Flagellar Glycosylation: Current Advances

Jumpei Hayakawa and Morio Ishizuka

1. Introduction

In this chapter, we present the current advances in flagellar glycosylation. Glycosylation is well-known as one of the most frequent posttranslational protein modification. Glycosylation is well studied in eukaryotes as the superficial and secretory proteins are mostly glycosylated in the eukaryotic cell. Protein glycosylation was considered to be a eukaryotic organism specific modification for many years. However, reports of bacterial glycosylation have increased since the discovery of surface layer glycosylation on the cell envelope in archaea and hyperthermophiles in the mid-1970's (Mescher & Strominger, 1976; Sleytr, 1975; Sleytr & Thorne, 1976).

1.1. Protein glycosylation

Protein glycosylation is largely classified as N-linked or O-linked while C-mannosylation is rarely identified (Furmanek & Hofsteenge, 2000). Glycan structures are enzymatically transferred to amino acid residues where they can covalently conjugate via the amino group of asparagine residues (N-glycosylation) or the hydroxyl group of serine or threonine residues (O-glycosylation). Both linkage types are distributed in eukaryotes and prokaryotes. The N-linkage glycosylation pathway is characterized in all three domains of life (eukarya, archaea, and bacteria) (Calo et al., 2010; Haeuptle & Hennet, 2009; Szymanski & Wren, 2005; Weerapana & Imperiali, 2006). The carbohydrate chain is synthesized at the membrane (endoplasmic reticulum (ER) in eukarya, or on the cytoplasmic side of the plasma membrane in archaea and bacteria) via a specific glycosyltransferase which transfers a nucleotide-activated sugar precursor onto the lipid carrier (dolichol-phosphate in eukarya and archaea, or undecaprenyl-phosphate in bacteria). The synthesized carbohydrate chain (oligosaccharide) is flipped across the membrane using a specific flippase, and is transferred to the asparagine residue in the nascent protein en bloc by an oligosaccharyltransferase (OST), which is composed of nine subunits in eukarya. Archaeal and bacterial OST are encoded by the aglB (Abu-Qarn et al., 2007) or pglB (Wacker et al., 2002) gene to yield a...
single protein, respectively. In eukarya and archaea, the asparagine residues of the N-linkage glycosylation site are conserved in a sequon, the N-X-S/T motif (where X is any amino acid except proline). Recently, the bacterial N-linkage sequon was characterized from *Campylobacter jejuni* (Kowarik et al., 2006). The two N terminal extended residues, D/E-X1-N-X2-S/T (where X1 and X2 are any amino acid except proline), were required for the bacterial OST recognition of the glycosylation site.

In eukaryote, glycoproteins typically possess a pentasaccharide core carbohydrate structures which consist of (Man)_2-Man-GlcNAc-GlcNAc with one or more glycan chain (N-glycosylation) and di-, or trisaccharide core structure based on GalNAc attached to serine or threonine (O-glycosylation). However, recent reports on *C. jejuni*, *Haemophilus influenzae*, and *Desulfovibrio desulfuricans* N-glycosylations indicate that there is no conserved core structure in bacteria (Gross et al., 2008; Ielmini & Feldman, 2011; Young et al., 2002). Thus, in prokaryotes many different glycoprotein structures have been observed that display much more variation than those observed in eukaryotes.

Glycoproteins have many biological functions: one example is recognition and adhesion among cells (Varki, 1993). The interactions between cells are mediated by the glycan structures on the cell surfaces. Therefore, the different glycan moieties on the cell surfaces serve as markers for cell recognition events, and modifications of the glycan structures can render several biological functions to the protein in eukarya. In recent years, accumulating studies for glycosylated bacterial proteins indicate that glycan structures mainly participate in the virulence of the mucosal pathogen (Szymanski & Wren, 2005). Most bacterial glycoproteins appear to be associated with the surface of the organism as in pili or flagella. Flagellin is one of the extensively studied glycosylated bacterial proteins and it is suggested that the flagellin glycosylation is responsible for their virulence, adherence, filament assembly, and filament stability (Arora et al., 2005; Goon et al., 2002; Szymanski et al., 2002; Taguchi et al., 2008).

