*, *wlaA* and *pglG* homologues elsewhere on the chromosome. However, unlike the organisms with *N*-linked glycosylation pathways described above, the glycosylation genes in the draft genome of the δ-proteobacterium *Desulfovibrio desulfuricans* are not

O-ANTIGEN
A polysaccharide antigen extending from the outer membrane of some Gram-negative bacterial cell walls that is a part of the LPS; this repeat is also the serodeterminant for the classical heat-stable O-typing scheme.

FRAMESHIFT MUTATION
A mutation arising from the loss or gain of a base or DNA segment leading to a change in the codon reading frame and thus a change in the amino acids incorporated into the protein.

PHASE VARIATION
Variable expression of a structure that is governed by random frameshift mutations within genes responsible for the biosynthesis of the structure or through changes in the regulation of structure synthesis.

HIGH-RESOLUTION MAGIC-ANGLE SPINNING NMR (HR-MAS NMR). An adaptation of solid-state NMR where samples are spun rapidly around an axis inclined at an angle of 54.7°, the 'magic angle', with the direction of the magnetic field.

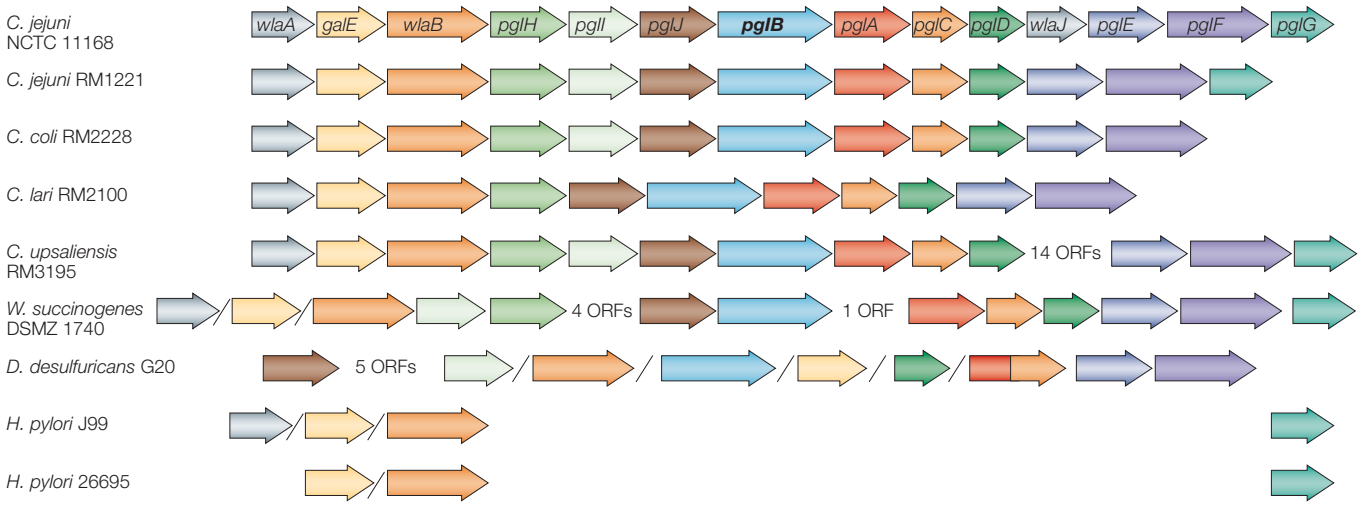

Figure 3 | **Gene schematic comparing bacterial *N*-linked protein glycosylation loci.** Schematic representation of the conserved *N*-linked protein glycosylation gene clusters in representative δ- and ε-proteobacteria. The gene encoding the essential oligosaccharyltransferase PglB is shown in bold. Slashed lines indicate homologues that are found elsewhere on the chromosome. Gene homologues for the *N*-linked pathway in the genomes of *Helicobacter hepaticus* and *Helicobacter mustelae* are similar to those shown for *H. pylori*. In *H. pylori*, *jhp0139* (in strain J99) and *Hp0151* (in strain 26695) show limited homology to Cj*pglB*, but translation of the genes indicates an absence of the catalytic domain WWDYG that is conserved in PglB orthologues in all three domains of life (see text). The CampyDB database for *Campylobacter* genome analysis created by Roy Chaudhuri and Mark Pallen (see *Campy*DB in the Online links box) was used to generate the gene schematic shown. The genome numbers for the proteobacterial homologues, together with the protein identity relative to *C. jejuni* NCTC 11168, are summarized in online supplementary information S1 (Table).

clustered (see *Desulfovibrio desulfuricans* genome in the Online links box). Also in contrast to the organisms described above, the catalytic motif in the oligosaccharyltransferase in both *D. desulfuricans* and *Desulfovibrio vulgaris* contains one amino acid substitution — from WWDYG to WWDWG. Interestingly, in *Campylobacter* the *pglA*, *pglC* and *pglD* genes are separate open reading frames (ORFs) in a linear arrangement, whereas in *D. desulfuricans pglA* and *pglC* are fused and in *Neisseria* species *pglC* and *pglD* have been fused. The gene homologues whose roles in the *N*-linked glycosylation pathway remain to be determined — *wlaA*, *wlaJ* and *pglG* — are not present in the *D. desulfuricans* genome.

There are reports in the literature that other bacteria can synthesize *N*-linked sugars (see summary in REF. 51); however, further studies are required to demonstrate that these are genuine *N*-linked sugars and that the organisms express an orthologue of the oligosaccharyltransferase that has been shown to be conserved throughout all three domains of life.

Biological roles of microbial protein glycosylation

Much of our knowledge on microbial protein glycosylation has been obtained from studies on S-layers of archaea and bacteria. During the past decade, glycosylated surface appendages such as flagella and pili have been described for bacteria[6–9]. As many of the described bacterial glycoproteins are surface-exposed, these modified proteins might have important roles in pathogenicity. Recent studies have shown that protein glycosylation can have roles in adhesion[52–54], protection against proteolytic cleavage[55], solubility[56], protein assembly[5,38], antigenic variation[11] and protective immunity[13,57].

