Doctoral Dissertation

Construction of an *Escherichia coli* System for the Production of *N*-linked Glycoproteins

Srichaisupakit Akkaraphol

November 2014

Laboratory of Applied Microbiology
International Center for Biotechnology

Department of Biotechnology
Graduate School of Engineering,
Osaka University

Content

page

List of abbreviations

Twenty (20) natural amino acids are denoted in standard 1-letter code or 3-letter code.

Alg	Asparagine-linked glycosylation
DEAE	Diethylaminoethyl
diNAcBac	Di-*N*-acetyl bacillosamine, diacetamido bacillosamine
Dol	Dolichol
DsbA	Disulfide oxidoreductase
E. coli	*Escherichia coli*
ER	Endoplasmic reticulum
GalNAc	*N*-Acetyl-D-galactosamine
GDP	Guanosine diphosphate
Glc	Glucose
GlcNAc	*N*-Acetyl-D-glucosamine
gmd	GDP-D-mannose dehydratase gene
GTase	Glycosyltransferase
h	Hour
HPLC	High-performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
L-ara	L-Arabinose
LB	Lysogeny broth
LLO	Lipid-linked oligosaccharide
M	Molar
Man	Mannose
MBP	Maltose binding protein
min	Minute(s)
ml	Milliliter
mM	Millimolar
OD	Optical density
OST	Oligosaccharyltransferase
P	Monophosphate

PA, or 2-PA	2-aminopyridine
P_{BAD}	Arabinose-inducible promoter
pgl	Protein glycosylation
PP	Diphosphate, pyrophosphate
P_{T7}	T7 promoter
RP-	Reversed-phase
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulfate
SF-	Size-fractionation
TFA	Trifluoroacetic acid
TMHMM	Trans Membrane prediction using Hidden Markov Models
Tris	Tris-(hydroxymethyl)-aminomethane
Trx	Thioredoxin
UDP	Uridine diphosphate
Und	Undecaprenol

List of figures

List of tables

Chapter 1
General Introduction

1.1 Protein *N*-glycosylation in domains of life

Asparagine-linked (*N*-linked) protein glycosylation is an abundant protein post-translational modification mechanism by addition of sugar moiety (generally described as glycan) onto designated amino acid sequence on the nascent polypeptide chain. The glycan structure varies depending on the corresponding hosts. This process is sophisticated, yet shares common mechanisms amongst eukaryotes and is available throughout domains of life. This chapter provides a general overview of the protein glycosylation essential in eukaryotes, archaea and bacteria, with a focus on *Saccharomyces cerevisiae* and *Campylobacter jejuni*.

1.1.1 Eukaryotic *N*-glycosylation

Eukaryotic *N*-glycosylation was identified in rat serum (1), and later was determined to possess a variety of glycosyl linkages, glycan branches and complex structures. The mature *N*-glycan structure is generally viewed as a heterogeneous population of glycans modified by glycosyltransferases (GTases) and glycanases. The mechanisms and principles of *N*-glycosylation are distinct processes that have been recently reviewed (2). The non-templated diversity of *N*-glycosylation allows the fine-tuning of the functions of glycans (3). Numerous biological implications of protein *N*-glycosylation including with respect to processes such as protein folding, oligomerization, and structural stability, control of intracellular trafficking, secretion, protease protection, and cell-to-cell adhesion (3-6). For example, in viral glycoproteins such as GP120 in human immunodeficiency virus (HIV) (7) and hemagglutinin of the influenza virus (8, 9), *N*-glycans serve to modulate antigenicity, evade the immune system, help in infectivity and help to support virion production.

In eukaryotes, before *N*-glycans underwent maturation and became a widely heterogeneous glycan population, they shared a similar assembly pathway. Kornfeld and Kornfeld (10) reviewed the pathway of protein glycosylation to contain the synthesis of $Glc_3Man_9GlcNAc_2$-PP-Dol (Glc is glucose, Man is mannose, GlcNAc is *N*-acetylglucosamine, and Dol is dolichol). Glycosylation onto designated asparagine residues of proteins is carried out by the multi-subunit enzyme oligosaccharyltransferase (OST) protein complex (11). Co-translational glycosylation is a recently proposed mechanism (12, 13) and not all asparagine residues are glycosylated (14). Later, in the ER and Golgi, diversity of *N*-glycan structures are generated by GTases and glycanases (15). This thesis focuses on early-stage biosynthesis of the eukaryotic *N*-glycan structure, so the later-stage glycan trimming and the fate of glycopeptides will not be extensively discussed.

1.1.1.1 Dolichol pathway conferring to *N*-glycosylation

The early stage of the eukaryotic *N*-glycosylation process involves of assembly of an oligosaccharide donor on dolichol, the lipid precursor that localize between cytosol and the ER lumen. The glycans linked to the lipid via phosphate or pyrophosphate linkage are generally referred to as lipid-linked oligosaccharides (LLOs). In the dolichol pathway, a series of GTases are required for the synthesis of specific LLO, $Glc_3Man_9GlcNAc_2$-PP-Dol (10, 16).

Proteins participating in this dolichol pathway are called Alg proteins [after the process known as asparagine-linked glycosylation (17)], and their predicted activities are shown in Figure 1-1. GTases on the cytosolic side of the ER facilitate the assembly of $Man_5GlcNAc_2$-PP-Dol. The molecular mechanism of the translocation or flipping of this heptasaccharide is unclear; to date only one candidate gene, *RFT* has been identified (18).

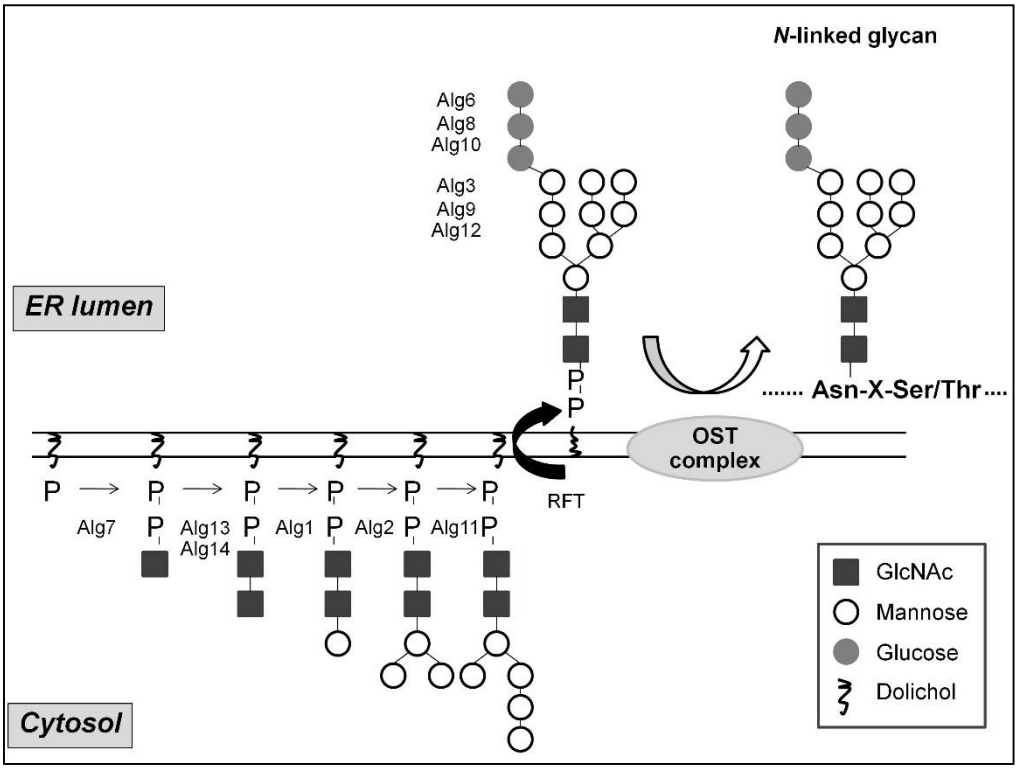

Figure 1-1. Glycosyltransferases of the dolichol pathway. Proposed enzymes are indicated for each step (OST is oligosaccharyltransferase).

On the lumenal side of the ER, after LLO translocation, GTases participate in the synthesis of dolichol-tetradecasaccharide (14 saccharide). Step-by-step addition with tight quality control of substrates has been observed. OST complex is considered to demonstrate a strong preference for $Glc_3Man_9GlcNAc_2$-PP-Dol (19). This recognition must occur to ensure the proper assembly of the $Glc_3Man_9GlcNAc_2$-PP-Dol before *N*-glycosylation by the OST complex can take place. If the assembly of the full-length 14 saccharide is not complete, the resulting underglycosylated proteins might be misfolded since lectins (the carbohydrate-binding property of proteins) in the quality control pathways might not correctly recognize the underglycosylated structure.

1.1.1.2 Overview of protein-folding glycosylation in quality control

Protein quality control commences upon the emergence of glycosylated polypeptides, and is followed by protein folding. This process strongly correlated with protein glycosylation. As an example in eukaryotes, the initial steps include deglucosylation by glucosidase I and glucosidase II (3, 20). As a result of glucosidase I digestion, the $Glc_2Man_9GlcNAc_2$-protein is recognized by the chaperones calnexin (CNX) and calreticulin (CRT), according to their carbohydrate-binding (lectin) properties. It is considered that the interaction between lectins and glycans helps to protect the proteins from premature folding. Furthermore, these chaperones recruit foldase protein, aiding protein-folding process (20). Later in this pathway, glucosidase II completely removes all Glc residues, thereby lowering the lectin affinity and causing the release of glycoprotein. The proteins released from CNX and CRT are transported from the ER to Golgi. However, misfolded proteins can reenter the CNX/CRT cycle by means of UDP-Glc: glycoprotein glycosyltransferase (UGGT), which recognizes the misfolded protein and thereby re-glucosylates the $Man_9GlcNAc_2$ structure. UGGT might recognize the hydrophobic patches on improperly folded proteins (21). Using chemical synthesis approaches, details of the molecular mechanisms underlying the specificities and interactions of the lectin domains in UGGT, CRT and ubiquitin ligase have been elucidated (22, 23). Furthermore, if the protein is not recognized by CNX/CRT, there are also peptide *N*-glycosidases (PNGase), ubiquitination enzymes, ER mannosidases and additional glycan modification enzymes to accommodate the different protein substrates.

1.1.1.3 Glycosyl donors and glycosyltransferases

Glycosyl donors of the dolichol pathway can be separated into two groups: donors in the lumen of the ER and donors in the cytosol. GTases in the cytosol utilize the nucleotide-activated sugars UDP-GlcNAc and GDP-Man. In the ER, the substrates for glycosylation are

dolichylphosphate-monosaccharide donors, Man-P-Dol and Glc-P-Dol (16). These monosaccharide donors are synthesized by Dpm1 and Alg5, respectively. However, the transport mechanism of Man-P-Dol and Glc-P-Dol is still unclear.

GTases participating in the dolichol pathway include Alg7, Alg13/14, Alg1, Alg2 and Alg11for cytosolic LLO assembly, and Alg3, Alg9, Alg12, Alg6, Alg8 and Alg10 for extension of the saccharides of the LLO moiety in the ER (by yeast nomenclature, listed stepwise, and partially displayed in Figure 1-1) (16). Some key GTases are mentioned in the paragraphs below.

Alg7 is an enzyme transferring the GlcNAc moiety from UDP-GlcNAc to P-Dol to generate GlcNAc-PP-Dol, and this transferring activity can be inhibited by tunicamycin (24, 25). *alg7* was found in *S. cerevisiae*, and its homologues have been identified in the archaea *Methanococcus voltae* and *Homo sapiens* (26). Alg7 is a complex multi-transmembrane spanning protein which hampers purification for biochemical analysis. Furthermore, there is evidence showing that Alg7 is associated with the Alg13/14 complex in *S. cerevisiae* (27).

The second GlcNAc is transferred by the Alg13/14 protein complex, generating GlcNAcβ1,4-GlcNAc-PP-Dol. This protein complex was identified by a protein structural approach, and was predicted to occur with a structural conformation similar to that of the *E. coli* MurG protein, which catalyzes the addition of MurNAc-pentapeptide to GlcNAc-PP-Und (Und is undecaprenol, which serves as a saccharide anchor of LLOs in *E. coli*) (28, 29). Functional similarities of Alg13/14 complex and MurG are presumed, but so far there have been no protein complementation analyses. Furthermore, physical contact between the small proteins Alg13 and Alg14 has been identified by a co-fractionation technique (30). A recent study showed that Alg14 recruits Alg13 to the cytoplasmic face of the ER membrane (31), and currently even the catalytic site is presumed to exist on soluble Alg13, such that the Alg14

protein ensures the cell viability and maintain the interaction with Alg7 (27). In brief, presence of Alg13 and Alg14 are necessary for the transfer of the GlcNAc from UDP-GlcNAc to GlcNAc-PP-Dol, forming the disaccharide intermediate in the dolichol pathway.

After GlcNAcβ1,4-GlcNAc-PP-Dol has formed, the first mannosylation in the dolichol pathway is carried out by Alg1. Using GDP-Man as a donor, Manβ1,4-GlcNAcβ1,4-GlcNAc-PP-Dol synthesis is catalyzed. Furthermore, Alg1 has been shown to oligomerize with Alg2 and Alg11 (32). Based on TMHMM (Trans Membrane prediction using Hidden Markov Models) (33) prediction, Alg1 has at least one predicted transmembrane domain at the N-terminal.

Alg2, a bifunctional mannosyltransferase, further adds two Man residues to $ManGlcNAc_2$-PP-Dol (34). Based on TMHMM (33) prediction, this mannosyltransferase has four transmembrane domains—i.e., two at the far N-terminus and two at the far C- terminus—and is responsible for α1,3- and α1,6-mannosylation of $ManGlcNAc_2$-PP-Dol, synthesizing LLO-pentasaccharide, $Man_3GlcNAc_2$-PP-Dol. Alg11 then modifies $Man_3GlcNAc_2$-PP-Dol by adding another 2 Man residues. The resulting $Man_5GlcNAc_2$-PP-Dol is flipped from the cytosol into the lumen of the ER. By disruption of Alg11, $Man_3GlcNAc_2$-PP-Dol was found to be accumulated, and thus its function was elucidated. With the $Man_5GlcNAc_2$-PP-Dol flipped into the lumen of the ER (18), mannosyltransferases and glucosyltransferases complete the synthesis of $Glc_3Man_9GlcNAc_2$-PP-Dol prior to the transfer of saccharide moieties to Asn of the N-X-S/T sequon by the OST protein complex.

1.1.1.4 Oligosaccharyltransferase (OST) of the eukaryotes

Unlike the single-subunit assembly of prokaryotic OST (35), eukaryotic OST exists in multi-subunit form. The eukaryotic OST complex is comprised of membrane-bound components. In yeast, 9 subunits have been identified, and Stt3, Swp1, Wbp1, Ost1 and Ost2

are essential (11, 36). Stt3 was identified as a catalytic subunit and is the most conserved subunit in terms of amino acid sequence. The eukaryotic OST recognizes N-X-S/T (where X is not Pro). All possible glycosylation sequons are not always glycosylated (14). Furthermore, the research revealed that glycosylation is not limited to N-X-S/T; more recently, N-X-C, N-X-V, and N-G-X glycosylations were statistically observed in the mouse glycoproteome (37).

The PglB encoded on the *pgl* operon of *C. jejuni* is found to be a homologue of Stt3 subunit of OST (38). Both enzymes possess a highly conserved WWDYG motif for OST catalysis (39). While the eukaryotic OST subunits form a functional protein complex, glycosylation by PglB can be carried out alone both in its native host and in heterologous expression system in *E. coli* (40). Moreover, the glycosylation sequon recognized by PglB are more extended than the eukaryotic one, D/E-X_1-N-X_2-S/T, where X_1 and X_2 are not Pro (41). Homologues of Swp1, Wbp1, Ost1 and Ost2 were identified in eukaryotes but was not found in *C. jejuni*.

1.1.2 archaeal *N*-glycosylation

The archaea are a domain of organisms that possess characteristics distinct from those of the eukarya and bacteria (42). "Archaea" have different cellular structures than those of other domains. Their membranes are composed of branched hydrocarbon chains that are attached to glycerol by ether linkages. Their cell walls contain no peptidoglycans. Usually, archaea inhabit extreme environments, such as highly sulfurous lakes, undersea hydrothermal vents, or hot geysers.

1.1.2.1 Discovery of *N*-linked glycoprotein in archaeal organisms

N-glycosylation was once believed to be restricted to eukaryotes. However, it has since become apparent that archaea and bacteria also possess this post-translational modification. Variety of glycan structures were found attached to asparagines of glycosylation sequon (43).

The first archaeal *N*-linked glycosylation was discovered in *Halobacterium salinarum* in 1976. Mescher and Strominger (44) described that the surface (*S*)-layer glycoprotein is modified by Asn-linked oligosaccharides, which possess a structure of repeating units of sulfated pentasaccharide.

The *N*-linked GlcNAc glycosylation was discovered in *Methanococcus voltae* (45). The glycan observed was trisaccharide, GlcNAc-(2,3diacetamido-2,3-dideoxy-β-glucuronic acid)-(2-acetamido-2-deoxy-β-mannuronic acid), which was linked to Asn residues.

1.1.2.2 LLO of archaea

The archaea core glycan is assembled on dolichol, unlike the case of bacteria for which the assembly of glycan takes place on undedaprenol (Und) (46). Archaea utilize both Dol-P and Dol-PP as carriers. An example of Dol-P as a carrier is seen in the *Halobacterium S*-layer glycoprotein, where bacitracin was unsuccessful at inhibiting the addition of GlcNAc-linked pentasaccharide, suggesting the use of Dol-P as a carrier of this glycan (47). Recently, *Haloferax volcanii S*-layer glycan-component biosynthesis was described (48), and a series of archaeal glycosylation (Agl) GTases were identified (49-52).

1.1.2.3 AglB is the archaeal oligosaccharyltransferase

The OST in eukaryotes is a complex of multi-subunit enzyme containing the Stt3 catalytic subunit, whereas in archaea and bacteria, OSTs exist as a single component. AglB OST (archaeal glycosylation oligosaccharyltransferase) was identified in *M. voltae* as a homologue to Stt3 (53). The Stt3 homologue was also identified in archaea *H. volcanii* (54).

1.1.3 Bacterial *N*-glycosylation

Protein glycosylation, once thought to be exclusive to eukaryotes, is now known to occur in all domains of life (43-46, 55). According to recent discoveries, many bacterial species

are able to modify protein substrates by *N*- (56) or *O*-glycosylation (57). Among these protein glycosylation-competent species, *C. jejuni*, a Gram-negative, spiral-shaped microaerophilic bacterium whose infection is one of the common causes of bacterial diarrhea (58, 59), is the most extensively studied bacterium in terms of the *N*-linked glycosylation pathway (60), which is quite different from its eukaryotic counterpart.

1.1.3.1 *Campylobacter jejuni N*-glycosylation

1.1.3.1.1 *C. jejuni* protein glycosylation operon

In 1998, Fry et al. isolated the 16 kb protein glycosylation (*pgl*) locus from *C. jejuni* NCTC 81116 (61). Previously, this locus was identified as the lipoposaccharide (LPS) biosynthesis operon by comparing it to the operons of other bacteria. Fry et al. observed a total of 11 LPS biosynthetic gene homologues, which consisted of three sugar biosynthesis genes, six GTases and two transporter genes. In their study, the *pgl* operon was subcloned, introduced, and expressed in *E. coli*. LPS profiling confirmed the putative glycosylation activity in *E. coli* harboring the cloned *C. jejuni* NCTC 81116 *pgl* operon. Later, Wacker et al. (56) and Linton et al. (40) further characterized the *pgl* operon, and their findings will be discussed below.

Orthologues of the *pgl* operon are also observed in other *Campylobacter* species. In some strains, such as *Campylobacter curvus* 525.92, *Campylobacter concisus* 13826 and *Campylobacter gracilis* RM3268, multiple copies of the PglB OST genes are present. The strains found to contain a *pgl* operon were summarized in the 2010 review of Nothaft and Szymanski (60).

1.1.3.1.2 Biosynthesis of the *C. jejuni N*-glycan

In *C. jejuni*, heptasaccharide (GalNAcα1,4-GalNAcα1,4-[Glcβ1,3-]GalNAcα1,4-GalNAcα1,4-GalNAcα1,3-diNAcBac, where GalNAc is *N*-acetylgalactosamine and diNAcBac is di-*N*-acetylbacillosamine) is assembled onto an undecaprenyl pyrophosphate

(Und-PP) lipid carrier. Biosynthetic genes encoding GTases, diNAcBac synthases and OST are located on the *pgl* operon.

Biosynthesis of *C. jejuni* LLO is initiated with UDP-GlcNAc modifications by PglF, PglE and PglD to form UDP-diNAcBac at the cytoplasmic side of the inner membrane (62). PglF, a C6 dehydratase, generates UDP-2-acetamido-2,6-dideoxy-D-xylo-4-hexulose. Then PglE, an aminotransferase, catalyzes the pyridoxal-dependent transfer of an amino group from L-glutamate to C4 of the UDP-4-keto-sugar to form UDP-2-acetamido-4-amino-2,4,6-trideoxy-α-D-glucose (a UDP-4-amino-sugar). PglD acetylates the C4 group on the UDP-4-amino-sugar to form UDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucose (also called UDP-2,4-diacetamido bacillosamine; UDP-diNAcBac), using acetyl CoA as an acetyl donor.

PglC, the first GTase, adds UDP-diNAcBac to Und-P, forming diNAcBac-PP-Und (63). PglA transfers UDP-GalNAc to form GalNAc-diNAcBac-PP-Und, disaccharide. PglJ adds a single α1,4-GalNAc residue, and PglH, working as a polymerase, then adds another 3 GalNAc residues, not by block transfer but one at a time, resulting in $GalNAc_5$-diNAcBac-PP-Und. PglI GTase adds a glucose branch to complete the common heptasaccharide LLO synthesis (40, 64). The ATP-dependent flippase PglK (65) then flips this assembled LLO *en bloc* (66) into the periplasmic space, where PglB OST transfers the glycan moiety of LLO onto Asn residue of D/E-X_1-N-X_2-S/T (where X_1 and X_2 are not Pro) acceptor site of the polypeptide chain (56). The general scheme of this section is shown in Figure 1-2.

In eukaryotes, systematic *N*-glycan modification by glycan-modifying enzymes in the corresponding protein secretion pathway was observed. However, in *C. jejuni*, no such activity has been described to date, although the addition of phosphoethanolamine to *N*-glycans was demonstrated (67).

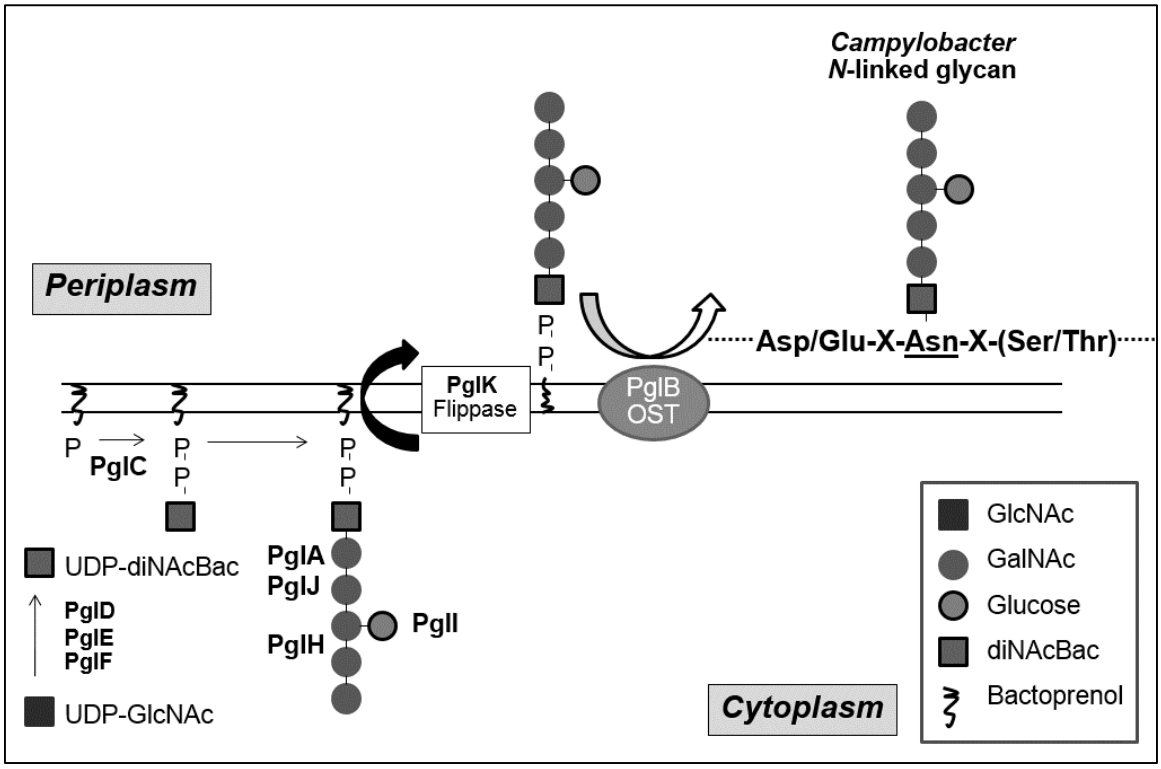

Figure 1-2. Summary of the *Campylobacter jejuni N*-glycan biosynthesis pathway and the corresponding cellular compartments where the bacterial glycosylation takes place. Please note that the glycosylation recognition site is extended compared to eukaryotes.

1.1.3.1.3 PglB substrate recognition and glycosylation mechanism

PglB is an enzyme encoded from the *pgl* operon. The presence of a single *pglB* gene in the *pgl* operon suggests that the PglB protein functions as a single-subunit OST (56). In contrast to eukaryotic OSTs, PglB recognized the extended glycosylation sequon D/E-X_1-N-X_2-S/T (where X_1 and X_2 are not Pro) (41, 68). Evidence suggested that PglB prefers the folded substrate *in vitro* (69). Recently, the X-ray structures of the globular domain of *C. jejuni* RM1221 PglB and the full-length *Campylobacter lari* PglB, in complex with an acceptor peptide, were elucidated (35, 70). In addition, the *in vitro* quantification of sequon binding and catalysis (71), and the mechanism of carboxamide activation (72) were studied in *C. lari* PglB. It is anticipated that these data will help clarify the mechanism of bacterial *N*-linked glycosylation.

1.1.3.1.4 Functions of *C. jejuni N*-glycan

In *C. jejuni*, numerous *N*-glycoproteins have been identified (73). Disruption of its *N*-glycosylation pathway has pleiotropic effects for the bacterium. Effects include reduced protein immunoreactivity with both human and animal sera (46, 74), a reduced ability to adhere to and invade human epithelial cells *in vitro* (75), and a decrease in mouse and chicken colonization *in vivo* (76, 77). However, disruption of either *pglI*, or *pglG* (which its function has not been determined), had no effect on chicken colonization. This absence of an effect on colonization ability was supported by the observation that mutation of these genes did not affect the *N*-glycosylation profile (66). Disruption of *pglD* resulted in the loss of chicken colonization, even though low levels of *N*-glycosylation could still be detected (66, 78).

The *C. jejuni N*-glycan has immunogenic properties. It was observed that the heptasaccharide is recognized by human macrophage galactose-type lectin (MGL) (79). This macrophage captures glycosylated antigens via its lectin activity, providing signals that modulate immune cell function. MGL recognizes *C. jejuni* glycoproteins, as well as *C. jejuni* LLOs with terminal GalNAc residues. Exposure of human dendritic cells to the *C. jejuni pglA* mutant resulted in increased production of the cytokine interleukin-6 (IL-6), because the *C. jejuni pglA* mutant does not possess *N*-linked GalNAc residues, the putative MGL ligand. Consequently, it could be considered that *C. jejuni N*-glycosylates its proteins in order to modulate the host immune response, which could help to limit cellular cytokine production (79).

1.1.3.2 Unusual protein glycosylation in other bacteria

Apart from *C. jejuni*, the *N*-glycosylation system was recently described in the Gram-negative, γ-proteobacterium *Haemophilus influenzae*, and was shown to involve the sequential transfer of sugars (80). HMW1 (the high-molecular-weight adhesion protein of *H. influenza*e) possesses 31 *N*-glycosylation sites, and interacts with sialylated *N*-glycoproteins on the host

cell surface (81). To resist degradation and maintain bacterial cell surface interaction with its cognate outer-membrane translocator protein, HMW1B, glycosylation of HMW1 is required. The GTase HMW1C seems to associate with HMW1A in the cytoplasm (82), where monohexoses and dihexoses are added to the conventional eukaryotic *N*-glycosylation sequon. An *in vitro* experiment demonstrated that HMW1C can transfer glucose, galactose, and small amounts of mannose residues. Surprisingly, HMW1C exhibits two functions: first, a distinct GTase activity, adding *N*-linked sugars to HMW1, and second, an ability to generate hexose–hexose bonds (83). These information help to characterize the new type of bacterial GTase in *H. influenza*e.

Remarkably, the *N*-glycosylation machinery that was identified in *Actinobacillus plueropneumoniae* exhibited *N*-GTase activity towards various donor substrates, mainly the Glc, but not the GlcNAc (84). The authors thus proposed an alternative route to synthesize *N*-glycosylation in bacteria.

1.2 Significances of protein glycosylation

1.2.1 Congenital disorders of glycosylation

Glycosylation is a complex process. Therefore, defects in the pathway can cause lethal phenotypes or severe medical symptoms. In human patients, congenital disorders of glycosylation (CDG) can be classified into two groups, CDG I and CDG II (85, 86). CDG caused by problematic biosynthesis of glycan results in aberrant glycosylation of intermediate glycan structures instead of the mature form. There are a wide range of clinical manifestations, from intestinal disorders to liver failure or severe mental retardation. To date, 11 human genes whose mutations can cause these defects have been identified. They are listed in Table 1-1, which was adapted from Aebi and Hennet 2001 (86). New CDGs symptoms conferring genes are also being proposed (87).

Table 1-1. Congenital disorders of glycosylation

Name	Gene defect	OMIM	Activity
CDG-Ia	*PMM2*	212065	Phosphomannomutase (Man-6-P → Man-1-P)
CDG-Ib	*PMI*	602579	Phosphomannose isomerase (Fru-6-P → Man-6-P)
CDG-Ic	*ALG6*	603147	α1-3 Glucosyltransferase
CDG-Id	*ALG3*	601110	α1-3 Mannosyltransferase
CDG-Ie	*DPM1*	603503	Dolichyl-phosphate-mannose synthase (GDP-Man → Dol-P-Man)
CDG-If	*LEC35*	604041	Unknown
CDG-IIa	*MGAT2*	202066	β1-2 *N*-acetylglucosaminyltransferase II
CDG-IIb	*GLS1*	601336	α1-2 Glucosidase
LADII/CDG-IIc	GDP-Fuc transporter	266265	Import of GDP-Fuc into Golgi and export of GMP
Ehlers–Danlos syndrome (progeroid form)	*XGPT*	130070	Xylose β1-4 galactosyltransferase
Galactosemia I	*GALT*	230400	Gal-1-P-uridyltransferase (Gal-1-P+UDP-Glc ↔UDP-Gal+Glc-1-P)
Galactosemia I	*GALE*	230350	Galactose epimerase (UDP-Gal ↔ UDP-Glc)
Galactosemia II	*GALK*	604313	Galactokinase (Gal → Gal-1-P)

Congenital Disorders of Glycosylation. Online Mendelian Inheritance in Man (OMIM) compendium catalog numbers are listed along with the gene names. Adapted from Aebi and Hennet 2001 (86).

