Understanding Protein Glycosylation Pathways in Bacteria

Abstract
Through advances in analytical methods to detect glycoproteins and to determine glycan structures, there have been increasing reports of protein glycosylation in bacteria. In this review, we summarize the known pathways for bacterial protein glycosylation: 1) lipid carrier-mediated “en bloc” glycosylation; and 2) cytoplasmic stepwise protein glycosylation. The exploitation of bacterial protein glycosylation systems, especially the "mix and match" of three independent but similar pathways (oligosaccharyltransferase-(OST)-mediated protein glycosylation, lipopolysaccharide (LPS) and peptidoglycan biosynthesis) in Gram-negative bacteria for glycoengineering recombinant glycoproteins is also discussed.

Abbreviations:
Gal, galactose; Glc, glucose; GlcNAc, N-acetylglucosamine; diNAcBac, 2,4-diacetamido-2,4,6-trideoxy-α-D-glucopyranose; GalNAc, N-acetylgalacosamine; UndP, undecaprenyl pyrophosphate; Hex, hexose; FucNAc, N-acetylfucosamine; Xyl, xylose; DATDH, 2,4-diacetamido-2,4,6-trideoxyhexose; GATDH, 2-glyceramido-4-acetamido-2,4,6-trideoxyhexose; α-5NβOHC47NFmPse, 5-N-β-hydroxybutyryl-7-N-formyl-pseudaminic acid; Man, mannose; PP, polyprenol phosphate. Pse, pseudaminic acid; Leg, 5,7-diacetamido-3,5,7,9-tetradeoxy-D-glycero-D-galacto-nonulosonic acid.

Key words
Protein glycosylation, glycosylation pathway, glycoengineering
1. Introduction

Bacteria have evolved complex biosynthetic pathways to generate an array of mono- or oligosaccharide moieties to produce capsules, lipopolysaccharides (LPS), peptidoglycans, and also to glycosylate lipids (including membrane cholesterols) and proteins. Glycosylation requires the action of enzymes known as glycosyltransferases (GTs) to recognize both the activated-sugar donor and the corresponding acceptor to form a new glycosidic bond between the sugar and the acceptor. The donors are phosphate-activated sugars, which can be linked to nucleotides (UDP-, GDP-, ADP- and CMP-) or dolichol-related lipid groups.

Protein glycosylation is a complex enzymatic process that involves GTs and activated substrates. Once thought to be exclusively present in eukaryotes, it is now evident that protein glycosylation is a common feature in all three domains of life (Eukarya, Bacteria, and Archaea) [1, 2]. Even though the mass content of glycans in a glycoprotein is often small (5-10%), glycans represent a significant part of the hydrodynamic volume of the overall glycoprotein, as glycans have a much lower density than proteins [3]. Glycans can have a profound influence on the physiochemical properties of a protein such as proper folding, aggregation, solubility, and stability [4, 5]. Recent studies have shown that glycans on bacterial proteins play key functions in biological processes including adhesion, immune evasion and host colonization [6, 7]. In this review, we summarize the well-described bacterial protein glycosylation pathways, and discuss its exploitation for glycoengineering.

2. An Overview of Bacterial Protein Glycosylation

The first confirmed report of N-linked glycosylation pathway was described in 2002 in the human gastrointestinal pathogen Campylobacter jejuni [8]. Since then a large number of bacterial glycoproteins have been identified including 1) the flagellins from C. jejuni [9, 10], Campylobacter coli [11], Helicobacter pylori [12-14], Aeromonas caviae [15], Pseudomonas aeruginosa [16], Listeria monocytogenes [17,
Based on the glycosidic linkage, protein glycosylation can be categorized into two major types: 1) O-glycosylation, in which the glycans are attached to the hydroxyl oxygen of serine or threonine (Ser/Thr) residues and 2) N-glycosylation, in which the glycans are attached to the amide nitrogen of asparagine (Asn) residues in targeted proteins. In eukaryotes and archaea, the Asn residues are present in the tripeptide consensus sequence: Asn-X-Ser/Thr (X can be any amino acid except proline), whereas in bacteria this consensus sequon is extended to be Asp/Glc-X1-Asn- X2-Ser/Thr (X1 and X2 represent any amino acid except proline) [43].

Depending on how the mono- or oligosaccharides are transferred to the targeted proteins, protein glycosylation can also be categorized into two main pathways: 1) lipid carrier-mediated “en bloc” glycosylation and; 2) stepwise protein glycosylation.
En bloc glycosylation involves the preassembly of the glycan onto a lipid carrier to form a lipid-linked oligosaccharide (LLO), which is then transported by a flippase from the cytoplasm to the periplasm, where the oligosaccharide is transferred as whole from the LLO to target proteins by an oligosaccharyltransferase (OST) [1]. In Gram-negative bacteria, the glycans are first preassembled onto an undecaprenyl pyrophosphate (UndP), while in the Gram-positive bacteria Actinomycetes, protein O-mannosylation utilizes polyrenol phosphate (PP) as the lipid carrier [44]. Stepwise protein glycosylation involves the sequential addition of nucleotide-activated sugars onto Asn and Ser/Thr residues of targeted proteins by GTs.