LIPOCHITOOLIGOSACCHARIDE Plant bacteria signalling molecules, also known as Nod factors, that consist of a backbone of 2–6 β-(1→4)-linked GlcNAc residues with an amide-bound fatty acyl residue (saturated or unsaturated) on the non-reducing terminal GlcN residue. This basic structure has variations that are dependent on each strain or species and determine the host specificity.

Roles for *O*-linked glycosylation

Host interactions. Recent studies have shown that the *Campylobacter* flagellar glycosylation loci are extremely variable, contain numerous phase-variable genes and encode enzymes that are involved in the biosynthesis of a variety of complex carbohydrates. The potential for generating alternative glycoforms on the flagellin protein due to *O*-linked glycosylation suggests a mechanism for antigenic diversity and immune evasion for this surface-exposed and immunodominant protein. In *N. gonorrhoeae*, glycoform variation on the pilin is due in part to *pgtA*, which is homologous to *C. jejuni pglA* from the *N*-linked pathway. It has been shown that many *N. gonorrhoeae* disseminated strains contain a phase-variable version of *pgtA*, whereas uncomplicated isolates predominantly lack a variable version[58]. This may be due to the observation that in *N. meningitidis* anti-Gal IgA antibodies bind to the neisserial pilus, blocking complement-mediated killing and potentially promoting disease[59]. So, strains with the ability to modify a structure that may be beneficial during one stage of pathogenesis but unnecessary or detrimental for another stage of their life cycle might be associated with different clinical outcomes. This is reminiscent of phosphocholine (ChoP) expression in multiple mucosal pathogens, including *Neisseria*. ChoP increases neisserial adherence and invasion of human epithelial cells, but also causes increased susceptibility to serum killing[60]. Interestingly, this modification is added to the pili of pathogenic *Neisseria* and to the LPS of commensal *Neisseria*[60]. Alternatively, the host specificity of plant pathogens is dependent on the variable sugar decorations found on their LIPOCHITOOLIGOSACCHARIDES[61].

Takeuchi *et al.* recently provided the first example of variable glycan expression on a glycoprotein affecting host recognition of the modified flagella of *P. syringae*[43]. This observation could also have implications for *P. aeruginosa* strains in which the polymorphic glycosylation island (see above) resembles part of the flagellin glycosylation island found in *P. syringae*[43].

In the first report of a naturally occurring glycosylated protein from *E. coli*, Lindenthal and Elsinghorst demonstrated that ETEC strain H10407 expresses a 104-kDa outer-membrane glycoprotein (TibA) that has been directly correlated with adherence and invasion of human epithelial cells[53,62]. Glycan detection with periodate suggested that TibA is the only glycoprotein produced by H10407. TibA has amino acid identity with AIDA-I from diffuse-adhering *E. coli* (26% identity) and with the pertactin precursor from *Bordetella pertussis* (24% identity). Analysis of the *tib* locus revealed an ORF, *tibC*, encoding an enzyme with similarity to the glycosyltransferase proposed to be responsible for TibA modification. Recently, the plasmid-encoded autotransproter adhesin, which is encoded by *aidA* on the plasmid of *E. coli* ETEC strain 2787, has also been shown to be glycosylated[39]. Carbohydrate-composition analysis indicates that this protein is modified with approximately 19 heptose residues per protein. Upstream of *aidA* is a gene with 70% sequence similarity to *tibC* of ETEC H10407 and to various heptosyltransferases, and which has been annotated *aah* (autotransporter adhesin heptosyltransferase). Mutation of *aah* results in reduced AIDA-I mass and loss of function, indicating that glycosylation is essential for adherence. This glycosylation modification is absent from most *E. coli* strains and is indicative of a specific adaptation of selected pathogenic *E. coli* such as ETEC. Heptoses are commonly found on bacterial LPS and CPS[33], but have only recently been shown to be attached to archaeal proteins in the S-layer of *A. thermoaerophilus*[32]. So, AIDA-I modification is the first example of heptoses being attached to bacterial proteins. Interestingly, the nucleotide-activated heptose precursors are shared with the LPS pathway (see above), yet mutations in the heptosyltransferases I and II, which are encoded by *waaC* and *waaF*, respectively, did not affect glycosylation, further indicating that the *aidA* gene product is the heptosyltransferase required for protein modification. Moormann *et al.* have demonstrated that recombinant TibC proteins can substitute for Aah in introducing the heptosyl modification to AIDA-I and that this modification can restore the adhesive function of AIDA-I[63]. So, like Aah, TibC is another example of a novel class of heptosyltransferases that specifically transfer heptose residues onto serines at multiple sites along the protein backbone.

Complex protein assembly. In *Campylobacter*, it was recently demonstrated that *O*-linked flagellar glycosylation is necessary for the formation of the filament and therefore for motility[4,5]. *O*-linked flagellar glycosylation is not unique to *Campylobacter*. Although the extensive flagellar glycosylation locus has no counterpart in the enterobacterial system, evidence in the literature indicates that flagellar glycosylation might be common to a number of organisms with polar flagella[64–68]. Similar genes from the *Campylobacter* flagellar glycosylation locus (TABLE 1) are found in *Helicobacter*, *Aeromonas*, *Clostridium*, *Caulobacter* and *Methanococcus* species. Schirm *et al.* recently demonstrated that the flagella of *Helicobacter* are also glycosylated with pseudaminic acid[30] (see above). Mutation of putative glycosylation genes in *H. pylori* resulted in a non-motile phenotype and loss of flagellar filament assembly, which indicates that, like *Campylobacter*, glycosylation is required for complex protein assembly[30]. By contrast, the absence of flagellar glycosylation in *P. aeruginosa* type A flagellins does not cause loss of flagellar filament assembly or motility[69]. However, the authors note that the flagella of *Campylobacter* and *H. pylori* are more complex, requiring two proteins, FlaA and FlaB, and having a higher level of glycosylation per monomer, indicating potentially different roles for the flagellar glycans[69] (see above). In *Haemophilus*, Grass *et al.* have demonstrated that glycosylation of HMW1 is required for stability and for 'tethering' the protein onto the cell surface[38], which is a requirement for HMW1-mediated adherence. Thus in this study, glycosylation seems to protect against premature degradation, similar to one of the many roles of glycans in eukaryotic glycoproteins.