1.2.2 Pharmaceutical glycoprotein functions and applications: recognition and cellular uptake mediated by mannosylated β-glucocerebrosidase

Glycosylation of certain therapeutic proteins can influence their physiological properties, such as intracellular targeting, protein-binding and molecular stability (88, 89). In brief, the purpose of glycosylation is to achieve the optimum therapeutic efficacy.

β-glucocerebrosidase (bGC) is an enzyme that is absent in patients with Gaucher disease. This lysosomal disease leads to accumulation of glucosylceramide substrate in the lysosome of various tissues. Symptoms of Gaucher disease include disorders of the liver, spleen, blood, bone, and lung (90). Treatment can be carried out by enzyme replacement therapy (ERT), with an enzyme purified from human placenta (91). Previously, the discovery of mannose receptors on macrophages by Stahl et al. (92) suggested that glycan modeling of bGC might be used for uptake purposes (93). However, human placental bGC with such oligomannose was not effectively utilized. In 1981, Furbish et al. (94) studied the effects of deglycosylation on bGC uptake, and found that the removal of complex glycans and maintenance of the core mannose improved macrophage targeting. The *N*-glycan structure on human bGC was later elucidated to contain one chain of oligomannose in its 4 glycosylation positions (95). The post-production glycan modification strategy is being used in production of imiglucerase, Cerezyme. Cerezyme is a recombinant human bGC produced in Chinese Hamster Ovary (CHO) cells, where the complex-type glycan on the respective glycosylation sites was modified to contain mannose terminals. Cerezyme has been reported to have clinical efficacy (96). Recently, the effect of the mannose chain length on targeting of bGC for ERT of Gaucher disease was examined (97), and the results indicated that increased binding to mannose-binding lectin occurred when the mannose core saccharide length of bGC was extended. However, the serum half-life and macrophage uptake were similar in all forms of bGC generated in the stated work. Considering that the interaction between lectin and the long mannoses chain, which might affect clearance and pharmacokinetics, production of bGC with extended mannoses is not required. This result underscores the importance of specific glycan composition when glycan-containing pharmaceutical proteins are being produced.

1.3 Production of glycoproteins in *Escherichia coli* and approach of glycoengineering

E. coli is a well characterized and robust protein expression host (98). Rapid and economical production of pharmaceutical proteins (99) are among the most attractive applications of this organism. To optimize this valuable and cost-effective host, strategies have been developed for maximizing the expression of soluble recombinant proteins (100, 101). However, the *E. coli* protein expression system lacked the glycosylation mechanism which is available in higher eukaryotic hosts. Recently, *E. coli* was engineered to be equipped with the *C. jejuni* protein glycosylation mechanism (56). Many types of engineered glycans could be transferred onto recombinant periplasmic- and subcellular- or extracellular-targeted proteins (102-104).

However, the unique glycan moiety, GalNAcα1-4GalNAcα1-4(Glcβ1-3)GalNAcα1-4GalNAcα1-4GalNAcα1-3Bacβ1-Asn (73), is totally different from the eukaryotic *N*-glycans found in mammals. Also, bacterial *N*-glycans are immunogenic (105, 106) and the Asn residues are linked by an unusual sugar, diNAcBac. Therefore, if the glycan portion is not for immunizing purposes, the bacterial glycosylation limits the use of the resulting glycoprotein in terms of cellular recognitions and functions (3, 4). This has led to the requirement of glycan structure modeling, namely glycoengineering.

Glycoengineering can be performed in two different ways: *in vitro* glycopeptide synthesis followed by chemical ligation, or *in vivo* glycan engineering. The first method has been successfully performed in numerous reports (107-109). However, this tedious process limits the synthesis of folded, complex, or long proteins. In this thesis, therefore, the term glycoengineering generally refers to the latter method.

The idea of glycoengineering in *E. coli* has been exploited since the transfer of the protein glycosylation mechanism from *C. jejuni* into *E. coli* (56, 61). To identify the function

of enzymes involved in *pgl* operon activity, mutational studies of operon components were performed (40). Later, by removing genes conferring synthesis and lipid-transfer of diNAcBac from *pgl* operon, the HexNAc residue, suggestively GlcNAc, at the reducing ends of *N*-glycan was obtained. With a relaxed substrate specificity of PglB OST, the glycan structures containing GlcNAc at the reducing ends could be transferred onto the respective acceptor sites of target proteins in the *E. coli* periplasmic space (103). Furthermore, it was clearly shown that PglB could transfer not only the *C. jejuni*-type, linear or branched glycan (110), but also *E. coli O*-antigens onto the model proteins (102, 103). These *E. coli O*-antigens are synthesized on the common LLO building block, the Und-P, initiated by *E. coli* native WecA (111-113). The resulting GlcNAc-PP-Und should be readily available in *E. coli*, since the second GTase of the *O*-polysaccharide biosynthesis pathway, *wbbL*, was missing in *E. coli* lab strains, and when *wbbL* was complemented, *O*-antigen was synthesized in previously lacking *E. coli* (114).

These insights have opened up the engineering of humanized glycan structures, bottom-up from GlcNAc-PP-Und and later, GlcNAc-Asn. Engineering of glycosylation pathway in *E. coli* is recognized as bottom-up because the *E. coli* constructed *N*-glycosylation pathway is novel. Taking into account that there was no previous protein glycosylation pathway installed in *E. coli*, the desired *N*-glycan to be obtained from the constructed pathway is rationally homogeneous.

Engineering recombinant glycan structures in *E. coli* for both research use and industrial application is a current goal of glycoengineering. A bacterial *N*-glycosylation glycoengineering approach requires 3 key steps: [1] the LLO glycan donor must be assembled onto the *E. coli* WecA product, GlcNAc-PP-Und; [2] the engineered LLO glycan must be flipped across the inner membrane into the periplasmic space; and [3] the engineered glycan is transferred onto the secreted protein by PglB OST. Figure 1-3 provides a general outline of this concept. The resulting engineered glycan, both on LLO and glycoprotein, can serve as a

material for functional studies and biomedical applications. The ultimate goal is to establish an *E. coli* platform for the production of homogeneous eukaryotic *N*-glycoproteins, which will be further discussed in Chapter 3 of this thesis.

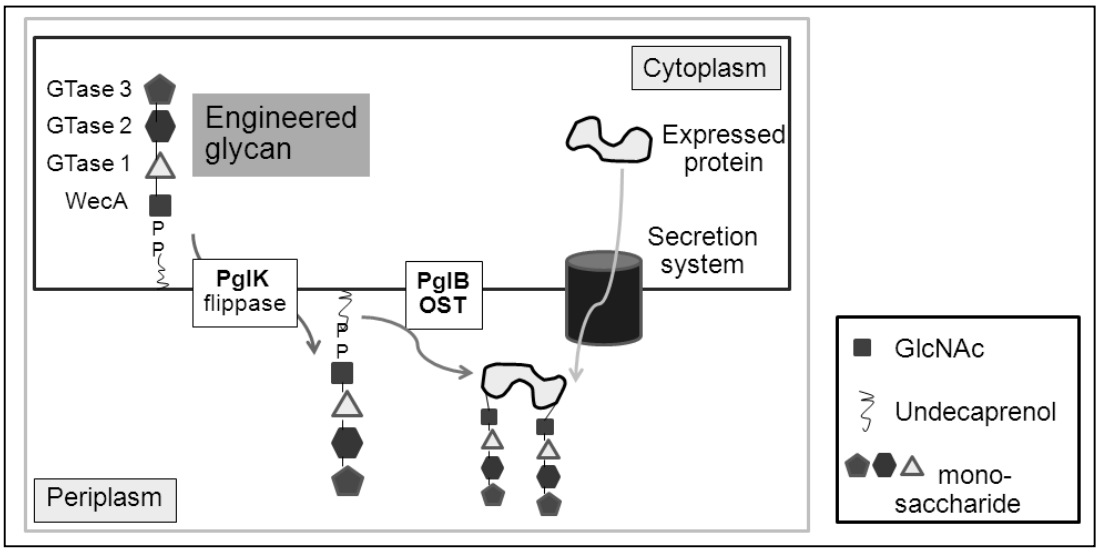

Figure 1-3. The bacterial *N*-glycosylation glycoengineering approach in *E. coli*

1.4 Summary

Protein glycosylation, once thought to be exclusive to eukaryotes, is observed in all domains of life. In bacteria, the *N*-glycosylation has been identified in *C. jejuni*. The pathway was found to be comprised of genes located on a 16 kb *pgl* operon. When this *pgl* operon was transferred into *E. coli*, protein glycosylation was observed. After the *pgl* operon was extensively examined, the possibility of bacterial *N*-glycan glycoengineering arose. Using property of relaxed substrate specificity of the PglB, glycans assembled from the installed recombinant pathway could be glycosylated onto the desired target protein.

1.5 Objective of the study

CHO cells are the most popular host in terms of production of human therapeutic glycoproteins. The CHO produced sialylated glycans are suitable for interaction with human cells (115), moreover the high titers of the secreted glycoprotein have reached the order of 10 g/l (116). However, there are drawbacks in using CHO cell lines for production of therapeutic protein productions, such as slower growth rate, expensive growth media, and risk of viral contamination (117). These factors have kept the production costs high. Recently, the production of glycoprotein in *E. coli* has been made possible by the characterization and successful transfer of *pgl* operon from *C. jejuni* to *E. coli* (56). Considered as a promising glycoprotein production platform with relatively low costs, yet the remaining challenges are such as engineering of glycan structure, and to avoid integration of unwanted bacterial glycan in the constructed *E. coli* system. Approaching the application of *E. coli N*-glycosylation, the biosynthetic pathway of eukaryotic *N*-glycan structure is to be constructed.

This thesis comprises of four chapters. Chapter 1, general introduction, describes protein glycosylation in domains of life, differences of their processes and approach of glycoengineering in *E. coli.* A *pgl* operon conferring *N*-glycosylation from *C. jejuni* strain JCM 2013 was identified in Chapter 2. Heterologous expression of the cloned *pgl* operon resulted in GlcNAc-Asn glycosylation of proteins. In Chapter 3, an *N*-glycosylation pathway containing *S. cerevisiae* GTases, connecting with native *E. coli O*-polysaccharide biosynthetic pathway, was constructed and introduced into *E. coli*. The eukaryotic *N*-glycan structure, as observed in the early dolichol pathway, was found to be assembled and glycosylated onto the model protein as a result. This work has opened up the possibility of engineering more complex, humanized *N*-glycan structures and *N*-glycoproteins that could be totally synthesized in *E. coli*, which the perspective is described in Chapter 4.

Chapter 2

Cloning and Characterization of the Protein Glycosylation Operon from *Campylobacter jejuni* JCM 2013

2.1 Introduction

Campylobacter is a group of spiral-shaped, microaerophilic, pathogenic bacteria. *C. jejuni* is the strain in which the protein *N*-glycosylation mechanism was found. *C. jejuni* NCTC 81116 and 11168 are type strains in which *pgl* operon was identified and characterized (40, 56, 61). The *pgl* operon consists of GTases, OST, and translocase (flippase), and the resulting glycan structure is a heptasaccharide, as generally described in Chapter 1. Furthermore, diverse *N*-glycan structures have been observed in *Campylobacter* other than *C. jejuni* (118).

C. jejuni strain JCM 2013 is available in the Japan Collection of Microorganisms (JCM) of the RIKEN Bioresource Center (Saitama, Japan), but there has been no report focusing on its protein *N*-glycosylation. In this study, identification of the *C. jejuni* JCM 2013 *pgl* operon was performed.

In *C. jejuni*, numerous *N*-glycoproteins have been identified (73). In previous reports, Peb3 and CmeA, which respectively function as an antigenic peptide (119) and a component of a multidrug efflux pump (120), were shown to be abundantly expressed and glycosylated in native *C. jejuni* (40). CmeA was chosen as a glycan acceptor protein in this work.

To assess the function of the cloned *pgl* operon, it was co-expressed with a glycan acceptor protein, CmeA, which possesses 2 prokaryotic *N*-glycosylation sites, in the constructed *E. coli* BL21 Δ*waaL* mutant. Protein expression was successful, and putative protein glycosylation activity was observed. The detailed glycan structure was analyzed by mass spectrometry (MS) and the fluorophore-labeling method.

Operon minimization, protein expression requirements, and induction conditions were studied. In addition, the maltose binding protein (MBP) and DsbA were C-terminally fused

with a synthetic glycosylation sequon and expressed as glycan acceptor proteins, alternative to CmeA. MBP was chosen because it can be expressed at a high-level in *E. coli* (data not shown), can carry a foreign protein following its C-terminal, and can be periplasmically directed and purified on amylose resin or metal affinity resin (121, 122). DsbA was chosen because it is a small periplasmic protein of *E. coli* with an efficient periplasmic secretion signal (123, 124).

2.2 Materials and methods

Chemicals were obtained from Wako Pure Chemical Industries (Osaka, Japan) unless mentioned otherwise. Tryptone and yeast extract were purchased from BD Biosciences (Franklin Lakes, NJ). Restriction enzymes were obtained from TaKaRa Bio Inc. (Kyoto, Japan). General molecular biology techniques were conducted following a laboratory manual (125).

2.2.1 Strains, medium and cultivation conditions

Genomic DNA of *C. jejuni* JCM 2013 (ATCC 29428) was obtained from the Japan Collection of Microorganisms (JCM). *E. coli* DH5α (126) and BL21 (DE3) (Novagen, WI) were used for genetic manipulation and protein expression, respectively. *E. coli* strains were grown in LB medium (1% tryptone, 0.5% yeast extract and 1% NaCl). Antibiotics of 100 mg ampicillin (Ap), 25 mg chloramphenicol (Cm) or 25 mg kanamycin (Km) per liter were added to the media for selection as needed.

2.2.2 DNA cloning

KOD-Plus DNA polymerase (TOYOBO, Osaka, Japan) was used for DNA amplifications. pGEM-T Easy (Promega, Madison, WI) was used for cloning of operon fragments. Sequence analysis was performed using a BigDye Terminator version 3.1 cycle sequencing kit on a 3130xl genetic analyzer (Applied Biosystems, Foster City, CA). DNA contigs were assembled using the ATGC application in the GENETYX 10 software package

(GENETYX Corp., Tokyo, Japan), and sequence identification was performed using the BLASTN and BLASTP facilities of NCBI.

The whole region of the *pgl* operon was separated into 5 regions, and each region was amplified separately. The fragments ranged from 3.5 to 4.5 kb in size, with intra-overlapping at the respectively predicted *pglK*, *pglI*, *pglB*, and *pglE* ORFs. Amplicons were generated by amplification of *C. jejuni* JCM 2013 genomic DNA with the following primer pairs: waaC-R and pglK-R; pglK-F and pglI-R; pglI-F and pglB-R; pglB-F1300 and pglE-R; and pglE-F and cheY-R. The oligonucleotide primers are listed in Table 2-1.

2.2.3 *E. coli* BL21 (DE3) *waaL* mutant generation

The *waaL* gene was disrupted according to the method reported for *Salmonella* species (127). Briefly, the *waaL* locus with a 3.5 kb flanking region was amplified using primers gumD-F and waaT-R. An amplified product with a size of 7.7 kb was ligated to pGEM-T Easy. Inverse-PCR was performed at the *waaL* locus on the cloned plasmid using the primers inv-waaL-F and inv-waaL-R in order to delete *waaL* and introduce *BamH*I cutting sites. The bacterial kanamycin-resistance cassette (*kanR*) was amplified from plant binary vector pBI121 (128) with terminal *BamH*I restriction sites using the primers kanR-F and kanR-R. Later, this cassette was ligated into a previously constructed and cut plasmid to replace the *waaL* ORF with *kanR*. This yielded a kanamycin-resistance cassette with flanking genes of *waaL*. This constructed *waaL* disruption cassette on pGEM-T Easy was amplified with M13 primers, and the amplified product was cut with *Not*I, purified and self-ligated using a TaKaRa Ligation kit version 2.1. The ligation product was ethanol-precipitated and then electrotransformed (0.1 cm cuvette, on Bio-Rad (Hercules, CA) Gene Pulser set at 2.4 kv, 400Ω, and 25μF) into *E. coli* BL21 (DE3). The transformants were plated on LB agar (Km) and incubated for 20 hrs at 37˚C, and were subsequently transferred onto three fresh agar plates: LB (Ap), LB (Km) and LB, respectively. The transformants which only showed resistance to Km but not Ap were chosen

as the candidates for the *waaL* mutant. Genomic integration of the *kanR* gene at the *waaL* locus was confirmed by PCR analysis. The obtained mutant was designated as the *E. coli* BL21 Δ*waaL* and used for the protein expression.

2.2.4 Protein expression of the *pgl* operon and CmeA

The *pgl* operon and the *cmeA* gene were amplified from *C. jejuni* JCM 2013 chromosomal DNA using KOD-Plus DNA polymerase. The amplified *pgl* operon and *cmeA* gene were digested with *Nco*I and *Xho*I and ligated to the *Nco*I and *Xho*I sites of pACYCDuet-1 (Novagen) and pET-22b (Novagen). The resulting plasmids were designated as pPgl and pCmeA, respectively.

Plasmids pPgl and pCmeA were co-introduced into the *E. coli* BL21 Δ*waaL* by electroporation. Freshly transformed cells were inoculated into LB media (Ap, Cm, Km) and incubated overnight with shaking as a preculture. The preculture was diluted 3% into a 500 ml Erlenmeyer flask containing 100 ml LB medium (Ap, Cm, Km). The culture was then incubated at 37˚C and 120 rpm on a reciprocal shaking incubator. After the cell density reached an OD_{600} of 0.5, the culture flask was transferred onto ice and kept for 15 min. Protein expression was initiated by shaking at 37˚C and 120 rpm for another 4 hrs. The expressed cells were harvested by centrifugation at 10,000 × *g* for 10 min, washed with 20 ml of 10 mM Tris-Cl pH 8.0, then pelleted again.

2.2.5 Protein extraction and purification

Periplasmic material extraction was performed using a single-step incubation. Briefly, the cell pellet was resuspended in 20% sucrose, 30 mM Tris-Cl pH 8.5, 1 mM EDTA and 1 mg/ml of lysozyme (Sigma). After incubation on ice for 30 min, the mixture was centrifuged at 10,000 × *g* for 15 min. The resulting supernatant was collected as periplasmic extract. The extract was then dialyzed against CmeA purification buffer (30 mM Tris-Cl pH 8.0, 300 mM NaCl, 1 mM phenylmethylsulfonyl fluoride (Sigma)) using a dialysis membrane (Wako). The

dialyzed extract was loaded onto 1 ml of a nickel-nitrilotriacetic acid (Ni-NTA) affinity resin (Bio-Rad), equilibrated with CmeA purification buffer. The unbound proteins were washed out with 50 ml CmeA purification buffer containing 5 mM imidazole (MP Biomedicals, Santa Ana, CA). The bound protein was eluted with CmeA purification buffer containing 20, 50 and 200 mM imidazole.

2.2.6 Protein electrophoresis, Western- and lectin-blotting

SDS-PAGE was performed with a discontinuous system as previously described (129) using 10% acrylamide (Nacalai Tesque, Kyoto, Japan) with 3.3% *N*, *N*′-methylenebisacrylamide. The protein was stained with Coomassie Brilliant Blue (CBB) following a general protocol (125) or by immunoblotting. For immunoblotting, the protein was transferred onto a polyvinylidene fluoride (PVDF) membrane (Immobilon-P; Millipore, Bedford, MA) by a semi-dry system with 25 mM Tris, 192 mM glycine and 20% methanol as buffer. The blotted membrane was blocked in 5% skim milk in phosphate buffered saline with 0.05% v/v Tween-20 (PBS-T) for 1 hr at room temperature. The blocked membrane was incubated with 1:2000-diluted anti-Penta His antibody (Qiagen, Hilden, Germany) as the primary antibody and 1:5000-diluted horseradish peroxidase-conjugated anti-Mouse IgG (GE Healthcare, Piscataway, NJ) as the secondary antibody. The immunoreacted signal was developed with Immobilon Luminata Forte Western HRP Substrate (Millipore). Chemiluminescent signals were visualized by exposure to RX-U Medical X-ray film (Fuji Film, Tokyo, Japan) and developed with an FPM100 medical film processor (Fuji Film). For soybean agglutinin (SBA) lectin blotting, the blotted membrane was blocked in 1% skim milk in PBS-T, then incubated with 1 μg/ml biotinylated SBA lectin (J-Oils Mills, Tokyo, Japan) in PBS-T for 1 hr with shaking and 1 μg/ml HRP-conjugated streptavidin (Vector Laboratories, Burlingame, CA) in PBS-T. After extensive washing in PBS-T, the membrane was developed with a POD-Staining kit (Wako).

2.2.7 Glycopeptide analysis by nanoLiquid chromatography/mass spectrometry (nanoLC/MS)

The purified putatively glycosylated CmeA was separated by SDS-PAGE and stained with CBB staining solution. Bands of interest were excised by scalpel, destained with 50 mM NH_4HCO_3 in 50% acetonitrile, dehydrated with acetonitrile, and in-gel digested with Trypsin Gold (Promega) in ProteaseMAX™ surfactant (Promega) at 37˚C for 3 hrs. After centrifugation at 15,000 ×g for 15 sec, the resulting solution containing tryptic peptides was subjected to mass spectrometry (MS) analysis on a nanoLC/MS 1200 System (Agilent, CA) equipped with a micrOTOF-QII instrument (Bruker Daltoniks, Bremen, Germany). For the LC analysis, ZORBAX 300SB-C_{18} columns were used as the trapping column (5 μm, 0.3 × 5 mm) and analytical column (3.5 μm, 0.075 × 150 mm). For the nanoLC separation, the mobile phase was composed of 0.1% formic acid (solvent A) and 0.1% formic acid in acetonitrile (solvent B). For peptide trapping in the column, 0.1% trifluoroacetic acid (solvent C) was used. After injection of the sample, the flow was directed to the trapping column in solvent C at a flow rate of 10 μl/min. The tryptic peptides were separated by linearly increasing the solvent B concentration from 2% to 8% over 5 min, and then to 30% over 30 min at a flow rate of 0.3 μl/min, followed by washing with 95% solvent B for 5 min, and then returned to the initial conditions. The MS data were deconvoluted using ESI Compass DataAnalysis 4.0 software (Bruker Daltoniks).

2.2.8 *N*-glycan analysis by fluorophore labeling

The sample for *N*-glycan structural analysis was prepared using a previously described method (130) with modifications. Briefly, glycan was released from 10 mg of lyophilized powder of the purified CmeA protein by hydrazinolysis at 100˚C for 10 hrs. Re-*N*-acetylation was performed using saturated $NaHCO_3$ and acetic anhydride. The resulting solution was desalted using a Dowex 50x2 resin (Muromachi Technos, Tokyo, Japan), then vacuum

concentrated and lyophilized. The obtained powder was pyridylaminated using purified 2-aminopyridine (PA). The excess of free PA was removed with 3 rounds of phenol:chloroform (1:1) extraction followed by 3 rounds of chloroform extraction. The aqueous phase, containing PA-glycan, was lyophilized and dissolved in 50 μl of Milli-Q water (Millipore) prior to analysis.

Size-fractionation (SF)–HPLC was performed using a 4.6 × 250 mm Asahipak NH2P-50 4E column (Showa Denko, Tokyo) on a Jasco HPLC apparatus, and PA-glycans were detected at the excitation and emission wavelengths of 310 and 380 nm, respectively. The chromatographic solvents were acetonitrile (SF solvent A) and 0.01% TFA (SF solvent B). After sample injection, the percentage of SF solvent B was held at 15% for 5 min as an initial condition. PA-glycans were eluted by increasing the percentage of the SF solvent B mixture linearly from 15% to 50% over 30 min at a flow rate of 0.7 ml/min. The column temperature was set at 40˚C.

To analyze the oligosaccharide structure, digestion of PA-glycan was performed using α-*N*-acetylgalactosaminidase (α-GalNAcase; NEB, Beverly, MA) according to the manufacturer's protocol. After boiling to inactivate the reactions and centrifugation, the resulting supernatants were subjected to SF-HPLC analysis.

The HPLC-eluted PA-glycan samples were collected and subjected to subsequent digestion and LC-MS/MS analysis on an Agilent Technologies 1200 series instrument equipped with an HCT Plus MS unit.

2.2.9 Transcriptional analysis of bacillosamine synthase and transferase genes under the constructed *E. coli* system

Total RNA was prepared from 500 μl culture of the *E. coli* BL21 Δ*waaL* mutant expressing pPgl and pCmeA using an RNeasy mini kit (Qiagen) according to the manufacturer's protocol. The cDNA was synthesized using a PrimeScript RT reagent kit (TaKaRa) and pglG-R primer. Primers amplifying the internal regions of *pglF, pglE, pglD,* and

pglC were designed by Primer3 software (131), yielding amplicons of 470, 495, 497 and 451 bp, respectively. The PCR reaction was performed for 35 cycles with a 25 µl reaction of GoTaq polymerase (Promega) using 10 ng of cDNA, and 10 µl of PCR product was separated on TBE agarose gel.

2.2.10 Elucidation of glycan composition and the glycan residue at the reducing end of *N*-linked glycan

First, the *pglI* gene (putative glucosyltransferase) was disrupted from the *pgl* operon. Primers pglH(down)-R and pglJ(ups)-F were used in inverse-PCR amplification, leaving approximately 100 bp at both ends of *pglI* ORF. The resulting PCR products were ethanol precipitated, phosphorylated with polynucleotide kinase (TaKaRa), ligated overnight with a DNA Ligation kit and transformed into *E. coli* DH5α. The resulting plasmid was named pPgl Δ*pglI*.

pPgl Δ*pglI* and pCmeA were introduced into *E. coli* BL21 Δ*waaL*. The protein expression and following preparations were performed as described above. PA-glycans were prepared and the corresponding sample was subjected to α-GalNAcase digestion. The SF-HPLC conditions were changed when monosaccharide was analyzed. PA-monosaccharide standards were obtained from TaKaRa. Using a SF-HPLC column, PA-monosaccharides were eluted by increasing the percentage of the SF solvent B mixture linearly from 5% to 50% over 30 min at a flow rate of 0.7 ml/min.

2.2.11 Construction of *pgl* operon derivatives and expression under the Ara promoter

To construct a *pgl* operon of minimal length, the partial *pgl* operon fragment from *wlaA* to *pglA* was amplified (primer wlaA-F x pglA-R). Fragments were ligated to the *Nco*I and *BamH*I sites of pACYCDuet-1. Subsequently, to remove *pglI* glucosyltransferase, inverse-PCR was performed as described above. The plasmids were named pPgl mini and pPgl mini Δ*pglI*,

accordingly. The T7 promoter (P_{T7}) was removed from pACYCDuet-1 using primers pACYC(T7down)-F x pACYC(T7up)-R, and the resulting plasmid was named pACYCDuet-1 ΔT7.

To construct an inducible protein glycosylation operon, an alternative to P_{T7}, arabinose-inducible promoter (P_{BAD}), was employed. P_{BAD} was cloned from genomic DNA of *E. coli* BL21 (DE3) and replaced P_{T7} and the *lac* operator on pACYCDuet-1. The obtained plasmid was named pAra-o. Subsequently, the *pgl* operon on pPgl was cut at *Nco*I and *Xho*I and ligated to the respective cutting sites on pAra-o. Designations were P_{BAD} (*name of previous plasmid*).

2.2.12 Construction of plasmids expressing glycan acceptor proteins

A Glycosylation Tag (GT) containing GGDQNAT as a prokaryotic glycosylation motif (104) was introduced onto plasmid pET-22b. By inverse-PCR, GT and GT_4 were introduced after the *Xho*I site and in front of the 6xHis sequence using primers pET22bGT-F x pET22bGT-R and pET22bGT4-F x pET22bGT4-R pairs, respectively. pET-22bGT and pET-22bGT_4 were then obtained. Next, *E. coli* MBP encoded by the *malE* gene, and disulfide isomerase encoded by the *dsbA* gene were amplified with primers NdeI-MalE-F x XhoI-MalE-R and NdeI-DsbA-Fw x NcoI-DsbA-Rv, and were ligated in-frame to pET-22bGT. The resulting constructs were designated MBP-GT and DsbA-GT, respectively. By ligation of *malE* and *dsbA* into pET-22bGT_4, constructs carrying 4 GT repeats were obtained, and the resulting proteins were named MBP-GT_4 and DsbA-GT_4.

Table 2-1. Oligonucleotide primers used in Chapter 2. Restriction sites are underlined.

Name	Sequence (5')
waaC-R	CTATTTTTCATTAAGTAAGCC
pglK-R	TCATTTCTCCTCTTTAAGC
pglK-F	ATGTTAAAAAAACTTTTTTTTATTTTAAG
pglI-R	CTAATTTTTGCATAAAGCCACC
pglI-F	ATGCCTAAACTTTCTGTTATAGTGC
pglB-R	TTAAATTTTAAGTTTAAAAACTTTAGC
pglB-F1300	CAACTATAAAGCGCCAACAG
pglE-R	TTAAGCCTTTATGCTCTTTAAGATC
pglE-F	ATGAGATTTTTTCTTTCTCCTCC
cheY-R	GTGAAATTGTTAGTTGTTGATG
wlaA-F	GTTTCCATGGCAAAAAATGAAGGTTATATTTG
pglG-R	GTTTCTCGAGTCAAATATTATTATTTTTATACTTAACCGTAAATAAA
cmeA-F	GTTTCCATGGATAGCAAAGAAGAAGCACC
cmeA-R	GTTTCTCGAGTTGTGCTCCAATTTCTTTAAC
gumD-F	ATGATCATCGTTACCGGC
waaT-R	ATACTTTAATATCGCGCGTG
inv-waaL-F	GGATCCATATTTTCAACCTATGCTAC
inv-waaL-R	GGATCCTGAATATGTGAAATAAAATC
kanR-F	GGATCCCAGTGAGAGCAGAGATAGC
kanR-R	GGATCCCGTCGATCAGTTCTTGCC
M13-F	GTTTTCCCAGTCACGAC
M13-R	CAGGAAACAGCTATGAC
pglH(down)-R	CCACACTTTTAGTATCATCGCTAGAATTATCATCGC
pglJ(ups)-F	GCCTAAGTATAAAAATCAACGCTCCTGCTTTG
pglA-R	GTTTGGATCCTCATACATTCTTAATTACCCTATCATAAAGTTTT
pET22bGT-F	AACGCGACCCACCACCACCACCACCACTGAGAT
pET22bGT-R	CTGATCGCCGCCCTCGAGTGCGGCCGCAAGC
pET22bGT4-F	GGCGGCGATCAGAACGCGACCGGCGGCGATCAGAACGCGACCCACCACCACCACCACCAC
pET22bGT4-R	GGTCGCGTTCTGATCGCCGCCGGTCGCGTTCTGATCGCCGCCCTCGAGTGCGGCCGCAAG
pACYC(T7down)-F	CCTGTAGAAATAATTTTGTTTAACTTTAATAAGGAGATATACC
pACYC(T7up)-R	ATTTCCTAATGCAGGAGTCGCATAAGG

2.3 Results and discussion

2.3.1 Identification of the *pgl* operon from *C. jejuni* JCM 2013

The nucleotide sequence of the *pgl* operon of *C. jejuni* JCM 2013 has never been reported, and therefore there was no information about the operon organization, including insertion, deletion or variation of operon components. To elucidate the *pgl* operon organization in the *C. jejuni* JCM 2013 strain, 15,597 nucleotides of the putative *pgl* locus were cloned. This *pgl* operon contained 14 ORFs, [*wlaA, gne, pglK, pglH, pglI, pglJ, pglB, pglA, pglC, pglD, wlaJ, pglE, pglF, pglG*] (Figure 2-1), with 98% identities for overall nucleotide sequence, and 99% for overall amino acid sequence, compared to that of *C. jejuni* NCTC 11168 (GenBank: AL111168.1). There were no insertions, deletions, or alterations of order or direction of any components. The sequence data and annotations can be accessed under accession number (AB872218). The sequence of CmeA can be found in the Appendix.