### 1.2. Flagellar structures

Most bacterial species swim by means of rotating flagella that are powered by the monovalent cation (H⁺ or Na⁺) influx. Many bacteria have extracellular flagellar structures and the pattern of flagellar arrangement is an identification tool in bacteria. A variety of flagella structures and swimming patterns have been discussed in previous works (Armitage & Macnab, 1987; Bardy et al., 2003; Charon & Goldstein, 2002; Macnab, 1977, McCarter, 2001, 2004; Shigematsu et al., 2005). There are classified in-to four flagella arrangements as follows: multi-flagella are randomly distributed on the overall cell surface (peritrichous; e.g. *Escherichia coli*, *Salmonella typhimurium*, *Bacillus subtilis*), several flagella are present at one end of the cell (lophotrichous; e.g. *Pseudomonas fluorescens*), a single polar flagella is projected from the cell end (monotrichous; e.g. *Vibrio cholerae*, *Rhodobacter sphaeroides*), and a single flagellum is present at each pole of the bacterium (amphitrichous; e.g. *Spirillum serpens*). Although bacteria possess different flagella arrangement, a basic flagella structure is common among many bacterial species. The study of bacterial flagella
has been intensively investigated in *Escherichia coli* and *Salmonella typhimurium* by using genetic and biochemical approaches. A typical flagella structure is shown in Figure 1. About 50 genes, which are related to the bacterial flagellar assembly, have been identified. More than 20 distinct proteins make up the flagella structure and it consist of three main parts: a basal body, a hook, and a filament. The basal body consists of four rings (L, P, MS, and C-ring) and rod which is located just above the MS-ring and connected to the proximal region of the hook. However, the L and P ring are not observed in gram-positive bacterial species, because of the difference in the cell wall architecture (DePamphilis & Adler, 1971; Francis et al., 1995). The C-ring which is a part of flagella rotor consists of three proteins, FliG, FliM, and FliN (in gram-positive bacteria FliY correspond to FliN). In particular, FliG is the most directly involved in the rotation of the flagella motor among the C-ring proteins, as it interact with the motor complex (MotA/B or PomA/B) component of the force-generating unit in the flagellar motor. The motor complex acts as the stator and the rotation energy is generated by the monovalent cation influx from the periplasmic space across the inner membrane. The C-ring is also called the “switch complex” as it can switch the direction of flagellar motor (Irikura et al., 1993; Kihara et al., 1996; Sockett et al., 1992). The switching event is caused by the binding of phosphorylated CheY (chemotaxis related gene) to the FliM of switch complex, and clockwise (CW)/counterclockwise (CCW) switching enable the bacterial cell to change swimming direction (Mathews et al., 1998; Sockett et al., 1992).

![Figure 1. The model structure of the flagellar motor in gram-negative bacteria.](image)

The motor consists of the Mot complex (MotA/B) and rotor (MS- and C-ring). The L- and P-ring do not exist in the gram-positive bacterial flagellar structure. The Mot complex is supposed to function as the force-generating unit via proton conduction while the C-ring functions as the switch. The phosphorylated form of CheY (CheY-P) interacts with FliM and
promotes CW rotation. When CheY-P is not bound to it, the motor rotates CCW. Flagellin subunits are transported from the cytoplasm and are delivered into a central channel in the basal body–hook filament structure (the diameter of the central channel is only 2 nm). OM, Outer membrane; PG, peptidoglycan layer; CM, cytoplasmic membrane (Irikura et al., 1993; Mathews et al., 1998).

The most impressive structure of the bacterial flagella motor is an extracellular long helical filament. In general, the flagella filament is composed of 20,000 ~ 30,000 subunits of a single protein called flagellin, and it reaches to more than 10 μm in length (Namba & Vonderviszt, 1997). The flagellar specific export apparatus is located on the inside of the C-ring, and most of the flagellar components proteins are translocated across the cytoplasmic membrane by this apparatus, and then the proteins are diffused in the narrow nascent lumen structure and self-assemble at the distal end of the flagellar structure (Aizawa, 1996).

Flagellar-based motility is also common to archaea, but its structural features are quite distinct from bacteria. Archaeal flagella are closely related to bacterial type IV pili in their structure and assembly and the origin of bacterial flagella is considered to be a type III secretion system (Bardy et al., 2004). In bacteria, the flagellar filament is composed of a single flagellin subunit, in contrast, two or more distinct flagellin subunits are require for production of the flagellar filament. The other notable differences include that archaeal flagella rotation is powered by ATP, the flagellin subunit has a signal peptide which is cleaved by a specific peptidase for the secret matured flagellin subunit from the cell, and the flagella is grown from the proximal end of the cell surface by the addition of subunits to the base.

1.3. Flagellin glycosylations

Although flagella structures from both eubacteria and archaea are different, flagellin glycosylation is reported in both organisms. Eubacterial flagellin glycosylations are classified as either N- or O-linkage in a single subunit, and to date, most reports are about O-linkages. The O-linked glycan positions of bacterial flagellin proteins appear to be limited to the central region of the primary flagellin structure. The amino acid sequence alignment indicates that flagellin proteins are well conserved in the N- and C-terminal regions, while the central region is highly variable (Beatson et al., 2006). Although the intensively studied peritrichous flagella from *Salmonella typhimurium* do not have a glycosylated flagellin, a complete atomic model of its flagellin protein was resolved by excellent X-ray crystallography analysis and electron cryomicroscopic observation of the intact flagella filament (Samatey et al., 2001; Yonekura et al., 2003, 2005). These structural data revealed that the *S. typhimurium* flagellin protein consists of four major domains, D0, D1, D2, and D3 (Figure 2). The N- and C-terminal regions of flagellin correspond to D0 and D1, which are composed mainly of α-helices, and they form the core part of a flagellar filament. D0 has two α-helices (ND0 and CD0), whereas D1 has two long α-helices (ND1a and CD1), one short α-helix (ND1b) and one short β-sheet. These domains are positioned in the filament core. Gugolya et al. suggested that the alpha-helical terminal regions that correspond to the D0 domains are important for the coiled-coil model of flagellar filament formation (Gugolya et al., 2003). The central variable region of flagellin forms the outside surface-exposed
domain (D2 and D3) in the assembled filament. Studies of the variable region have focused on its role in H antigenicity, the effect of deletions on filament formation and motility, and the insertion of foreign peptides for extracellular display on bacterial flagella (Malapaka et al., 2007; Reid et al., 1999; Westerlund-Wikström, 2000; Woodset al., 2007; Yoshioka et al., 1995). Thus, flagellin glycan structures are restricted in the D2 and D3 domains (a few glycans were located at the D2 proximal end of the D1 domain), and it is considered that these glycan moieties are exposed to environmental conditions (Figure 3).