3. Lipid Carrier-Mediated en bloc Protein Glycosylation

Three different categories of lipid carrier-mediated en bloc protein glycosylation in bacteria are described here: 1) the OST-mediated protein N-glycosylation in Gram-negative bacteria; 2) the OST-mediated protein O-glycosylation in Gram-negative bacteria and; 3) the O-mannosylation in Gram-positive Actinomycetes (Figure 1).

3.1 OST-Mediated Protein N-glycosylation in Campylobacter and Helicobacter spp. The protein N-glycosylation pathway in C. jejuni (Figure 1A) is the best-studied glycosylation system in bacteria and has been considered as a prototype for studying and appreciating the importance of protein glycosylation in host-microbe interactions. This well-characterized system is encoded by a unique 17-kb N-linked protein glycosylation locus that contains 12 open reading frames (ORF). The enzymes encoded within this locus are responsible for glycosylating more than 60 different proteins with the heptasaccharide, GalNAc-GalNAc-Glc-GalNAc-GalNAc-GalNAc-diNAcBac [39]. The sugar donor for diNAcBac in Campylobacter is UDP-diNAcBac, which is synthesized from the universal substrate UDP-GlcNAc (Figure 2) [2, 5, 45, 46]. The assembly of the heptasaccharide onto the lipid carrier involves five GTs: the first GT PglC (Cj1124c)
is responsible for the transfer of diNAcBac to the lipid carrier UndP, forming UndP-linked diNAcBac [47]; the second GT PglA (Cj1125c) adds the second GalNAc residue [48]; Subsequently, the third GT PglJ (Cj1127c) adds the third GalNAc residue [48]; the fourth GT PglH (Cj1129c) then acts as a polymerase to transfer another three GalNAc residues to form a UndP-linked hexasaccharide [49]; finally, the fifth GT PglI (Cj1128c) adds the branching Glc residue to complete the assembly of lipid-linked heptasaccharide [50]. After the assembly in the cytoplasm, the LLO is then translocated by the flippase PglK (Cj1130c) into the periplasm, where the LLO is transferred en bloc by the OST PglB (Cj1126c) to acceptor proteins [39] (Figure 1A).

Like *C. jejuni*, all the other *Campylobacter* species possess a similar *N*-glycosylation gene cluster and therefore can produce a range of structurally distinct *N*-linked glycans with variations in the number of sugar residues (penta-, hexa-, and heptasaccharides) in the presence of branching sugars and in the overall monosaccharide content [51].

There are several other bacterial genera, including *Helicobacter* and *Wolinella*, that belong to the same phylogenetic grouping as *Campylobacter*, the ε-proteobacteria [1], and therefore it is not surprising that similar *N*-glycosylation pathways have also been confirmed in *Helicobacter* [52] and *Wolinella spp.* [1, 51]. Three closely related *Helicobacter* species (*H. pullorum*, *H. canadenis* and *H. winghamensis*) contain two unrelated *pglB* genes (*pglB1* and *pglB2*) [52]. It has been shown in *H. pullorum* that PglB1 is the OST responsible for the *N*-linked protein glycosylation with a pentasaccharide glycan comprised of two HexNAc residues and three unknown sugar residues with MWs of 217 Da and 216 Da [52]. Of note is that except for the above three *Helicobacter* species, most other *Helicobacter* species, including *H. pylori*, lack the *pglB* gene [5], and thus it is very likely that *N*-linked protein glycosylation is absent in these bacteria. The *N*-glycosylation pathway found in *Wolinella succinogenes* has been shown to produce a hexasaccharide comprising one Hex residue, one DATDH and four unknown sugar residues with MWs of 216 Da,
216 Da, 216 Da and 232 Da [1, 51].

In addition to the ε-proteobacteria mentioned above, Desulfovibrio vulgaris and Desulfovibrio desulfuricans, which belong to δ-proteobacteria, have also been found to possess an N-linked protein glycosylation pathway [2].

3.2 OST-Mediated Protein O-Glycosylation in Gram-Negative Bacteria

Similar to the general protein N-glycosylation system in C. jejuni, several bacteria including N. meningitidis [20] and N. gonorrhoeae [21] (Figure 1B), P. aeruginosa (Figure 1C), Acinetobacter baumannii [37] (Figure 1D), and also B. fragilis [27] and Francisella tularensis [36] are found to possess a general O-linked glycosylation system which encodes enzymes including GTs, flippases and OSTs required for the assembly, translocation and ligation of oligosaccharides to target proteins.

In N. meningitidis, pilin glycosylation genes (pglB or pglB2, pglC, pglD, pglA, pglE and pglF) are responsible for the glycosylation of the major subunit of type IV pili PilE and the surface-exposed nitrite reductase AniA with a trisaccharide composed of Gal-Gal-DATDH/GATDH) [41]. Of note, is that the structure of DATDH was later unequivocally shown to be diNAcBac [53]. The sugar donor for diNAcBac/GATDH in N. meningitidis is UDP-diNAcBac/GATDH, which is synthesized from UDP-GlcNAc [2] (Figure 3). The first step is catalyzed by PglD (homologous to the C. jejuni PglF) to mediate the dehydration of UDP-GlcNAc; PglC (homologous to the C. jejuni PglE) in the second step transfers the amino group; the third step is then catalyzed by the acetyltransferase function of PglB or PglB2 to form UDP-DATDH or UDP-GATDH, respectively. Subsequently, the bifunctional enzyme PglB or PglB2 utilizes its GT function to transfer the activated diNAcBac or GATDH to the lipid carrier. The Gal transferases PglA and PglE are responsible for the transfer of Gal residues to the diNAcBac or GATDH to form a lipid-linked trisaccharide [54]. The flippase PglF then translocates the lipid-linked trisaccharide to the periplasm, where PglL (OST) transfers the oligosaccharide to glycosylate PilE or AniA [55] (Figure
Very recently, it was shown that the PilE of class II pilins in *N. meningitidis* is heavily coated with glycans, having up to five glycosylation sites, thus limiting the access of antibodies to the pilin polypeptide chain and mediating immune evasion from the host [56]. This is an excellent example demonstrating the importance of protein glycosylation in bacterial pathogenesis.