Figure 2-1. Identified *pgl* operon of *C. jejuni* JCM 2013

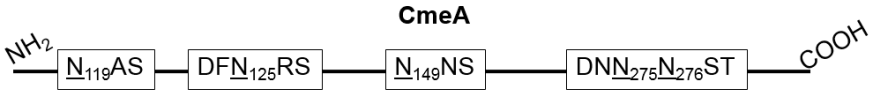

Figure 2-2. Landmark of *N*-glycosylation sites available on CmeA

2.3.2 Heterologous expression of the *pgl* operon and *C. jejuni* glycoprotein CmeA in *E. coli*

To examine the *N*-glycosylation activity of the putative *pgl* operon from *C. jejuni* JCM 2013, the cloned *pgl* operon was co-expressed with CmeA as a glycan acceptor protein. CmeA cloned from *C. jejuni* JCM 2013 has 2 extended *N*-glycosylation sites at amino acid N_{125} and N_{275} or N_{276} (Figure 2-2). To avoid the integration of foreign *Campylobacter* glycan into the *E. coli O* polysaccharide structure, the chromosomal *waaL* gene encoding *O*-antigen ligase was disrupted by insertion of a bacterial kanamycin-resistance gene (Figure 2-3). The obtained mutant, designated as the *E. coli* BL21 Δ*waaL*, was used in protein expression.

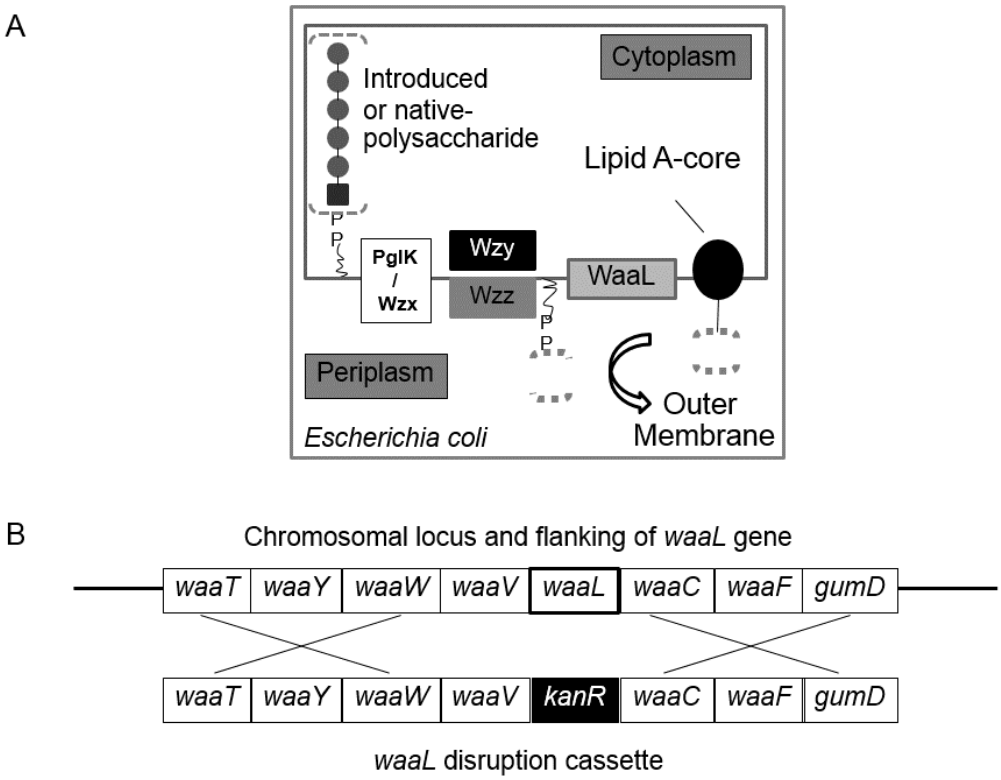

Figure 2-3. Construction of the *waaL* disruption cassette. A: Possible function of WaaL as an *O*-antigen ligase that inhibits the accumulation of assembled polysaccharide by ligating polysaccharides to the cell wall structure. B: Chromosomal locus of *waaL* in the genome of *E. coli* BL21 (DE3) and the corresponding disruption cassette.

The glycan acceptor protein CmeA was expressed with the N-terminal PelB leader signal peptide to facilitate periplasmic secretion, where the *C. jejuni*-type glycosylation is expected to occur. The recombinant proteins encoded in the *pgl* operon and CmeA were produced without IPTG induction. Use of IPTG in the constructed system (pPgl and pCmeA) yielded the CmeA protein product, but failed to generate glycosylated CmeA, as shown on Western blot. The use of the described expression plasmids (pPgl and pCmeA) in the *E. coli* BL21 (DE3) wild type resulted in abolished glycosylation compared to expression in the *E. coli* BL21 Δ*waaL* (Figure 2-4).

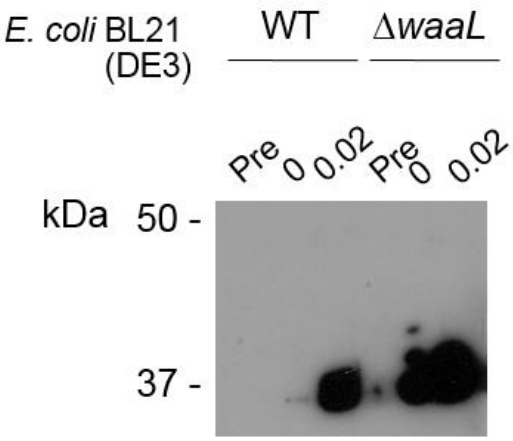

Figure 2-4. The putative CmeA glycosylation activity by the cloned *pgl* operon. Expressions in *E. coli* BL21 (DE3) and the Δ*waaL* mutant were compared by loading total protein extracts on the Western blot. Protein glycosylation was only observed in the non-IPTG-induced Δ*waaL* strain expression. *Pre* indicates the pre-induced sample, and the numbers indicate the IPTG concentrations.

The periplasmic-localized recombinant CmeA was purified by Ni-NTA and analyzed by SDS-PAGE (Figure 2-5 A) and Western blotting (Figure 2-5 B). SDS-PAGE analysis showed that the CmeA protein was purified to at least 95% homogeneity in the 50 mM imidazole-eluted fraction. In both analyses, two additional protein bands with slower mobility were observed along with the main protein band. These additional bands probably represented the glycosylated CmeA, with one and two *N*-glycan chains.

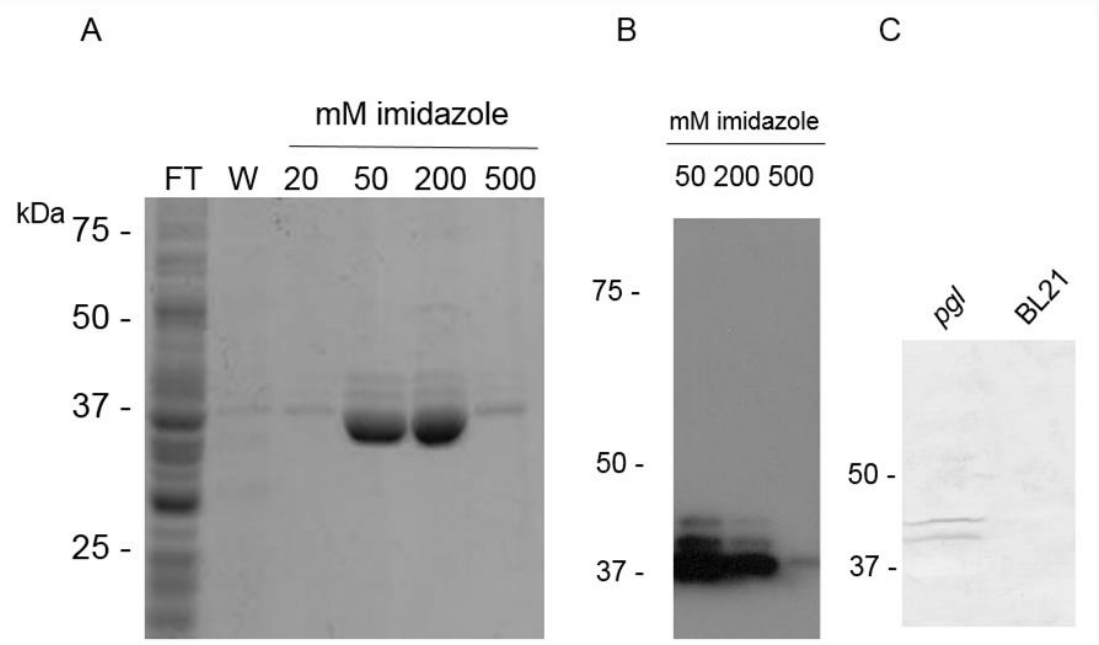

Figure 2-5. CmeA purification. A: Coomassie-stained SDS-PAGE of Ni-NTA purified CmeA. B: Western blot of purified CmeA where His-tagged CmeA was detected with anti-His antibody. C: SBA lectin blot of CmeA carrying glycan synthesized from pPgl. BL21 was *E. coli* BL21 Δ*waaL*, transformed with blank plasmids.

The *C. jejuni* JCM 2013 *pgl* operon identification revealed that the structure of this operon was similar to that of the previously reported *C. jejuni* NCTC 11168 *pgl* operon (40). Therefore, *C. jejuni* JCM 2013 was predicted to possess a similar *C. jejuni N*-glycan structure (GalNAcα1,4-GalNAcα1,4-[Glcβ1,3-]GalNAcα1,4-GalNAcα1,4-GalNAcα1,3-diNAcBac).

To examine glycosylation on the recombinant CmeA expressed in the constructed system using *C. jejuni N*-glycan synthesized by *pgl* operon components, lectin blotting with soybean agglutinin (SBA), which reacts to the terminal GalNAc residue (66, 132), was performed. Crude periplasmic extracts of the *E. coli* BL21 Δ*waaL* co-expressing pPgl and pCmeA were subjected to SDS-PAGE and SBA lectin blotting. The two slowly migrating bands above the molecular standard of 37 kDa reacted to SBA lectin (Figure 2-5 C). This migration pattern corresponded to SDS-PAGE and anti-Penta His Western blotting of the purified CmeA. These results suggested that CmeA was glycosylated with a glycan structure that possessed a GalNAc residue at the non-reducing end.

2.3.3 nanoLC/MS/MS analysis of the recombinant glycosylated CmeA

To analyze the detailed glycan structure on CmeA, a nanoLC/MS/MS analysis was performed. The Ni-NTA purified protein band corresponding to the expected glycosylated CmeA was excised, destained, in-gel digested with trypsin and subjected to analysis with the nanoLC/MS Q-TOF system. After the chromatogram was integrated and deconvoluted, the tryptic glycopeptide of interest containing a glycosylation motif with the putative glycosylated Asn underlined [AIFDNNNSTLLPGAFATITSEGFIQK], as depicted in Fig. 2-6 A, was identified as a major glycosylated peptide signal at *m/z* 4149.84. Due to the technical limitations of the nanoLC/MS/MS, it was still unclear which Asn residue of the DNNNST glycosylation motif was glycosylated. In addition, the MS/MS fragmentation analysis did not show the MS signal of the predicted diNAcBac-containing glycopeptide (*m/z* 2997), but the glycopeptide consisted of 6 HexNAc residues and 1 Hex residue (*m/z* 2972.33) (Fig. 2-6 B). This analysis revealed that the *N*-linked glycan at the reducing end was HexNAc, not the predicted diNAcBac. Together with the results of SBA lectin blotting, these results suggested that the glycan structure in the recombinant CmeA co-expressed with the *pgl* operon of *C. jejuni* JCM 2013 was GalNAc-($HexNAc_5$Hex)-Asn, which can be inferred to have the structure of GalNAc-HexNAc-(Hex-)HexNAc-HexNAc-HexNAc-HexNAc-Asn, which in turn corresponds to the previously reported structure of GalNAc-GalNAc-[Glc]GalNAc-GalNAc-GalNAc-HexNAc-Asn (40).

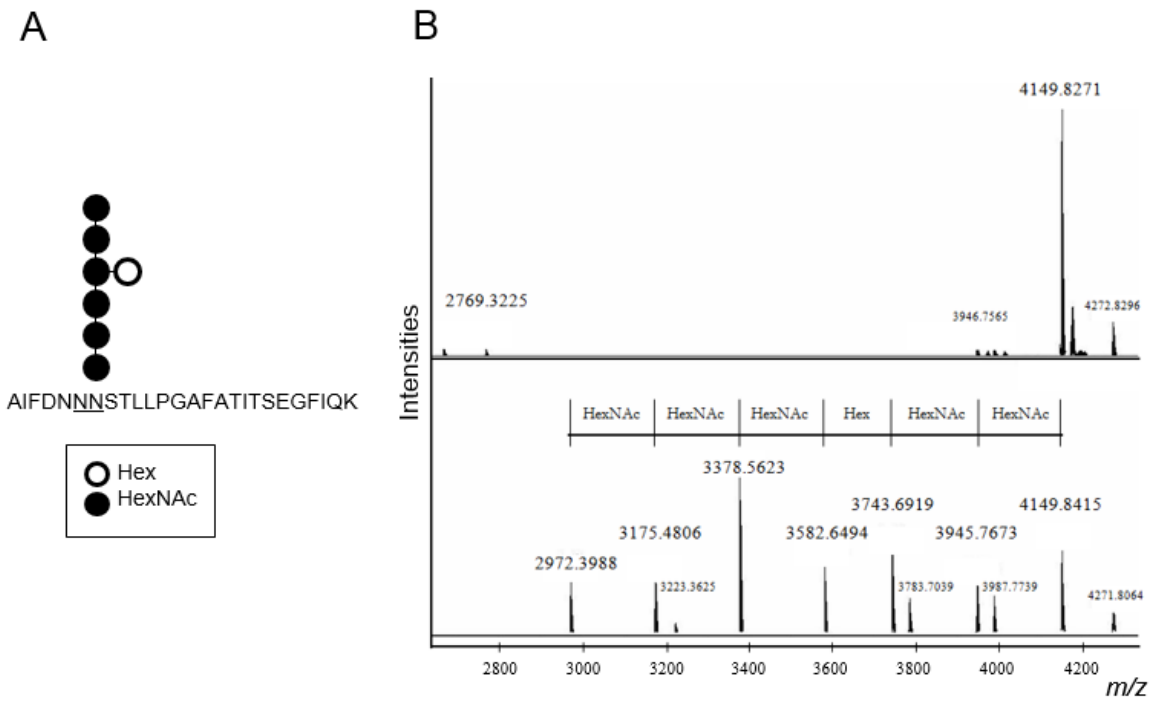

Figure 2-6. CmeA glycopeptide analysis. A: Predicted glycopeptide. B: Mass spectrometric analysis of glycopeptide.

2.3.4 Transcriptional analysis of the diNAcBac synthase and diNAcBac transferase genes under protein expression condition in the constructed *E. coli* system

Although the diNAcBac moiety was predicted to be present at the reducing end of the *N*-linked glycan in the recombinant CmeA, it was not observed. This result differed from the previous observation of *C. jejuni N*-linked glycan production in *E. coli* expressing the *C. jejuni* NCTC 11168 *pgl* operon (56). To explore the reason for the absence of the diNAcBac residue in the *N*-glycan, the expression of the genes involved in diNAcBac synthesis – *pglF, pglE, pglD* and *pglC* – on the *pgl* operon was analyzed by reverse transcription (RT) PCR. *pglA* was selected as a control transcript since its function has been elucidated (40), and was positioned adjacent to *pglC*. The background of DNA contamination in the RNA extract was examined, together with verification of successful cDNA synthesis in the RT-PCR reaction by *pglA* amplification as a positive control (Figure 2-7 A). All four genes were transcribed under the

protein expression conditions (Figure 2-7 B), although their expected glycan product was absent.

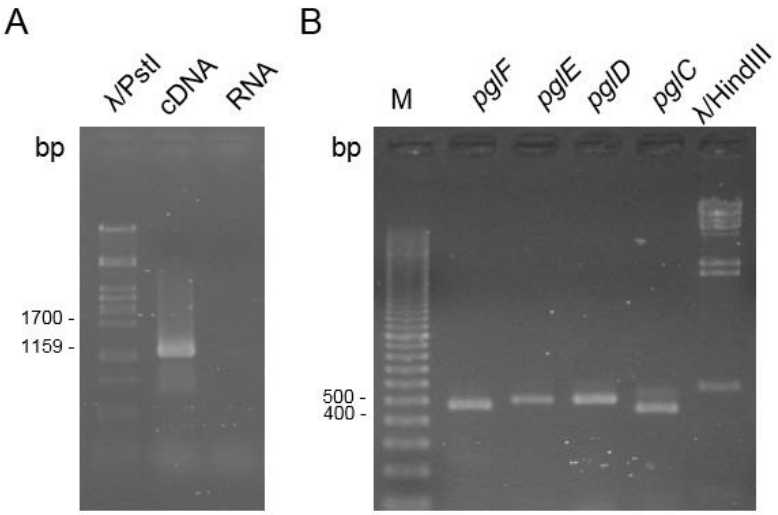

Figure 2-7. Reverse-transcriptional analysis. A: The control reaction consisting of *pglA* amplification in cDNA but not RNA. B: Transcriptional analysis of *pglF*, *pglE*, *pglD* and *pglC* genes conferring diNAcBac synthesis. M indicates a 100 bp ladder from Bio-Rad.

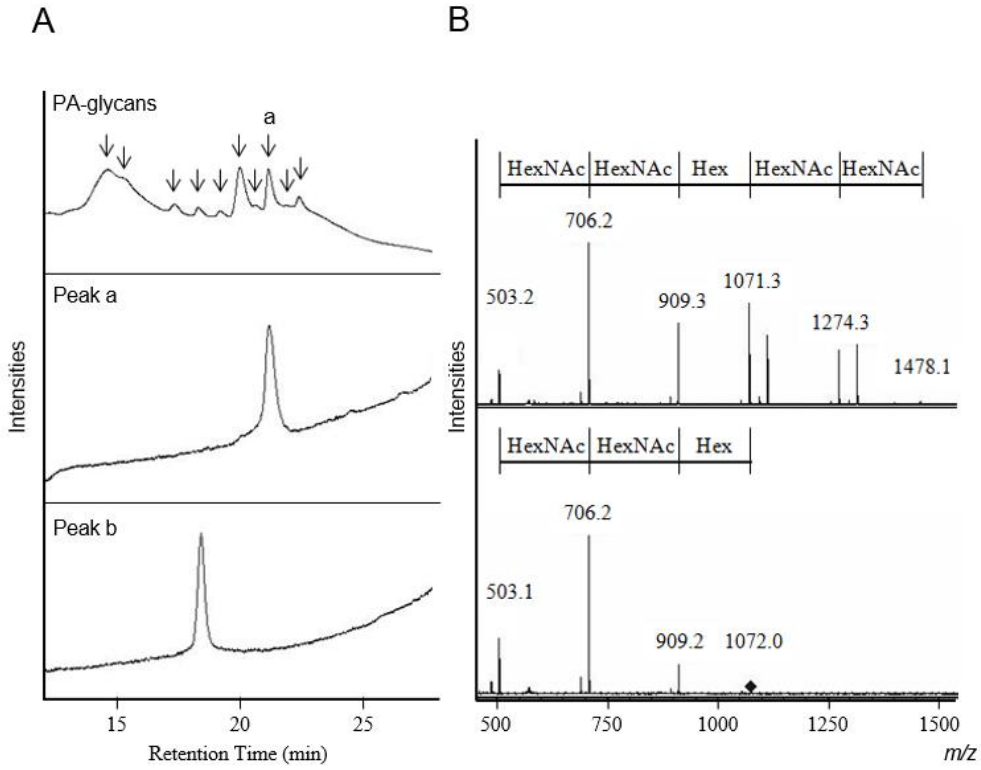

Figure 2-8. PA-glycans analysis. A: PA-glycans from CmeA. Peak a is the corresponding glycan from glycopeptide MS analysis and Peak b is the α-GalNAcase digest of Peak a. B: MS chart of the sample corresponding to Peaks a (upper) and b (lower).

2.3.5 *N*-glycan analysis of glycosylated CmeA

In order to analyze the *N*-linked glycan structure on CmeA in detail, including assessment of the monosaccharide constituent and linkage manner, a conventional *N*-glycan analysis by PA labeling with glycosidase digestion was performed. Using SF-HPLC, the PA-glycan prepared from the purified CmeA was separated. A total of 10 peaks were collected, concentrated and analyzed by LC-MS. The results showed that only the peak a eluted on SF-HPLC at 21.8 min corresponded to PA-glycan of CmeA (Figure 2-8 A). The other 9 peaks were analyzed and found not to correspond to signal of any PA-oligosaccharide, since they lacked the characteristic fragmentation pattern in MS/MS analyses (data not shown). Peak a possessed an *m/z* of 1478.1. By MS/MS sequencing, this signal was fragmented to *m/z* signals of 1274.3, 1071.3, 909.3, 706.2 and 503.2, which corresponded to Hex-$HexNAc_5$-PA, Hex-$HexNAc_4$-PA, $HexNAc_4$-PA, $HexNAc_3$-PA, and $HexNAc_2$-PA, respectively (Figure 2-8 B). This structure agreed with the findings of the nanoLC/MS/MS analysis. Peak a was then digested with α-*N*-acetylgalactosaminidase (α-GalNAcase). The α-GalNAcase digest was analyzed on SF-HPLC, eluted at 18.7 min and named peak b (Figure 2-8 A). Peak b was subjected to MS analysis and showed an *m/z* of 1072.0, which was fragmented to *m/z* signals of 909.2, 706.2 and 503.1; the peak was found to correspond to Hex-$HexNAc_4$-PA (Figure 2-8 B). This result showed that 2 α-GalNAc residues were digested, and the digestion was stopped at the HexNAc residue possessing a Hex-branch. Collectively, these results suggested that the proposed glycan structure attached to the glycosylated recombinant CmeA was GalNAcα-GalNAcα-(Hex-)HexNAc-HexNAc-HexNAc-HexNAc-Asn.

2.3.6 Expression of *pgl* operon derivatives

In an attempt to minimize the length the *pgl* operon, plasmids containing different versions of a minimized *pgl* operon (Figure 2-9) were constructed. For example, pPgl mini consisted of ORFs on pPgl but with omission of the ORFs following *pglA*, which contained

genes conferring synthesis of diNAcBac, as well as several unidentified ORFs. From pPgl Δ*pglI* and pPgl mini Δ*pglI*, an inverse-PCR was performed to remove the ORF of *pglI*. The obtained plasmids were introduced into *E. coli* Δ*waaL* for co-expression with pCmeA and expressed as described above. Glycosylated CmeA proteins were found in the periplasmic space and all of them were sensitive to α-GalNAcase digestions (Figure 2-10).

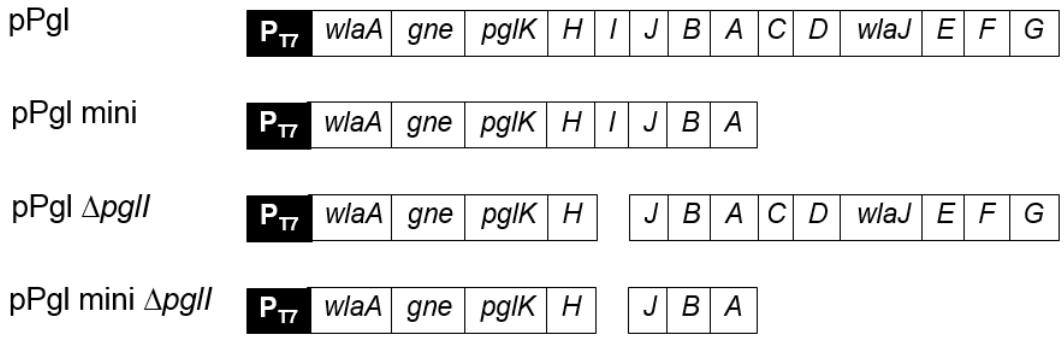

Figure 2-9. Construction of pPgl and its derivatives. Pgl mini was Pgl without the ORFs following *pglA*. The operons Pgl with *pglI* deletion and Pgl mini with *pglI* deletion were designated Pgl Δ*pglI* and Pgl mini Δ*pglI*, respectively.

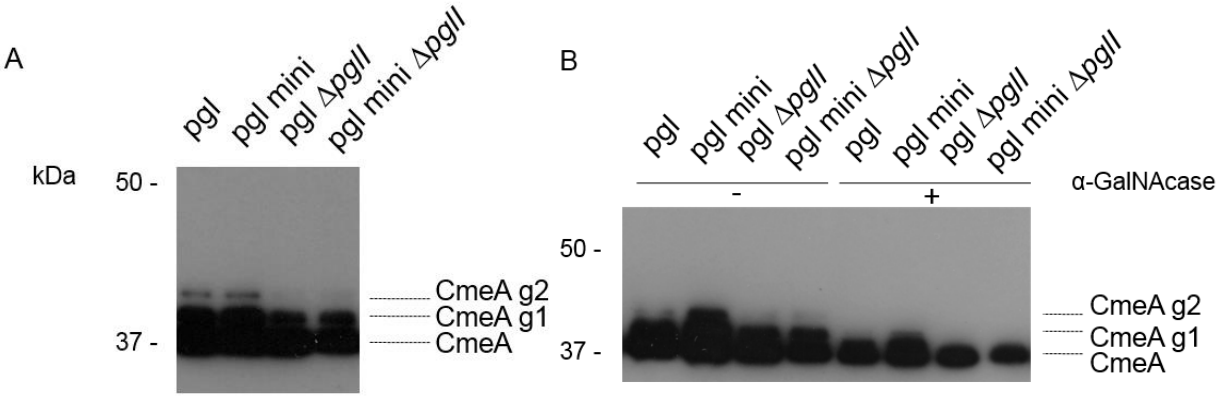

Figure 2-10. CmeA glycosylated with pPgl and the derivative glycosylation plasmids. A: Western blot of CmeA, co-expressed with *pgl* operon derivatives. B: Deglycosylation by α-GalNAcase digestions. Complete digestion was observed in the case of the *pglI*-deleted *pgl* operons. CmeA g1 and g2 were putative glycosylated CmeA bands.

MS analysis of CmeA putative glycopeptide resulted from glycosylation activity of pPgl Δ*pglI* protein expression is shown in Figure 2-11. A summary of the operon structures and their resulting glycan structures is given in Table 2-2. In brief, the *pgl* operon and *pgl* mini, yielded the same detailed glycan structure as seen in the constructed *E. coli* BL21 Δ*waaL*.

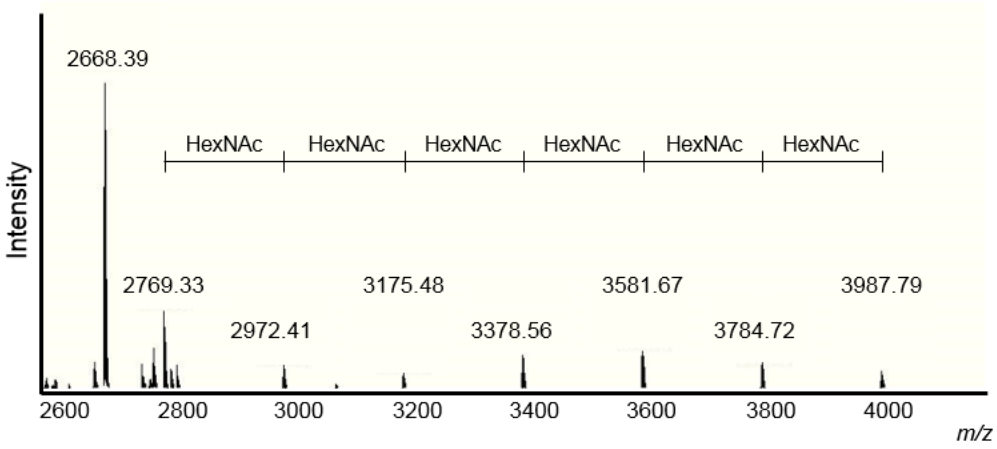

Figure 2-11. Mass spectrometry of a CmeA derived peptide, glycosylated with pPgl Δ*pglI*

Table 2-2. Summary of *pgl* the operon structures and their resulting glycan structures when expressed in *E. coli* BL21 Δ*waaL*

Construct Number	Plasmid Description	Observed *N*-glycan structure
1	pPgl	(Hex)$HexNAc_6$-Asn
2	pPgl mini	(Hex)$HexNAc_6$-Asn
3	pPgl Δ*pglI*	$HexNAc_6$-Asn
4	pPgl mini Δ*pglI*	$HexNAc_6$-Asn

2.3.7 Elucidation of glycan composition and the glycan residue at the reducing end of *N*-linked glycan

To analyze the glycan composition and the glycan moiety at the reducing-ends of *N*-glycans, which were synthesized in *E. coli* BL21 Δ*waaL* co-expressing pPgl Δ*pglI* and pCmeA, cells were cultivated. The results of MS analysis of the glycopeptide are shown in Figure 2-11. PA-glycans resulting from the protein glycosylation of pPgl Δ*pglI* was prepared. The PA-glycans were resolved on SF-HPLC, and a peak corresponding to $HexNAc_6$-PA was identified (Figure 2-12 A). The mass spectrum of $HexNAc_6$-PA was obtained and is shown in Figure 2-12 B. The $HexNAc_6$-PA sample was subjected to α-GalNAcase digestion. The Digest was resolved on SF-HPLC with a more moderate gradient than that used for analysis of longer glycan chains. The retention time of the PA-monosaccharide produced by the digest was compared against Glc-PA, GlcNAc-PA and GalNAc-PA standards (Figure 2-13). The digest possesses PA-monosaccharide, which was eluted at a retention time of 20 min, just as in the case of GlcNAc-PA. Thus, it can be inferred that the glycan moiety at the reducing-ends of *N*-glycans synthesized in this constructed system was GlcNAc. Furthermore, *N*-glycan structure (Hex)-α$GalNAc_5$-GlcNAc, could be inferred.

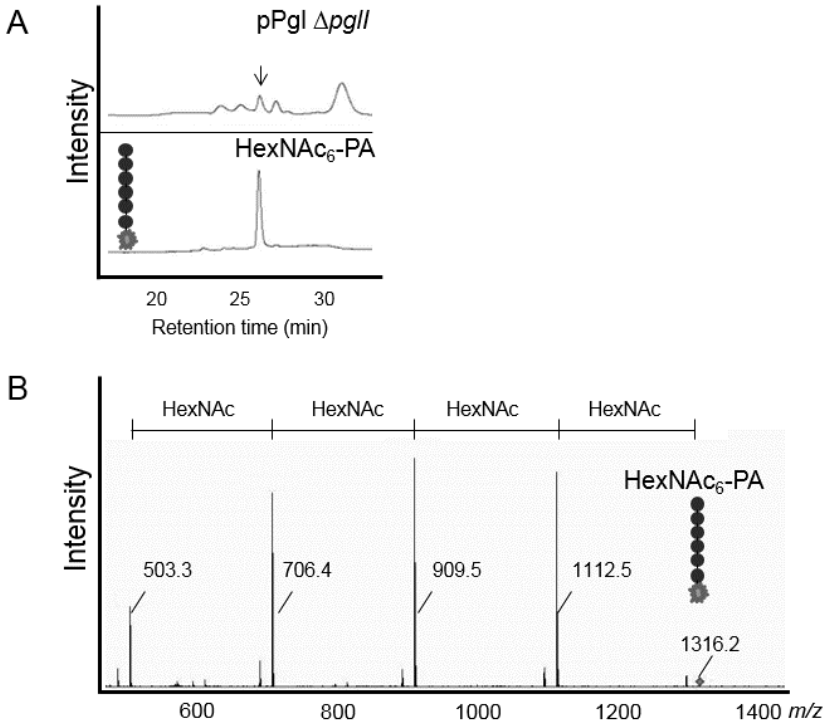

Figure 2-12. PA-glycans analysis. A: PA-glycan preparation of CmeA glycosylated with pPgl Δ *pglI* and resolved on SF-HPLC. The purified $HexNAc_6$-PA was eluted at 26.5 min. B: $HexNAc_6$-PA was subjected to MS analysis. The results showed 4 fragmented HexNAc ions, along with loss of 1 Hex ion that was observed in the pPgl expression.