Figure 2. Bacterial flagellar filament.

Flagellar filament structure and complete flagellin 3D model of Salmonella typhimurium (Samatey et al., 2001; Yonekura et al., 2003, 2005). A flagellar filament consists of substantial amount of flagellin subunit and takes on a tubular structure.

Glycosylated flagellins are schematically aligned with the S. typhimurium flagellin domain. Light and dark grey boxes indicate 100 highly conserved terminal amino acid residues which constitute the alpha-helical structures. The corresponding amino acid size of the flagellin proteins is indicated at the C-terminal. Positions of glycosylation on each flagellin are indicated (●).

Figure 3. O-linked glycosylation sites of bacterial flagellin.
Glycosylation

In contrast with eubacteria, three archaeal flagellin glycosylations are reported as N-linkage (Chaban et al., 2007; Voisin et al., 2005; Wieland & Sumper, 1985). Work on the most extensively studied flagellin from Methanococcus voltae demonstrated that the glycan attached positions are not limited to anywhere specific in the flagellin primary structure and the N-linked asparagine residues seem to follow the classic eukaryotic type consensus sequon (N-X-S/T) rather than the recently identified bacterial N-linkage sequence (D/E-X-N-X-S/T).

2. Flagellar glycosylation

There have been many reports on flagella glycosylation since it was discovered about 20 years ago. Flagellin glycosylation is mainly found in gram-negative pathogenic bacterial species, and has been identified in about 30 microorganism strains including the archaea and gram-positive species (Logan, 2006). The distribution of flagellin glycosylation among several species is shown in Table 1, and a gene cluster which is potentially involved in the posttranslational modification of flagellin glycosylation is shown in Figure 4.

**Figure 4.** Organization of the glycosylation island located around the flagellin gene.

The location of glycosylation islands was not restricted to directly upstream or downstream of the flagellin gene, and the component genes were highly-diverse. A glycosyltransferase, which is responsible for the glycan attachment to a flagellin protein is usually included in the proximal glycosylation island of a flagellin gene, whereas, it was not identified to date in this region in *C. jejuni.*
<table>
<thead>
<tr>
<th>Organism</th>
<th>linkage type</th>
<th>Glycan characterization</th>
<th>Function</th>
<th>GTase*5 (OST)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gram-negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeromonas caviae Sch3</td>
<td>O</td>
<td>Pse5Ac7Ac</td>
<td>Motility</td>
<td>n.d.</td>
<td>Tabei et al., (2009)</td>
</tr>
<tr>
<td>Campylobacter jejuni 81-176</td>
<td>O</td>
<td>Pse5Ac7Ac, PseAm, Pse8OAc, PseAmGlnAc</td>
<td>Assembly</td>
<td>n.d.</td>
<td>Thibault et al., (2001)</td>
</tr>
<tr>
<td>Campylobacter jejuni 11168</td>
<td>O</td>
<td>Leg5Am7Ac, Leg5AmNMe7Ac</td>
<td>n.d.</td>
<td>n.d.</td>
<td>Zampronio et al., (2011)</td>
</tr>
<tr>
<td>Campylobacter coli VC167</td>
<td>O</td>
<td>Pse5Ac7Ac, PseAm, Pse/PseAm-deoxypentose, Leg5Ac7Ac, Leg5Am7Ac, Leg5AmNMe7Ac</td>
<td>Assembly</td>
<td>n.d.</td>
<td>Logan et al., (2002)</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa PAK</td>
<td>O</td>
<td>Rha-(2-7 variable oligosaccharide chain)-deoxyhexosamine (dHexN)-deoxyhexose (dHex)</td>
<td>Virulence</td>
<td>FgtA</td>
<td>Schirm et al., (2004a)</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa PAO1</td>
<td>O</td>
<td>dHex(P0)-192Da</td>
<td>Virulence</td>
<td>FgtA</td>
<td>Verma et al., (2006)</td>
</tr>
</tbody>
</table>
### Table 1. Flagellin glycosylations