Roles for *N*-linked glycosylation

Unlike loci that are involved in pilin and flagellin biosynthesis, the *N*-linked glycosylation locus is highly conserved among the Proteobacteria[19,44,45,47,48] and lacks putative phase-variable genes. Disruption of *N*-linked glycosylation in *Campylobacter* results in cells that have a reduced ability to adhere to and invade *in vitro* cultured eukaryotic cells and a reduced ability to colonize the intestines of chickens and mice[70–73]. However, these results are difficult to interpret due to alterations in the glycosylation of several proteins, leading to multiple pleiotropic effects. Recently, Larsen *et al.* demonstrated a specific function for the *N*-linked glycan in complex protein assembly[74]. In *C. jejuni* 81-176, the type IV secretion system (T4SS) contains at least one glycoprotein, VirB10. Amino acid substitution experiments demonstrated that VirB10 is modified at two Asn-X-Ser/Thr sites, but that modification of only one residue, Asn97, results in the loss of T4SS assembly and competence. Examination of the identified *N*-linked *C. jejuni* glycoproteins indicates that many of these proteins can also exist in complexes — for example, KpsD, which is a putative capsule transport protein[2,75], and AcrA, which is a multidrug efflux pump component[1,76]. It is of note that *N*-linked glycosylation takes place in the periplasm so that only proteins secreted across the inner membrane will become *N*-glycosylated. Indeed, nearly all *N*-linked glycoproteins that have been identified so far contain predicted signal peptides, indicating they have an extracytoplasmic location. Thus, possible functions of *N*-linked glycosylation include protection against proteolytic cleavage, enhancement of protein stability or signals for cellular sorting similar to eukaryotic *N*-linked glycans[56,77].

Comparisons with archaea and eukaryotes

Although both archaea and bacteria lack the cellular organelles that are found in eukaryotes, similar processes for glycoconjugate biosynthesis have been conserved. These include the segregation of important steps in the respective pathways by cellular membranes, the use of nucleotide-activated sugars and/or lipid-bound intermediates, and the transfer of sugars to similar amino acid sequons.

As mentioned above, the *N*-linked glycosylation pathway in *C. jejuni* is proposed to be located at the inner membrane. Nucleotide-activated sugars from the cytoplasm are sequentially assembled onto a lipid carrier, presumably undecaprenylpyrophosphate, which is anchored in the inner membrane. The heptasaccharide is then 'flipped' as a block across the membrane into the periplasm. This assembly process is similar to other bacterial glycoconjugate biosynthetic pathways, such as those described for CPS and LPS. The transfer of the heptasaccharide onto the targeted protein is proposed to occur in the periplasm for two reasons — the proteins that are modified by this pathway are secreted across the inner membrane and the conserved catalytic domain of the oligosaccharyltransferase PglB is located in the periplasmic space[78]. Thus, the bacterial periplasm is equivalent to the eukaryotic ER during *N*-linked protein glycosylation.

The most common modification of secretory proteins in eukaryotes is *N*-linked protein glycosylation. In this well-characterized system, nucleotide-activated sugars are added to a dolichol pyrophosphate carrier and then flipped across the ER membrane. Interestingly, at this point in both the bacterial and eukaryotic systems, the glycan is a lipid-linked heptasaccharide containing one branching sugar. However, in eukaryotes, the heptasaccharide is further processed to a tetradecasaccharide ($Glc_3Man_9GlcNAc_2$). In the central step of the process, the oligosaccharyltransferase complex transfers an assembled oligosaccharide as a block from the lipid carrier to asparagine residues in the emerging polypeptide. This stage requires the conserved PglB orthologue STT3 (FIG. 2). The glycans that are now covalently attached to proteins are further modified in the ER and Golgi complex by trimming, further glycosylation, sulphation and epimerization. By contast, *C. jejuni* PglB seems to be the only component that is necessary for transfer of the glycan to the common *N*-linked sequon and there is no evidence for further processing of the heptasaccharide.

In archaea, *N*-linked glycan synthesis proceeds through lipid-linked intermediates on the extracellular cell surface and further trimming reactions have also been described[79]. Interestingly, reports have demonstrated the use of both polyprenyl carriers and dolichol phosphate carriers in *Haloferax volcanii* for glycan synthesis[80,81]. The *N*-linked glycosylation sequon recognized by the oligosaccharyltransferase — Asn-X-Ser/Thr — is conserved in eukaryotes, bacteria and archaea. However, in archaea, the sequence requirement is not as strict as that in eukaryotes and bacteria because replacement of the serine residue in one glycosylation site in *Halobacterium halobium* did not prevent *N*-glycosylation of the S-layer, although alteration of a second site did[82]. In eukaryotes, not every sequon is glycosylated by STT3, but rather only those polypeptides containing the proposed Asx turn[83]. It remains to be determined whether bacterial and archaeal glycosylation is also restricted by the recognition of protein structural domains. Also, in eukaryotes, the range of sugars that can be linked to Asn through a GlcNAc moiety is defined, whereas bacteria and archaea are capable of synthesizing glycoproteins with multiple glycan structures and linkage types, resulting in a highly diverse assortment of glycoproteins. By contrast, Schirm *et al.* recently demonstrated that the peritrichous flagella of *Listeria monocytogenes* were glycosylated with six *O*-linked β-GlcNAc moieties[84], a modification common to numerous eukaryotic nuclear and cytoplasmic proteins[85].