Compared to only two glycoproteins PilE and AniA in *N. meningitidis*, eleven membrane-associated proteins have been demonstrated to be glycosylated in *N. gonorrhoeae* [21]. Homologous to PglL in *N. meningitidis*, PglO in *N. gonorrhoeae* is the OST responsible for the transfer of glycans to proteins [21] (Figure 1B). Both PglO and PglL contain the signature Wzy_C motif present in LPS *O*-antigen ligase WaaL, which transfers an undecaprenyl-linked *O*-antigen to the LPS lipid A-core [57]. Interestingly, the OST PilO (contains the Wzy_C motif) in *P. aeruginosa* 1244 transfers a single *O*-antigen unit, a trisaccharide comprised of 5NβOHC47NFmPse-Xyl-FucNAc, generated from the LPS *O*-antigen biosynthetic pathway to glycosylate the pilin subunit PilA [58, 59] (Figure 1C). WbpL initiates the assembly of the single *O*-antigen unit by the transfer of FucNAc from UDP-FucNAc to UndP carrier [60, 61]. The GTs for the transfer of Xyl and 5NβOHC47NFmPse are unknown and once assembly of the *O*-antigen unit is complete, it is translocated by the *O*-antigen flippase Wzx to the periplasm, where it is used by the *O*-antigen ligase WaaL for the ligation of *O*-antigen with the LPS lipid A-core, and it is also used by the OST PilO for the glycosylation of PilA [22], suggesting an evolutionary collection between LPS biosynthesis and OST-mediated pilin glycosylation in Gram-negative bacteria.

Very recently, the general *O*-linked glycosylation system has been identified and characterized in the opportunistic nosocomial pathogen *A. baumannii* [37, 62] (Figure 1D). Multiple proteins in *A. baumannii* strain ATCC 17978 were found be modified with a branched pentasaccharide comprising of GalNAc, Glc, Gal, GlcNAc and a glucuronic acid derivative, GlcNAc3NAcA4OAc [37, 62]. Interestingly, this
pentasaccharide also serves as the building block for capsular polysaccharides, suggesting a common pathway for \( O \)-linked protein glycosylation and capsule biosynthesis in \textit{A. baumannii} [62]. As this common pathway is dependent on the activity of PglC (the initiating enzyme for the synthesis of the pentasaccharide), it has been proposed that the assembly of the pentasaccharide is initiated by the transfer of GalNAc to the UndP carrier, by PglC. Four GTs (A1S_0058, A1S_3482, A1S_0059, and A1S_0060) transfer corresponding sugars to complete the pentasaccharide. Wzx (A1S_0056) flips the completed pentasaccharide structure to the periplasm and subsequently coupled to target proteins by the OST PglL (A1S_3176)[62] (Figure 1D).

Other examples of general OST-mediated protein \( O \)-glycosylation systems are found in \textit{Bacteroides} [27] and \textit{F. tularensis} [36]. \textit{Bacteroides} comprise one of the most abundant genera of commensals in the human colon, and it has been shown that \( O \)-glycosylation is central not only to the physiology of \textit{B. fragilis} but is also required for competitive colonization of its mammalian host [27]. The \( O \)-glycan in \textit{B. fragilis} has been demonstrated to be an oligosaccharide comprising of nine sugar units, three Hex residues (one reduced), two hexuronic acids, two HexNAc residues, one methylated deoxyhexose and one deoxyhexose residue [63]. Five putative GTs (BF4299, BF4300, BF4301, BF4305, BF4306) and a fucosyltransferase are proposed to be involved in the assembly of the glycan, and the \( O \)-antigen flippase Wzx (BF4298) is thought to be responsible for the translocation of lipid-linked glycan from the cytoplasm to the periplasm [27]. However, the OST in \textit{B. fragilis} has yet to be identified.

Pilin protein PilA (encoded by \textit{FTS_0381}) [36] and a lipoprotein encoded by gene \textit{FTS_0065} [35] in \textit{F. tularensis} FSC200 strain, and a virulence factor DsbA protein in \textit{F. tularensis} SchuS4 strain [64], were found to be glycosylated with a phosphate-containing hexasaccharide (HexNAc-unknown sugar-HexNAc-Hex-Hex-HexNAc). The OST involved in the \textit{en bloc} glycosylation
was identified as PglA (encoded by *FTL_0425* in FCS200 strain, and *FTT_0905* in SchuS4 strain) [36]. The 12.5 kB gene cluster *FTT0789-FTT0800* in the SchuS4 strain was shown to be involved in the assembly and translocation of the hexasaccharide [64, 65]. Interestingly, the mutation of *wbtC* gene from LPS biosynthetic locus resulted in not only altered LPS but also reduced protein glycosylation [66]. Similarly the mutation of the putative glycan flippase gene *FTS_1402* in the FCS200 strain lead to alteration in protein glycosylation and also a decrease in LPS production [65], suggesting a cross-talk between LPS biosynthesis and protein glycosylation in *F. tularensis*.