<table>
<thead>
<tr>
<th>Organism</th>
<th>linkage type</th>
<th>Glycan characterization</th>
<th>Function</th>
<th>GTase</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Clostridium difficile</em></td>
<td>O</td>
<td>HexNAc</td>
<td>Assembly</td>
<td>n.d.</td>
<td>Twine et al., (2009)</td>
</tr>
<tr>
<td><em>Clostridium botulinum</em></td>
<td>O</td>
<td>αLeg5GluNMe7Ac</td>
<td>n.d.</td>
<td>n.d.</td>
<td>Twine et al., (2008)</td>
</tr>
<tr>
<td><em>Geobacillus stearothermophilus</em></td>
<td>O</td>
<td>PAS, β-elimination</td>
<td>Assembly</td>
<td>n.d.</td>
<td>Hayakawa et al., (2009a)</td>
</tr>
<tr>
<td><em>Bacillus sp.</em> PS3</td>
<td>O</td>
<td>PAS, β-elimination</td>
<td>Assembly</td>
<td>n.d.</td>
<td>Hayakawa et al., (2009a)</td>
</tr>
<tr>
<td><strong>Archaea</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Methanococcus voltae</em></td>
<td>N</td>
<td>β-ManpNAcA6Thr-(1-4)-β-</td>
<td>Assembly</td>
<td>AglA</td>
<td>Voisin et al., (2005)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GlcNAc3NAcA-(1-3)-β-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GlcNAc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Methanococcus maripaludis</em></td>
<td>N</td>
<td>Sug-4-β-ManpNAc3NAAcA6Thr-4-β-GlcNAc3NAcA-3-β-GalNAc</td>
<td>Motility</td>
<td>AglB</td>
<td>VanDyke et al., (2009)</td>
</tr>
</tbody>
</table>

*1 not determined; *2 Periodic acid stain; Carbohydrate specific staining method; *3 Removal of Ser/Thr binding carbohydrate structure; *4 Protein anomalous migration on SDS-PAGE; *5 Responsible for glycan attachment to flagellin.

#### 2.1. Gram-negative

**2.1.1. Pseudomonas spp.**

*Pseudomonas* spp. are ubiquitous in nature and frequently isolated as opportunistic pathogen of both plant and animal. *P. aeruginosa* has a single polar flagellum which is classified by its type of flagella filament (type-a and type-b) by flagellin subunit size, amino acid sequence, and antigenicity (Allison et al., 1985; Lanyi, 1970). Both types of *P. aeruginosa* flagellin are known to contain O-linked glycosylated proteins. *P. aeruginosa* PAK and *P. aeruginosa* JJ692 produce glycosylated type-a flagellin protein, which is modified with a rhamnose (Rha) residue based on the glycan attached at two sites of each flagellin monomer (Schirm et al., 2004a). The glycan structure of PAK flagellin is a complex oligosaccharide which is composed of Rha-(2-7 variable oligosaccharide chain)-deoxyhexosamine (dHexN)-deoxyhexose (dHex), whereas JJ692 flagellin has only a single Rha glycosylation on both glycosylation site. The glycan form of PAO type-b flagellin is simpler than that of the PAK type-a flagellin, as it has a dHex linked sugar containing a
phosphate moiety at two sites of the flagellin monomer (Verma et al., 2006). With regards to the *Pseudomonas* spp. of plant pathogens, flagellin glycosylation from *Pseudomonas syringae* pv. *glycinea*, *Pseudomonas syringae* pv. *toma*, and *Pseudomonas syringae* pv. *tabaci* 6605 have been identified (Taguchi et al., 2003, and Takeuchi et al., 2003). The structural characterization of the flagellin protein from *P. syringae* pv. *tabaci* 6605 revealed that six sites of the flagellin were modified with a novel trisaccharide, which was composed of two rhamnosyl (Rha) residues and one modified 4-amino-4,6-dideoxyglucosyl (Qui4N; trivial name, viosamine; Vio) residue, \( \beta\)-D-Quip4N(3-hydroxy-1-oxobutyl)2Me-(1-3)-\( \alpha\)-L-Rhap-(1-2)-\( \alpha\)-L-Rhap (Takeuchi et al., 2007). The flagella glycosylation island of these *Pseudomonas* spp. have been identified and located in the upstream region of their flagellin gene (Arora et al., 2001; Taguchi et al., 2006) (Figure 4). The PAK glycosylation island is composed of 14 ORFs (~16 kb) containing putative carbohydrate synthesis related genes and glycosyltransferase (orfN). In contrast, both the PAO and *syringae* pv. *tabaci* glycosylation island are more simple (only 4 and 3 genes, respectively), and a putative glycosyltransferase is encoded for each. Functional analysis of these glycosyltransferases demonstrated that they are essential for the addition of glycan structure to flagellin protein (Schirm et al., 2004a; Verma et al., 2006; Taguchi et al., 2006). Recently, a flagellin glycan biosynthesis gene cluster was newly identified from *P. syringae* pv. *tabaci* (Nguyen et al., 2009). The gene cluster is related to viosamine biosynthesis (viosamine island) and these genes are homologous to a part of the PAK glycosylation island (orfA-E and orfG) (Chiku et al, 2011). Mutagenesis analysis of glycosyltransferases and flagellin subunits demonstrated that flagellin glycosylation of PAK and PAO was not require for flagella biosynthesis and motility, but a remarkable reduction of virulence was observed upon mutation (Montie et al., 1982; Arora et al., 2005). Whereas, in *P. syringae* pv. *tabaci*, loss of flagellin glycosylation reduce not only virulence but also motility, in addition, mutations which resulted in a loss of glycosylation showed differences in the bundle formation of flagella, i.e. flagella bundles on the wild-type cell were loose, and in contrast mutant filaments seemed to be tightly interacting with each other. These results indicated that glycosylation stabilizes the filament structure and lubricates the rotation of the bundle (Taguchi et al, 2008, 2010). A similar conclusion was drawn for the glycan function of flagellin in the marine magnetotactic ovoid bacterium MO-1. Flagella bundles of MO-1 were enclosed with in sheaths structure and its glycosylation was required for smooth swimming (Lefèvre et al., 2010). Flagellin proteins were also glycosylated and each flagella bundle consisted of seven individual flagella filament, which were organized in a hexagon with a seventh in the middle. Considering the compact arrangement of the seven flagella filaments in the bundle, flagellin glycosylation might function as a lubricant (Zhang et al., 2012). Recently, an fgt2 inactivation mutant from the biosurfactant producing species *P. syringae* pv. *syringae* B728a demonstrate upregulation of the late-stage flagellar genes (class IV), and increase surfactant production (Burch et al., 2012; Xu et al., 2012). The authors suggested that over-production of the biosurfactant helps smooth cell migration and minimize flagella breakage on sticky surfaces, such as a leaf surface.
2.1.2. *Campylobacter* spp.