In eukaryotes, *O*-linked glycosylation occurs sequentially at serine or threonine residues, where the product of one glycosyltransferase becomes the acceptor for the next glycosyltransferase and nucleotide-activated sugar (FIG. 2). This is similar to the *O*-linked process in bacteria such as *N. meningitidis*, where disruption of *galE* prevents incorporation of the digalactose moieties but does not affect addition of DATDH to the pilins[36]. By contrast, *N*-linked glycosylation proceeds by assembly of sugars onto lipid anchors and their subsequent transfer as blocks. So, disruption of *N*-linked biosynthetic pathways results in proteins that are not modified. The process of *O*-linked glycosylation occurs mainly in the Golgi complex in eukaryotes, whereas in archaea the plasma membrane must perform many functions that are normally carried out by the ER and Golgi. In bacteria, *O*-linked glycosylation is proposed to occur in the cytoplasm or at the interface between the cytoplasm and surface appendages such as pili and flagella (FIG. 2). For example, the glycosyltransferase HMW1C, which is necessary for modifying the HMW1 adhesin in *Haemophilus*, lacks a signal sequence yet influences the glycosylation and secretion of the adhesin[38]. To demonstrate that HMW1C interacts with and modifies HMW1 in the cytoplasm, Grass *et al.* created a HMW1 fusion protein and expressed the protein in *E. coli* in the presence of HMW1C. They then purified HMW1 from *E. coli* by affinity chromatography and demonstrated that the adhesin was glycosylated. This not only confirmed that the interaction took place in the cytoplasm, but the study also demonstrated that *E. coli* produces glycans that can be recognized by HMW1 as additional genes required for glycosylation in *Haemophilus* were not expressed in *E. coli*.

Possible exploitation of bacterial glycosylation

The importance of *N*-glycosylation in humans is demonstrated in rare human congenital disorders of glycosylation where mutations affecting different parts of the biosynthetic pathway have pleiotropic effects. Patients with these disorders suffer a range of clinical symptoms, including variable degrees of neurological

alterations, liver dysfunction and intestinal disorders[86]. Several elegant studies have mapped these mutations in humans and in simple eukaryotic systems, such as *Saccharomyces cerevisiae*, thereby allowing the glycosylation pathway to be elucidated (FIG. 2). Remarkably, this knowledge has led to the development of dietary mannose supplements as an effective treatment in some patients to compensate for deficiencies in these pathways[87,88]. The discovery of the *C. jejuni* Pgl pathway and the recent demonstration that this system can be functionally cloned into *E. coli*[1] might provide the opportunity to study some of these glycosylation processes in a comparatively simple organism such as *E. coli* (and *C. jejuni*), and should provide a useful model for understanding the function and biosynthetic mechanism of more complex glycosylation systems in higher organisms. Of particular significance is the observation that the important *C. jejuni* oligosaccharyltransferase PglB can be mutated without being lethal to the *E. coli* or *C. jejuni* cell. By contrast, STT3 (the PglB orthologue) is essential for viability in eukaryotes such as *S. cerevisiae*. Thus, a more detailed characterization of PglB and the respective glycosylation pathway is possible.

Many proteins in biological systems are glycosylated. Comparisons of well-characterized protein sequence database entries indicate that more than half of all proteins in nature will eventually be identified as glycoproteins[89]. It is also well known that glycosylation is the fine-tuning mechanism that is used to determine the precise function of a protein[90]. The field of glycobiology has been hampered by the inability to clone general glycosylation systems into a simple host such as *E. coli* and by the resulting recombinant proteins that lack the suitable modifications[91]. This could now change. The recent finding that the heterologous *Campylobacter N*-linked glycosylation cluster can glycosylate proteins in *E. coli* is important. This could provide new possibilities to produce recombinant glycoproteins and to engineer numerous permutations of sugar structures, thereby forming the basis of a new era of glycoengineering. Glycoengineering — the synthesis of novel glycan structures — is essential for studies to elucidate the crucial role of glycosylation in a myriad of biological phenomena. This might also be important for the biotechnological synthesis of novel glycoconjugates and potential immunostimulating agents. For example, *in vitro* galactosylation and sialylation of therapeutic glycoproteins have been suggested to increase serum half-life[92], and increasing the amount of glycosylation on proteins that are used as therapeutics has been shown to increase and prolong *in vivo* activity[93]. But, before this potential can be realized, the bacterial *N*-linked glycosylation systems must be fully elucidated at the genetic and structural levels. In particular, the key bacterial component in this system — PglB, which fulfils the oligosaccharyltransferase function in *C. jejuni* and *E. coli* — has to be thoroughly characterized. The characterization of additional bacterial pathways provides a unique opportunity to dissect the genetics of related, but distinct, general glycosylation systems. Identifying and characterizing the variations in the pathways and relating this to glycan structure will be important in terms of providing basic knowledge for engineering complex biosynthetic glycosylated pathways in a relatively simple *E. coli* host. This will also allow for the synthesis of homogeneous glycans, which has only recently been exploited in an artificial yeast glycosylation system (see note added in proof)[94].

Bacterial *O*-linked glycosylation systems also offer the potential for exploitation. DiGiandomenico and co-workers explored the finding that the *O*-linked glycan structure on the *P. aeruginosa* pilin was the same as the structure present on the LPS[40]. They demonstrated that the pilin glycan was a product of the O-antigen biosynthetic pathway and that the glycosyltransferase PilO could not only modify pilin with the O-antigen from that organism but could also attach foreign O-antigens to pilin. Therefore, PilO is a unique and attractive enzyme for engineering *O*-linked glycoproteins because it demonstrates low substrate specificity and, uniquely, it also has the ability to add a glycan block, rather than needing to build the glycan sequentially as is normally observed for *O*-linked glycosylation pathways. Recently, it has been shown that *C. jejuni* PglB also has a relaxed specificity and can add foreign carbohydrate antigens onto asparagine residues (M. Feldman and M. Aebi, personal communication). The remarkable lack of substrate specificity of PilO and PglB indicates that these block transferases have potential for further exploitation, especially in the context of glycoconjugate vaccines against LPS that lack endotoxin. In a broader context, understanding the similarities and differences between bacterial glycosylation systems will be important in developing the field of glycoengineering into a defined science.