3.3 Protein O-Mannosylation in Gram-Positive Actinomycetes

A general protein O-mannosylation pathway has been found in Gram-positive bacteria Actinomycetes, which includes the genera of *Mycobacterium* [67, 68], *Corynebacterium* [69], and *Streptomyces* [70]. It has been shown that the membrane-bound phosphate binding protein PstS in *Streptomyces coelicolor* is glycosylated with a Man-linked trihexose [70]. The polyprenylphosphate Man synthase Ppm1 is known to be responsible for catalyzing the transfer of Man from GDP-Man to the lipid carrier polyprenyl phosphate, forming polyprenyl monophosphomannose [70]. However, how the lipid-linked Man is further extended to the trihexose moiety is not clear. The O-mannosyltransferase Pmt catalyzes the glycosylation of the Ser or Thr residues of PstS with the Man-linked trihexose [2, 70].

4. Stepwise Protein Glycosylation

As mentioned earlier, bacterial stepwise protein glycosylation is OST-independent and is catalysed by GTs, which sequentially transfer monosaccharides directly to an acceptor protein. Some representative examples of stepwise protein glycosylation include the N-glycosylation of adhesin protein HMW1 in *H. influenzae* [29, 30], the O-glycosylation of flagellins in *Campylobacter* [40] and *Helicobacter* [12, 13], and also the heptosylation of autotransporters [6].
4.1 Stepwise Protein N-glycosylation in *H. influenzae*.

*N*-linked glycosylation of the *H. influenzae* HMW1 adhesion occurs in the cytoplasm and involves the direct transfer of Hex or Hex-Hex residues to 31 sites on the HMW1 protein [29, 30] (Figure 4A). The Hex is either Gal or Glc, and all but one of the glycosylation sites are Asn residues within the eukaryotic-like sequon Asn-X-Ser/Thr [29, 30]. The GT responsible for this process has been confirmed as HMW1C, which has been assigned as a member of CAZy (Carbohydrate Active Enzyme database) family GT41 [30, 71, 72]. Interestingly, homology analysis reveals 42–68% identity and 58–83% similarity between the full-length HMW1C sequence and proteins in a number of other Gram-negative bacteria, suggesting a GT41 family of bacterial HMW1C-like proteins with *N*-glycosylation glycosyltransferase activity [30]. Indeed, the HMW1C-like protein in *Actinobacillus pleuropneumoniae* has been shown to be able to transfer Gal or Glc to known Asn glycosylation sites in HMW1 [73].

4.2 Stepwise Protein O-Glycosylation of Bacterial Flagellins.

Bacterial flagellins are the most well studied proteins with stepwise cytoplasmic *O*-glycosylation. Flagellins from *C. jejuni* and *C. coli* [40], *H. pylori* [12, 13], and *A. caviae* [15] are modified with the 9-carbon sialic acid-like sugar Pse or legionaminic acid (Leg) and it has been demonstrated that flagellin glycosylation is essential for motility and flagella assembly in these bacteria [74, 75]. In contrast, the glycosylation of flagellin structural protein FliC in *P. aeruginosa* is not required for flagellar motility and assembly. FliC in *P. aeruginosa* PAK is modified with a heterogeneous *O*-linked glycan comprising up to 11 additional monosaccharides, linked to the protein backbone via a rhamnose residue [16].

*C. jejuni* strain 81-176 flagellins have been shown to be predominantly modified with Pse5Ac7Ac with additional heterogenous derivatives, including Pse5Am7Ac (Figure 4B). This imparts an approximate 6 kDa shift in the molecular mass of the predicted protein [40, 75]. Similarly for *C. coli* VC 167 T2, the flagellins are extensively
modified with Pse5Ac7Ac and its derivatives [40]. In addition to Pse, the flagellins of
*C. jejuni* 11168 and *C. coli* VC167 are also found to be glycosylated with Leg and its
derivatives [76].

Compared to the *Campylobacter* flagellin, which are glycosylated with numerous
Pse/Leg and related derivatives, the glycosylation of *H. pylori* flagellin is far simpler,
and only a single sugar species (Pse5Ac7Ac) is present on *H. pylori* flagellin [40]. It
has been demonstrated that *H. pylori* strain 1061 flagellin are glycosylated with
Pse5Ac7Ac at 7 sites in FlaA and 10 sites in FlaB [13] (Figure 4C).

A common feature of flagellin glycosylation is that the sites of glycosylation are
located in the central, surface-exposed domain of flagellin when it is assembled in a
filament, and the mechanism of attachment seems to be unrelated to a consensus
peptide sequence, but is rather related to the surface accessibility of the Ser/Thr
residues in the folded protein [13, 16, 40].

The sugar donor for Pse in *H. pylori* and *C. jejuni* has been shown to be CMP-Pse
[13]. Six Pse biosynthesis enzymes (PseB, C, H, G, I, F) have been found to constitute
the complete CMP-Pse biosynthetic pathway starting from UDP-GlcNAc [14]
(Figure 5).