*Campylobacter jejuni* have polar flagellum at one or both ends of the cell. The flagellin proteins are extensively O-glycosylated with structural analogues of the nine-carbon sugar pseudaminic acid (Pse), legionaminic acid (Leg), and their derivatives. Flagellin glycosylation is well characterized in three species of *Campylobacter*, i.e. *Campylobacter jejuni* 81-176, *C. jejuni* NCTC 11168, and *C. coli* VC167. Flagellin modification of *Campylobacter* species were identified at 19 serine or threonine residues, 16 sites, and at least 4 sites, in *C. jejuni* 81-176, *C. coli* VC167, and *C. jejuni* NCTC 11168, respectively (Thibault et al., 2001; Logan et al., 2002; Zampronio et al., 2011). The flagellin molecular weight from *C. jejuni* 81-176 is predicted by its amino-acid sequence to be 59.5 kDa, however flagellin from this strain is actually approximately 65 kDa. The additional 10% mass is attributed to attachment of substantial glycan structure (Thibault et al., 2001). In strain 81-176, the probable flagellin glycosylation related genes are largely involved in pseudaminic acid and acetamidino pseudaminic acid biosynthesis (Pse family), and lie downstream of the flagellin gene (about 27 kb) (Guerry et al., 2006). Similarly, Pse family glycosylation islands were identified in both *C. coli* VC167 and *C. jejuni* NCTC 11168, and in addition, in *C. coli* VC167 other flagellin modification genes which are involved in the synthesis of legionaminic acid and its derivatives (ptm family) were identified (McNally et al., 2007). However, the glycosyltransferase which catalyzes the addition of sugar residues to the protein backbone has not been identified. A mutation in the first step Pse biosynthesis gene leads to intercellular accumulation of unglycosylated flagellin protein (Goon et al., 2003). The biological roles of *Campylobacter* flagellin glycosylation have been mentioned in many reports. In 2007, an excellent review of the functions of flagellin glycosylation from *Campylobacter* species was published (Guerry, 2007).

2.1.3. *Helicobacter pylori*

*Helicobacter pylori* is a human gastric pathogen associated with gastric and duodenal ulcers as well as gastric cancer. Flagella and motility are important for colonization onto the mucosal of the human stomach. Two distinct flagellin subunits (FlaA and FlaB) were identified as glycosylated, and their glycan structures were characterized in a similar manner to that of *C. jejuni*, with Pse5Ac7Ac found at seven sites on FlaA and ten sites on FlaB, in addition flagellin glycosylation is required for functional filament assembly (Schirm et al., 2003, 2005; Josenhans et al., 2002). The mutagenesis analysis of four genes (HP0178, HP0326A, HP0326B, and HP0114) previously reported to be involved in flagellar glycosylation and polysaccharide biosynthesis demonstrated a non-motile phenotype with no structural flagella filament and only minor amounts of flagellin protein (Schirm et al., 2003). In contrast, inactivation of HP0518 resulted in altered motility and an increased level of flagellin glycosylation (Asakura et al., 2010). Complementation of a *H. pylori* HP0518 mutant and a recombinant HP0158 protein assay demonstrated the decreased glycosylation level of *H. pylori* flagellin in vivo and in vitro suggesting that HP0518 functions in the deglycosylation of flagellin. The *H. pylori* HP0518 mutant showed an increased colonization capability for the gastric tissues of mice. These results indicate that HP0518 is involved in the deglycosylation of flagellin, thereby regulating pathogen motility.
2.1.4. Burkholderia spp.