Conclusions

The precise roles of glycosylation systems in mucosal bacterial pathogens have yet to be resolved, but as many of these organisms dedicate a significant proportion of their relatively small genomes to glycosylate proteins, these modifications are likely to have crucial roles similar to their eukaryotic counterparts. There are still several questions that remain. For example, why have these mucosal pathogens evolved to express such elaborate glycosylation machineries? How do bacteria know which pathways to use without any obvious compartmentalization? Why are only some proteins specifically targeted to be glycosylated by the general glycosylation systems? In eukaryotes, only proteins containing *N*-linked sequons and the proposed Asx turn are modified[95]. The structural requirements for *N*-linked glycosylation in bacteria remain unknown. Understanding bacterial protein folding, trafficking, chaperone use and regulation of these glycosylation pathways will give further clues about methods that could be used to optimize these systems and to provide the appropriate machineries for glycoengineering. Examination of bacterial glycosylation systems will also lead to the development of new analytical

techniques for carbohydrate analyses that can be applied generally to the field of glycobiology. The direct study of the eukaryotic *N*-linked glycosylation system has been limited by the requirement of this pathway for cell viability. The enormous repertoire of novel enzymes with alternate sugar specificities and potentially enhanced stabilities will also have biotechnological benefits. Thus, with the proper sets of glycosylation enzymes and essential sugar building blocks, glycoengineering in *E. coli* on a practical scale may become a reality.

We have only begun to identify and understand bacterial glycosylation systems. The next few years promise to be a voyage of discovery in terms of comprehending and exploiting this universal biological phenomenon.

Note added in proof

The humanization of *N*-linked protein glycosylation in yeast has recently been summarized in an excellent review by Stefan Wildt and Tillman Gerngross in the February issue of *Nature Reviews Microbiology*[105].