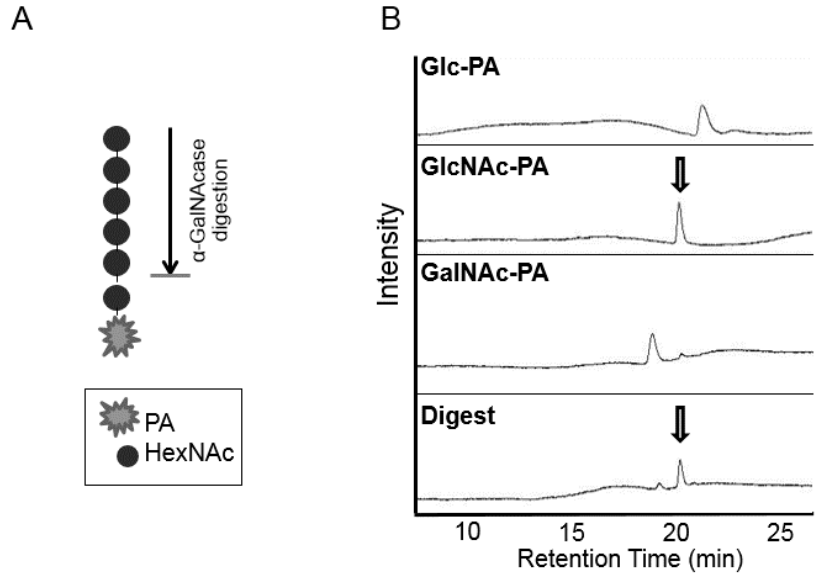

Figure 2-13. Analysis of the glycan moiety at the reducing ends. A: The scheme of α-GalNAcase digestion of $HexNAc_6$-PA. B: The resulting PA-monosaccharide from the digest was compared against standard PA-monosaccharides on a gradual SF-HPLC run program. Because its elution time of PA-digest was identical to that for GlcNAc-PA (arrows), the reducing end of glycan synthesized in the constructed system was determined to be the GlcNAc moiety.

2.3.8 Promoter and requirements pertaining to protein expression of the *pgl* operon

The identified *pgl* operon contained 14 putative ORFs. Till this work, there has been no evidence that the external promoter required for protein expression, nor indeed that the presence of *wlaA* is particularly important. To elucidate the essential component for protein expression of the *pgl* operon, either P_{T7} or *wlaA* was deleted from pPgl. The resulting constructs are shown in Figure 2-14.

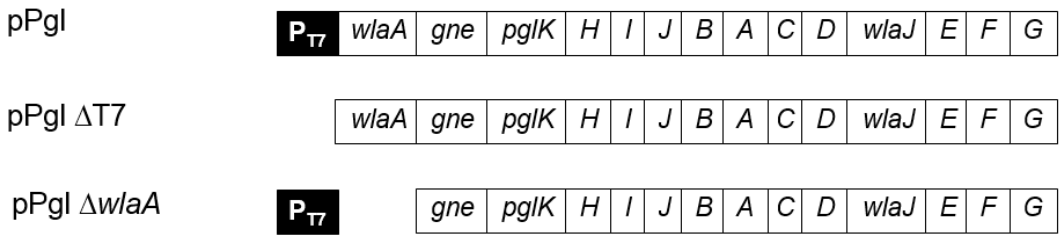

Figure 2-14. Construction of *pgl* expression vector with P_{T7} deletion or *wlaA* deletion.

A *pgl* operon was cloned into plasmid pACYCDuet-1 ΔT7 for which inverse-PCR had been performed beforehand to disrupt P_{T7} and the *lac* operator from pACYCDuet-1. The resulting pPgl ΔT7 was introduced into *E. coli* BL21 Δ*waaL* along with pCmeA. Glycosylation of periplasmic CmeA was probed by Western blotting against anti-His. CmeA protein glycosylation was eliminated (Figure 2-15 A).

To elucidate the function of *wlaA*, *wlaA* was deleted from pPgl. The resulting pPgl Δ*wlaA* (Figure 2-14) was co-introduced with pCmeA to *E. coli* BL21 Δ*waaL*. No protein glycosylation was observed as a result of this treatment (Figure 2-15 B). In brief then, based on the observations in the present work, an external promoter is required for protein expression of *pgl* operon, and *wlaA* is also required even its function has not been elucidated.

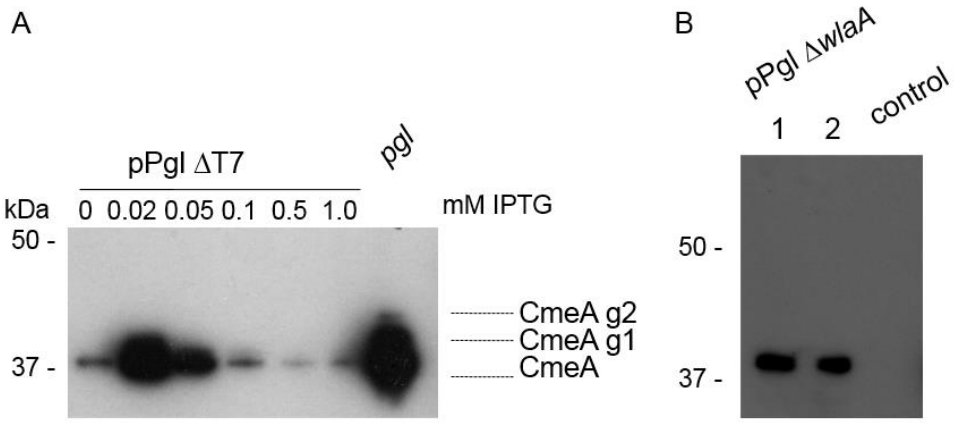

Figure 2-15. CmeA glycosylation by pPgl with a promoter deletion or with deletion of *wlaA*. A: CmeA glycosylation was not observed when pPgl ΔT7 was used. B: *wlaA* ORF, which has an undetermined function, was required for protein glycosylation. CmeA g1 and g2 were putative glycosylated CmeA bands.

As described above, when the *pgl* operon was placed under P_{T7}, IPTG induction was detrimental for protein glycosylation. Even at a concentration as low as 0.02 mM IPTG, induction was not successful. Only constitutive expression was allowed, and basal transcription of P_{T7} was required for protein expression. This situation would not allow higher expression of the *pgl* operon or of the model protein, which was also expressed under P_{T7}. In an attempt to resolve this limitation, we designed a model protein that would continue to be expressed under P_{T7}, while the *pgl* operon was designed to be expressed under mildly-inducible P_{BAD} (133). P_{T7} of pACYCDuet-1 was exchanged with P_{BAD} of native *E. coli* BL21 (DE3). The resulting plasmid was named pAra-o. *pgl* operon fragments derived from pPgl and pPgl mini were subcloned into pAra-o at *Nco*I and *Xho*I, resulting in P_{BAD} pgl and P_{BAD} pgl mini Δ*pglI*, respectively (Figure 2-16).

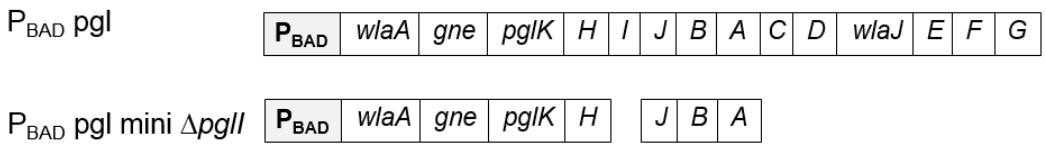

Figure 2-16. Plasmid construction of arabinose-inducible *pgl* operons.

The P_{BAD} pgl and P_{BAD} pgl mini Δ*pglI* constructs were tested for induction conditions. Glycosylation plasmid was induced by L-ara at concentrations of 0.2%, 0.02%, 0.002%, 0.0002%, and 0.00002%. Periplasmic materials were prepared from cells expressed at different inductions. The L-ara concentrations of 0.002% and 0.02% seemed to be optimal for inducing protein glycosylation using P_{BAD} pgl and P_{BAD} pgl mini Δ*pglI*, respectively. At a concentration of 0.2% L-ara, the operon appeared to be overinduced and glycosylation was diminished as a result (Figure 2-17). The results suggested that the *pgl* operon could be mildly induced by an arabinose promoter.

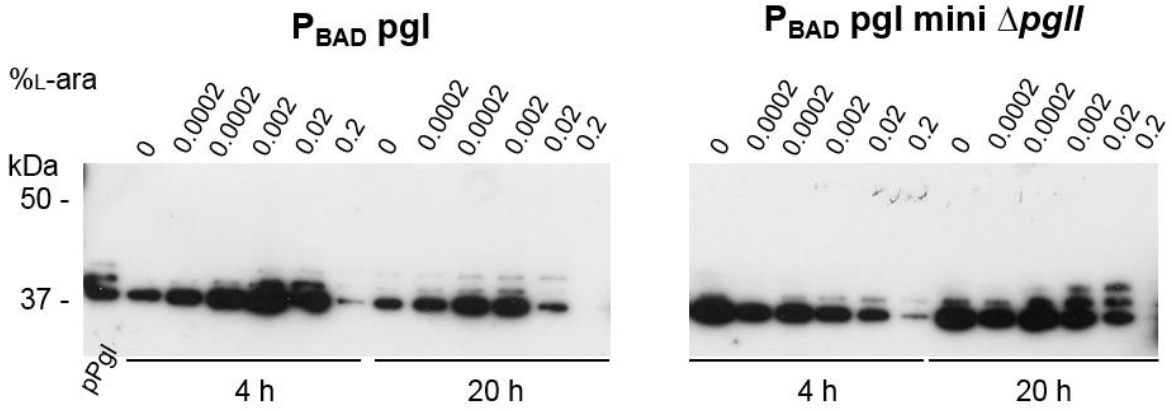

Figure 2-17. CmeA glycosylation by P_{BAD} pgl and P_{BAD} pgl mini Δ*pglI*. Different L-ara concentrations were used, and cells were harvested at 4 h and 20 h. The most intense glycosylation could be observed when the *pgl* operon was induced at 0.02% L-ara. pPgl was sample taken from *E. coli* BL21 Δ*waaL* expressing pPgl and pCmeA.

To test that the promoters P_{BAD} for the *pgl* operon and P_{T7} for the model protein are independent and self-sufficient for the protein expression of dual plasmids, IPTG induction was performed in a system of *E. coli* BL21 Δ*waaL* harboring P_{BAD} pgl and pCmeA, and P_{BAD} pgl mini Δ*pglI* and pCmeA. The L-ara concentration was maintained at 0.02% while varying IPTG concentrations were used (Figure 2-18). At 0.01 or 0.02 mM IPTG, the expression and glycosylation of CmeA were optimal in both transformants. The results showed that more IPTG induction does not always led to greater amounts of either protein or glycosylated protein. However, a certain concentration of IPTG should be used. Furthermore, the induction conditions might be adapted to the fermentation mode of each of the proteins being used.

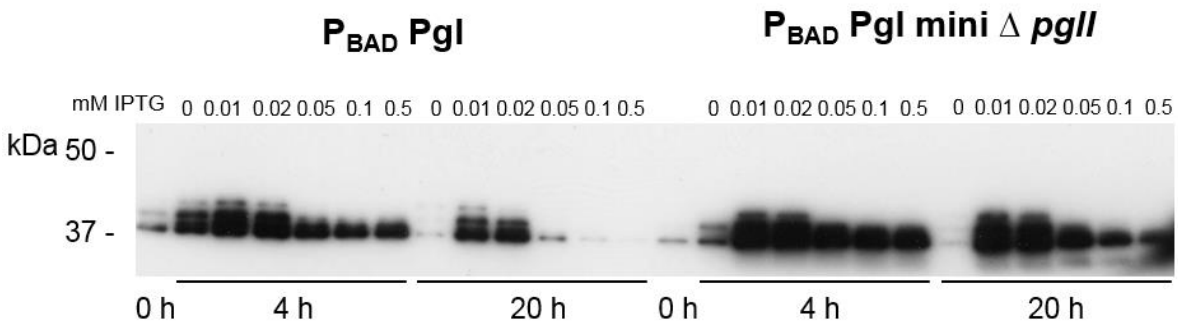

Figure 2-18. Induction of the *pgl* operon under P_{BAD} and of CmeA under P_{T7}. Concentrations of IPTG were tested to screen for optimum protein production while the L-ara level was kept at 0.02%. The results showed that an IPTG concentration of 0.02 mM was optimal for glycosylated CmeA expression.

2.3.9 Glycosylation of other proteins containing a Glycosylation Tag

DQNAT sequence, a theoretically strong glycosylation motif towards PglB, has been shown to initiate glycosylation both *in vitro* (134) and *in vivo* (104). To construct a glycan acceptor protein, inverse-PCR was performed using a nucleotide primer coding the amino acid GGDQNAT inserted at the front of the 6xHis sequence on pET-22b. Plasmid pET-22bGT was generated. *malE* and *dsbA* were subsequently ligated at *Nde*I and *Xho*I, and *Nde*I and *Nco*I, respectively. These constructs are outlined in Figure 2-19. A plasmid pET-22bGT$_4$, coding

(GGDQNAT)$_4$ - GT$_4$, was also constructed. Finally, glycan acceptor proteins including MBP-GT, MBP-GT$_4$, DsbA-GT and DsbA-GT$_4$ could be produced.

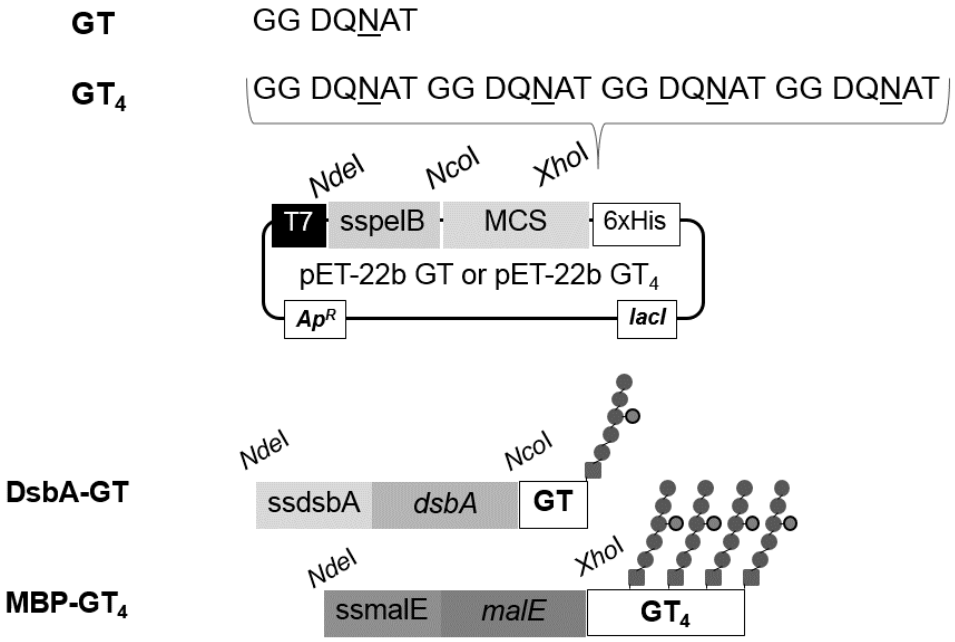

Figure 2-19. Constructions schemes for the plasmids pET-22b GT and pET-22b GT$_4$. Examples of cloned model proteins and hypothetic glycosylation on GT and GT$_4$ motifs are depicted.

Protein-GT plasmids were introduced into *E. coli* BL21 (DE3) and BL21 Δ*waaL*, in the latter case with co-introduction of pPgl. Periplasmic material was prepared. Glycosylated proteins, which migrated slower than similar proteins expressed in the non-glycosylating (wild type BL21) *E. coli*, could be observed (Figure 2-20). Their calculated tryptic mass values of peptides containing GT or GT$_4$ and generated by trypsin digestion are displayed in Table 2-3.

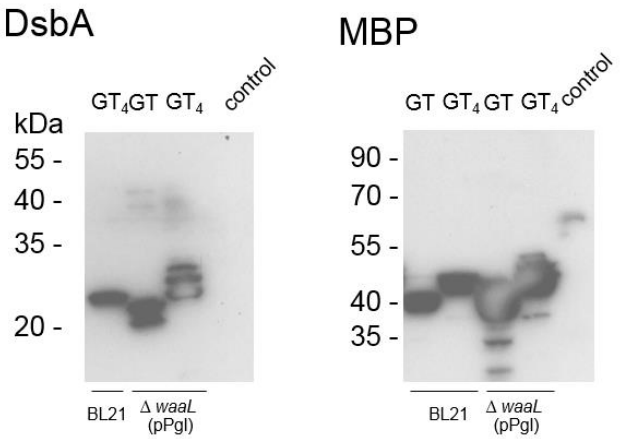

Figure 2-20. Protein expressions of GT and GT_4 MBP and DsbA. Glycosylation was performed in *E. coli* BL21 Δ*waaL* expressing pPgl. The periplasmic fraction was prepared and analyzed by anti-His Western blot. *E. coli* BL21-expressing blank plasmids were used as a control.

Table 2-3. Tryptic peptides carrying glycosylation motifs in DsbA and MBP on GT plasmids, shown with their calculated masses and predicted pI values.

Name	Tryptic Peptide	Mass value (Da)	pI
MBP-GT	LEGGDQNATHHHHHH	1726.75	6.27
MBP-GT_4	LEGGDQNATGGDQNATGGDQNATGGDQNATHHHHHH	3657.58	5.41
DsbA-GT	LAAALEGGDQNATHHHHHH	2053.14	6.27
DsbA-GT_4	LAAALEGGDQNATGGDQNATGGDQNATGGDQNATHHHHHH	3983.93	5.40

MS analysis was performed and revealed the glycan structures on the designed GT motifs (Figures 2-21 and 22). MBP-GT mass analysis showed a GT glycosylated peptide at *m/z* 3107.17, which fragmented 3 HexNAc ions and 1 Hex ion, where *m/z* 2335.93 was the predicted peptide carrying 3 HexNAc ions. The unglycosylated peptide was also shown at *m/z* 1726.69, with a mass value similar to the estimated value. The peptide from MBP-GT_4 was shown to be glycosylated with a glycan structure derived from pPgl at *m/z* 5036.89, which fragmented 3 HexNAc ions and 1 Hex ion.

Surprisingly, when DsbA-GT was co-expressed with pPgl, glycopeptide signals containing Asn-HexNAc (precursor *m/z* at 3433.37) and a minor population of Asn-228 Da residue (precursor *m/z* 3458.40) were observed. This might have represented glycosylation of a small amount of diNAcBac (theoretical mass~ 228 Da) containing glycan. It should be mentioned that DsbA is a strong disulfide isomerase (135), and thus it might create an oxidizing periplasmic environment and thereby affect protein glycosylation. Furthermore, the periplasmic extract of DsbA-expressing cells was very sticky, and was obtained in a smaller amount compared to when MBP was expressed, which might indicate stress in the periplasmic space. Possibly it might also be asserted that DsbA properties or a specific glycosylation motif of DsbA must have yielded sensitive protein glycosylation, even to a minor glycan population.

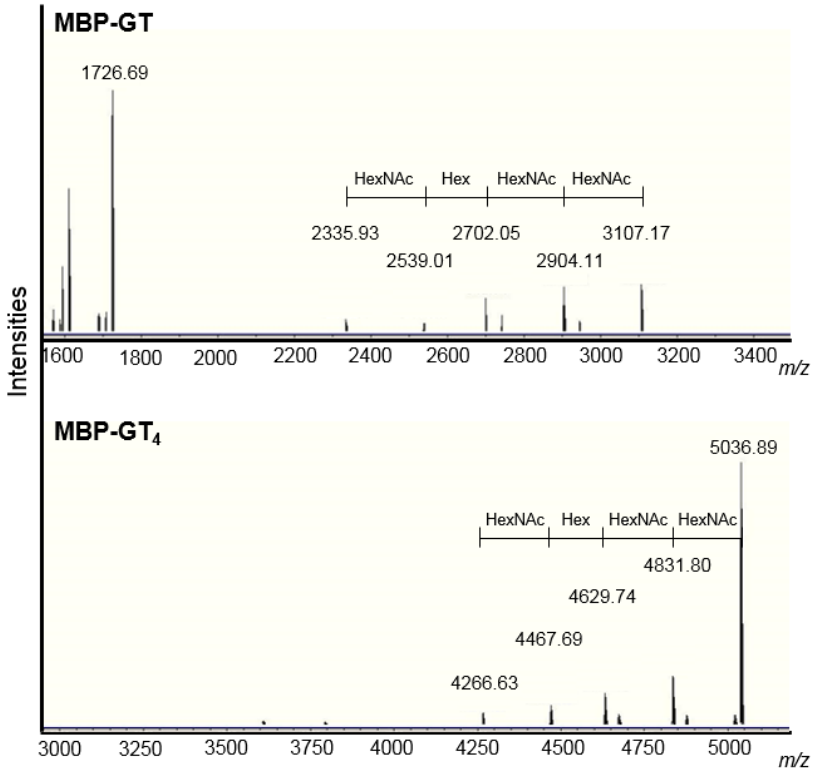

Figure 2-21. Mass spectrometry of the protein glycosylation of MBP-GT and MBP-GT_4 by pPgl.

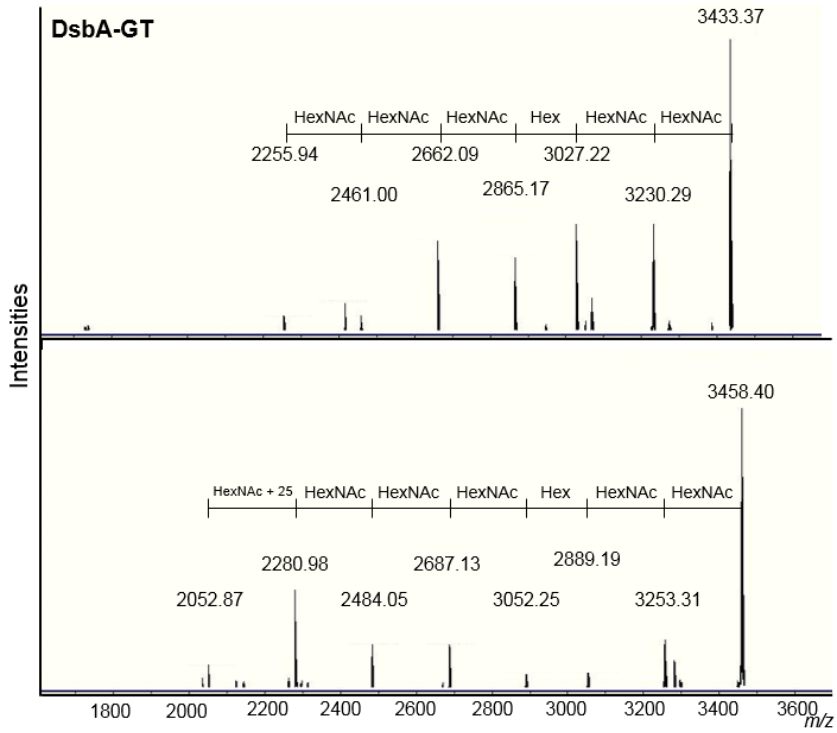

Figure 2-22. Mass spectrometry of DsbA-GT glycosylated by pPgl activity. The mass spectrum in the lower panel represents a minor glycosylated portion of DsbA exhibiting putative glycosylation of a diNAcBac-containing glycan structure. The predicted diNAcBac (228 Da) was listed as the HexNAc+25 fragment.

2.4 Summary

In this chapter, the *pgl* operon from *C. jejuni* JCM 2013 was cloned and identified. Protein glycosylation on the CmeA model protein was observed when the plasmid containing the *pgl* operon (pPgl) was co-introduced with pCmeA into *E. coli* BL21 Δ*waaL*. Glycan structural analysis by mass spectrometry of glycopeptide and by flourophore labeling revealed the Hex-$HexNAc_6$ structure, which was further identified by α-GalNAcase digestion to be $GalNAc_2$-Hex-$HexNAc_4$. When the gene conferring a Glc branching enzyme was disrupted from pPgl, the GalNAc residues were completely digested until PA-monosaccharide was obtained, and the resulting reducing-end saccharide moiety was therefore revealed to be GlcNAc.

The glycan structure of the minimized *pgl* operon was analyzed and found to be similar to the full-length *pgl* operon. This constructed system synthesized 2 kinds of glycan structures, one with a Glc branches, and one without, which were designated Hex-$HexNAc_6$ and $HexNAc_6$, respectively. By consolidating the data presented here along with the findings from a previous publication (36), it could be inferred that the observed structures were Glc-$\alpha GalNAc_5$-GlcNAc and $\alpha GalNAc_5$-GlcNAc.

Next, a promoter and the important components for *pgl* operon expression were determined. A promoter, such as example P_{T7}, is required for *pgl* operon expression. When placed under an arabinose promoter, the *pgl* operon became inducible at certain L-ara concentrations. A *wlaA* ORF at leftmost of *pgl* operon was required for protein expression of the *pgl* operon, even though the detailed function of this *wlaA* was not elucidated.

Glycan acceptor proteins containing a glycosylation tag (GT) were constructed. Periplasmic-expressed DsbA-GT, DsbA-GT_4, MBP-GT and MBP-GT_4 were obtained. Mass spectrometric analysis was used to confirm that glycan was transferred onto the constructed GT motifs. In other words, model proteins that could be used as alternatives for glycan acceptors were generated.

Chapter 3

Biosynthesis of an Initial-Stage Eukaryotic *N*-Glycan and its Protein Glycosylation in *Escherichia coli*

3.1 Introduction

N-glycosylation is an abundant protein post-translational modification. First elucidated in eukaryotes, *N*-glycosylation has since been observed in all domains of life; but the resulting glycan structures are varying. Eukaryotic oligosaccharide assembly in the dolichol pathway is shared from yeasts to humans (16), but in the prokaryotic glycosylation counterpart, the assembly of "block" saccharide takes place on undecaprenol lipid (136), which is a common constituent of bacterial oligosaccharide assemblies (103, 113). Owing to the relaxed substrate specificity of *Campylobacter* PglB OST (102, 103), it could be used for transfer of various saccharide moieties onto a desired periplasmic-targeted protein.

$Man_3GlcNAc_2$ was selected as the desired structure. It was selected because it is the smallest branched glycan unit observed in eukaryotes, and could accommodate further modifications to obtain complex-type or hybrid-type glycans. Furthermore, the $Man_3GlcNAc_2$ structure was tested and found minimally required for clinical efficacy in delivering of therapeutic glycoprotein, bGC (97).

To obtain $Man_3GlcNAc_2$ in *E. coli*, GTases Alg13/Alg14, Alg1, Alg2 (yeast nomenclature) are required. Some GTases from *S. cerevisiae* were shown to be solubly expressed in *E. coli* with observed *in vitro* activity (34, 137), so *S. cerevisiae* was chosen as a source of GTase genes. The *E. coli* endogenous lipid-GTase WecA (111) helps establish the GlcNAc-PP-Und, which is required for the subsequent glycosyltransfer steps listed above. In an engineered *E. coli*, glycosylation occurs in the periplasmic space. To exploit the native

glycan-flipping activity, the activity of *E. coli* Wzx is expected (138, 139). Finally, PglB derived from *C. jejuni* JCM 2013 (140), which was identified in Chapter 2, was used as an OST. The glycan acceptor protein was MBP-GT, which contains a synthetic prokaryotic glycosylation motif recognized by PglB. This system was transferred into the constructed *E. coli* BL21 Δ*waaL*Δ*gmd* mutant. The *gmd* gene was disrupted with the expectation of accumulating GDP-Man (141) which is a substrate of mannosyltransferases Alg1 and Alg2. The basic synthetic pathway is depicted in Figure 3-1.

Protein expression and the LLO biosynthesis in the constructed system were analyzed by Western blotting and HPLC, respectively. By subjecting fluorophore-labeling to glycan released from MBP-GT, the glycan structure was identified. Glycosylation of the glycan structure synthesized from the constructed pathway was observed on a constructed glycosylation motif by mass spectrometric analysis.

3.2 Materials and methods

For general methods regarding DNA cloning, Western blotting, mass spectrometric analysis and *N*-glycan analysis, see the information in Chapter 2.

3.2.1 Strain, medium and cultivation conditions

S. cerevisiae BY4741 was obtained from the National Bio-Resource Project (NBRP) of MEXT, Japan. Genomic DNA of *S. cerevisiae* BY4741 was extracted and used as a source of GTase genes. Genomic DNA of *Xanthomonas campestris* NBRC 13303 and of *Sulfolobus acidocaldarius* NBRC 15157 were obtained from NBRP. *E. coli* DH5α (126) and BL21 (DE3) mutants were used for genetic manipulation and protein expression, respectively. *E. coli* strains were grown in LB medium, and antibiotics of 100 mg ampicillin (Ap), 25 mg chloramphenicol (Cm), 25 mg kanamycin (Km), or 5 mg tetracycline (Tet) per liter were added to the medium for selection as required.

3.2.2 Gene amplification and plasmid construction

DNA was amplified using KOD-Plus Neo (Toyobo, Osaka) or PrimeSTAR Max (TaKaRa). The *ALG13, ALG14* and *ALG2* genes were amplified from genomic DNA of *S. cerevisiae* BY4741. The *ALG1* and *pglB* genes (original amino acid sequence from *S. cerevisiae* BY4741 and *C. jejuni* JCM 2013, respectively) were codon-optimized for *E. coli* expression and synthesized by GeneArt AG (Regensburg, Germany). N-terminal T7 tag was added to Alg13 and PglB. Alg1 has a C-terminal T7 tag. A C-terminal S Tag was added to Alg14. Thioredoxin-fused Alg2 (TrxAlg2) and MBP were expressed with a C-terminal hexahistidine (6xHis) tag. All open reading frames (ORFs), except that of the *TrxALG2* gene, contained a synthetic ribosome binding sequence (RBS), 5'-AAGAAGGAGATATATAT, in front of the ATG start codon, introduced in oligonucleotide primers. The primer sequences are listed in Table 3-1. PvuI-KpnI-T7-PglB-Fw and XhoI-PglB-Rv were used to exchange and remove the *pglK* gene from the respective plasmids. Furthermore, ALG1opt, ALG1dTMopt and *pglB*opt were amplified using the primers NotI-ALG1opt-Fw, NotI-ALG1dTMopt-Fw and PvuI-T7-ALG1opt-Rv, and PvuI-KpnI-T7-PglBopt-Fw was amplified using XhoI-PglBopt-Rv, respectively.