The disruption of any of the Pse genes in this pathway results in a non-motile
phenotype, no structural flagella filament and severely reduced levels of flagellin
protein [13, 45, 77, 78]. The absence of glycosyated flagellins in the cytoplasm of the
*H. pylori* wild-type or pse mutants suggests that flagellin glycosylation might be
coupled with their secretion by the flagella Type III secretion system [77]. The sugar
donor for Leg in *Campylobacter* is CMP-Leg, and eleven enzymes have been
successfully reconstituted *in vitro* for the biosynthesis of CMP-Leg starting from
Fru-6P [79] (Figure 6).
Although the biosynthesis of CMP-Pse and CMP-Leg have been elucidated, the key GTs involved in the transfer of Pse/Leg and related derivatives to flagella are yet to be identified and confirmed. In *H. pylori*, HP0114 is not directly involved in the biosynthesis of CMP-Pse, but its inactivation resulted in a non-motile phenotype, and therefore, it has been suggested that *HP0114* might encode the pseudaminyltransferase [13]. A recent study showed that the motility-associated factor gene *maf1* in *A. caviae* might be the putative pseudaminyltransferase responsible glycosylation of the flagellin in this microbe [15].

Flagellin O-glycosylation has also been demonstrated in the Gram-positive bacteria *L. monocytogenes* [17], *Clostridium botulinum* [80] and *C. difficile* [19]. Up to six sites of the central surface-exposed region of the flagellin monomer of *L. monocytogenes* is glycosylated with β-O-linked GlcNAc, which is transferred by the GT protein Lmo0688 [12, 17]. The flagellins of *C. botulinum* and *C. difficile* have been shown to be modified with Leg derivatives and HexNAc residues, respectively [19, 80].

### 4.3 Stepwise Heptosylation of Autotransporters

Autotransporters constitute the largest number of secreted virulence factors in Gram-negative bacteria. The modular structure of autotransporters consist of a signal peptide, an *N*-terminal functional passenger domain, and a *C*-terminal β-barrel domain which mediates the translocation of the passenger domain across the bacterial outer membrane [6, 81]. The passenger domains of autotransporters TibA from enterotoxigenic *E. coli* (ETEC) strain H10407 [82, 83], AIDA-I from diarrheagenic *E. coli* (DAEC) clinical isolate 2787 [26, 82], and Ag43 from pathogenic as well as commensal *E. coli* strains [23, 84] have long been known to be extensively glycosylated with *O*-linked Hep moieties. Very recently, it has been demonstrated that heptosylation of AIDA-I is catalyzed by the heptosyltransferase AAH and its paralog AAH2, and this modification is essential for adhesion to host cells [6]. Structural
analysis of the heptosylated TibA passenger domain revealed that TibA was modified by the heptosyltransferase TibC with 35 Hep moieties, which are exclusively on Ser residues [6].

AAH/AAH2 and TibC are members of the newly-defined bacterial autotransporter heptosyltransferase (BAHT) family that contain a ferric ion and adopt a dodecamer assembly [6]. The gene encoding BAHT is always located upstream of its substrate autotransporter in the same operon [6] (Figure 7A). CARC, an AIDA-like autotransporter from *C. rodentium* is hyper-heptosylated by its cognate heptosyltransferase BAHTCr (Figure 7A), and this modification is essential for colonization in mice [6]. In addition, BAHTs have also been identified in *Salmonella enterica* serovar Urbana R8-2977, *Shigella sp.* D9, *Laribacter hongkongensis* HLHK9, *Cronobacter sakazakii* ATCC BAA-894, and *Burkholderia spp.*, and these BAHT proteins share 40%–80% homology to AAH [6].

The sugar donors utilized by BAHT for the heptosylation of autotransporters are recruited from the LPS biosynthetic pathway, in which ADP-DD-Hep and ADP-LD-Hep are synthesized (Figure 7B). Both ADP-DD-Hep and ADP-LD-Hep could serve as the sugar donors for the glycosylation of autotransporters [6]. The pathway for the stepwise heptosylation of autotransporters by BAHT is summarized in (Figure 7C).

5. The "Mix and Match" of Three Independent Glycosylation Pathways for Glycoengineering

In Gram-negative bacteria, the biosynthetic pathways of OST-mediated protein glycosylation, LPS and peptidoglycan synthesis are independent, but all follow a similar route: 1) using UndP as a lipid carrier to form LLO; 2) the translocation of LLO to the periplasm by flippases and; 3) the ligation of glycans to acceptors by glycosyltransferases. An evolutionary connection between LPS synthesis and
OST-mediated protein glycosylation has been suggested [57, 85]. For example, the OST-mediated \(N\)-glycosylation flippase PglK in \(C.\) \(jejuni\) shares 37% identity with the LPS \(O\)-antigen flippase Wzk in \(H.\) \(pylori\), and both enzymes are functionally exchangeable [85]. More importantly, both LPS \(O\)-antigen ligases (WaaL) and \(O\)-OSTs contain the periplasmic domain Wzy_C motifs, however based on current knowledge it is impossible to precisely differentiate based on protein sequences alone whether a Wzy_C motif containing peptide functions as a ligase or \(O\)-OST [57]. Remarkably, the lipid-linked \(O\)-antigen 5\(\beta\)OH-C47NFmPse-Xyl-FucNAc trisaccharide unit in \(P.\) \(aeruginosa\) 1244 is utilized in both the LPS and protein glycosylation pathways [20, 57]. In the LPS pathway, the ligase WaaL is responsible for the addition of the \(O\)-antigen to the lipid A-core, whereas in the protein glycosylation pathway, the OST PilO is responsible for the addition of a single \(O\)-antigen unit to pilin protein PilA [20, 57].