1. Wacker, M. *et al. N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**, 1790–1793 (2002).
Demonstration of the functional transfer of a glycosylation system into *E. coli*.
2. 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 determination of the structure of the *Campylobacter N*-linked glycan and confirmation of the glycosylation of multiple proteins by the *N*-linked general glycosylation pathway in *Campylobacter*.
3. Thibault, P. *et al.* Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. *J. Biol. Chem.* **276**, 34862–34870 (2001).
Determination of the structure of the *Campylobacter O*-linked glycan.
4. Logan, S. M., Kelly, J. F., Thibault, P., Ewing, C. P. & Guerry, P. Structural heterogeneity of carbohydrate modifications affects serospecificity of *Campylobacter* flagellins. *Mol. Microbiol.* **46**, 587–597 (2002).
Detailed demonstration of the complexity of *O*-linked sugar modification on the surface-exposed flagella of *Campylobacter*.
5. 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).
6. Power, P. M. & Jennings, M. P. The genetics of glycosylation in Gram-negative bacteria. *FEMS Microbiol. Lett.* **218**, 211–222 (2003).
7. Benz, I. & Schmidt, M. A. Never say never again: protein glycosylation in pathogenic bacteria. *Mol. Microbiol.* **45**, 267–276 (2002).
8. Schmidt, M. A., Riley, L. W. & Benz, I. Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol.* **11**, 554–561 (2003).
9. Messner, P. & Schäffer, C. in *Progress in the Chemistry of Organic Natural Products* (eds Herz, W., Falk, H. & Kirby, G. W.) 51–124 (Springer–Verlag, Wien, 2003).
10. Szymanski, C. M., Logan, S. M., Linton, D. & Wren, B. W. *Campylobacter* — a tale of two protein glycosylation systems. *Trends Microbiol.* **11**, 233–238 (2003).
11. Doig, P., Kinsella, N., Guerry, P. & Trust, T. J. Characterization of a post-translational modification of *Campylobacter* flagellin: identification of a sero-specific glycosyl moiety. *Mol. Microbiol.* **19**, 379–387 (1996).
12. Parkhill, J. *et al.* The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**, 665–668 (2000).
13. Guerry, P. *et al.* Identification and characterization of genes required for post-translational modification of *Campylobacter coli* VC167 flagellin. *Mol. Microbiol.* **19**, 369–378 (1996).
14. Linton, D. *et al.* Phase variation of a β-1,3 galactosyltransferase involved in generation of the ganglioside GM1-like lipo-oligosaccharide of *Campylobacter jejuni*. *Mol. Microbiol.* **37**, 501–514 (2000).
15. Macnab, R. M. How bacteria assemble flagella. *Annu. Rev. Microbiol.* (2003).
16. Karlyshev, A. V., Linton, D., Gregson, N. A. & Wren, B. W. A novel paralogous gene family involved in phase-variable flagella-mediated motility in *Campylobacter jejuni*. *Microbiology* **148**, 473–480 (2002).
17. Korolik, V., Fry, B. N., Alderton, M. R., van der Zeijst, B. A. & Coloe, P. J. Expression of *Campylobacter hyoilei* lipo-oligosaccharide (LOS) antigens in *Escherichia coli*. *Microbiology* **143**, 3481–3489 (1997).
18. Fry, B. N. *et al.* The lipopolysaccharide biosynthesis locus of *Campylobacter jejuni* 81116. *Microbiology* **144**, 2049–2061 (1998).
19. 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).
First report of the *Campylobacter* general glycosylation pathway.
20. Zufferey, R. *et al.* STT3, a highly conserved protein required for yeast oligosaccharyl transferase activity *in vivo*. *EMBO J.* **14**, 4949–4960 (1995).
21. Knauer, R. & Lehle, L. The oligosaccharyltransferase complex from yeast. *Biochim. Biophys. Acta* **1426**, 259–273 (1999).
22. Silberstein, S. & Gilmore, R. Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *FASEB J.* **10**, 849–858 (1996).
23. Yan, A., Ahmed, E., Yan, Q. & Lennarz, W. J. New findings on interactions among the yeast oligosaccharyl transferase subunits using a chemical cross-linker. *J. Biol. Chem.* **278**, 33078–33087 (2003).
24. Yan, Q. & Lennarz, W. J. Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J. Biol. Chem.* **277**, 47692–47700 (2002).
First evidence that STT3p is directly involved in glycosylation.
25. Virji, M. in *Glycomicrobiology* (ed. Doyle,R. J.) 31–65 (Kluwer Academic/Plenum, New York, 2000).
26. Hegge, F. T. *et al.* Unique modifications with phosphocholine and phosphoethanolamine define alternate antigenic forms of *Neisseria gonorrhoeae* type IV pili. *Proc. Natl Acad. Sci. USA* **101**, 10798–10803 (2004).
27. Endo, T. *et al.* α-1,4-*N*-acetylgalactosamine transferase gene and process for producing this enzyme and *N*-acetylgalactosamine-containing complex carbohydrate. Japan patent WO 0177337 (2002).
28. Suzuki, S., Perry, M., Thibault, P. & Honda, S. in *Capillary Electrophoresis of Carbohydrates*. (eds Thibault, P. & Honda, S.) 285–305 (Humana Press, New Jersey, 2003).
29. Castric, P., Cassels, F. J. & Carlson, R. W. Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. *J. Biol. Chem.* **276**, 26479–26485 (2001).
30. Schirm, M. *et al.* Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol. Microbiol.* **48**, 1579–1592 (2003).
31. Soo, E. C. *et al.* Selective detection and identification of sugar nucleotides by CE-electrospray-MS and its application to bacterial metabolomics. *Anal. Chem.* **76**, 619–626 (2004).
Novel method for the analysis of sugar pathway intermediates, leading to a new era of metabolomics.
32. Kneidinger, B., Graninger, M., Puchberger, M., Kosma, P. & Messner, P. Biosynthesis of nucleotide-activated D-*glycero*-D-*manno*-heptose. *J. Biol. Chem.* **276**, 20935–20944 (2001).
33. Karlyshev, A. V. *et al.* Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the generation of structural diversity and the ability to form complex heptoses. *Mol. Microbiol.* **55**, 90–103 (2005).
34. Bernatchez, S. *et al.* A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in *Campylobacter jejuni*. *J. Biol. Chem.* 27 Oct 2004 (doi:10.1074/jbc.M407767200).
35. Fry, B. N. *et al.* The *galE* gene of *Campylobacter jejuni* is involved in lipopolysaccharide synthesis and virulence. *Infect. Immun.* **68**, 2594–2601 (2000).
36. Stimson, E. *et al.* Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol. Microbiol.* **17**, 1201–1214 (1995).
Identification of a bacillosamine derivative on meningococcal pilin.
37. Fischer, W. & Haas, R. The RecA protein of *Helicobacter pylori* requires a post-translational modification for full activity. *J. Bacteriol.* **186**, 777–784 (2004).
38. 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).
39. 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).
40. DiGiandomenico, A. *et al.* Glycosylation of *Pseudomonas aeruginosa* 1244 pilin: glycan substrate specificity. *Mol. Microbiol.* **46**, 519–530 (2002).
Demonstration that the *Pseudomonas O*-glycan transferase, PilO, can attach foreign O-antigens to pilin, providing new possibilities for glycoengineering.
41. Kahler, C. M. *et al.* Polymorphisms in pilin glycosylation locus of *Neisseria meningitidis* expressing class II pili. *Infect. Immun.* **69**, 3597–3604 (2001).
42. 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).
43. Takeuchi, K. *et al.* Flagellin glycosylation island in *Pseudomonas syringae* pv. *glycinea* and its role in host specificity. *J. Bacteriol.* **185**, 6658–6665 (2003).
Evidence that flagellin glycosylation has a role in host specificity.
44. Dorrell, N. *et al.* Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res.* **11**, 1706–1715 (2001).
45. Leonard, E. E. *et al.* Use of an open-reading-frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates. *J. Infect. Dis.* **187**, 691–694 (2003).
46. 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).
47. Taboada, E. N. *et al.* Large-scale comparative genomics meta-analysis of *Campylobacter jejuni* isolates reveals low level of genome plasticity. *J. Clin. Microbiol.* **42**, 4566–4576 (2004).
48. Pearson, B. M. *et al.* Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Lett.* **554**, 224–230 (2003).
49. Fouts, D. E. *et al.* Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. *PLoS Biol.* **3**, 1–14 (2005).
50. Baar, C. *et al.* Complete genome sequence and analysis of *Wolinella succinogenes*. *Proc. Natl Acad. Sci. USA* **100**, 11690–11695 (2003).
51. Schaffer, C., Graninger, M. & Messner, P. Prokaryotic glycosylation. *Proteomics* **1**, 248–261 (2001).
52. Kuo, C., Takahashi, N., Swanson, A. F., Ozeki, Y. & Hakomori, S. An *N*-linked high-mannose type oligosaccharide, expressed at the major outer membrane protein of *Chlamydia trachomatis*, mediates attachment and infectivity of the microorganism to HeLa cells. *J. Clin. Invest.* **98**, 2813–2818 (1996).