Amplified fragments were ligated one-by-one into pACYCDuet-1 (Novagen), whose original P_{T7} was exchanged with P_{BAD}, an arabinose inducible promoter (123) cloned from genomic DNA of *E. coli* BL21 (DE3), by using the primers pACYCT7up-araC-Fw and pACYC::PBAD-Rv (described as pAra-o in Chapter 2); the original RBS and nucleotide sequence encoding the His-tag were also removed from the polylinker region of pAra-o by the primers pACYC-SacI-F and pACYC-(-RBS&NcoI)-R. The organization of the cloned genes was as follows: [*Sac*I-*ALG13-Pst*I-*ALG14-Not*I-*ALG1-Kpn*I-*pglB-Xho*I-*T7pALG1dTM-Xho*I]. The T7pALG1 dTM is a cassette conferring the expression of an optimized Alg1 with a 35 amino acid deletion at the N-terminal, under inducible P_{T7}, inserted in the direction opposite

that of the arabinose promoter expression. The resulting plasmid was designated as pM1 (Figure 3-8).

The *ALG2* gene was amplified by primers ALG2-CDS-Fw and ALG2-no stop-Rv, and cloned using a pBAD/TOPO ThioFusion Expression Kit (Invitrogen). Subsequently, the T7 expression cassette containing the *malE* gene with C-terminal GT (104) and 6xHis, amplified from pET-22b MBP-GT (refer to Chapter 2) by primers KpnI-T7pro Fw and PstI-T7ter Rv, was ligated following the original *rrn* terminator of pBADTrxALG2 at *Pst*I and *Kpn*I (previously, inverse-PCR using primers pBADter(PstI) Fw and pBADter(KpnI) Rv was performed to introduce *Pst*I and *Kpn*I sites into pBADTrxALG2). This plasmid expresses arabinose-inducible Trx-ALG2 and T7 inducible MBP-GT, and is thus named pBADTrxALG2 MBP-GT. Furthermore, GumI and Saci_1262 were amplified by primers NotI-RBS-GumI and PvuI-T7-GumI Rv and, SacI-RBS-Saci_1262 and NotI-Stag-Saci_1262 Rv from genomic DNA of *X. campestris* pv. *campestris* NBRC 13303 and *Sulfolobus acidocaldarius* NBRC 15157, respectively. DNA sequencing of plasmids was performed prior to protein expression, using a BigDye Terminator version 3.1 cycle sequencing kit on a 3130xl genetic analyzer (Applied Biosystems).

3.2.3 Construction of *E. coli* BL21 (DE3) *waaL* and *gmd* mutant

A method similar to that used in the genomic disruption of *Salmonella* species (127) was adopted. Briefly, the *gmd* locus with a 3.5 kb flanking region was amplified by PCR using primers cpsB-Rv and wcaC-Fw and ligated into pGEM-T Easy (Promega). The ORF of the *gmd* gene on the resulting plasmid was exchanged with a tetracycline-resistance gene amplified from the pBR322 vector (142) by PCR primers tet+350-Rv and tet-350-Fw. The *gmd*-disruption cassette was amplified with an M13 primer pair, and then transformed into the previously constructed *E. coli* BL21 *waaL* mutant (refer to Chapter 2) (140). Transformants

were selected on LB agar (Km and 12.5 mg/l Tet). The resulting mutant was designated as *E. coli* BL21Δ*waaL*Δ*gmd*.

3.2.4 Protein expression, SDS-PAGE and Western blotting

Plasmids pM1 and pBADTrxALG2 MBP-GT were co-introduced into *E. coli* BL21 Δ*waaL*Δ*gmd* by electrotransformation. Freshly transformed cells were used for protein expression. To screen for protein expression conditions, 5 ml cultivation in a 50 ml screw-cap tube was performed. Different concentrations of L-ara were added for induction. For protein purification and further analysis, a culture of 1 l in a 3 l Sakaguchi flask at an OD_{600} equal to 1.0 was induced with 0.2% L-ara and a final concentration of 0.05 mM IPTG at 30˚C for 16 hrs with 100 rpm reciprocal shaking. After harvesting at 10,000 × *g* for 10 min, the cells were washed with 10 mM Tris-Cl pH 8.0, pelleted and stored at -80˚C.

After SDS-PAGE and semi-dry transfer, the blocked membrane was incubated with 1:2,000 diluted anti-Penta His antibody (Qiagen, Hilden, Germany), 1:10,000 diluted anti-T7 antibody (Novagen) or 1:10,000 diluted anti-S Tag (Novagen) as the primary antibody. Then 1:10,000-diluted horseradish peroxidase-conjugated anti-Mouse IgG (GE Healthcare, Piscataway, NJ) was used as the secondary antibody. 3-Color Prestained XL-Ladder (APRO Science, Tokushima) was included as a protein standard.

3.2.5 LLO analysis

To analyze the composition of LLO, lipid extraction was conducted essentially as described by Gao and Lehrman (143) and Folch et al. (144) Subsequently, 2-aminopyridine (PA) labeling of acid-hydrolysis-released glycan was performed by a previously described method (130). *E. coli* BL21 Δ*waaL*Δ*gmd* expressing pM1 and pBADTrxALG2 MBP-GT, induced with 0.2% L-arabinose at 30˚C for 16 h, was harvested and frozen. To the frozen cell pellet, chloroform:methanol, CM (2:1), was added in a volume 20 times that of the wet weight

of the pellet. Lipid was extracted twice by sonication and occasional vortexing for 5 min. This CM extract was washed with MilliQ water, and evaporated to dryness. The resulting CM extracted pellet was resuspended in 2 ml CM and vacuum dried. The water soluble content was washed twice with 10 ml MilliQ water. Evaporation of the pellet in methanol was performed to drive out residual water, and then the dried pellet was extracted twice with chloroform:methanol:water, CMW (10:10:3). The CMW extracts were pooled, and the CM extract was dissolved in 1 ml CMW. DEAE-cellulose was used to trap LLOs while allowing free oligosaccharide to pass through. DEAE-cellulose-bound LLOs were eluted with CMW containing 300 mM NH_4OAc. The eluate was extracted against 4 parts of chloroform and 1 part of MilliQ water. After phase separation, water was removed but care was taken not to disrupt the white residual interphase. The extract was then dried and subjected to acid-hydrolysis (0.1 N HCl in 50% isopropanol, 50˚C, 1 h). Released oligosaccharide was extracted by using a butanol-water mixture. The water phase was collected and subjected to cation-exchange chromatography prior to pyridylamination as described in Chapter 2 (140). PA-glycan was analyzed by high performance liquid chromatography (HPLC) and LC-MS analysis was performed on an Agilent Technologies 1200 HPLC equipped with an HCT mass spectrometer.

3.2.6 Protein purifications by TALON and ConA chromatography

E. coli BL21 Δ*waaL*Δ*gmd* harboring pM1 and pBADTrxALG2 MBP-GT was cultivated. Cells were harvested and periplasmic extract was prepared by osmotic release as described in Chapter 2. Dialyzed periplasmic extract supplemented with 5 mM imidazole (MP Biomedicals) was loaded onto a 5 ml bed volume of TALON cobalt affinity resin (Clontech), the resin slurry was washed with 125 ml TALON-purification buffer (30 mM Tris-Cl pH 8.0 and 300 mM NaCl) containing 5 mM imidazole, and then the bound MBP-GT was eluted with buffer containing 500 mM imidazole. The purified protein was extensively diafiltrated on a Vivaspin 20, MWCO 10,000 (Sartorius Stedim) against MilliQ water. The resulting protein

solution was lyophilized and subjected to hydrazinolysis prior to fluorophore labeling and *N*-glycan analysis.

To prepare the protein sample for mass spectrometric analysis, the glycosylated protein was trapped by concanavalin A (ConA) lectin chromatography. Cells were resuspended in ConA buffer (20 mM Tris-Cl pH 7.4, 500 mM NaCl, 1 mM $MgCl_2$, 1 mM $MnCl_2$ and 1 mM $CaCl_2$) with 1 mg/ml lysozyme, then sonicated. After centrifugation at 10,000 × g for 10 min, the supernatant was loaded onto 1 ml ConA-agarose (J-Oil Mills) prewashed with 1 M NaCl, 5 mM $MgCl_2$, 5 mM $MnCl_2$ and 5 mM $CaCl_2$ and equilibrated with ConA buffer. After the slurry was washed with 25 ml ConA buffer, elution was performed using 5 ml of 1 M methyl-α-D-glucopyranoside (Sigma). The ConA eluate was concentrated on a Vivaspin 20 and diafiltrated 3 times against TALON-purification buffer, 20 ml each time. ConA-derived, His-tagged MBP-GT was purified on TALON resin and subsequently analyzed using nanoLiquid chromatography-mass spectrometry (nanoLC/MS) after tryptic digestion.

3.2.7 Glycan structural analysis by fluorophore labeling and HPLC analysis

N-glycan structural analysis was performed as described in Chapter 2. Briefly, glycan was released from lyophilized powder of the purified MBP-GT protein by hydrazinolysis. Re-*N*-acetylation, desalting of the sample and lyophilization were performed. The obtained powder was pyridylaminated using purified PA.

Analyses were performed on a Jasco HPLC apparatus, and PA-glycans were detected at the excitation and emission wavelengths of 310 and 380 nm, respectively. Size-fractionation (SF)–HPLC was performed using a 4.6 × 250 mm Asahipak NH2P-50 4E column (Showa Denko, Tokyo) The chromatographic solvents were acetonitrile (SF solvent A) and water (SF solvent B). After sample injection, the percentage of SF solvent B was held at 15% for 5 min as an initial condition. PA-glycans were eluted by increasing the percentage of the SF solvent

B linearly from 15% to 50% over 30 min at a flow rate of 0.7 ml/min. Reversed-phase (RP)-HPLC was performed using a Cosmosil 5C18-AR II column (Nacalai Tesque, Kyoto), and PA-glycans were eluted by increasing the acetonitrile concentration in 0.02% trifluoroacetic acid (TFA) linearly from 0% to 20% over 40 min at a flow rate of 1.2 ml/min.

3.2.8 Glycopeptide mass analysis

Following a ConA chromatographic step and TALON in order to trap the MBP-GT protein glycosylated with Man residues, SDS-PAGE was performed and the MBP-GT protein band was excised from the Coomassie-stained gel. Tryptic digestion and peptide mass analysis were performed on a nanoLC-MS with micrOTOF-QII as described in Chapter 2 (140). The data processing was carried out using DataAnalysis 4.0 and Biotools 3.2 software (Bruker Daltonics). MS signals corresponding to the peptide moieties in the predicted glycopeptides (LEGGDQNATHHHHHH, calculated *m/z* 1726.75) and predicted glycopeptides (calculated *m/z* 2619.08) were sought based on the combination of the expected mass of the tryptic digests and accurate mass measurement of precursor ion spectra. The glycosylation yield was estimated by dividing the intensity of the MS signal corresponding to the predicted glycopeptide by the sum of those of the MS signals corresponding to the peptide moiety and the glycopeptide.

Table 3-1. Oligonucleotide primers used in Chapter 3.
Restriction sites are underlined.

Name	Sequence (5')
pACYCT7up-araC-Fw	CGACTCCTGCATTAGGAAATTGCATAATGTGCCTGTCAAATGG
pACYC::PBAD-Rv	CCAAAAAAACGGGTATGGAGAAAC
pACYC-SacI-F	GAGCTCGGCGCGCCTGCAGG
pACYC-(-RBS&NcoI)-R	ATTAAAGTTAAACAAAATTATTTCTACAGG
KpnI-T7pro Fw	CCTTGGTACCCGATCCCGCGAAATTAATAC
PstI-T7ter Rv	ACGACTGCAGCAAAAAACCCCTCAAGACCC
pBADter(PstI) Fw	TTTGCTGCAGTCGTTTTATCTGTTGTTTG
pBADter(KpnI) Rv	ATCGGGTACCAAGGCCCAGTCTTTCGACTG
SacI-RBS-Saci_1262	GGTTTGAGCTCAAGAAGGAGATATATATATGATCGACAACCCATTACTG
NotI-Stag-Saci_1262 Rv	GGTTTGCGGCCGCTTACGAGTCCATGTGCTGGCGTTCAAATTTCGCAGCAGCGGTTTCTTTCTCTCTAAGAAATTCTGCTAATTCAC
NotI-RBS-GumI	GGTTTGCGGCCGCAAGAAGGAGATATATATATGAGCGCGTCTGCTTCGCTGCC
PvuI-T7-GumI Rv	GGTTTCGATCGCTAACCCATTTGCTGTCCACCAGTCATGCTAGCCATCAGCGCGGCGTCTCCATCCTTGC
SacI-RBS-T7-ALG13-Fw	AATGAGCTCAAGAAGGAGATATATATATGGCTAGCATGACTGGTGGACAGCAAATGGGTATGGGTATTATTGAAGAAAAGGCTC
PstI-ALG13-Rv	TCTTCTGCAGCTAGCTGTATATAGTTTCAACTAGCAATCG
PstI-RBS-S-ALG14	GGTTTCTGCAGAAGAAGGAGATATATATATGAAAGAAACCGCTGCTGCGAAATTTGAACGCCAGCACATGGACTCGAAAACGGCCTACTTGGCGTCATTGGTGC
NotI-ALG14-Rv	GGTTTGCGGCCGCTTAAACAAGGATGCCGAACCACTTGGATCTTGG
NotI-RBS-ALG1-Fw	GGTTTGCGGCCGCAAGAAGGAGATATATATATGTTTTTGGAAATTCCTCGGTGG
PvuI-T7-ALG1-Rv	GTTTCGATCGTCAACCCATTTGCTGTCCACCAGTCATGCTAGCCATATGAATTAGCTTCAAATCTCTCAT
NotI-RBS-ALG1dTM-Fw	GTTTGCGGCCGCAAGAAGGAGATATATATATGTCGACCAAAAAAAGGATCATCATATTTG
ALG2-CDS-Fw	ATGATTGAAAAGGATAAAAGAACGATTGC
ALG2-no stop-Rv	TATTTCTTCATAAGGGTAGGAGAGAGCG
NotI-ALG1opt-Fw	TTTAAGCGGCCGCAAGAAGGAGATATATATATGTTCCTGGAAATTCCGCG
PvuI-T7-ALG1opt-Rv	TTAAACGATCGTTAACCCATCTGCTGACCACCGGTCATGCTTGCCATATGAATCAGTTTCAGATCACGCATG

NotI-RBS-ALG1dTMopt-Fw	TTTAAGCGGCCGCAAGAAGGAGATATATATATGAGCACCAAAAAACGCATCATTATCTTC
PvuI-KpnI-T7-PglBopt-Fw	TAACGATCGTTTGGGGTACCAAGAAGGAGATATATATATGGCAAGCATGACCGGTGGTCA
XhoI-PglBopt-Rv	ACCAGACTCGAGTTAGATTTTCAGTTTAAACACTTTGGCATCACG
T7 Rv	GCTAGTTATTGCTCAGCGG
cpsB-Rv	AAGAACCGACATCACTCCAGCCCGCATC
wcaC-Fw	ATCGCACCATTACTCGTACACCGGGTCC
tet+350-Rv	GACTTTACGAAACACGGAAACCG
tet-350-Fw	AAGGCAAAATGCCGCAAAAAAG
PvuI-KpnI-T7-PglB-Fw	GTTTCGATCGGGTACCAAGAAGGAGATATATATATGGCTAGCATGACTGGTGGACAGCAAATGGGTTTGAAAAAAGAGTATTTAAAAAACCC
XhoI-PglB-Rv	AGACTCGAGTTAAATTTTAAGTTTAAAAACTTTAGC

3.3 Results and discussion

3.3.1 Plasmid construction, mutant host construction and protein expression

In order to synthesize $Man_3GlcNAc_2$ in *E. coli*, some GTases are required. Figure 3-1 provides an overview of the construction and resulting synthetic pathway.

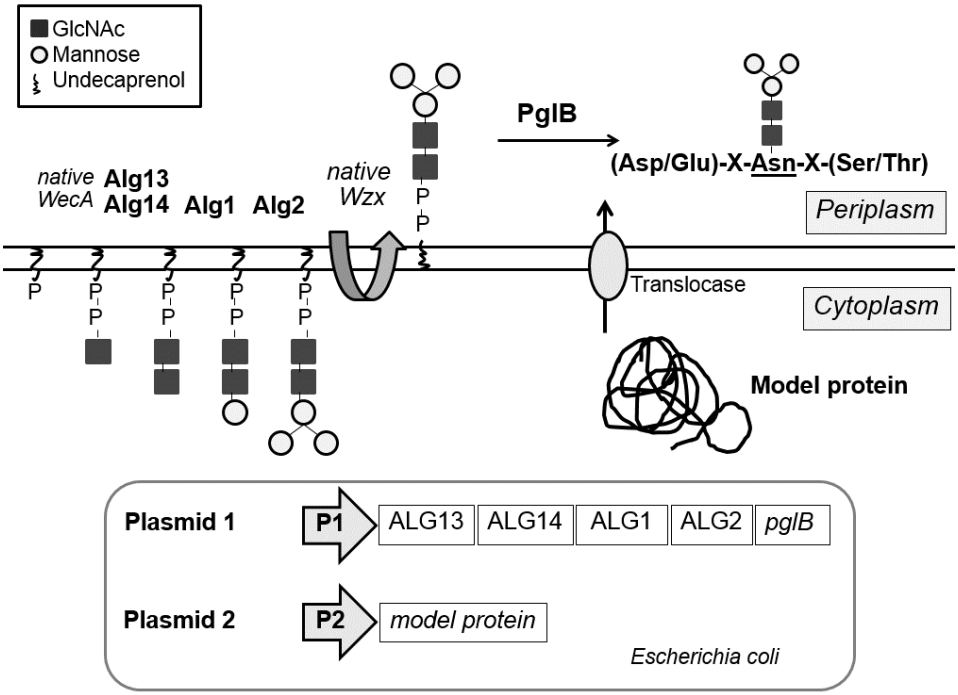

Figure 3-1. Overview of the protein expression system and the general concept of plasmid constructs conferring $Man_3GlcNAc_2$ biosynthesis and glycosylation. Plasmid a was used for OST and GTases expression and Plasmid b for model protein expression. P1 and P2 indicated that the two expression plasmids might be driven by different *E. coli* functional promoters.

3.3.2 Construction of plasmid pM1

According to general outline presented in Figure 3-1, a series of plasmids was constructed. Plasmid 1 contained the *ALG13, ALG14, ALG1 pglK* and *pglB* expression cassettes. Plasmid 2 was similar to Plasmid 1, but the 35 amino acid-encoding codon at the N-terminal of ALG1 was omitted. The β-mannosyltransferase is required to extend $GlcNAc_2$-PP-Und; however, the potential sources for this enzyme are limited. Apart from the characterized *S. cerevisiae* Alg1, GumI is a GDP-Man:glycolipid 4-β-D-mannosyltransferase from *X.*

campestris pv. *campestris*. This GumI mannosyltransferase was identified to be involved in the xanthan polysaccharide biosynthesis pathway, using GDP-Man as a donor substrate and glucuronic acid-β-1,2-Man-α-1,3-Glc-β-1,4-Glc-PP-polyisoprenol as an acceptor (145). The *gumI* expression cassette was constructed into Plasmid 3 and 6 in order to explore the use of this alternative mannosyltransferase, even though its substrate preference towards dolichol pathway intermediates was not determined.

In 2008, Gao et al. used a bioinformatics technique to identify the Alg13/14 complex, which possesses structural homology to *E. coli* MurG (MurG is an UDP-GlcNAc undecaprenyl-PP-MurNAc pentapeptide: *N*-acetylglucosaminyltransferase that has role in bacterial peptidoglycan biosynthesis) (28). To find a candidate for a β-GlcNAc transferase using predicted secondary structural homology, the MurG amino acid sequence was submitted to the Protein Homology/analogY Recognition Engine (PHYRE) database server (146, 147). The protein coded from the *S. acidocaldarius* Saci_1262 ORF was listed as a structural homologue to MurG. Saci_1262 possesses 15% structural homology to MurG, while that of Alg13/14 complex is 11%. To test the applicability of Saci_1262, its expression cassette was constructed and was cloned into Plasmids 4, 5 and 6, to substitute ALG13 and ALG14, while other components were maintained.

The expression of Alg1 in *E. coli* was problematic. The native, full-length ALG1 and an N-terminal 35 amino acid-deleted mutant (ALG1dTM) were cloned into pET-23b at *Bam*HI and *Xho*I for expression under P_{T7}. Only the pET-23b-ALG1dTM construct expressed the protein, which migrated approximately at 40 kDa on the Western blot (Figure 3-5). Alg1dTM encoded from pET-23b-ALG1dTM had an N-terminal T7 tag and C-terminal 6xHis tag. ALG1dTM could be detected in both anti-His (data not shown) and anti-T7 blot, indicating intact protein expression without degradation at either terminal, although the migration was faster than the calculated mass (predicted at 48.9 kDa). To explore the feasibility of thioredoxin

(Trx) as a potential expression fusion partner (148) with the goal of improving Alg1 expression, Plasmid 7 harboring TrxALG1 and *pglB* was constructed. The constructs are listed in Figure 3-2.

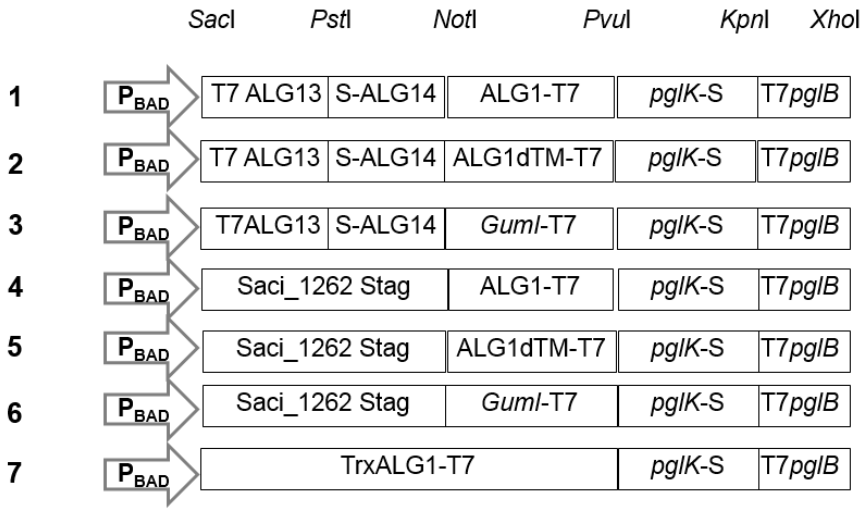

Figure 3-2. Plasmid constructs containing GTases, flippase, and OST. T7 indicates a T7 tag, and S or Stag indicate an S Tag. Construction was based on pAra-o, in which the *Nco*I site and His-tag sequence were removed. P_{BAD} indicates an arabinose-inducible promoter.

Plasmid 1 through Plasmid 7 were co-introduced into *E. coli* BL21 Δ*waaL*Δ*gmd* along with blank pET-22b. Protein expression was induced with 0.2% L-ara at 30˚C. Cells were harvested and total cell extract was subjected to Western blotting with corresponding antibodies. Blots of protein expression are shown in Figure 3-3.

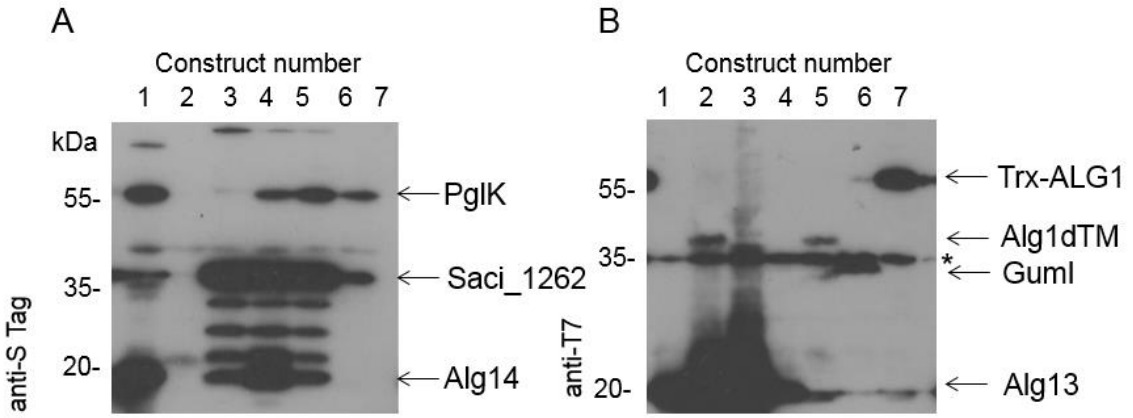

Figure 3-3. Protein expression of components in Plasmids 1 through 7 (the numbers are listed above). A: Blot of anti-S Tag where expression of PglK, Saci_1262 and Alg14 was observed. B: Anti-T7 blot where TrxAlg1, Alg1dTM, GumI and Alg1 was observed expressed. The asterisk (*) indicates a T7 non-specific interaction.

Unfortunately, the full-length Alg1 in Plasmid 1 and Plasmid 4 was not expressed (Figure 3-3 B). PglK was found to be expressed in Plasmids 1, 4, 5 and 6 but not in Plasmids 2 and 7. However, in Plasmid 3 a faint band of PglK was detected. Unfortunately, PglB was not expressed in any constructs. Expression of PglK (Plasmids 1-6) and TrxAlg1 (Plasmids 7) might have inhibited the PglB expression.

Based on the expression data and in order to narrow down to the potential yeast GTases for use in the preliminary study, Plasmids 2, 3 and 7 were chosen for further modification. To examine the effect of the gene located upstream of *pglB* on the expression of *pglB*, *pglK* was designated to be omitted from Plasmids 2, 3 and 7. Cassette [*Pvu*I-*Kpn*I-RBS-*pglB*-*Xho*I] was constructed and ligated to Plasmids 2, 3 and 7 at *Pvu*I and *Xho*I, Plasmid 2 Δ*pglK*, 3 Δ*pglK* and 7 Δ*pglK* were then generated. The obtained plasmids (constructs depicted in Figure 3-4 A) were re-introduced into *E. coli* BL21 Δ*waaL*Δ*gmd*. Protein expression was induced and Western blotting was performed (Figure 3-4 B).

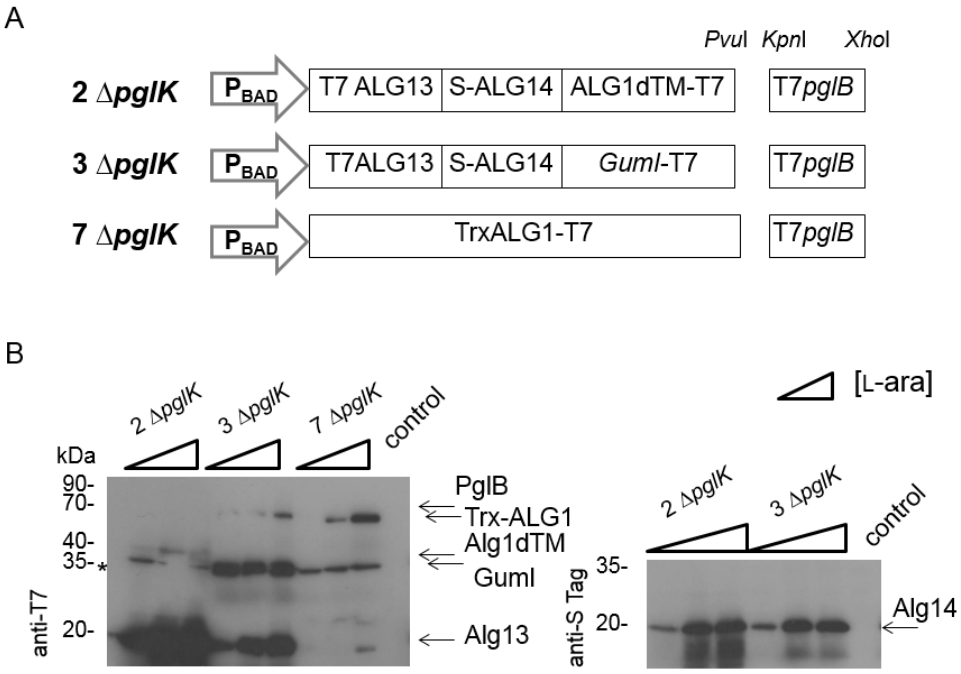

Figure 3-4. A: Outline of the construction of Plasmids 2 Δ*pglK*, 3 Δ*pglK* and 7 Δ*pglK*. B: Western blot of the constructed components. The triangles indicate decreasing increments of L-ara concentration, each representing a tenfold reduction, with the rightmost being 0.2%. The asterisk (*) indicates a T7 non-specific interaction.

The Western blot analysis revealed the expression of Alg13 and Alg14 in Plasmids 2 Δ *pglK* and 3 Δ *pglK*; GumI in Plasmid 3 Δ *pglK*; Alg1dTM in Plasmid 2 Δ *pglK* and TrxAlg1 in Plasmid 7 Δ *pglK*. PglB in Plasmid 3 Δ *pglK*, positioned after *gumI*, was expressed. However, in Plasmids 2 Δ *pglK* and 7 Δ *pglK*, where *pglB* was followed *ALG1* derivatives, the PglB expression was abolished. It might be presumed that if *pglB* is positioned after *ALG1*, away from *pglK*, the *pglB* expression will be inhibited.

To overcome the protein expression problem, codon-optimization of genes was employed. The optimized DNA of *ALG1* and *pglB*, for *E. coli* expression (denoted as ALG1opt and *pglB*opt), were ligated at *Not*I and *Xho*I of Plasmid 1, just after the ALG13 and ALG14 cassettes. ALG1dTMopt, the N-terminal 35 amino acid-deleted mutant was also generated and

cloned back into *Not*I and *Pvu*I. Using a method similar to that for the preliminary constructs, the resulting plasmids were named Plasmid1$_{OPT}$ and Plasmid2$_{OPT}$ (Figure 3-5 A).

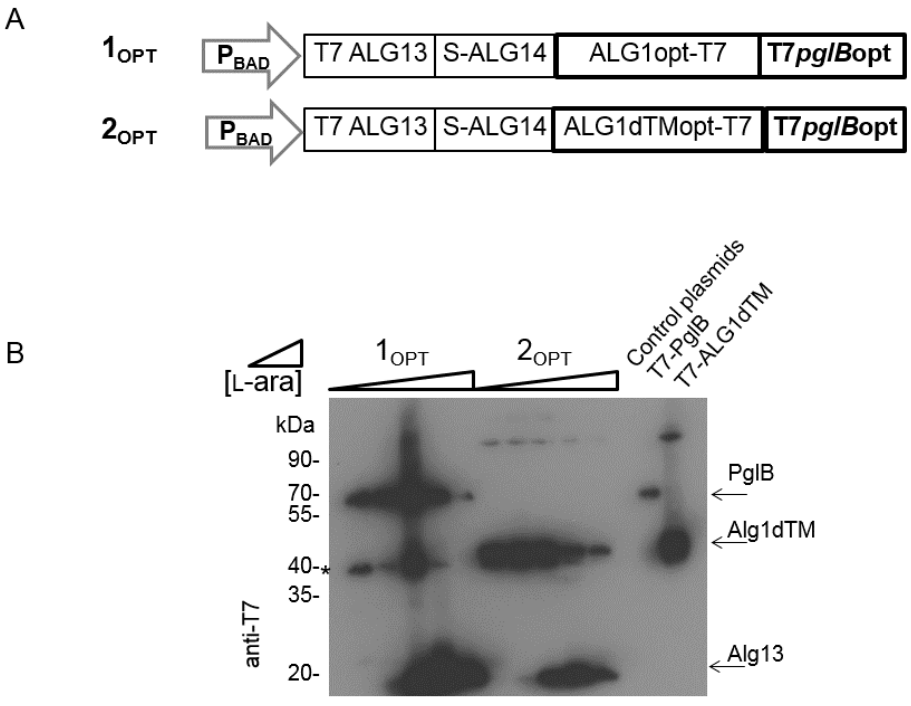

Figure 3-5. Plasmid construction and protein expression. A: Construction of Plasmid1$_{OPT}$ and Plasmid2$_{OPT}$. B: Protein expression of Plasmid1$_{OPT}$ and Plasmid2$_{OPT}$ monitored on anti-T7 blot. The triangles indicate increments of L-ara concentration. Control plasmids were blank plasmids transformant. T7-PglB and T7-ALG1dTM were generated by ligating *pglB* and ALG1dTM into pET-23b. The asterisk (*) indicates a T7 non-specific interaction, and bold-lined boxes indicated the use of optimized DNA.