By taking advantage of the relaxed substrate specificities of the OST, the three independent glycosylation pathways, protein glycosylation, LPS biosynthesis and peptidoglycan biosynthesis, have been "mixed and matched" for glycoengineering a number of recombinant glycoproteins in \(E.\) \(coli\) and \(S.\) \(enterica\) [86-90]. It has been shown that \(C.\) \(jejuni\) PglB (\(N\)-OST) is capable of transferring various LPS \(O\)-antigens from \(E.\) \(coli\) O16, \(P.\) \(aeruginosa\) O11, and \(Shigella\) \(dysenteriae\) O1 onto the protein acceptor AcrA in \(E.\) \(coli\) cells [87, 88]. The commonality among these \(O\)-antigen oligosaccharides is the \(N\)-acetylhexosamine moiety of their reducing sugars (GlcNAc, FucNAc, GalNAc). In contrast, \(O\)-antigen oligosaccharides with Gal as the reducing sugar could not be transferred by PglB to AcrA [87]. Recently, the X-ray structure of PglB has been resolved and it is thought that the \(N\)-acetylhexosamine moiety plays a role in the glycosylation mechanism [91]. The relaxed substrate specificities of \(O\)-OSTs PilO from \(P.\) \(aeruginosa\) 1244 and PglL from \(N.\) \(meningitidis\) have also been demonstrated in \(E.\) \(coli\) [22, 89]. In contrast to PglB, PglL does not require the 2-acetamido group at the reducing end for activity [74, 89] and so is able to transfer virtually any glycan (including \(C.\) \(jejuni\) heptasaccharide, \(E.\) \(coli\) O7 antigen, \(E.\) \(coli\)
K30 capsular structure, *S. enterica* O-antigen and *E. coli* O16 peptidoglycan subunits) to pilin in both *E. coli* and *Salmonella* [22, 89].

6. Conclusions and Future Perspectives

Through advances in genetics, genomics and glycobiology in the past decade, tremendous progress has been made in the field of bacterial protein glycosylation. The identification and elucidation of the complexities of bacterial glycosylation mechanisms have enabled the successful reconstitution of OST-mediated N-glycosylation systems of *C. jejuni*, and OST-mediated O-glycosylation systems of *P. aeruginosa* and *N. meningitidis* in *E. coli* and *S. enterica*, leading to the production of different recombinant glycoproteins. It is expected that with the ever-increasing repertoire of GTs and OSTs, together with the further understanding and optimization of bacterial glycoengineering systems, the large-scale and cost-effective production of glycoconjugate therapeutics or vaccines in bacteria is not far away from reality.
Executive Summary

- Protein glycosylation in bacteria plays key functions in biological processes including adhesion, immune evasion and host colonization.
- Both bacterial protein N-glycosylation and O-glycosylation can proceed via lipid carrier-mediated “en bloc” pathway or cytoplasmic stepwise pathway.
- The glycans of bacterial glycoproteins are more complex than their eukaryotic counterparts and many bacterial protein glycans are not fully characterized.
- OST-mediated protein glycosylation, LPS biosynthesis and peptidoglycan biosynthesis pathways (although independent but all proceed via the lipid carrier-mediated “en bloc” pathway), have been "mixed and matched" for glycoengineering a number of recombinant glycoproteins in bacteria.
References

*of interest

**of considerable interest


8. Young NM. Structure of the N-Linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni.* *J. Biol. Chem.* 277(45), 42530-42539 (2002). **Determination of the N-linked glycan structure as a diNAcBac-containing heptasaccharide present on multiple proteins by the N-linked general glycosylation pathway in C. jejuni.**


**H. pylori.**


27. Fletcher CM, Coyne MJ, Villa OF, Chatzidaki-Livanis M, Comstock LE. A
General O-glycosylation system important to the physiology of a major human intestinal symbiont. *Cell.* 137(2), 321-331 (2009). **Demonstration of a general OST-mediated protein O-glycosylation systems in Bacteroides.**


38. Harding CM, Nasr MA, Kinsella RL *et al.* *Acinetobacter* strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple proteins. *Mol.*


49. Troutman JM, Imperiali B. Campylobacter jejuni PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. Biochemistry. 48(12), 2807-2816 (2009).


Pathogens. 1(1), 12-29 (2012).


Elucidation of biosynthetic pathway of sugar donor CMP-Leg utilized for flagellin glycosylation in *C. jejuni*.