53. Lindenthal, C. & Elsinghorst, E. A. Identification of a glycoprotein produced by enterotoxigenic *Escherichia coli*. *Infect. Immun.* **67**, 4084–4091 (1999).
First detailed report of protein glycosylation in *Escherichia coli*.
54. Miron, J. & Forsberg, C. W. Characterisation of cellulose-binding proteins that are involved in the adhesion mechanism of *Fibrobacter intestinalis* DR7. *Appl. Microbiol. Biotechnol.* **51**, 491–497 (1999).
55. Herrmann, J. L., O'Gaora, P., Gallagher, A., Thole, J. E. & Young, D. B. Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19-kDa antigen from *Mycobacterium tuberculosis*. *EMBO J.* **15**, 3547–3554 (1996).
56. Marceau, M. & Nassif, X. Role of glycosylation at Ser63 in production of soluble pilin in pathogenic *Neisseria*. *J. Bacteriol.* **181**, 656–661 (1999).
57. Romain, F. *et al.* Deglycosylation of the 45/47-kilodalton antigen complex of *Mycobacterium tuberculosis* decreases its capacity to elicit *in vivo* or *in vitro* cellular immune responses. *Infect. Immun.* **67**, 5567–5572 (1999).
58. Ghosh, S. K. *et al.* Pathogenic consequences of *Neisseria gonorrhoeae* pilin glycan variation. *Microbes. Infect.* **6**, 693–701 (2004).
59. Hamadeh, R. M., Estabrook, M. M., Zhou, P., Jarvis, G. A. & Griffiss, J. M. Anti-Gal binds to pili of *Neisseria meningitidis*: the immunoglobulin A isotype blocks complement-mediated killing. *Infect. Immun.* **63**, 4900–4906 (1995).
60. Serino, L. & Virji, M. Genetic and functional analysis of the phosphorylcholine moiety of commensal *Neisseria* lipopolysaccharide. *Mol. Microbiol.* **43**, 437–448 (2002).
61. Schultze, M. & Kondorosi, A. The role of lipochitooligosaccharides in root nodule organogenesis and plant cell growth. *Curr. Opin. Genet. Dev.* **6**, 631–638 (1996).
62. Lindenthal, C. & Elsinghorst, E. A. Enterotoxigenic *Escherichia coli* TibA glycoprotein adheres to human intestine epithelial cells. *Infect. Immun.* **69**, 52–57 (2001).
63. Moormann, C., Benz, I. & Schmidt, M. A. Functional substitution of the TibC protein of enterotoxigenic *Escherichia coli* strains for the autotransporter adhesin heptosyltransferase of the AIDA system. *Infect. Immun.* **70**, 2264–2270 (2002).
64. Leclerc, G., Wang, S. P. & Ely, B. A new class of *Caulobacter crescentus* flagellar genes. *J. Bacteriol.* **180**, 5010–5019 (1998).
65. Gryllos, I., Shaw, J. G., Gavin, R., Merino, S. & Tomas, J. M. Role of *flm* locus in mesophilic *Aeromonas* species adherence. *Infect. Immun.* **69**, 65–74 (2001).
66. 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).
67. Schmitz, A., Josenhans, C. & Suerbaum, S. Cloning and characterization of the *Helicobacter pylori flbA* gene, which codes for a membrane protein involved in coordinated expression of flagellar genes. *J. Bacteriol.* **179**, 987–997 (1997).
68. Bedouet, L. *et al.* Evidence for an heterogeneous glycosylation of the *Clostridium tyrobutyricum* ATCC 25755 flagellin. *Microbios* **94**, 183–192 (1998).
69. Schirm, M. *et al.* Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *J. Bacteriol.* **186**, 2523–2531 (2004).
70. Szymanski, C. M., Burr, D. H. & Guerry, P. *Campylobacter* protein glycosylation affects host cell interactions. *Infect. Immun.* **70**, 2242–2244 (2002).
71. 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).
72. Szymanski, C. M., Goon, S., Allan, B. & Guerry, P. in *Campylobacter: Molecular and Cellular Biology* (eds Ketley, J. M. & Konkel, M. E.) 259–273 (Horizon Bioscience, Norwich, 2005).
73. 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).
74. 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).
75. Whitfield, C. & Roberts, I. S. Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol. Microbiol.* **31**, 1307–1319 (1999).
76. Lin, J., Michel, L. O. & Zhang, Q. CmeABC functions as a multidrug efflux system in *Campylobacter jejuni*. *Antimicrob. Agents Chemother.* **46**, 2124–2131 (2002).
77. Helenius, A. & Aebi, M. Intracellular functions of *N*-linked glycans. *Science* **291**, 2364–2369 (2001).
Review of the functions of eukaryotic *N*-linked glycans in the endoplasmic reticulum.
78. Wacker, M., Nita-Lazar, M. & Aebi, M. PglB, an oligosaccharyltransferase in the eubacterium *Campylobacter jejuni*? (Abstracts of the 11th International Workshop on *Campylobacter*, *Helicobacter* and related organisms. September 1–5, 2001. Freiburg, Germany). *Int. J. Med. Microbiol.* **291** (Suppl.), 81 (2001).
79. Wieland, F. Structure and biosynthesis of prokaryotic glycoproteins. *Biochimie* **70**, 1493–1504 (1988).
80. Zhu, B. C., Drake, R. R., Schweingruber, H. & Laine, R. A. Inhibition of glycosylation by amphomycin and sugar nucleotide analogs PP36 and PP55 indicates that *Haloferax volcanii* β-glucosylates both glycoproteins and glycolipids through lipid-linked sugar intermediates: evidence for three novel glycoproteins and a novel sulfated dihexosyl-archaeol glycolipid. *Arch. Biochem. Biophys.* **319**, 355–364 (1995).
81. Kuntz, C., Sonnenbichler, J., Sonnenbichler, I., Sumper, M. & Zeitler, R. Isolation and characterization of dolichol-linked oligosaccharides from *Haloferax volcanii*. *Glycobiology* **7**, 897–904 (1997).
82. Zeitler, R., Hochmuth, E., Deutzmann, R. & Sumper, M. Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn–Ala–Ser does not prevent *N*-glycosylation of the cell surface glycoprotein from *Halobacterium halobium*. *Glycobiology* **8**, 1157–1164 (1998).
83. Imperiali, B. & Hendrickson, T. L. Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. *Bioorg. Med. Chem.* **3**, 1565–1578 (1995).
84. Schirm, M. *et al.* Flagellin from *Listeria monocytogenes* is glycosylated with β-*O*-linked *N*-acetylglucosamine. *J. Bacteriol.* **186**, 6721–6727 (2004).
85. Zachara, N. E. & Hart, G. W. *O*-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim. Biophys. Acta* **1673**, 13–28 (2004).
86. Aebi, M. & Hennet, T. Congenital disorders of glycosylation: genetic model systems lead the way. *Trends Cell Biol.* **11**, 136–141 (2001).
Review of the use of *Saccharomyces cerevisiae* as a model to study inherited human diseases involving congenital disorders of *N*-linked glycosylation.
87. Alton, G. *et al.* Direct utilization of mannose for mammalian glycoprotein biosynthesis. *Glycobiology* **8**, 285–295 (1998).
88. Davis, J. A. & Freeze, H. H. Studies of mannose metabolism and effects of long-term mannose ingestion in the mouse. *Biochim. Biophys. Acta* **1528**, 116–126 (2001).
89. 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).
90. Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130 (1993).
91. Betenbaugh, M. J., Tomiya, N., Narang, S., Hsu, J. T. & Lee, Y. C. Biosynthesis of human-type *N*-glycans in heterologous systems. *Curr. Opin. Struct. Biol.* **14**, 601–606 (2004).
92. Raju, T. S., Briggs, J. B., Chamow, S. M., Winkler, M. E. & Jones, A. J. Glycoengineering of therapeutic glycoproteins: *in vitro* galactosylation and sialylation of glycoproteins with terminal *N*-acetylglucosamine and galactose residues. *Biochemistry* **40**, 8868–8876 (2001).
93. Elliott, S. *et al.* Enhancement of therapeutic protein *in vivo* activities through glycoengineering. *Nature Biotechnol.* **21**, 414–421 (2003).
94. Bobrowicz, P. *et al.* Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: production of complex humanized glycoproteins with terminal galactose. *Glycobiology* **14**, 757–766 (2004).
95. Imperiali, B. & O'Connor, S. E. Effect of *N*-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* **3**, 643–649 (1999).
Review of the effect of *N*-linked glycosylation on protein folding or the stability of the native glycoprotein conjugates.
96. Gemmill, T. R. & Trimble, R. B. Overview of *N*- and *O*-linked oligosaccharide structures found in various yeast species. *Biochim. Biophys. Acta* **1426**, 227–237 (1999).
97. Luo, Y., Luther, K. B. & Haltiwanger, R. S. Protein *O*-fucosyltransferase 1 (*O*-FucT-1) is a soluble enzyme localized in the ER. *Glycobiology* **14**, 1059 (2004).
98. Hart, G. W. Dynamic *O*-linked glycosylation of nuclear and cytoskeletal proteins. *Annu. Rev. Biochem.* **66**, 315–335 (1997).
99. Hanover, J. A. Glycan-dependent signaling: *O*-linked *N*-acetylglucosamine. *FASEB J.* **15**, 1865–1876 (2001).
100. Creuzenet, C. Characterization of Cj1293, a new UDP-GlcNAc C6 dehydratase from *Campylobacter jejuni*. *FEBS Lett.* **559**, 136–140 (2004).
101. Creuzenet, C., Schur, M. J., Li, J., Wakarchuk, W. W. & Lam, J. S. FlaA1, a new bifunctional UDP-GlcNAc C6 Dehydratase/ C4 reductase from *Helicobacter pylori*. *J. Biol. Chem.* **275**, 34873–34880 (2000).
102. Linton, D. *et al.* Multiple *N*-acetyl neuraminic acid synthetase (*neuB*) genes in *Campylobacter jejuni*: identification and characterization of the gene involved in sialylation of lipo-oligosaccharide. *Mol. Microbiol.* **35**, 1120–1134 (2000).
103. Bleumink-Pluym, N. M., Verschoor, F., Gaastra, W., van der Zeijst, B. A. & Fry, B. N. A novel approach for the construction of a *Campylobacter* mutant library. *Microbiology* **145**, 2145–2151 (1999).
104. Guerry, P,. Logan, S. M., Thornton, S. & Trust, T. J. Genomic organization and expression of *Campylobacter* flagellin genes. *J. Bacteriol.* **172**, 1853–1860 (1990).
105. Wildt, S. & Gerngross, T. U. The humanization of *N*-glycosylation pathways in yeast. *Nature Rev. Microbiol.* **3**, 119–128 (2005).