The protein expression of Plasmid1$_{OPT}$ and Plasmid2$_{OPT}$ in *E. coli* BL21 Δ*waaL*Δ*gmd* was analyzed by anti-T7 Western blotting (Figure 3-5 B). The results showed that PglBopt expression was recovered in Plasmid1$_{OPT}$, but Alg1opt was not expressed. On the other hand, Plasmid2$_{OPT}$ showed expression of Alg1dTMopt, but PglBopt was not expressed. According to these results, if ALG1 and *pglB* were positioned next to each other, only one or the other would be expressed, even if optimized genes were used.

To solve the Alg1 expression problem, an additional expression cassette of Alg1dTMopt under P_{T7} was constructed and ligated onto the *Xho*I site downstream of *pglB* of Plasmid 1_{OPT}. TrxAlg1opt was also constructed in the same manner. The insertion direction was verified by DNA sequencing using the primer T7 Rv. Representative clones of each direction were picked. Constructs are shown in Figure 3-6 A. The end-to-end ligations were emphasized by (N) affix.

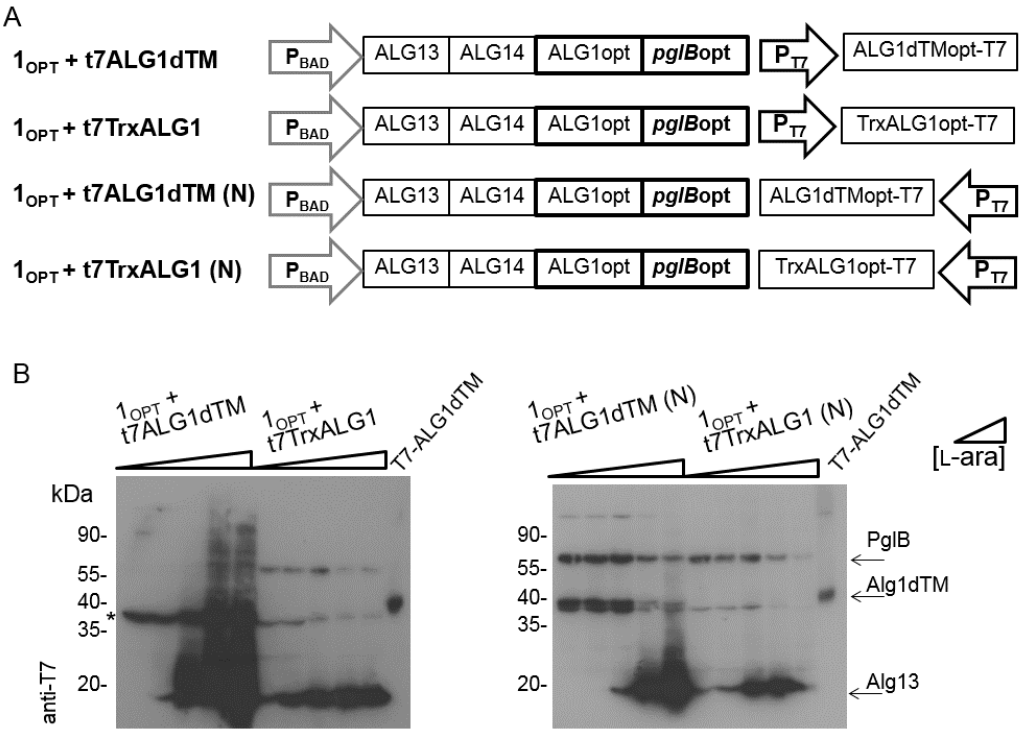

Figure 3-6. Construction of Plasmid 1_{OPT} derivatives. A: Plasmid construction of Plasmid 1_{OPT} with expression cassettes of ALG1dTM and TrxALG1. B: Protein expression of the shown constructs, probed by anti-T7 Western blot. The asterisk (*) indicates a T7 non-specific interaction, and bold-lined boxes indicate the use of optimized DNA.

The results of the protein expression measurements showed that the inserted t7ALG1dTM cassette recovered the Alg1dTM expression in Plasmid 1_{OPT} + t7ALG1dTM (N). However, a similar ligation of the constructed expression cassette's direction to the downstream ORFs of P_{BAD} caused inhibition of its upstream-located *pglB*, as observed in Plasmid 1_{OPT} + t7ALG1dTM. Furthermore, none of ligation strategies recovered the expression

of t7TrxALG1. TrxALG1 might only be expressed following P_{BAD} (Figure 3-2 and 3-3). Finally, the Alg1dTM expression was solved. The Plasmid 1_{OPT} + t7ALG1dTM (N) was designated as pM1, as the construct seemed to confer synthesis of the $ManGlcNAc_2$ structure.

3.3.3 Plasmid construction of pBADTrxALG2 MBP-GT

In the previously constructed protein expression system, Alg2 could not be expressed without a thioredoxin fusion partner. Moreover, the TrxAlg2 construct was not expressed unless it was in juxtaposition to P_{BAD}. To construct the dual expression plasmid, *ALG2* was amplified and ligated into pBAD/TOPO ThioFusion (Figure 3-7 A). Subsequently, to express the model glycan acceptor protein, the expression cassette of periplasm-targeted MBP with synthetic glycosylation tag, MBP-GT, under P_{T7} was cloned by amplification of pET-22b MBP-GT (refer to Chapter 2). The dual expression vector, pBADTrxALG2 MBP-GT, was then obtained and used. The sequence of genes used in this work can be found in the Appendix. The protein expressions of pBADTrxALG2 and pBADTrxALG2 MBP-GT are shown in Figure 3-7 B and C, respectively.

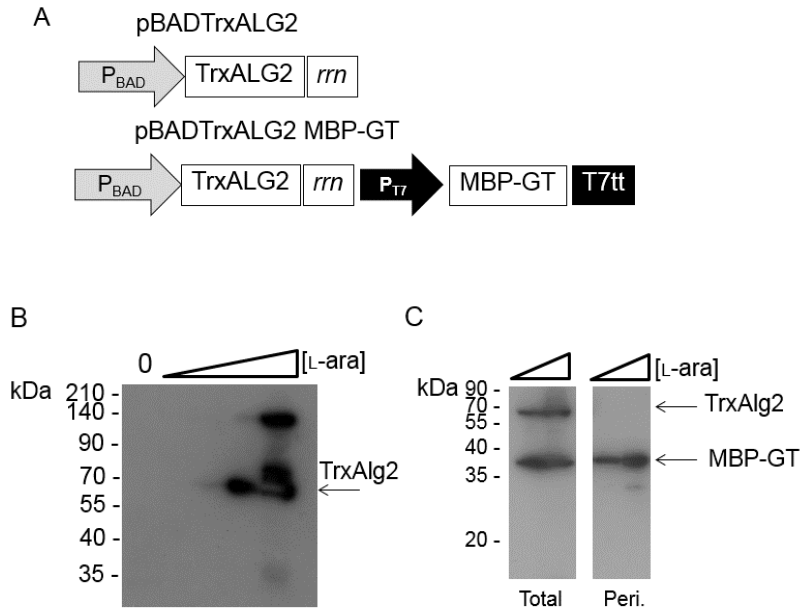

Figure 3-7. A: Plasmid construction of pBADTrxALG2 and pBADTrxALG2 MBP-GT. *rrn* and T7tt were Rrn and T7 transcriptional terminators, respectively. B: Protein expression of pBADTrxALG2, monitored by anti-His. TrxAlg2 was found to be expressed. In the last lane, the band migrated at 140 kDa might be a dimer of TrxAlg2, and a smear band can also be observed. C: Protein expression of the dual vector, pBADTrxALG2 MBP-GT. Both proteins were expressed and MBP-GT was secreted to the periplasmic space (Peri.), while TrxAlg2 was retained in the cytoplasm.

To express genes conferring synthesis and transferring of $Man_3GlcNAc_2$ and the acceptor protein, the plasmids pM1 and pBADTrxALG2 MBP-GT were co-introduced into *E. coli* BL21 Δ*waaL*Δ*gmd*. The construction and expression system is summarized in Figure 3-8 A. The pM1 is the expression vector of Alg13, Alg14 and PglB under an arabinose inducible promoter and ALG1dTM under P_{T7}. The pBADTrxALG2 MBP-GT is a plasmid for expressions of Trx-ALG2 under P_{BAD}, and the glycan acceptor protein MBP-GT under P_{T7}.

Protein expressions in *E. coli* BL21 Δ*waaL*Δ*gmd* harboring pM1 and pBADTrxALG2 MBP-GT were monitored by Western blotting with the corresponding antibodies. At an L-ara concentration of 0.2%, Alg13, Alg14, PglB and Trx-Alg2 were expressed (Figure 3-8 B). In addition to genomic *waaL* deletion, the *gmd* gene was disrupted in order to block GDP-fucose

synthesis (141), thereby retaining more GDP-Man to be used as a substrate by the Alg1 and Alg2 mannosyltransferases, respectively.

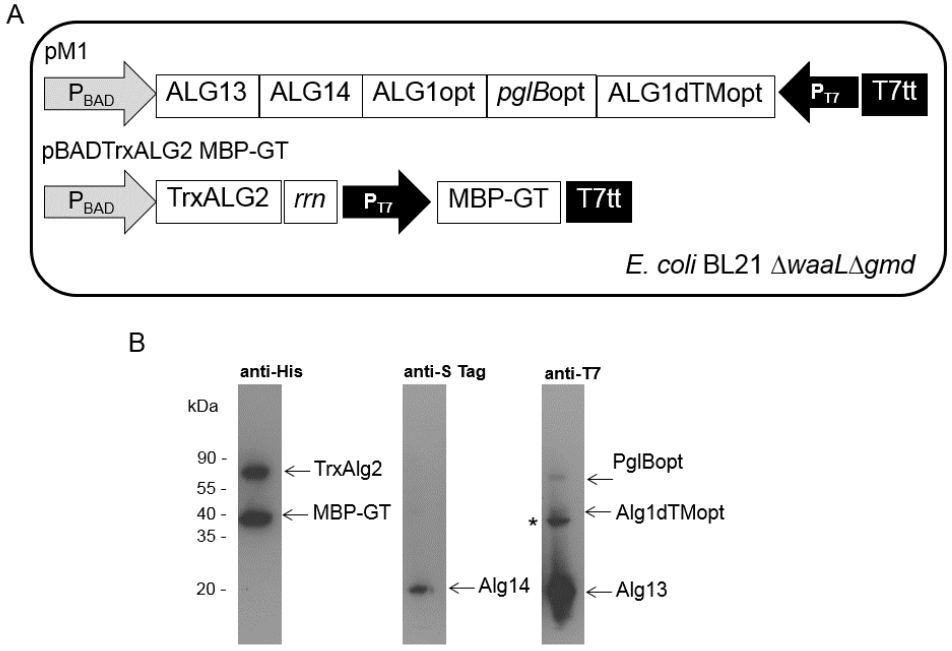

Figure 3-8. Consolidation of the protein expression system. A: Outline of the *E. coli* system for production of $Man_3GlcNAc_2$-Asn. *rrn* and T7tt are Rrn and T7 transcriptional terminators, respectively. B: Protein expression data as visualized by staining with 3 antibodies. All constructed components were expressed. An asterisk (*) indicates a T7 non-specific interaction.

3.3.4 LLO analysis

Prior to the analysis of protein glycosylation, to examine whether the expected glycan was assembled on *E. coli* Und, the structure of the glycan moiety on LLO was analyzed. *E. coli* BL21 Δ*waaL*Δ*gmd* cells expressing pM1 and pBADTrxALG2 MBP-GT were subjected to total lipid extraction. Hydrophobic LLO is expected to be contained in chloroform:methanol (CM) extracts, while less hydrophobic LLO is expected in chloroform:methanol:water (CMW) extracts, according to the protocol described by Gao and Lehrman (143). LLO extracts were captured on a DEAE-cellulose column. The free saccharides passed through the column, while

the LLOs were concentrated and eluted. The glycan moiety was released from LLO by acid-hydrolysis, prior to pyridylamination. Most of the lipid extracted by CM contained target LLOs.

The PA-glycan derived from LLOs was analyzed on SF-HPLC. The analysis showed 2 distinct peaks at retention times of 19 and 24 min (Figure 3-9 A). The peaks were collected and analyzed on an MS system, resulting in signals of *m/z* 665.6 and 989.9, which fragmented and corresponded to chitobiose extended with one or three hexose residues (Figure 3-9 B). In consideration of the constructed recombinant genes, these could be related to the structures of $ManGlcNAc_2$ and $Man_3GlcNAc_2$, respectively. Based on the analyses, the PA-glycans derived from LLO and eluted at 19 and 24 min were designated as PA-LLO M1 and PA-LLO M3, respectively. A broad peak ranging from 21 to 22 min of SF-HPLC did not correspond to the diagnostic mass value of glycans (data not shown).

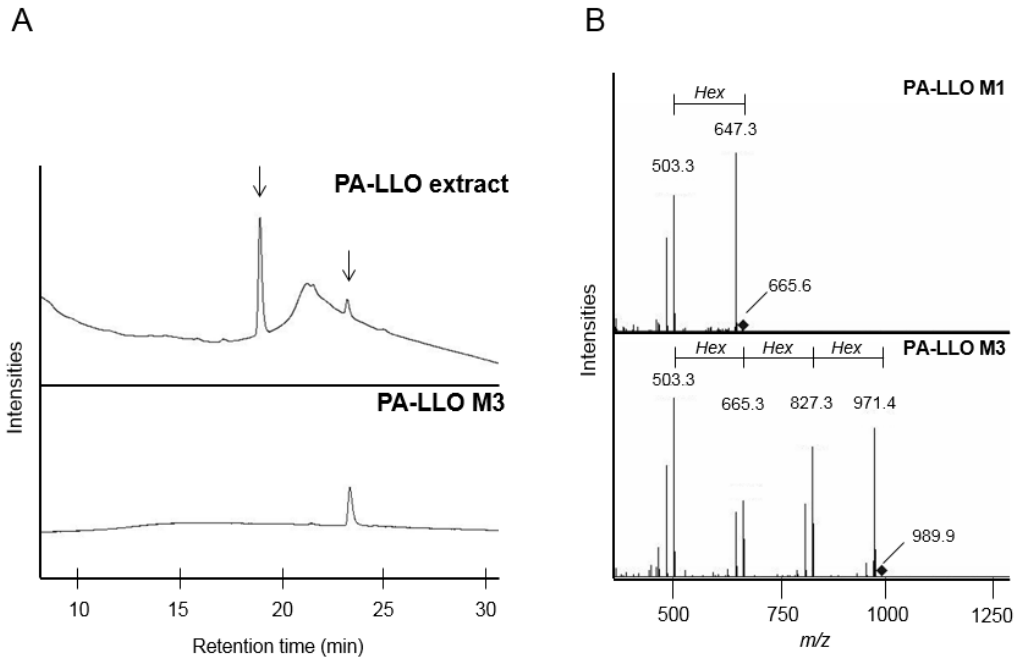

Figure 3-9. LLOs analysis. A: LLOs from recombinant *E. coli* Δ*waaL*Δ*gmd* co-expressing pM1 and pBADTrxALG2 MBP-GT were prepared. Lipid extract was subjected to acid-hydrolysis and the released glycans were labeled with 2-aminopyridine (PA). The sample was thus named PA-LLO extract. Using an SF-HPLC system equipped with a fluorescence detector, PA-labeled sugar was detected. Two distinct peaks were observed at retention times of 19 and 24 min. In the lower chart, the less abundant peak which eluted at the same retention time as the PA-sugar chain 016 (TaKaRa Bio), 24 min on SF-HPLC, was concentrated and shown. The later mass spectrometry analysis revealed a specific glycan corresponding to the structure of $Hex_3HexNAc_2$-PA, with a peak at a retention time of 24 min, which was therefore designated PA-LLO M3.

B: Mass spectra of purified PA-glycans from LLO. The upper and lower chromatograms show the results for samples previously collected from SF-HPLC at retention times of 19 and 24 min, respectively. The upper chromatogram showed a major signal of *m/z* 665.6, which fragmented one Hex (162 Da) ion and yielded an *m/z* of 503.3. This sample corresponded to $HexHexNAc_2$-PA and was thus named PA-LLO M1. The peak retention time at 24 min has a primary ion of *m/z* 989.9 and fragmented 3 Hex residues. This analysis corresponded to $Hex_3HexNAc_2$-PA and thus the sample was designated as PA-LLO M3. Diagonal squares indicate the precursor ions in each mass spectrum.

3.3.5 Protein purification and *N*-glycan analysis

To observe protein glycosylation activity, periplasmic material from cells expressing pM1 and pBADTrxALG2 MBP-GT was prepared. The periplasmic material was loaded onto equilibrated TALON resin. After the contaminants were washed, the concentrated proteins were eluted. MBP-GT was purified to near homogeneity and migrated closely to the 40 kDa protein standard on SDS-PAGE gel (Figure 3-10).

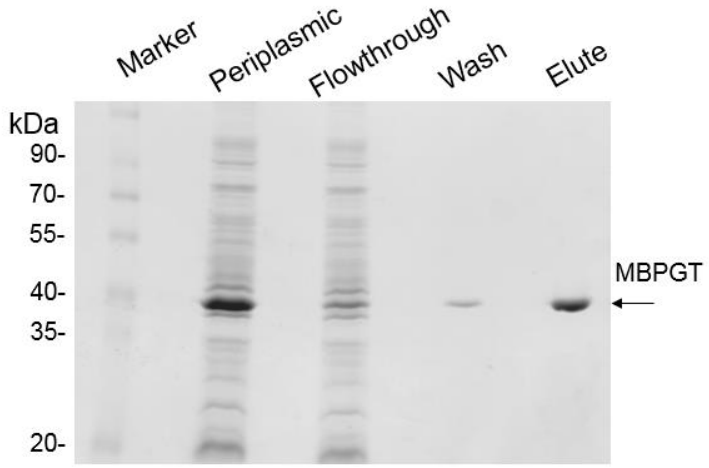

Figure 3-10. Coomassie-stained SDS-PAGE gel of the TALON-purified MBP-GT. The marker is 3-Color Prestained XL-Ladder (APRO Science). Periplasmic is the periplasmic sample before loading onto the TALON chromatography. Flowthrough means the fraction of proteins not bound to the column. Wash indicates the washing fractions. Elute is the fraction containing the eluted and purified MBP-GT.

PA-glycan was prepared from purified MBP-GT and named MBP-PA glycans. After being resolved on SF-HPLC, the peaks were collected and the one eluted at a retention time of 24 min (Figure 3-11 A upper) was identified as PA-MBP M3; its mass spectrum is shown in Figure 3-12. The other peaks of PA-glycan eluted on SF-HPLC were examined by mass spectrometry, and neither glycan intermediates nor other glycan structures were observed. When PA-MBP M3 was co-injected with the PA glycan standard (M3B, PA-016, TaKaRa), the peak was co-eluted at the same retention time on SF-HPLC (Figure 3-11 A lower). To further elucidate the identity of the glycan structure and composition, PA-MBP-M3 was

resolved on reversed-phase HPLC. The results showed that PA-MBP M3 and the PA sugar chain standard M3B were eluted at the same retention time (Figure 3-11 B). Consolidating the results for SF-HPLC and RP-HPLC, it was suggested that the structure of the PA-MBP M3 synthesized and glycosylated in this constructed system was Manα1-6(Manα1-3)-Manβ1-4GlcNAcβ1-4GlcNAc-PA.

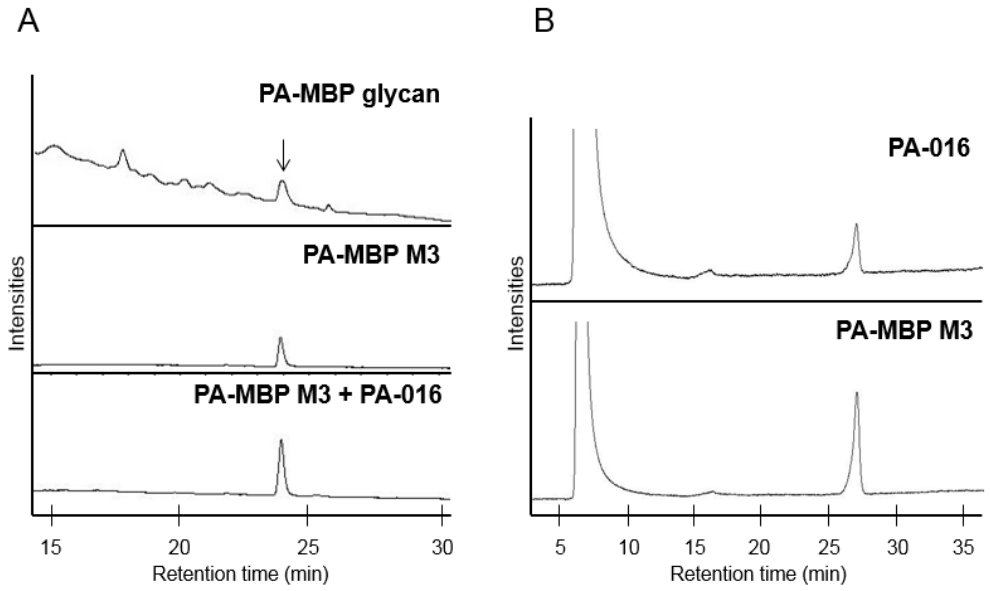

Figure 3-11. Glycan analysis of MBP-GT. A: SF-HPLC analysis. Periplasmic-localized MBP-GT was purified and subjected to hydrazinolysis and PA labeling. PA-MBP glycans was resolved. The candidate peak of PA-MBP M3 (marked with arrow) was purified and shown in the middle chart. In the lower chart, PA-MBP M3 was coinjected with the PA-sugar chain 016 (PA-016); a single peak in SF-HPLC was observed. The PA-MBP M3 rationally possesses a glycan structure similar to the standard PA-sugar chain 016.

B: RP-HPLC analysis. PA-MBP M3 and the PA-sugar chain 016 (PA-016) were subjected to RP-HPLC analysis. One and the other, standard PA-016 and sample PA-MBP M3 eluted at same retention time. This result showed that PA-MBP M3 exists in identical isomerism of standard PA-016, as Manα1-6(Manα1-3)-Manβ1-4GlcNAcβ1-4GlcNAc-PA.

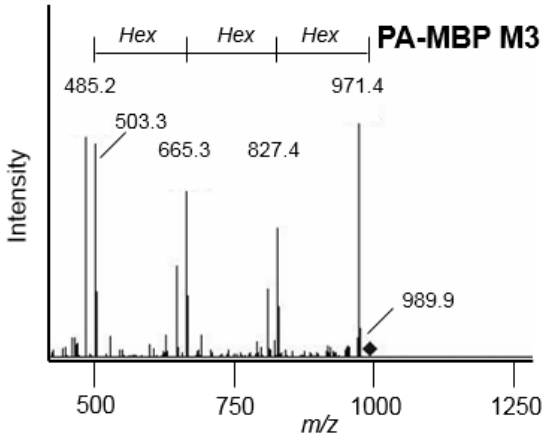

Figure 3-12. Mass spectrometry of purified PA-MBP M3. PA-MBP M3 was subjected to mass spectrometry. The subject PA sample showed a precursor *m/z* of 989.9 (diagonal square), fragmented to 971.4, 827.4, 665.3 and 503.3, which corresponded to $Hex_3HexNAc_2$-PA.

3.3.6 MS analysis of glycopeptide

The glycan assembly and its identity were successfully analyzed. To confirm the protein glycosylation on MBP-GT, nanoLC-MS analysis was performed. Glycan analysis of MBP-PA glycans suggested protein glycosylation containing α-Man residues, and therefore ConA chromatography was employed. Following the ConA purification, TALON was used to purify the Man-glycosylated MBP-GT protein. The purified Man-glycosylated MBP-GT protein was digested with trypsin, and the resulting tryptic peptides was subjected to a nanoLC-MS Q-TOF analytical system. The peptide containing the engineered glycosylation motif (LEGGDQNATHHHHHH, calculated *m/z* 1726.75; *N*-glycosylation site underlined) and the glycopeptide were sought and found to elute at 8.9 min on nanoLC. The signals of the putative glycopeptide and the peptide were observed at m/z 2619.97 and 1726.67, respectively (Figure 3-13). The mass difference between the peptide and glycopeptide (893.3 Da) suggested glycosylation of the $Man_3GlcNAc_2$ structure. This mass analysis data supported the results of the hypothetical glycan structure synthesized from the constructed pathway, LLO analysis and *N*-glycan analysis. Three technical replicates were performed (where glycosylation yields of

1.5%, 3.5% and 4.6% were observed), and the average yield was 3.2±1.5% glycosylation on the observed peptide (not the total yield); other possible tryptic glycopeptides were sought but were not found. The different yields may have been caused by different batches of cultivation.

In brief, glycosylation of Manα1-6(Manα1-3)-Manβ1-4GlcNAc-β1-4GlcNAc ($Man_3GlcNAc_2$) to Asn on the introduced glycosylation motif, DQNAT, of MBP-GT was observed.

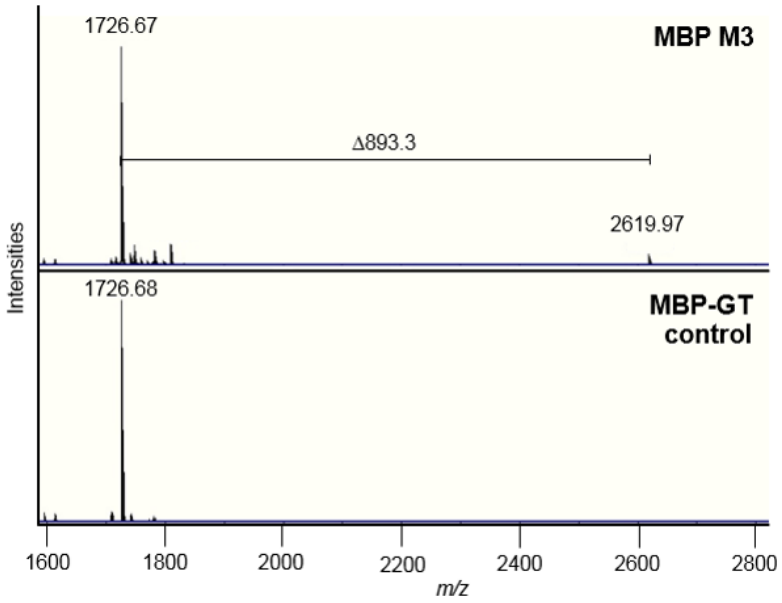

Figure 3-13. Mass spectrometry of the glycopeptide. The trypsin-digested glycosylated MBP-GT (ConA and TALON-purified) was subjected to nanoLC-MS analysis. The mass spectrum of the nanoLC eluate at 8.9 min is shown. The *m/z* values of 2619.97 and 1726.67 corresponded to the *m/z* values of a hypothetical $Man_3GlcNAc_2$-linked peptide and a tryptic peptide of LEGGDQNATHHHHHH, respectively. In the lower chart, TALON-purified MBP-GT produced from the pET-22b backbone in *E. coli* BL21 (DE3) was subjected to tryptic digestion and nanoLC-MS analysis, and shown as a control.

In this chapter, plasmids conferring a glycan assembly of the yeast initial protein glycosylation pathway were constructed. The *E. coli* BL21 Δ*waaL*Δ*gmd* mutant was obtained and used. Protein expressions were monitored, and all of the recombinant proteins were found to be expressed. To confirm the glycan biosynthesis, LLO was prepared from cells expressing pM1 and pBADTrxALG2 MBP-GT. LLO $ManGlcNAc_2$ and LLO $Man_3GlcNAc_2$ were both observed, the former in abundance. In addition, structural analysis by fluorophore labeling of the glycan released from MBP-GT revealed an *N*-glycan structure of $Man_3GlcNAc_2$, but not a population of $ManGlcNAc_2$. Peptide MS analysis further confirmed the presence of the $Man_3GlcNAc_2$ structure on the designated glycosylation motif.

The production of glycoproteins in *E. coli* is an emerging topic in biotechnology. This approach has opened up since the successful transfer of protein glycosylation machinery from *C. jejuni* into *E. coli* (56). It is important to note that, when the PglB protein was used as the OST in previous studies, it showed relaxed substrate specificities toward various glycan structures, including immature native *C. jejuni N*-glycans, *E. coli O*-polysaccharides, and synthetic glycan structures (39, 40, 102, 103, 136). In addition, WecA is an enzyme that initiates synthesis of Und-PP-GlcNAc using Und-P and UDP-GlcNAc (111-113). These contributions benefit the production of *N*-glycoprotein in *E. coli* by establishing the GlcNAc-PP-Und, and later GlcNAc-Asn, which is fundamental to eukaryotic *N*-glycosylation.

To facilitate biosynthesis of the cell wall structure, *E. coli* and other bacteria possess a polysaccharide membrane translocator, generally identified as Wzx (138). To exploit the native *E. coli* Wzx flippase (65, 139), the *pglK* gene for flipping assembled LLO from *C. jejuni* (66) was not included in the construction. Nonetheless, protein glycosylation of MBP-GT was observed. Even though this flipping mechanism seemed to be interchangeable, the substrate specificities toward engineered glycan structures would not be easily determined. Still, there might be a chance that overexpression of either Wzx or PglK could have improved the flipping

of LLOs and thus the protein glycosylation yield, by saturating the donor substrate for glycosylation in the periplasmic space.

In this constructed expression system, LLO $ManGlcNAc_2$ was abundant, but only $Man_3GlcNAc_2$ was found to be glycosylated to MBP-GT based on analysis using the PA-glycan technique. As $Man_{1\text{-}3}GlcNAc_2$ is not natural glycan substrate for PglB, so the substrate preference, specificity and selectivity of *C. jejuni* PglB protein may be discussed that *C. jejuni* PglB protein might prefers branched eukaryotic *N*-glycan structures. In order to improve the glycosylation yield, mutants of PglB might be generated, or an alternative OST from one of various sources may be implemented. In contrast to a previous report (149), the $Man_4GlcNAc_2$ was not observed in the present study. This discrepancy might have been related to the different substrate specificities of the participating enzymes. In particular, Trx fusion to Alg2, which might have increased solubility and specificity of Alg2, resulted in the observed glycan assembly and following glycosylation. The previously reported *in vitro* activity of Trx-fused Alg2 showed that glycan elongation of $ManGlcNAc_2$-PP-Dol substrate was rigorously stopped at $Man_3GlcNAc_2$-PP-Dol (34), which supports the interpretation of the results presented herein. At this moment, however, no clear explanation can be made.

More glycosylation efficiency is desired. Simple production optimization does not seem to suffice. To resolve this problem, an improvement in the expression and solubility of GTases, a reduction of the metabolic-bottleneck and burden of protein expression, or the use of an alternative OST may be implemented. Furthermore, the stringent requirement of a PglB glycosylation recognition motif, D/E-X_1-N-X_2-S/T, is not attractive, and a PglB replacement, which should recognize N-X-S/T, might be identified in the future.