87. Wacker M. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc Nati Acad Sci U S A.* 103(18), 7088-7093 (2006). **Demonstration of the use of bacteria for glycoengineering a number of recombinant proteins.**


89. Faridmoayer A, Fentabil MA, Haurat MF et al. Extreme substrate promiscuity of the neisseria oligosaccharyl transferase involved in protein O-glycosylation. *J. Biol. Chem.* 283(50), 34596-34604 (2008). **Demonstration of the OST PgL from *N. meningitidis* is able to transfer virtually any glycan including the *C. jejuni* N-linked glycan, the LPS O-antigen and peptidoglycan subunits to pilin in engineered *E. coli* and *Salmonella* cells.**


Figure Legends

Figure 1. Lipid Carrier-Mediated *en bloc* Protein Glycosylation Pathways

A. *En bloc* protein N-glycosylation pathway in *C. jejuni*. Three enzymes PglF, PglE and PglD are responsible for the conversion of UDP-GlcNAc to UDP-diNAcBac, which is the first sugar residue transferred to UndP carrier by PglC to form UndPP-linked diNAcBac. Further actions of GTs PglA, PglJ and PglH add five GalNAc residues and one branched Glc residue to complete the UndPP-linked heptosaccharide. Once completed, the LLO is flipped by the flippase PglK into the periplasm, where the OST PglB transfers the glycan to the Asn residue present in the protein consensus sequence: Asp/Glc-X1-Asn-X2-Ser/Thr(D/E-X1-N-X2-S/T) (X1 and X2 represent any amino acid except proline).

B. *En bloc* protein O-glycosylation pathway in *N. meningitidis* and *N. gonorrhoeae*. The sequential actions of PglD, PglC, and PglB or PglB2 are involved in the conversion of UDP-GlcNAc to UDP-DATDH (by PglB) or UDP-GATDH (by PglB2). The bifunctional PglB and PglB2 also function to transfer DATDH and GATDH to the lipid carrier, respectively. Further actions of GTs PglA and PglE add Gal residues to complete the UndPP-linked trisaccharide. The LLO is then flipped by the flippase PglF into the periplasm, where the OST PglL in *N. meningitidis* or the OST PglO in *N. gonorrhoeae* transfers the glycan to Ser/Thr (S/T) residues of target proteins.

C. *En bloc* protein O-glycosylation pathway in *P. aeruginosa* 1244. It is proposed that WbpM and WbpK are responsible for the conversion of UDP-GlcNAc to UDP-FucNAc, which is the first sugar transferred to the lipid carrier by WbpL. Further actions of unknown GTs add Xyl and α-5NβOH47NFmPse residues to complete the UndPP-linked trisaccharide. The LLO is then flipped by the O-antigen flippase Wzx into the periplasm, where the OST PilO transfers the glycan to the Ser/Thr (S/T) residues of PilA.

D. *En bloc* protein O-glycosylation pathway in *A. baumannii* ATCC 17978. It is proposed that PglC transfers the first sugar GalNAc to the lipid carrier. Further actions of four putative GTs (A1S_0058, A1S_3482, A1S_0059, and A1S_0060) add Glc, GlcNAc, Gal and UDP-GlcNAc3NAcA4OAc to complete the UndPP-linked pentasaccharide. The LLO is then flipped by the O-antigen flippase Wzx (A1S_0056) into the periplasm, where the OST PglL (A1S_3176) transfers the glycan to the Ser/Thr (S/T) residues of target proteins. The biosynthesis of UDP-GlcNAc3NAcA4OAc from UDP-GlcNAc is proposed to be

**Figure 2. The Biosynthetic Pathway for UDP-diNAcBac in *C. jejuni***
The biosynthesis of UDP-diNAcBac involves three steps starting from UDP-GlcNAc. Step 1: the conversion of UDP-GlcNAc to UDP-2-acetamido-2,6-dideoxy-α-D-xylo-4-hexulose by dehydratase PglF (Cj1120c); Step 2: the transfer of an amino group by aminotransferase PglE (Cj1121c), forming UDP-4-amino-4,6-dideoxy-α-D-GlcNAc; Step 3 is catalyzed by acetyl transferase PglD (Cj1123c).

**Figure 3. The Biosynthetic Pathway for UDP-DATDH and UDP-GATDH in *N. meningitidis* and *N. gonorrhoeae***.
The biosynthesis of UDP-DATDH and UDP-GATDH involves three steps starting from UDP-GlcNAc. Step 1 and Step 2 are the same as the first two steps of the biosynthesis of UDP-diNAcBac in *C. jejuni* (Figure 2), by PglD (homologous to the *C. jejuni* PglF) and PglC (homologous to the *C. jejuni* PglE), respectively; Step 3 is catalyzed by the acetyltransferase function of PglB or PglB2 to form UDP-DATDH or UDP-GATDH, respectively.*the structure of DATDH has been shown to be diNAcBac [53].

**Figure 4. Stepwise Protein Glycosylation Pathways**
A. Stepwise protein N-glycosylation pathway in *H. influenzae*. The nucleotide activated sugars UDP-Gal and UDP-Glc are individually transferred to the eukaryotic-like sequon: Asn-X-Ser/Thr of HMW1 by the GT HMW1C.

B. Stepwise protein O-glycosylation of flagellins in *C. jejuni*. The nucleotide activated sugars CMP-Pse and CMP-Leg, and their derivatives are sequentially added to the serine or threonine residues of FlaA of *C. jejuni*.

C. Stepwise protein O-glycosylation of flagellins in *H. pylori*. The nucleotide activated sugar CMP-Pse is individually transferred to the serine or threonine residues of FlaA and FlaB of *H. pylori*.