*Burkholderia pseudomallei* is also known as *Pseudomonas pseudomallei* and is important as a human and animal pathogen. In contrast, *Burkholderia thailandensis* is closely related to *B. pseudomallei* but is a nonpathogenic bacterium. Top-down and bottom-up mass spectrometry (MS) analyses of both flagellin proteins identified that there were posttranslationally modified with novel glycans (Scott et al., 2011). MS analysis of the flagellin carbohydrate moiety suggested that *B. pseudomallei* flagellin was modified with a glycan with a mass of 291 Da, while *B. thailandensis* flagellin protein was modified with related glycans with a mass of 300 or 342 Da which included an acetylated hexuronic acid. A mutagenesis analysis of the lipopolysaccharide (LPS) O-antigen biosynthetic cluster demonstrated that it was important for flagellin glycosylation and motility in *B. pseudomallei*.

2.2. Gram-positive

2.2.1. Clostridium

*Clostridium* spp., a gram-positive spore-forming anaerobic bacterium, is an emerging opportunistic pathogen towards humans and plants, and includes *Clostridium botulinum*, *Clostridium difficile*, and *Clostridium glutumae*. The genus *Clostridium* provides the most examples of gram-positive bacterium flagellin glycosylation which has been known since the discovery of *Clostridium tyrobutyricum* (Bédouet et al., 1998, Arnold et al., 1998). Structural characterization of the carbohydrate moiety from *C. botulinum* flagellin has been achieved, and it was shown to be composed of the Leg derivative, 7-acetamido-5-(N-methyl-glutam-4-yl)-amino-3,5,7,9-tetrahydro-d-glycero-α-d-galacto-nonulosonic acid (αLeg5GluNMe7Ac) (Twine et al., 2008). For the *C. botulinum* strain Langeland, a bioinformatic analysis of the flagella glycosylation island was completed between *flgB* and *fliD* as a large gene cluster (~48 kb), many of which appeared to be involved in carbohydrate biosynthesis (Sebaihia et al., 2007). This glycosylation island could be divided into two regions, a variable region which was located immediately downstream of the flagellin gene, and a subsequent conserved region. The carbohydrate biosynthesis genes, which are significantly related to the legionaminic acid biosynthesis genes (ptm family) in *Campylobacter coli*, were encoded in the variable region, whereas the conserved region also encoded the carbohydrate biosynthesis genes (McNally et al., 2007). The *C. botulinum* strain Langeland was found to have homologous proteins to the capsular biosynthetic proteins from *Streptococcus agalactia*, including those derived from a second set of the sialic acid biosynthetic genes, *neuA* and *neuB*.

2.2.2. Listeria monocytogenes

*Listeria monocytogenes* is a gram-positive bacterium responsible for listeriosis, and *Listeria* species are found throughout the food-processing environment. Flagellin subunits are covalently modified by monomeric β-O-linked N-acetylglucosamine (GlcNAC) residues at three to six sites per subunit (Schirm et al., 2004b). The functional consequence of flagellin
glycosylation in *L. monocytogenes* was investigated by modification of the O-GlcNAc transferase (Lmo0688 renamed to GmaR), which is located just upstream of the flagellin gene (Shen et al., 2006). An in-frame deletion mutant of lmo0688 (Δ688) resulted in a nonmotile bacteria similar to what was observed for *Campylobacter* species and *Helicobacter pylori* (Josenhans et al., 2002), but this phenotype differed from that reported for gram-negative species, as it was caused by a loss of flagellin expression. The point mutation analysis of the functional residues involved in the glycosyltransferase activity demonstrated full flagellin expression (without glycosylation) and motility. The authors concluded that GmaR is a bifunctional glycosyltransferase. However, glycosylation of flagellin is not required for any flagella functions and it remains to be determined what role glycosylation of the flagellin protein plays in *Listeria monocytogenes*.

### 2.2.3. Thermophilic Bacillus spp.

Thermophilic *Bacillus* species have been isolated from deepest sea mud, hot springs, and soil, and produce multiple peritrichous flagella. These thermophiles belong to genus *Geobacillus* and are not considered to be pathogens regardless of their flagellin glycosylation. In recent years, O-linked flagellin glycosylation was reported in two thermophilic *Bacillus* species, *Geobacillus stearothermophilus* NBRC 12550 and *Bacillus* sp. PS3 (Hayakawa et al., 2009a). These flagellin glycosylations were confirmed by PAS staining and beta-elimination. The analysis of the modification sites indicated that glycan structures were attached to at least 4 sites of the flagellin monomer in *Bacillus* sp. PS3, but the structural detail of the carbohydrate chains and total number of the modification sites is currently unknown. Although it was a partial sequence, the probable glycosylation islands from both thermophilic bacterial species were confirmed downstream of these flagellin genes (J. Hayakawa and M. Ishizuka, unpublished data). In *G. stearothermophilus*, a dTDP-L-rhamnose biosynthesis gene cluster (rml operon) was also identified immediately after GTases, which is highly homologous to the glycan biosynthesis genes of the S-layer glycoprotein from a closely related *G. stearothermophilus* strain (*G. stearothermophilus* NRS2004/3a) (Novotny et al., 2004 and Steine et al., 2007). The heterologous gene expression of these flagellin in a *Bacillus subtilis* flagellin deficient mutant demonstrated that unglycosylated flagellin proteins were intracellularly accumulated and phenotypically paralyzed (Hayakawa et al., 2009a), however amino acid substitutions could restore functional filament assembly and motility (Hayakawa et al., 2009b. described below). These results supported the proposal that flagellin glycosylation is important for filament assembly. However, the carbohydrate structure and more detail of the biological functions remain to be elucidated.