3.4 Summary

Plasmids encoding GTase genes from *S. cerevisiae*, *C. jejuni* PglB OST, and MBP-GT, which conferred synthesis and glycosylation of $Man_3GlcNAc_2$, were constructed and transformed into the obtained *E. coli* BL21 Δ*waaL*Δ*gmd*. LLO synthesis was confirmed and $Man_3GlcNAc_2$ was observed along with intermediate $ManGlcNAc_2$. Protein glycosylation of the model protein, MBP-GT, was analyzed by conventional *N*-glycan analysis and further confirmed by glycopeptide mass spectrometry. A homogeneous, pentasaccharide-glycosylated protein was obtained in this study.

Chapter 4

General Conclusions and Perspectives

The *E. coli* protein expression system has many appealing features, such as rapid expression, high protein yields, ease of culture and manipulation, and low cost with good cost-effectiveness (150). However, the inability to glycosylate proteins via this system has reduced its usefulness for the expression of glycosylated proteins. If glycosylation machinery could be made available in *E. coli*, the system would then become a useful host for recombinant glycoprotein expressions, which might be used for functional studies and in biomedical applications.

N-glycosylation is no longer restricted to eukaryotes, and indeed has been observed in all domains of life. Currently, *C. jejuni* is one of the bacteria with a well-characterized *N*-glycosylation pathway. *pgl* operon components enable the *C. jejuni* glycosylation function. The reconstitution of the glycosylation mechanism in *E. coli* (56), and characterization of the pathway (40), have opened up a pathway for engineering the glycan structure as a general concept discussed in terms of glycoengineering in Chapter 1.

Genomic DNA of *C. jejuni* JCM 2013 was obtained from JCM. Its protein glycosylation pathway had remained elusive. To study the operon conferring *N*-glycosylation of *C. jejuni* JCM 2013, the *pgl* operon was cloned and characterized in Chapter 2. The *pgl* operon was found to contain a total of 15,597 nucleotides, and putatively coded for 14 ORFs. Protein glycosylation of the co-expressed CmeA as a model protein was analyzed. The structures of $HexHexNAc_6$ and $HexNAc_6$, which corresponded to those of Glc-$GalNAc_5$-HexNAc and $GalNAc_5$-HexNAc, respectively, were elucidated. Later, the HexNAc moiety linked to Asn (at the glycan-reducing end) was demonstrated as the GlcNAc moiety putatively derived from the bacterial polysaccharide biosynthetic pathway via the activity of WecA (111, 112).

Furthermore, MBP and DsbA containing a glycosylation tag (GT) were constructed and confirmed to function as glycan acceptor proteins. In brief, GlcNAc-Asn glycosylation was demonstrated, and glycan acceptor proteins were obtained.

In Chapter 3, in order to synthesize a typical eukaryotic *N*-glycan $Man_3GlcNAc_2$, the corresponding pathway (including *ALG13, ALG14, ALG1, ALG2* and *pglB* genes) was constructed into *E. coli*. Some difficulties were encountered in the expression of Alg1, Alg2 and PglB, and later those protein expressions were solved by concluding gene arrangement in the expression plasmids. Next, the LLO assembly was observed and putative structures of $ManGlcNAc_2$ and $Man_3GlcNAc_2$ were found. Protein glycosylation was observed on MBP-GT, where the glycan identity was elucidated as Manα1-6(Manα1-3)-Manβ1-4GlcNAc-β1-4GlcNAc by means of co-elution in HPLC PA-glycan analysis. Finally, mass spectrometric analysis confirmed the glycosylation of the $Man_3GlcNAc_2$ structure onto the GT motif.

To conclude, Proof of Principle was conducted to show that the early-stage eukaryotic *N*-glycan structure could be synthesized and glycosylated onto a model protein in *E. coli*. For future application, the design and engineering of a glycan structure and protein secretion pathway will need to be determined. Following the $Man_3GlcNAc_2$ core structure, a complex glycan structure could be synthesized by integration with the constructed system of eukaryote- or human-derived *N*-acetylglucosaminyltransferase I (GlcNAcT), β-galactosyltransferase (GalT) and sialyltransferase (SiaT), since the former 2 human-derived enzymes have been functionally expressed in *E. coli* (151, 152). This concept is shown in Figure 4.

The culmination of this particular biotechnology research area will be the *E. coli*-synthesis of eukaryotic glycoproteins in high yield, and the engineering of *N*-glycan structures for applications of benefit to humans.

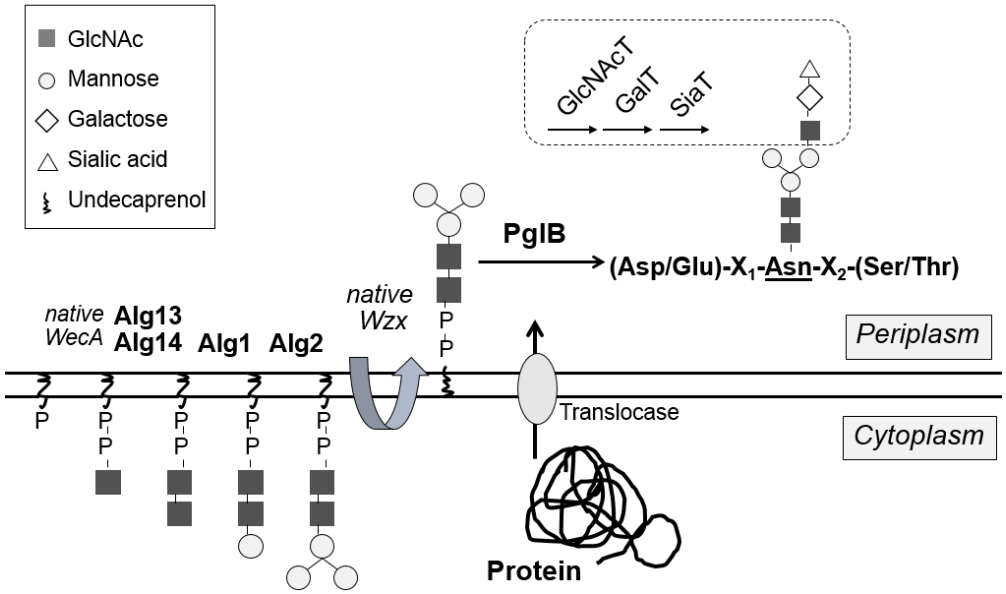

Figure 4. Summary of the constructed eukaryotic *N*-glycosylation pathway in *E. coli* presented in this thesis. The complex-type glycan structure shown is beyond the scope of this thesis and is a perspective work. The dashed rectangle denotes proteins GlcNAcT, GalT and SiaT, which were required for the transfer of GlcNAc, galactose and the sialic acid moiety, respectively.

References

1. **Kohno, M. and Yamashina, I.:** Identification of asparagine as the site of glycosylation in glycoprotein biosynthesis, The Journal of Biochemistry, **73**, 1089-1094 (1973).
2. **Schwarz, F. and Aebi, M.:** Mechanisms and principles of N-linked protein glycosylation, Curr. Opin. Struct. Biol., **21**, 576-582 (2011).
3. **Helenius, A. and Aebi, M.:** Intracellular functions of N-linked glycans, Science, **291**, 2364-2369 (2001).
4. **Dwek, R. A.:** Glycobiology: toward understanding the function of sugars, Chem. Rev., **96**, 683-720 (1996).
5. **Varki, A.:** Biological roles of oligosaccharides: all of the theories are correct, Glycobiology, **3**, 97-130 (1993).
6. **Varki, A., Freeze, H. H., and Gagneux, P.:** Evolution of glycan diversity, Cold Spring Harbor Laboratory Press (2009).
7. **Raska, M., Takahashi, K., Czernekova, L., Zachova, K., Hall, S., Moldoveanu, Z., Elliott, M. C., Wilson, L., Brown, R., Jancova, D., and other authors:** Glycosylation patterns of HIV-1 gp120 depend on the type of expressing cells and affect antibody recognition, J. Biol. Chem., **285**, 20860-20869 (2010).
8. **Mishin, V. P., Novikov, D., Hayden, F. G., and Gubareva, L. V.:** Effect of hemagglutinin glycosylation on influenza virus susceptibility to neuraminidase inhibitors, J. Virol., **79**, 12416-12424 (2005).
9. **Sun, X., Jayaraman, A., Maniprasad, P., Raman, R., Houser, K. V., Pappas, C., Zeng, H., Sasisekharan, R., Katz, J. M., and Tumpey, T. M.:** N-linked glycosylation of the hemagglutinin protein influences virulence and antigenicity of the 1918 pandemic and seasonal H1N1 influenza A viruses, J. Virol., **87**, 8756-8766 (2013).

10. **Kornfeld, R. and Kornfeld, S.:** Assembly of asparagine-linked oligosaccharides, Annu. Rev. Biochem., **54**, 631-664 (1985).

11. **Kelleher, D. J. and Gilmore, R.:** An evolving view of the eukaryotic oligosaccharyltransferase, Glycobiology, **16**, 47R-62R (2006).

12. **Yan, A. and Lennarz, W. J.:** Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons, Glycobiology, **15**, 1407-1415 (2005).

13. **Chavan, M. and Lennarz, W.:** The molecular basis of coupling of translocation and N-glycosylation, Trends Biochem. Sci., **31**, 17-20 (2006).

14. **Petrescu, A.-J., Milac, A.-L., Petrescu, S. M., Dwek, R. A., and Wormald, M. R.:** Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding, Glycobiology, **14**, 103-114 (2004).

15. **Dean, N.:** Asparagine-linked glycosylation in the yeast Golgi, Biochimica et Biophysica Acta (BBA) - General Subjects, **1426**, 309-322 (1999).

16. **Burda, P. and Aebi, M.:** The dolichol pathway of N-linked glycosylation, Biochim. Biophys. Acta, **1426**, 239-57 (1999).

17. **Huffaker, T. C. and Robbins, P. W.:** Temperature-sensitive yeast mutants deficient in asparagine-linked glycosylation, J. Biol. Chem., **257**, 3203-3210 (1982).

18. **Helenius, J., Ng, D. T. W., Marolda, C. L., Walter, P., Valvano, M. A., and Aebi, M.:** Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein, Nature, **415**, 447-450 (2002).

19. **Burda, P., Jakob, C. A., Beinhauer, J., Hegemann, J. H., and Aebi, M.:** Ordered assembly of the asymmetrically branched lipid-linked oligosaccharide in the endoplasmic reticulum is ensured by the substrate specificity of the individual glycosyltransferases, Glycobiology, **9**, 617-625 (1999).

20. **Hebert, D. N., Garman, S. C., and Molinari, M.:** The glycan code of the endoplasmic reticulum: asparagine-linked carbohydrates as protein maturation and quality-control tags, Trends Cell Biol., **15**, 364-370 (2005).

21. **Ito, Y., Hagihara, S., Matsuo, I., and Totani, K.:** Structural approaches to the study of oligosaccharides in glycoprotein quality control, Curr. Opin. Struct. Biol., **15**, 481-489 (2005).

22. **Arai, M. A., Matsuo, I., Hagihara, S., Totani, K., Maruyama, J.-i., Kitamoto, K., and Ito, Y.:** Design and synthesis of oligosaccharides that interfere with glycoprotein quality-control systems, Chembiochem, **6**, 2281-2289 (2005).

23. **Totani, K., Ihara, Y., Matsuo, I., Koshino, H., and Ito, Y.:** Synthetic substrates for an endoplasmic reticulum protein-folding sensor, UDP-glucose: glycoprotein glucosyltransferase, Angewandte Chemie International Edition, **44**, 7950-7954 (2005).

24. **Sharma, C. B., Lehle, L., and Tanner, W.:** Solubilization and characterization of the initial enzymes of the dolichol pathway from yeast, Eur. J. Biochem., **126**, 319-325 (1982).

25. **Rine, J., Hansen, W., Hardeman, E., and Davis, R. W.:** Targeted selection of recombinant clones through gene dosage effects, Proc. Natl. Acad. Sci., **80**, 6750-6754 (1983).

26. **Shams-Eldin, H., Chaban, B., Niehus, S., Schwarz, R. T., and Jarrell, K. F.:** Identification of the Archaeal *alg7* gene homolog (encoding N-Acetylglucosamine-1-phosphate transferase) of the N-linked glycosylation system by cross-domain complementation in Saccharomyces cerevisiae, J. Bacteriol., **190**, 2217-2220 (2008).

27. **Lu, J., Takahashi, T., Ohoka, A., Nakajima, K.-i., Hashimoto, R., Miura, N., Tachikawa, H., and Gao, X.-D.:** Alg14 organizes the formation of a

multiglycosyltransferase complex involved in initiation of lipid-linked oligosaccharide biosynthesis, Glycobiology, **22**, 504-516 (2012).

28. **Gao, X.-D., Moriyama, S., Miura, N., Dean, N., and Nishimura, S.-I.:** Interaction between the C termini of Alg13 and Alg14 mediates formation of the active UDP-N-acetylglucosamine transferase complex, J. Biol. Chem., **283**, 32534-32541 (2008).
29. **Chantret, I., Dancourt, J., Barbat, A., and Moore, S. E. H.:** Two proteins homologous to the N- and C-terminal domains of the bacterial glycosyltransferase MurG are required for the second step of dolichyl-linked oligosaccharide synthesis in *Saccharomyces cerevisiae*, J. Biol. Chem., **280**, 9236-9242 (2005).
30. **Bickel, T., Lehle, L., Schwarz, M., Aebi, M., and Jakob, C. A.:** Biosynthesis of lipid-linked oligosaccharides in *Saccharomyces cerevisiae*: Alg13p and Alg14p form a complex required for the formation of GlcNAc2-PP-Dolichol, J. Biol. Chem., **280**, 34500-34506 (2005).
31. **Gao, X.-D., Tachikawa, H., Sato, T., Jigami, Y., and Dean, N.:** Alg14 recruits Alg13 to the cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-N-acetylglucosamine transferase required for the second step of N-linked glycosylation, J. Biol. Chem., **280**, 36254-36262 (2005).
32. **Gao, X.-D., Nishikawa, A., and Dean, N.:** Physical interactions between the Alg1, Alg2, and Alg11 mannosyltransferases of the endoplasmic reticulum, Glycobiology, **14**, 559-570 (2004).
33. **Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L.:** Predicting transmembrane protein topology with a hidden markov model: application to complete genomes, J. Mol. Biol., **305**, 567-580 (2001).

34. **O'Reilly, M. K., Zhang, G., and Imperiali, B.:** In vitro evidence for the dual function of Alg2 and Alg11: essential mannosyltransferases in N-linked glycoprotein biosynthesis, Biochemistry (ACS Publications), **45**, 9593-9603 (2006).

35. **Maita, N., Nyirenda, J., Igura, M., Kamishikiryo, J., and Kohda, D.:** Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases, J. Biol. Chem., **285**, 4941-4950 (2010).

36. **Silberstein, S. and Gilmore, R.:** Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase, The FASEB Journal, **10**, 849-58 (1996).

37. **Zielinska, D. F., Gnad, F., Wiśniewski, J. R., and Mann, M.:** Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints, Cell, **141**, 897-907 (2010).

38. **Szymanski, C. M., Logan, S. M., Linton, D., and Wren, B. W.:** *Campylobacter* – a tale of two protein glycosylation systems, Trends Microbiol., **11**, 233-238 (2003).

39. **Glover, K. J., Weerapana, E., Numao, S., and Imperiali, B.:** Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*, Chem. Biol., **12**, 1311-1316 (2005).

40. **Linton, D., Dorrell, N., Hitchen, P. G., Amber, S., Karlyshev, A. V., Morris, H. R., Dell, A., Valvano, M. A., Aebi, M., and Wren, B. W.:** Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway, Mol. Microbiol., **55**, 1695-1703 (2005).

41. **Kowarik, M., Young, N. M., Numao, S., Schulz, B. L., Hug, I., Callewaert, N., Mills, D. C., Watson, D. C., Hernandez, M., Kelly, J. F., Wacker, M., and Aebi, M.:** Definition of the bacterial N-glycosylation site consensus sequence, EMBO J., **25**, 1957-1966 (2006).

42. **Madigan, M. T. M. M. T. B. T. D.:** Brock biology of microorganisms, Pearson/Benjamin Cummings (2009).

43. **Calo, D., Kaminski, L., and Eichler, J.:** Protein glycosylation in Archaea: sweet and extreme, Glycobiology, **20**, 1065-1076 (2010).

44. **Mescher, M. F. and Strominger, J. L.:** Purification and characterization of a prokaryotic glycoprotein from the cell envelope of *Halobacterium salinarium*, J. Biol. Chem., **251**, 2005-2014 (1976).

45. **Voisin, S., Houliston, R. S., Kelly, J., Brisson, J.-R., Watson, D., Bardy, S. L., Jarrell, K. F., and Logan, S. M.:** Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*, J. Biol. Chem., **280**, 16586-16593 (2005).

46. **Szymanski, C. M. and Wren, B. W.:** Protein glycosylation in bacterial mucosal pathogens, Nat. Rev. Microbiol., **3**, 225-237 (2005).

47. **Wieland, F., Dompert, W., Bernhardt, G., and Sumper, M.:** Halobacterial glycoprotein saccharides contain covalently linked sulphate, FEBS Lett., **120**, 110-114 (1980).

48. **Eichler, J., Arbiv, A., Cohen-Rosenzweig, C., Kaminski, L., Kandiba, L., and Konrad, Z.:** N-glycosylation in *Haloferax volcanii*: adjusting the sweetness, Frontiers in Microbiology, **4**, (2013).

49. **Abu-Qarn, M., Giordano, A., Battaglia, F., Trauner, A., Hitchen, P. G., Morris, H. R., Dell, A., and Eichler, J.:** Identification of AglE, a second glycosyltransferase involved in N glycosylation of the *Haloferax volcanii* S-layer glycoprotein, J. Bacteriol., **190**, 3140-3146 (2008).

50. **Yurist-Doutsch, S., Abu-Qarn, M., Battaglia, F., Morris, H. R., Hitchen, P. G., Dell, A., and Eichler, J.:** *aglF*, *aglG* and *aglI*, novel members of a gene island involved in

the N-glycosylation of the *Haloferax volcanii* S-layer glycoprotein, Mol. Microbiol., **69**, 1234-1245 (2008).

51. **Guan, Z., Naparstek, S., Kaminski, L., Konrad, Z., and Eichler, J.:** Distinct glycan-charged phosphodolichol carriers are required for the assembly of the pentasaccharide N-linked to the *Haloferax volcanii* S-layer glycoprotein, Mol. Microbiol., **78**, 1294-1303 (2010).

52. **Kaminski, L., Abu-Qarn, M., Guan, Z., Naparstek, S., Ventura, V. V., Raetz, C. R. H., Hitchen, P. G., Dell, A., and Eichler, J.:** AglJ adds the first sugar of the N-linked pentasaccharide decorating the *Haloferax volcanii* S-layer glycoprotein, J. Bacteriol., **192**, 5572-5579 (2010).

53. **Chaban, B., Voisin, S., Kelly, J., Logan, S. M., and Jarrell, K. F.:** Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in Archaea, Mol. Microbiol., **61**, 259-268 (2006).

54. **Abu-Qarn, M. and Eichler, J.:** Protein N-glycosylation in Archaea: defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation, Mol. Microbiol., **61**, 511-525 (2006).

55. **Messner, P.:** Prokaryotic protein glycosylation is rapidly expanding from “curiosity” to “ubiquity”, Chembiochem, **10**, 2151-2154 (2009).

56. **Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., and Aebi, M.:** N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*, Science, **298**, 1790-1793 (2002).

57. **Aas, F. E., Vik, A., Vedde, J., Koomey, M., and Egge-Jacobsen, W.:** *Neisseria gonorrhoeae* O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure, Mol. Microbiol., **65**, 607-24 (2007).

58. **Allos, B. M.:** *Campylobacter jejuni* infections: update on emerging issues and trends, Clin. Infect. Dis., **32**, 1201-1206 (2001).

59. **Blaser, M. J., Wells, J. G., Feldman, R. A., Pollard, R. A., and Allen, J. R.:** *Campylobacter* enteritis in the United States a multicenter study, Ann. Intern. Med., **98**, 360-365 (1983).

60. **Nothaft, H. and Szymanski, C. M.:** Protein glycosylation in bacteria: sweeter than ever, Nat. Rev. Microbiol., **8**, 765-778 (2010).

61. **Fry, B. N., Korolik, V., ten Brinke, J. A., Pennings, M. T. T., Zalm, R., Teunis, B. J. J., Coloe, P. J., and van der Zeijst, B. A. M.:** The lipopolysaccharide biosynthesis locus of *Campylobacter jejuni* 81116, Microbiology, **144**, 2049-2061 (1998).

62. **Olivier, N. B., Chen, M. M., Behr, J. R., and Imperiali, B.:** In vitro biosynthesis of UDP-N, N'-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system, Biochemistry (ACS Publications), **45**, 13659-13669 (2006).

63. **Glover, K. J., Weerapana, E., Chen, M. M., and Imperiali, B.:** Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the *Campylobacter jejuni* N-linked glycosylation pathway, Biochemistry (ACS Publications), **45**, 5343-5350 (2006).

64. **Troutman, J. M. and Imperiali, B.:** *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions, Biochemistry (ACS Publications), **48**, 2807-2816 (2009).

65. **Alaimo, C., Catrein, I., Morf, L., Marolda, C. L., Callewaert, N., Valvano, M. A., Feldman, M. F., and Aebi, M.:** Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides, EMBO J., **25**, 967-976 (2006).

66. **Kelly, J., Jarrell, H., Millar, L., Tessier, L., Fiori, L. M., Lau, P. C., Allan, B., and Szymanski, C. M.:** Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer, J. Bacteriol., **188**, 2427-2434 (2006).

67. **Scott, N. E., Nothaft, H., Edwards, A. V. G., Labbate, M., Djordjevic, S. P., Larsen, M. R., Szymanski, C. M., and Cordwell, S. J.:** Modification of the *Campylobacter jejuni* N-linked glycan by EptC protein-mediated addition of phosphoethanolamine, J. Biol. Chem., **287**, 29384-29396 (2012).

68. **Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S., and Aebi, M.:** The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation, Glycobiology, **15**, 361-7 (2005).

69. **Kowarik, M., Numao, S., Feldman, M. F., Schulz, B. L., Callewaert, N., Kiermaier, E., Catrein, I., and Aebi, M.:** N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase, Science, **314**, 1148-1150 (2006).

70. **Lizak, C., Gerber, S., Numao, S., Aebi, M., and Locher, K. P.:** X-ray structure of a bacterial oligosaccharyltransferase, Nature, **474**, 350-355 (2011).

71. **Lizak, C., Gerber, S., Zinne, D., Michaud, G., Schubert, M., Chen, F., Bucher, M., Darbre, T., Zenobi, R., Reymond, J.-L., and Locher, K. P.:** A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB, J. Biol. Chem., **289**, 735-746 (2014).

72. **Lizak, C., Gerber, S., Michaud, G., Schubert, M., Fan, Y.-Y., Bucher, M., Darbre, T., Aebi, M., Reymond, J.-L., and Locher, K. P.:** Unexpected reactivity and

mechanism of carboxamide activation in bacterial N-linked protein glycosylation, Nat. Commun., **4**, (2013).

73. **Young, N. M., Brisson, J.-R., Kelly, J., Watson, D. C., Tessier, L., Lanthier, P. H., Jarrell, H. C., Cadotte, N., St. Michael, F., Aberg, E., and Szymanski, C. M.:** Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*, J. Biol. Chem., **277**, 42530-42539 (2002).

74. **Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J., and Guerry, P.:** Evidence for a system of general protein glycosylation in *Campylobacter jejuni*, Mol. Microbiol., **32**, 1022-1030 (1999).

75. **Karlyshev, A. V., Everest, P., Linton, D., Cawthraw, S., Newell, D. G., and Wren, B. W.:** The *Campylobacter jejuni* general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks, Microbiology, **150**, 1957-1964 (2004).

76. **Szymanski, C. M., Burr, D. H., and Guerry, P.:** *Campylobacter* protein glycosylation affects host cell interactions, Infect. Immun., **70**, 2242-2244 (2002).

77. **Hendrixson, D. R. and DiRita, V. J.:** Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract, Mol. Microbiol., **52**, 471-484 (2004).

78. **Ding, W., Nothaft, H., Szymanski, C. M., and Kelly, J.:** Identification and quantification of glycoproteins using ion-pairing normal-phase liquid chromatography and mass spectrometry, Mol. Cell. Proteomics, **8**, 2170-2185 (2009).

79. **van Sorge, N. M., Bleumink, N. M. C., van Vliet, S. J., Saeland, E., van Der Pol, W. L., van Kooyk, Y., and van Putten, J. P. M.:** N-glycosylated proteins and distinct lipooligosaccharide glycoforms of *Campylobacter jejuni* target the human C-type lectin receptor MGL, Cell. Microbiol., **11**, 1768-1781 (2009).

80. **Gross, J., Grass, S., Davis, A. E., Gilmore-Erdmann, P., Townsend, R. R., and St. Geme, J. W.:** The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification, J. Biol. Chem., **283**, 26010-26015 (2008).

81. **St Geme, J. W.:** The HMW1 adhesin of nontypeable *Haemophilus influenzae* recognizes sialylated glycoprotein receptors on cultured human epithelial cells, Infect. Immun., **62**, 3881-3889 (1994).

82. **Grass, S., Buscher, A. Z., Swords, W. E., Apicella, M. A., Barenkamp, S. J., Ozchlewski, N., and St Geme, J. W.:** 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).

83. **Grass, S., Lichti, C. F., Townsend, R. R., Gross, J., and St Geme, J. W.:** The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin, PLoS Pathog., **6**, e1000919 (2010).

84. **Naegeli, A., Neupert, C., Fan, Y.-Y., Lin, C.-W., Poljak, K., Papini, A. M., Schwarz, F., and Aebi, M.:** Molecular analysis of an alternative N-glycosylation machinery by functional transfer from *Actinobacillus pleuropneumoniae* to *Escherichia coli*, J. Biol. Chem., **289**, 2170-2179 (2014).

85. **Freeze, H. H.:** Update and perspectives on congenital disorders of glycosylation, Glycobiology, **11**, 129R-143R (2001).

86. **Aebi, M. and Hennet, T.:** Congenital disorders of glycosylation: genetic model systems lead the way, Trends Cell Biol., **11**, 136-141 (2001).

87. **Scott, K., Gadomski, T., Kozicz, T., and Morava, E.:** Congenital disorders of glycosylation: new defects and still counting, J. Inherit. Metab. Dis., 1-9 (2014).

88. **Solá, R. and Griebenow, K.:** Glycosylation of therapeutic proteins, BioDrugs, **24**, 9-21 (2010).

89. **Solá, R. J. and Griebenow, K.:** Effects of glycosylation on the stability of protein pharmaceuticals, J. Pharm. Sci., **98**, 1223-1245 (2009).

90. **Grabowski, G. A.:** Enzyme therapy is not enough, The Lancet, **358, Supplement**, S29 (2001).

91. **Beutler, E.:** Gaucher disease, Blood Rev., **2**, 59-70 (1988).

92. **Stahl, P. D., Rodman, J. S., Miller, M. J., and Schlesinger, P. H.:** Evidence for receptor-mediated binding of glycoproteins, glycoconjugates, and lysosomal glycosidases by alveolar macrophages, Proc. Natl. Acad. Sci., **75**, 1399-1403 (1978).

93. **Achord, D. T., Brot, F. E., Bell, C. E., and Sly, W. S.:** Human β-glucuronidase: In vivo clearance and in vitro uptake by a glycoprotein recognition system on reticuloendothelial cells, Cell, **15**, 269-278 (1978).

94. **Furbish, F. S., Steer, C. J., Krett, N. L., and Barranger, J. A.:** Uptake and distribution of placental glucocerebrosidase in rat hepatic cells and effects of sequential deglycosylation, Biochimica et Biophysica Acta (BBA) - General Subjects, **673**, 425-434 (1981).

95. **Takasaki, S., Murray, G. J., Furbish, F. S., Brady, R. O., Barranger, J. A., and Kobata, A.:** Structure of the N-asparagine-linked oligosaccharide units of human placental beta-glucocerebrosidase, J. Biol. Chem., **259**, 10112-10117 (1984).

96. **Grabowski, G. A., Barton, N. W., Pastores, G., Dambrosia, J. M., Banerjee, T. K., McKee, M. A., Parker, C., Schiffmann, R., Hill, S. C., and Brady, R. O.:** Enzyme therapy in type 1 Gaucher disease: comparative efficacy of mannose-terminated glucocerebrosidase from natural and recombinant sources, Ann. Intern. Med., **122**, 33-39 (1995).

97. **Van Patten, S. M., Hughes, H., Huff, M. R., Piepenhagen, P. A., Waire, J., Qiu, H., Ganesa, C., Reczek, D., Ward, P. V., Kutzko, J. P., and Edmunds, T.:** Effect of

mannose chain length on targeting of glucocerebrosidase for enzyme replacement therapy of Gaucher disease, Glycobiology, **17**, 467-478 (2007).

98. **Baneyx, F.:** Recombinant protein expression in *Escherichia coli*, Curr. Opin. Biotechnol., **10**, 411-421 (1999).

99. **Swartz, J. R.:** Advances in *Escherichia coli* production of therapeutic proteins, Curr. Opin. Biotechnol., **12**, 195-201 (2001).

100. **Hannig, G. and Makrides, S. C.:** Strategies for optimizing heterologous protein expression in *Escherichia coli*, Trends Biotechnol., **16**, 54-60 (1998).

101. **Sørensen, H. P. and Mortensen, K. K.:** Advanced genetic strategies for recombinant protein expression in *Escherichia coli*, J. Biotechnol., **115**, 113-128 (2005).

102. **Wacker, M., Feldman, M. F., Callewaert, N., Kowarik, M., Clarke, B. R., Pohl, N. L., Hernandez, M., Vines, E. D., Valvano, M. A., Whitfield, C., and Aebi, M.:** Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems, Proc. Natl. Acad. Sci., **103**, 7088-7093 (2006).

103. **Feldman, M. F., Wacker, M., Hernandez, M., Hitchen, P. G., Marolda, C. L., Kowarik, M., Morris, H. R., Dell, A., Valvano, M. A., and Aebi, M.:** Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*, Proc. Natl. Acad. Sci., **102**, 3016-3021 (2005).

104. **Fisher, A. C., Haitjema, C. H., Guarino, C., Çelik, E., Endicott, C. E., Reading, C. A., Merritt, J. H., Ptak, A. C., Zhang, S., and DeLisa, M. P.:** Production of secretory and extracellular N-linked glycoproteins in *Escherichia coli*, Appl. Environ. Microbiol., **77**, 871-881 (2011).

105. **Comstock, L. E. and Kasper, D. L.:** Bacterial glycans: key mediators of diverse host immune responses, Cell, **126**, 847-850 (2006).

106. **LaTemple, D. C., Abrams, J. T., Zhang, S. Y., and Galili, U.:** Increased immunogenicity of tumor vaccines complexed with anti-Gal: studies in knockout mice for α1,3-galactosyltransferase, Cancer Res., **59**, 3417-3423 (1999).

107. **Wang, L. X.:** Toward oligosaccharide- and glycopeptide-based HIV vaccines, Curr. Opin. Drug Discovery Dev., **9**, 194-206 (2006).

108. **Wang, L.-X. and Amin, Mohammed N.:** Chemical and chemoenzymatic synthesis of glycoproteins for deciphering functions, Chem. Biol., **21**, 51-66 (2014).