Acknowledgements

We thank S. Logan for the critical reading of this review, J.-R. Brisson, N. M. Young and N. Khieu for constructing the space-filling models and M. Aebi for reviewing the manuscript and providing information prior to publication. We also thank R. Mandrell and W. Miller for providing genomic data for the *Campylobacter* species described in this review and for assembling online supplementary information S1 (Table). We gratefully acknowledge the NRC Genomics and Health Initiative for providing funding to C.M.S and the Leverhulme Trust and BBSRC for funding to B.W.W.

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez: http://www.ncbi.nlm.nih.gov/Entrez/
Campylobacter coli | *C. jejuni* NCTC 11168 | *C. jejuni* RM1221 | *Campylobacter lari* RM2100 | *Campylobacter upsaliensis* RM3195 | Cj1305c | Cj1306c | Cj1310c | Cj1318 | Cj1333 | Cj1334 | Cj1335 | Cj1336 | Cj1337 | Cj1340 | Cj1341 | Cj1342c | *galE* | *Helicobacter pylori* | *Neisseria gonorrhoeae* | *Neisseria meningitidis* | *neuA2* | *neuA3* | *neuB2* | *neuB3* | *neuC2* | *Pseudomonas syringae* | *Wolinella succinogenes*

SwissProt: http://www.expasy.org/sprot/
FlaA | FlaB | PglA | PglB | PglC | PglD | PglE | PglF | PilO | STT3 | TibA | WlaB

FURTHER INFORMATION

***Campy*DB:** http://campy.bham.ac.uk/

***Desulfovibrio desulfuricans* genome:**
http://genome.jgi-psf.org/draft_microbes/desde/desde.home.html

Christine M. Szymanski's laboratory: http://ibs-isb.nrc-cnrc.gc.ca/ourstories/cszymanskibio_e.htm

Brendan W. Wren's laboratory:
http://www.lshtm.ac.uk/pmbu/staff/bwren/bwren.html

SUPPLEMENTARY INFORMATION

See online article: S1 (table)

Access to this links box is available online.