### 2.3. Archaea

Archaeal flagellin glycosylation was first identified in *Halobacterium salinarum* (Wieland et al., 1985). Its flagellin subunit was glycosylated with sulfated glucuronic acid which is the same type as the cell surface S-layer glycoprotein. The detailed structural characterization of flagellin attached carbohydrate was accomplished for *Methanococcus voltae* (Voisin et al.,
Flagellar Glycosylation: Current Advances

2005). *M. voltae* flagellin proteins were modified with a novel trisaccharide, β-ManpNAcA6Thr-(1-4)-β-GlcNAc3NAcA-(1-3)-β-GlcNAc, N-linked to Asn. In addition, the peptide containing the N-linked sequence motif of the flagellin protein was Asn-X-Ser/Thr, which is identical to that observed for S-layer protein glycosylation. Recently, a tetrasaccharide glycan which was N-linked to the flagellin subunits in *M. maripaludis* was also characterized, with a reported structure of Sug-4-β-ManNAc3NAcA6Thr-4-β-GlcNAc3NAcA-3-β-GalNAc, where Sug is a (5S)-2-acetamido-2,4-dideoxy-5-O-methyl-α-l-erythro-hexos-5-ulo-1,5-pyranose, representing the first example of a naturally occurring diglycoside of an aldulose (Kelly et al., 2009, Jones et al., 2012). A deletion mutant analysis of three glycosyltransferases and an oligosaccharyltransferase (Stt3p homologue) from *M. maripaludis* revealed that these genes were responsible for flagellin glycosylation supported by the fact that glycan reduced flagellins were not assembled into the flagella filament (VanDyke et al., 2009). The structural and genetic analysis of archaeal flagellin glycosylation is frequently linked with S-layer protein glycosylation, and the reader is referred to a recent detailed review (Jarrell et al., 2010).

3. Glycosylation pathway

The complete pathway of bacterial flagellin glycosylation is still not clarified. There are two reviews which provide an overview of the O-linked flagellin glycosylation pathway (Logan et al., 2006; Nothaft & Szymanski, 2010). Bacterial flagella assembly occurs at the distal end of the basal body. The nascent flagellin protein is transported across the cytoplasmic membrane by a type three secretion system, and then proceeds through the narrow central channel of the flagella structure. Finally, the flagellin subunit associates with the tip of the filament structure which is elongated and reaches a length of about ten micrometers. In contrast to the archaean flagellin export pathway, bacterial flagellin protein is not exposed outside of the inner membrane containing the periplasmic space until assembled into the filament. In other words, if flagellin glycosylation occurred extracellularly, it must be achieved far away from the cell. Therefore, it is reasonable to assume that the flagellin glycosylation machinery is located in the vicinity of the flagella basal body. Recently, the *C. jejuni* O-linked flagellin glycosylation machinery was localized at the pole of the cell along with the flagella (Ewing et al., 2009). Three genes involved in pseudaminic acid biosynthesis (PseC, which is the enzyme involved in the second step of PseAc synthesis, PseE, the putative PseAc transferase, and PseD, the putative PseAm transferase) were labeled with GFP fusion and expressed in *C. jejuni* 81-176. The fluorescent microscopic observation demonstrated that some, but not all, of the enzymatic glycosylation machinery was localized at the poles of the cells, consistent with a possible association with the flagellar basal body/export apparatus. Further study indicated that O-linked glycan biosynthesis could be reconstructed in vitro (Schoenhofen et al., 2009). The flagellin monomers from *Campylobacter* species are predominantly glycosylated with pseudaminic acid (Pse) and legionaminic acid (Leg). The precursors of these glycans are utilized in the form of CMP-activated sugars (CMP-Pse, CMP-Leg, and their derivatives), and they are added to the serine or threonine residues of flagellin by a specific glycosyltransferase (Note that the glycosyltransferases
Glycosylation 140

responsible for O-glycan attachment to flagellin have yet to be identified). The eleven candidates of glycan biosynthetic enzymes (PtmF, PtmA, PgmL, PtmE, GlmU, LegB, LegC, LegH, LegG, LegL, and LegF) from Campylobacter jejuni have been individually purified and characterized. It was confirmed that Leg and its CMP-activated form were synthesized from fructose-6-phosphate. The authors also suggested that O-linked glycan biosynthesis was involved in the synthesis of the N-linked glycan.