**Figure 5. The Biosynthetic Pathway for CMP-Pse in *H. pylori***
The biosynthesis of CMP-Pse involves six steps starting from UDP-GlcNAc. Step 1: the dehydration and epimerization of GlcNAc by the bi-functional 4,6-dehydratase/5-epimerase PseB (HP0840), forming UDP-2-acetamido-2,6-dideoxy-β-L-arabino-hexose-4-ulose; Step 2: the aminotransfer by the aminotransferase PseC (HP0366), forming
UDP-4-amino-4,6-dideoxy-\(N\)-acetyl-\(\beta\)-L-altrosamine; Step 3: the acylation by the \(N\)-acetyltransferase PseH (HP0327), forming UDP-2,4-bis(acetamido)-2,4,6-trideoxy-\(\beta\)-L-altropyranose; Step 4: the removal of UDP by the nucleotidase PseG (HP0326B); Step 5: the synthesis of Pse by the Pse synthetase PseI (HP0178); Step 6: the production of CMP-Pse by the CMP-Pse synthetase PseF (HP0326A).

Figure 6. The Biosynthetic Pathway for CMP-Leg in *C. jejuni*

The biosynthesis of CMP-Leg involves ten steps starting from Fru-6\(P\) (I). Step 1: Fru-6\(P\) (I) is converted to GlcN-6\(P\) (II) by the actions of the glutaminase PtmA (Cj1332) and the isomerase PtmF (Cj1330); Step 2: the conversion of GlcN-6\(P\) to GlcN-1\(P\) (III) by the mutase PgmL (Cj1407c); Step 3: the formation of GDP-GlcN (IV) by the guanyltransferase PtmE (Cj1329); Step 4: the formation of GDP-GlcNAc (V) by the action of \(N\)-acetyltransferase GlmU (Cj0821); Step 5: the dehydration of V by the dehydratase LegB (Cj1319), forming GDP-2-acetamido-2,6-dideoxy-\(\alpha\)-D-xylo-hexos-4-ulose (VI); Step 6: the aminotransfer of VI by the aminotransferase LegC (Cj1320), forming GDP-4-amino-4,6-dIDEOXY-\(\alpha\)-D-GlcNAc (VII); Step 7: the \(N\)-acetyltransfer of VII by the \(N\)-acetyltransferase LegH (Cj1298), forming GDP-2,4-diacetamido-2,4,6-trideoxy-\(\alpha\)-D-glucopyranose (VIII); Step 8: the removal of NDP from VIII by the NDP-sugar hydrolase/2-epimerase LegG (Cj1328), forming 2,4-diacetamido-2,4,6-trideoxy-\(D\)-mannopyranose (IX); Step 9: the condensation of IX with pyruvate by the Leg synthase LegI (Cj1327), forming Leg (X); Step 10: the CMP-activation of Leg by the CMP-Leg synthetase LegF (Cj1331), producing CMP-Leg (XI).

Figure 7. Stepwise Heptosylation of Autotransporters

A. The genomic organization of autotransporters and their corresponding BAHTs in three representative bacterial strains. The gene encoding the BAHT is always located upstream of its substrate autotransporter in the same operon of the genome.

B. The biosynthetic pathway for ADP-LD-Hep and ADP-DD-Hep in Gram-negative bacteria. Four steps are involved for the production of ADP-DD-Hep starting from D-sedo-heptulose-7\(P\). Step 1: the conversion to DD-Hep-7\(P\) by isomerase GmhA; Step 2: the conversion to DD-Hep-1,7\(PP\) by the bi-functional RfaE (kinase/adenosyltransferase); Step 3: the conversion to DD-Hep-1\(P\) by phosphatase GmhB; Step 4: the production of ADP-DD-Hep by the bi-functional RfaE (kinase/adenosyltransferase). For the production of ADP-LD-Hep, it is
converted from ADP-DD-Hep by epimerase RfaD.

C. The pathway for the stepwise heptosylation of autotransporters. Prior to its secretion, the autotransporter is heptosylated by the heptosyltransferase BAHT on numerous Ser residues in the passenger domain. Both ADP-DD-Hep and ADP-LD-Hep derived from LPS biosynthetic pathway can serve as the sugar donors for the heptosylation of autotransporter by BAHT.
Figures

Figure 1
Figure 2
Figure 3

UDP-GlcNAc → PglD → UDP-2-acetamido-2,6-dideoxy-α-D-xylo-4-hexulose → PglC → UDP-4-amino-4,6-dideoxy-α-D-GlcNAc → PglB → UDP-DATDH* (UDP-diNAcBac) → UDP-GATDH

Figure 3
Figure 4
Figure 5

UDP-GlcNAc

\[ \text{PseB (HP0840)} \]

\[ \text{UDP-2-acetamido-2,6-dideoxy-\(\beta\)-L-arabino-hexos-4-ulse} \]

\[ \text{PseC (HP0366)} \]

\[ \text{UDP-4-amino-4,6-dideoxy-\(N\)-acetyl-\(\beta\)-L-altrosamine} \]

\[ \text{PseH (HP0327)} \]

\[ \text{UDP-2,4-bis(acetamido)-2,4,6-trideoxy-\(\beta\)-L-altropyranose} \]

\[ \text{PseG (HP0326B)} \]

\[ \text{2,4-bis(acetamido)-2,4,6-trideoxy-\(\beta\)-L-altropyranose} \]

\[ \text{PseI (HP0178)} \]

\[ \text{Pse} \]

\[ \text{PseF (HP0326A)} \]

\[ \text{CMP-Pse} \]
Figure 7