4. Amino acid substitutions of flagellin protein

Many attempts have been carried out to obtain insight into the significance of flagellin glycosylation. One of the most visible experiments is the disruption of glycosyltransferase activity which allows the evaluation of the flagella assembly, filament morphology, motility, and virulence (see above). In this section, we focus on the effects of amino acid substitution in glycosylated flagellin proteins.

4.1. Influence of loss of glycosylation to the motility and virulence

4.1.1. Campylobacter jejuni 81-176

The major flagellin of Campylobacter jejuni 81-176, FlaA, has been shown to be glycosylated at 19 serine or threonine residues, and this glycosylation is required for flagellar filament formation (Thibault et al., 2001; Goon et al., 2003). Mutants were constructed in which each of the 19 serine or threonines that are glycosylated in FlaA was converted to an alanine. Eleven of the 19 mutants displayed no observable phenotype, but the remaining 8 mutants had two distinct phenotypes. Five mutants (mutations S417A, S436A, S440A, S457A, and T481A) were fully motile but defective in autoagglutination. Three other mutants (mutations S425A, S454A, and S460A) were reduced in motility and synthesized truncated flagellar filaments (Ewing et al., 2009).

4.1.2. Pseudomonas syringae pv. tabaci

Flagellin glycosylation of Pseudomonas syringae pv. tabaci 6605 has been reported at six serine residues, positioned at amino acids 143, 164, 176, 183, 193 and 201 (Taguchi et al., 2006). Mutants where 6 serine residues were converted to alanine individually were compared with the mutant containing the flagellin specific glycosyltransferase, fgt1. All mutants displayed reduced swarming ability, swimming speed, filament stability, and virulence (Taguchi et al., 2006; 2008; Takeuchi et al., 2008). In addition, reduction of the molecular weight of each mutant flagellin protein corresponded to the loss of a single carbohydrate chain moiety, and the degree of reduced biological functions were smaller than that of an all glycosylation-serine-replacement mutant (6 S/A).

4.2. Restoration of filament formation without glycosylation

Flagellin glycosylation of a thermophilic bacillus species was recently reported for Bacillus sp. PS3. Although there was low coverage of the flagellin sequence, at least four serine and
threonine residues were identified as glycosylation sites (Hayakawa et al., 2009a). This potentially glycosylated flagellin protein was expressed in \textit{B. subtilis} Δ\textit{hag} (flagellin deficient mutant strain) for complementation. The resulting transformant was non-motile, and the produced flagellin protein derived from \textit{Bacillus} sp. PS3 was not glycosylated and accumulated intercellularly. However, spontaneously isolated flagellin mutants partially restored the motility and produced a truncated flagella filament without glycosylation (Hayakawa et al., 2009b and J. Hayakawa and M. Ishizuka, unpublished data). All characterized suppressing mutations contained single or double point mutations and about 30 residue intragenic duplications in the flagellin in the highly variable region (D2 and D3 domain) and the end of the α-helical structure (D1 domain). The positions of these mutations were in good accordance with the previously reported flagellin glycosylations sites from many other bacterial species. To our knowledge, this is the first report of a gain-of-function mutant of flagellin glycosylation.

5. Conclusions

Glycosylation is no longer a rare event regardless of whether bacteria or eukaryote are considered. Complete genomic information for several bacteria is now available and bioinformatic analyses demonstrated that bacterial flagellin glycosylation is widely spread over several genera. Many speculative functions of flagella glycosylation have been demonstrated, for example filament assembly (including flagellin export), filament stability, motility, virulence, gene regulation and mimicry with host-cell surface glycan structure. These glycosylation functions are similar regardless of the variety of eukaryote. In addition, the bacterial glycosylation pathway is becoming better defined; many genes which participate in flagellin glycosylation have been identified, but their number and loci are diverse in each bacterial species. Rapid increases in the knowledge of glycosyltransferases and glycan biosynthesis gene clusters will undoubtedly be achieved through glycoengineering with an aim to design a bacterial flagella motor for the development of a novel vaccine or drug-delivery-system.

Author details

Jumpei Hayakawa and Morio Ishizuka*
Department of Applied Chemistry, Faculty of Science and Engineering, Chuo University, Tokyo, Japan

Acknowledgement

The authors acknowledge In-Tech-Open Access Publisher for the kind invitation to contribute this chapter. This work was supported in part by Chuo University Grant for Special Research to M. I.

* Corresponding Author
6. References


Hayakawa, J., Kambe, T., & Ishizuka, M. (2009b). Amino acid substitutions and intragenic duplications of *Bacillus* sp. PS3 flagellin cause complementation of the *Bacillus subtilis* flagellin deletion mutant. *Bioscience, biotechnology, and biochemistry*, Vol. 73, No. 10, (October 2009), pp. 2348-5231, ISSN 0916-8451


acceptor sequence requirements. *Glycobiology*, Vol. 21, No. 6, (June 2011), pp. 734-742, ISSN 0959-6658