109. **Yang, Q., Li, C., Wei, Y., Huang, W., and Wang, L.-X.:** Expression, glycoform characterization, and antibody-binding of HIV-1 V3 glycopeptide domain fused with human IgG1-Fc, Bioconjug. Chem., **21**, 875-883 (2010).

110. **Schwarz, F., Lizak, C., Fan, Y.-Y., Fleurkens, S., Kowarik, M., and Aebi, M.:** Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo, Glycobiology, **21**, 45-54 (2011).

111. **Meier-Dieter, U., Barr, K., Starman, R., Hatch, L., and Rick, P. D.:** Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen, J. Biol. Chem., **267**, 746-753 (1992).

112. **Alexander, D. C. and Valvano, M. A.:** Role of the *rfe* gene in the biosynthesis of the *Escherichia coli* O7-specific lipopolysaccharide and other O-specific polysaccharides containing N-acetylglucosamine, J. Bacteriol., **176**, 7079-7084 (1994).

113. **Whitfield, C.:** Biosynthesis of lipopolysaccharide O antigens, Trends Microbiol., **3**, 178-185 (1995).

114. **Rubirés, X., Saigi, F., Piqué, N., Climent, N., Merino, S., Albertí, S., Tomás, J. M., and Regué, M.:** A gene (*wbbL*) from *Serratia marcescens* N28b (O4) complements the rfb-50 mutation of *Escherichia coli* K-12 derivatives, J. Bacteriol., **179**, 7581-6 (1997).

115. **Varki, A.:** Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins, Nature, **446**, 1023-1029 (2007).

116. **Kim, J., Kim, Y.-G., and Lee, G.:** CHO cells in biotechnology for production of recombinant proteins: current state and further potential, Appl. Microbiol. Biotechnol., **93**, 917-930 (2012).

117. **Noh, S. M., Sathyamurthy, M., and Lee, G. M.:** Development of recombinant Chinese hamster ovary cell lines for therapeutic protein production, Current Opinion in Chemical Engineering, **2**, 391-397 (2013).

118. **Jervis, A. J., Butler, J. A., Lawson, A. J., Langdon, R., Wren, B. W., and Linton, D.:** Characterization of the structurally diverse N-linked glycans of *Campylobacter species*, J. Bacteriol., **194**, 2355-2362 (2012).

119. **Pei, Z. H., Ellison, R. T., 3rd, and Blaser, M. J.:** Identification, purification, and characterization of major antigenic proteins of *Campylobacter jejuni*, J. Biol. Chem., **266**, 16363-9 (1991).

120. **Lin, J., Michel, L. O., and Zhang, Q.:** CmeABC functions as a multidrug efflux system in *Campylobacter jejuni*, Antimicrob. Agents Chemother., **46**, 2124-2131 (2002).

121. **di Guana, C., Lib, P., Riggsa, P. D., and Inouyeb, H.:** Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein, Gene, **67**, 21-30 (1988).

122. **Pryor, K. D. and Leiting, B.:** High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system, Protein Expr. Purif., **10**, 309-319 (1997).

123. **Schierle, C. F., Berkmen, M., Huber, D., Kumamoto, C., Boyd, D., and Beckwith, J.:** The DsbA signal sequence directs efficient, cotranslational export of passenger

proteins to the *Escherichia coli* periplasm via the Signal Recognition Particle pathway, J. Bacteriol., **185**, 5706-5713 (2003).

124. **Kadokura, H. and Beckwith, J.:** Detecting folding intermediates of a protein as it passes through the bacterial translocation channel, Cell, **138**, 1164-1173 (2009).

125. **Sambrook, J., Fritsch, E. F., and Maniatis, T.:** Molecular cloning : a laboratory manual Cold Spring Harbor Laboratory (2001).

126. **Hanahan, D.:** Studies on transformation of *Escherichia coli* with plasmids, Mol. Biol., **166**, 557-580 (1983).

127. **Yamada, M., Hakura, A., Sofuni, T., and Nohmi, T.:** New method for gene disruption in *Salmonella typhimurium*: construction and characterization of an *ada*-deletion derivative of *Salmonella typhimurium* TA1535, J. Bacteriol., **175**, 5539-47 (1993).

128. **Bevan, M.:** Binary Agrobacterium vectors for plant transformation, Nucleic Acids Res., **12**, 8711-21 (1984).

129. **Laemmli, U. K.:** Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature, **227**, 680-685 (1970).

130. **Fujiyama, K., Misaki, R., Katsura, A., Tanaka, T., Furukawa, A., Omasa, T., and Seki, T.:** N-linked glycan structures of a mouse monoclonal antibody produced from tobacco BY2 suspension-cultured cells, J. Biosci. Bioeng., **101**, 212-218 (2006).

131. **Rozen, S. and Skaletsky, H.:** Primer3 on the WWW for general users and for biologist programmers, Methods Mol. Biol., **132**, 365-86 (2000).

132. **Linton, D., Allan, E., Karlyshev, A. V., Cronshaw, A. D., and Wren, B. W.:** Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*, Mol. Microbiol., **43**, 497-508 (2002).

133. **Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J.:** Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter, J. Bacteriol., **177**, 4121-30 (1995).

134. **Chen, M. M., Glover, K. J., and Imperiali, B.:** From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in *C. jejuni*, Biochemistry (ACS Publications), **46**, 5579-5585 (2007).

135. **Bardwell, J. C. A., McGovern, K., and Beckwith, J.:** Identification of a protein required for disulfide bond formation in vivo, Cell, **67**, 581-589 (1991).

136. **Weerapana, E., Glover, K. J., Chen, M. M., and Imperiali, B.:** Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine, J. Am. Chem. Soc., **127**, 13766-7 (2005).

137. **Revers, L. W., Iain B.H., Webberley, M. C., and Flitsch, S. L.:** The potential dolichol recognition sequence of beta-1,4-mannosyltransferase is not required for enzymatic activity using phytanyl-pyrophosphoryl-alpha-N, N'- diacetylchitobioside as acceptor, Biochem J. , **299**, 23-27 (1994).

138. **Islam, S. T. and Lam, J. S.:** Wzx flippase-mediated membrane translocation of sugar polymer precursors in bacteria, Environ. Microbiol., **15**, 1001-1015 (2013).

139. **Feldman, M. F., Marolda, C. L., Monteiro, M. A., Perry, M. B., Parodi, A. J., and Valvano, M. A.:** The activity of a putative polyisoprenol-linked sugar translocase (Wzx) involved in *Escherichia coli* O antigen assembly is independent of the chemical structure of the O repeat, J. Biol. Chem., **274**, 35129-35138 (1999).

140. **Srichaisupakit, A., Ohashi, T., and Fujiyama, K.:** Identification of a protein glycosylation operon from *Campylobacter jejuni* JCM 2013 and its heterologous expression in Escherichia coli, J. Biosci. Bioeng., **doi: 10.1016/j.bisc.2014.02.011.**,

141. **Sturla, L., Bisso, A., Zanardi, D., Benatti, U., De Flora, A., and Tonetti, M.:** Expression, purification and characterization of GDP-D-mannose 4,6-dehydratase from *Escherichia coli*, FEBS Lett., **412**, 126-130 (1997).

142. **Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S.:** Construction and characterization of new cloning vehicle. II. A multipurpose cloning system, Gene, **2**, 95-113 (1977).

143. **Gao, N. and Lehrman, M. A.:** Non-radioactive analysis of lipid-linked oligosaccharide compositions by fluorophore-assisted carbohydrate electrophoresis, Methods Enzymol., **Volume 415**, 3-20 (2006).

144. **Folch, J., Lees, M., and Stanley, G. H. S.:** A simple method for the isolation and purification of total lipides from animal tissues, J. Biol. Chem., **226**, 497-509 (1957).

145. **Salinas, S. R., Bianco, M. I., Barreras, M., and Ielpi, L.:** Expression, purification and biochemical characterization of GumI, a monotopic membrane GDP-mannose: glycolipid 4-β-D-mannosyltransferase from *Xanthomonas campestris* pv. campestris, Glycobiology, **21**, 903-913 (2011).

146. **Kelley, L. A. and Sternberg, M. J. E.:** Protein structure prediction on the web: a case study using the Phyre server, Nat. Protocols, **4**, 363-371 (2009).

147. **Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J. E., and Kelley, L. A.:** Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre, Proteins: Struct., Funct., Bioinf., **70**, 611-625 (2008).

148. **LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., and McCoy, J. M.:** A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm, Nat Biotech, **11**, 187-193 (1993).

149. **Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y.-Y., Reading, C. A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M., and DeLisa, M. P.:** An engineered

eukaryotic protein glycosylation pathway in *Escherichia coli*, Nat. Chem. Biol., **8**, 434-436 (2012).

150. **Demain, A. L. and Vaishnav, P.:** Production of recombinant proteins by microbes and higher organisms, Biotechnol. Adv., **27**, 297-306 (2009).

151. **Fujiyama, K., Ido, Y., Misaki, R., Moran, D. G., Yanagihara, I., Honda, T., Nishimura, S.-I., Yoshida, T., and Seki, T.:** Human N-acetylglucosaminyltransferase I. Expression in *Escherichia coli* as a soluble enzyme, and application as an immobilized enzyme for the chemoenzymatic synthesis of N-linked oligosaccharides, J. Biosci. Bioeng., **92**, 569-574 (2001).

152. **Nakazawa, K., Furukawa, K., Narimatsu, H., and Kobata, A.:** Kinetic study of human β-1,4-galactosyltransferase expressed in *E. coli*, J. Biochem. (Tokyo), **113**, 747-753 (1993).

List of publications

1. Srichaisupakit, A., Ohashi, T., and Fujiyama, K.: Identification of a protein glycosylation operon from *Campylobacter jejuni* JCM 2013 and its heterologous expression in *Escherichia coli*, J. Biosci. Bioeng., doi: 10.1016/j.bisc.2014.02.011.

2. Srichaisupakit, A., Ohashi, T., Misaki, R., and Fujiyama, K.: Production of the Initial-Stage Eukaryotic *N*-Glycan and its Protein Glycosylation in *Escherichia coli*. J. Biosci. Bioeng.

Appendix

DNA sequences and their respective amino acid sequences

Tags are shown in bold, and putative glycosylation motifs were underlined.

(pelB) cmeA

(ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGC
CGGCGATGGCC)ATGGATAGCAAAGAAGAAGCACCAAAAATACAAATGCCGCCT
CAACCTGTAACAACCATGAGTGCTAAATCTGAAGATTTACCACTTAGTTTTACTT
ACCCTGCTAAACTTGTCAGTGATTATGATGTCATTATAAAGCCTCAAGTTAGTGG
CGTAATAGAAAATAAACTTTTTAAAGCTGGAGATAAAGTAAAAAAAGGACAAAC
ATTATTTATTATAGAACAAGACAAATTTAAAGCTAGTGTTGATTCGGCTTACGGA
CAAGCTTTGATGGCTAAGGCAACTTTCGAAAATGCAAGCAAGGATTTTAATCGTT
CTAAAGCTCTTTTTAGTAAAAGTGCAATCTCTCAAAAGGAATACGACTCTTCTCT
TGCTACATTTAACAATTCAAAAGCTAGTCTAGCAAGTGCTAGAGCACAGCTTGCA
AATGCAAGAATTGATCTAGATCATACCGAAATAAAAGCTCCTTTTGATGGTACTA
TAGGAGATGCTTTAGTTAATATAGGAGATTATGTAAGTGCTTCAACAACTGAACT
AGTTAGAGTTACAAATTTAAATCCTATTTACGCAGATTTCTTTATTTCAGATACAG
ATAAACTAAATTTAGTCCGCAATACTCAAAATGGAAAATGGGATTTAGACAGCA
TTCATGCAAATTTAAATCTTAATGGAGAAACCGTTCAAGGCAAACTTTATTTTAT
TGATTCTGTTATAGATGCTAATAGTGGAACAGTAAAAGCCAAAGCTATATTTGAC
AACAACAACTCAACACTTTTACCAGGTGCTTTTGCAACAATTACTTCAGAAGGTT
TTATACAAAAAAATGGCTTTAAAGTGCCTCAAATAGCTGTTAAACAAAATCAAA
ATGATGTTTATGTTCTTCTTGTTAAAAATGGAAAAGTAGAAAAATCTTCTGTACA
TATAAGCTACCAAAACAATGAATATGCCATTATTGACAAAGGATTACAAAATGG
CGATAAAATCATTTTAGATAACTTTAAAAAAATTCAAGTTGGTAGCGAAGTTAAA
GAAATTGGAGCACAACTCGAG**CACCACCACCACCACCAC**TGA

Translation:

MKYLLPTAAAGLLLLAAQPAMAMDSKEEAPKIQMPPQPVTTMSAKSEDLPLSFTYP
AKLVSDYDVIIKPQVSGVIENKLFKAGDKVKKGQTLFIIEQDKFKASVDSAYGQALM
AKATFENASKDFNRSKALFSKSAISQKEYDSSLATFNNSKASLASARAQLANARIDL
DHTEIKAPFDGTIGDALVNIGDYVSASTTELVRVTNLNPIYADFFISDTDKLNLVRNT
QNGKWDLDSIHANLNLNGETVQGKLYFIDSVIDANSGTVKAKAIFDNNNSTLLPGAF
ATITSEGFIQKNGFKVPQIAVKQNQNDVYVLLVKNGKVEKSSVHISYQNNEYAIIDK
GLQNGDKIILDNFKKIQVGSEVKEIGAQLE**HHHHHH***

MBP-GT

ATGAAAATAAAAACAGGTGCACGCATCCTCGCATTATCCGCATTAACGACGATG
ATGTTTTCCGCCTCGGCTCTCGCCAAAATCGAAGAAGGTAAACTGGTAATCTGGA
TTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGA
AAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAAT
TCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGA
CCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAA

GCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCA
AGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGA
TCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGA
ACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTT
CACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGC
AAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTG
ACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACT
CCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCC
CGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTAC
TGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGG
TATTAACGCCGCCAGTCCGAACAAAGAGCTGGCGAAAGAGTTCCTCGAAAACTA
TCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGC
CGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGC
CACCATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTC
CGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAG
ACTGTCGATGAAGCCCTGAAAGACGCGCAGACTCGTATCACCAAGCTCGAGGGC
GGCGATCAGAACGCGACCCACCACCACCACCACCACTGA

Translation:

MKIKTGARILALSALTTMMFSASALAKIEEGKLVIWINGDKGYNGLAEVGKKFEKDT
GIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDK
LYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKS
ALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKN
KHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKP
FVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAK
DPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTRIT
KLEGGDQNAT**HHHHHH***

The MBP-GT_4 has 4 times repeating of the GGDQNAT sequence

DsbA-GT

ATGAAAAAGATTTGGCTGGCGCTGGCTGGTTTAGTTTTAGCGTTTAGCGCATCGG
CGGCGCAGTATGAAGATGGTAAACAGTACACTACCCTGGAAAAACCGGTAGCTG
GCGCGCCGCAAGTGCTGGAGTTTTTCTCTTTCTTCTGCCCGCACTGCTATCAGTTT
GAAGAAGTTCTGCATATTTCTGATAATGTGAAGAAAAAACTGCCGGAAGGCGTG
AAGATGACTAAATACCACGTCAACTTCATGGGTGGTGACCTGGGCAAAGATCTG
ACTCAGGCATGGGCTGTGGCGATGGCGCTGGGCGTGGAAGACAAAGTCACTGTT
CCGCTGTTTGAAGGCGTACAGAAAACCCAGACCATTCGTTCAGCATCTGATATCC
GCGATGTATTTATCAACGCAGGTATTAAAGGTGAAGAGTACGACGCGGCGTGGA
ACAGCTTCGTGGTCAAATCTCTGGTCGCTCAGCAGGAAAAAGCTGCAGCTGACG
TGCAATTGCGTGGCGTTCCGGCGATGTTTGTTAACGGTAAATATCAGCTGAATCC
GCAGGGTATGGATACCAGCAATATGGATGTTTTTGTTCAGCAGTATGCTGATACA
GTGAAATATCTGTCCGAGAAAAAAGGATCAACTAGTGGTTCTGGTTCCGCGGGT
CTGGTGCCACGCGGTTCCATGGATATCGGAATTAATTCGGATCCGAATTCGAGCT

CCGTCGACAAGCTTGCGGCCGCACTCGAGGGCGGCGATCAGAACGCGACCCACC
ACCACCACCACCACTGA

Translation:

MKKIWLALAGLVLAFSASAAQYEDGKQYTTLEKPVAGAPQVLEFFSFFCPHCYQFE
EVLHISDNVKKKLPEGVKMTKYHVNFMGGDLGKDLTQAWAVAMALGVEDKVTVP
LFEGVQKTQTIRSASDIRDVFINAGIKGEEYDAAWNSFVVKSLVAQQEKAAADVQLR
GVPAMFVNGKYQLNPQGMDTSNMDVFVQQYADTVKYLSEKKGSTSGSGSAGLVPR
GSMDIGINSDPNSSSVDKLAAALEGGDQNAT**HHHHHH***

The DsbA-GT_4 has 4 times repeating of the GGDQNAT sequence

ALG13

ATGGCTAGCATGACTGGTGGACAGCAAATGGGTATGGGTATTATTGAAGAAAAG
GCTCTTTTTGTTACGTGTGGGGCAACGGTGCCATTTCCAAAGCTCGTCTCATGTGT
GCTAAGCGACGAATTCTGCCAAGAATTGATTCAATATGGATTCGTACGTCTAATC
ATTCAGTTTGGGAGAAACTACAGTTCTGAATTTGAGCATTTAGTGCAAGAACGCG
GGGGCCAAAGAGAAAGCCAAAAAATTCCAATTGACCAGTTTGGCTGTGGCGACA
CCGCAAGACAGTATGTCCTGATGAACGGGAAATTAAAGGTGATCGGGTTTGACT
TTTCGACCAAGATGCAAAGTATTATACGTGATTATTCAGATTTGGTCATATCACA
CGCTGGAACGGGCTCTATACTAGATTCTCTACGGTTGAATAAACCGTTGATAGTT
TGCGTAAACGATTCTTTGATGGATAACCACCAGCAGCAGATAGCAGACAAGTTT
GTAGAGTTGGGCTACGTATGGTCTTGTGCACCCACTGAAACAGGTTTGATAGCTG
GTTTACGTGCATCTCAAACAGAGAAACTCAAACCATTCCCAGTTTCTCATAACCC
GTCATTTGAGCGATTGCTAGTTGAAACTATATACAGCTAG

Translation:

MASMTGGQQMGMGIIEEKALFVTCGATVPFPKLVSCVLSDEFCQELIQYGFVRLIIQ
FGRNYSSEFEHLVQERGGQRESQKIPIDQFGCGDTARQYVLMNGKLKVIGFDFSTKM
QSIIRDYSDLVISHAGTGSILDSLRLNKPLIVCVNDSLMDNHQQQIADKFVELGYVWS
CAPTETGLIAGLRASQTEKLKPFPVSHNPSFERLLVETIYS*

ALG14

ATGAAAGAAACCGCTGCTGCGAAATTTGAACGCCAGCACATGGACTCGAAAACG
GCCTACTTGGCGTCATTGGTGCTCATCGTATCGACAGCATATGTTATTAGGTTGAT
AGCGATTCTGCCTTTTTTCCACACTCAAGCAGGTACAGAAAAGGATACGAAAGA
TGGAGTTAACCTACTGAAAATACGAAAATCGTCAAAGAAACCGCTCAAGATTTT
TGTATTCTTAGGATCGGGAGGTCATACTGGTGAAATGATCCGTCTTCTAGAAAAT
TACCAGGATCTTTTACTGGGTAAGTCGATTGTGTACTTGGGTTATTCTGATGAGG
CTTCCAGGCAAAGATTCGCCCACTTTATAAAAAAATTTGGTCATTGCAAAGTAAA
ATACTATGAATTCATGAAAGCTAGGGAAGTTAAAGCGACTCTCCTACAAAGTGT
AAAGACCATCATTGGAACGTTGGTACAATCTTTTGTGCACGTGGTTAGAATCAGA
TTTGCTATGTGTGGTTCCCCTCATCTGTTTTTATTGAATGGGCCTGGAACATGCTG
TATAATATCCTTTTGGTTGAAAATTATGGAACTTCTTTTGCCCCTGTTGGGTTCCT

CCCATATAGTTTATGTAGAATCGCTGGCAAGGATTAATACTCCTAGTCTGACCGG
AAAAATATTATATTGGGTAGTGGATGAATTCATTGTCCAGTGGCAAGAATTGAGG
GACAATTATTTACCAAGATCCAAGTGGTTCGGCATCCTTGTTTAA

Translation:

M**KETAAAKFERQHMDS**KTAYLASLVLIVSTAYVIRLIAILPFFHTQAGTEKDTKDG
VNLLKIRKSSKKPLKIFVFLGSGGHTGEMIRLLENYQDLLLGKSIVYLGYSDEASRQR
FAHFIKKFGHCKVKYYEFMKAREVKATLLQSVKTIIGTLVQSFVHVVRIRFAMCGSP
HLFLLNGPGTCCIISFWLKIMELLLPLLGSSHIVYVESLARINTPSLTGKILYWVVDEFI
VQWQELRDNYLPRSKWFGILV*

ALG1

ATG(TTTTTGGAAATTCCTCGGTGGTTACTTGCCTTAATAATATTATACCTTTCCAT
ACCGTTAGTGGTTTATTATGTTATACCCTACTTGTTTTATGGCAACAAG)TCGACC
AAAAAAAGGATCATCATATTTGTGCTGGGTGATGTAGGACACTCTCCAAGGATAT
GCTATCACGCTATAAGTTTCAGTAAGTTAGGTTGGCAAGTCGAGCTATGCGGTTA
TGTGGAGGACACTCTACCCAAAATTATTTCCAGTGATCCAAATATCACCGTCCAT
CATATGTCAAACTTGAAAAGAAAGGGAGGCGGAACATCAGTTATATTTATGGTA
AAGAAGGTGCTTTTTCAAGTTTTAAGTATTTTCAAATTACTTTGGGAATTGAGAG
GAAGCGATTACATACTAGTTCAAAATCCACCGAGCATACCCATTCTTCCGATTGC
TGTGCTATACAAGTTGACCGGTTGTAAACTAATTATTGATTGGCACAATCTAGCA
TATTCGATATTGCAACTAAAATTTAAAGGAAACTTTTACCATCCTTTAGTGTTGAT
ATCTTACATGGTAGAGATGATATTCAGCAAATTTGCTGATTATAACTTGACTGTT
ACTGAAGCAATGAGGAAATATTTAATTCAAAGCTTTCACTTGAATCCAAAGAGAT
GTGCTGTTCTCTACGACCGCCCGGCTTCCCAATTTCAACCTTTGGCAGGTGACATT
TCTCGTCAAAAAGCCCTAACTACCAAAGCCTTTATAAAGAATTATATTCGCGATG
ATTTTGATACAGAAAAAGGCGATAAAATTATTGTGACTTCAACATCATTCACCCC
TGATGAAGATATTGGTATTTTATTAGGTGCCCTAAAGATTTACGAAAACTCTTAT
GTCAAATTTGATTCAAGTTTGCCTAAGATCTTGTGTTTTATAACGGGTAAAGGAC
CACTAAAGGAGAAATATATGAAGCAAGTAGAAGAATATGACTGGAAGCGCTGTC
AAATCGAATTTGTGTGGTTGTCAGCAGAGGATTACCCAAAGTTATTACAATTATG
CGATTACGGAGTTTCCCTGCATACTTCAAGTTCAGGGTTGGACCTGCCAATGAAA
ATTTTAGATATGTTTGGCTCAGGTCTTCCTGTTATTGCAATGAACTATCCAGTGCT
TGACGAATTAGTACAACACAATGTAAATGGGTTAAAATTTGTTGATAGAAGGGA
GCTTCATGAATCTCTGATTTTTGCTATGAAAGATGCTGATTTATACCAAAAATTG
AAGAAAAATGTAACGCAGGAAGCTGAGAACAGATGGCAATCAAATTGGGAACG
AACAATGAGAGATTTGAAGCTAATTCATATGGCTAGCATGACTGGTGGACAGCA
AATGGGTTGA

Translation:

M(FLEIPRWLLALIILYLSIPLVVYYVIPYLFYGNK)STKKRIIIFVLGDVGHSPRICYHAI
SFSKLGWQVELCGYVEDTLPKIISSDPNITVHHMSNLKRKGGGTSVIFMVKKVLFQV
LSIFKLLWELRGSDYILVQNPPSIPILPIAVLYKLTGCKLIIDWHNLAYSILQLKFKGNF
YHPLVLISYMVEMIFSKFADYNLTVTEAMRKYLIQSFHLNPKRCAVLYDRPASQFQP

LAGDISRQKALTTKAFIKNYIRDDFDTEKGDKIIVTSTSFTPDEDIGILLGALKIYENSY
VKFDSSLPKILCFITGKGPLKEKYMKQVEEYDWKRCQIEFVWLSAEDYPKLLQLCDY
GVSLHTSSSGLDLPMKILDMFGSGLPVIAMNYPVLDELVQHNVNGLKFVDRRELHES
LIFAMKDADLYQKLKKNVTQEAENRWQSNWERTMRDLKLIH**MASMTGGQQMG***

The bracketed DNA sequence and amino acids were omitted in ALG1dTM

TrxALG1

(ATGGGATCTGATAAAATTATTCATCTGACTGATGATTCTTTTGATACTGATGTAC
TTAAGGCAGATGGTGCAATCCTGGTTGATTTCTGGGCACACTGGTGCGGTCCGTG
CAAAATGATCGCTCCGATTCTGGATGAAATCGCTGACGAATATCAGGGCAAACT
GACCGTTGCAAAACTGAACATCGATCACAACCCGGGCACTGCGCCGAAATATGG
CATCCGTGGTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACC
AAAGTGGGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTG
GCCGGCTCTGGGTCCGGTGATGACGATGACAAG)ATGTTTTTGGAAATTCCTCGG
TGGTTACTTGCCTTAATAATATTATACCTTTCCATACCGTTAGTGGTTTATTATGT
TATACCCTACTTGTTTTATGGCAACAAGTCGACCAAAAAAAGGATCATCATATTT
GTGCTGGGTGATGTAGGACACTCTCCAAGGATATGCTATCACGCTATAAGTTTCA
GTAAGTTAGGTTGGCAAGTCGAGCTATGCGGTTATGTGGAGGACACTCTACCCA
AAATTATTTCCAGTGATCCAAATATCACCGTCCATCATATGTCAAACTTGAAAAG
AAAGGGAGGCGGAACATCAGTTATATTTATGGTAAAGAAGGTGCTTTTTCAAGTT
TTAAGTATTTTCAAATTACTTTGGGAATTGAGAGGAAGCGATTACATACTAGTTC
AAAATCCACCGAGCATACCCATTCTTCCGATTGCTGTGCTATACAAGTTGACCGG
TTGTAAACTAATTATTGATTGGCACAATCTAGCATATTCGATATTGCAACTAAAA
TTTAAAGGAAACTTTTACCATCCTTTAGTGTTGATATCTTACATGGTAGAGATGAT
ATTCAGCAAATTTGCTGATTATAACTTGACTGTTACTGAAGCAATGAGGAAATAT
TTAATTCAAAGCTTTCACTTGAATCCAAAGAGATGTGCTGTTCTCTACGACCGCC
CGGCTTCCCAATTTCAACCTTTGGCAGGTGACATTTCTCGTCAAAAAGCCCTAAC
TACCAAAGCCTTTATAAAGAATTATATTCGCGATGATTTTGATACAGAAAAAGGC
GATAAAATTATTGTGACTTCAACATCATTCACCCCTGATGAAGATATTGGTATTTT
ATTAGGTGCCCTAAAGATTTACGAAAACTCTTATGTCAAATTTGATTCAAGTTTG
CCTAAGATCTTGTGTTTTATAACGGGTAAAGGACCACTAAAGGAGAAATATATG
AAGCAAGTAGAAGAATATGACTGGAAGCGCTGTCAAATCGAATTTGTGTGGTTG
TCAGCAGAGGATTACCCAAAGTTATTACAATTATGCGATTACGGAGTTTCCCTGC
ATACTTCAAGTTCAGGGTTGGACCTGCCAATGAAAATTTTAGATATGTTTGGCTC
AGGTCTTCCTGTTATTGCAATGAACTATCCAGTGCTTGACGAATTAGTACAACAC
AATGTAAATGGGTTAAAATTTGTTGATAGAAGGGAGCTTCATGAATCTCTGATTT
TTGCTATGAAAGATGCTGATTTATACCAAAAATTGAAGAAAAATGTAACGCAGG
AAGCTGAGAACAGATGGCAATCAAATTGGGAACGAACAATGAGAGATTTGAAG
CTAATTCATATGGCTAGCATGACTGGTGGACAGCAAATGGGTTGA

Translation:

(MGSDKIIHLTDDSFDTDVLKADGAILVDFWAHWCGPCKMIAPILDEIADEYQGKLT
VAKLNIDHNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSG

SGDDDDK)MFLEIPRWLLALIILYLSIPLVVYYVIPYLFYGNKSTKKRIIIFVLGDVGHS
PRICYHAISFSKLGWQVELCGYVEDTLPKIISSDPNITVHHMSNLKRKGGGTSVIFMV
KKVLFQVLSIFKLLWELRGSDYILVQNPPSIPILPIAVLYKLTGCKLIIDWHNLAYSILQ
LKFKGNFYHPLVLISYMVEMIFSKFADYNLTVTEAMRKYLIQSFHLNPKRCAVLYDR
PASQFQPLAGDISRQKALTTKAFIKNYIRDDFDTEKGDKIIVTSTSFTPDEDIGILLGAL
KIYENSYVKFDSSLPKILCFITGKGPLKEKYMKQVEEYDWKRCQIEFVWLSAEDYPK
LLQLCDYGVSLHTSSSGLDLPMKILDMFGSGLPVIAMNYPVLDELVQHNVNGLKFV
DRRELHESLIFAMKDADLYQKLKKNVTQEAENRWQSNWERTMRDLKLIH**MASMT**
GGQQMG*

Trx was bracketed.

ALG1opt

ATG(TTCCTGGAAATTCCGCGTTGGCTGCTGGCACTGATTATTCTGTATCTGAGCA
TTCCGCTGGTTGTGTATTATGTTATTCCGTACCTGTTCTACGGCAACAAAAGCACC
AAAAAACGCAT)CATTATCTTCGTTCTGGGTGATGTTGGTCATAGTCCGCGTATTT
GTTATCATGCAATCAGCTTTAGCAAACTGGGTTGGCAGGTTGAACTGTGTGGTTA
TGTTGAAGATACCCTGCCGAAAATTATCAGCAGCGATCCGAATATTACCGTGCAT
CACATGAGCAATCTGAAACGTAAAGGTGGTGGCACCAGCGTGATTTTTATGGTTA
AAAAAGTTCTGTTCCAGGTGCTGAGCATCTTTAAACTGCTGTGGGAACTGCGTGG
TAGCGATTATATTCTGGTTCAGAATCCGCCTAGCATTCCGATTCTGCCGATTGCA
GTTCTGTATAAACTGACCGGTTGTAAACTGATCATCGATTGGCATAATCTGGCCT
ATAGCATTCTcCAGCTGAAATTCAAAGGCAACTTTTATCATCCGCTGGTGCTGATT
AGCTATATGGTGGAAATGATCTTCAGCAAATTCGCCGATTATAACCTGACCGTTA
CCGAAGCAATGCGCAAATATCTGATTCAGAGCTTTCATCTGAATCCGAAACGTTG
TGCCGTTCTGTATGATCGTCCGGCAAGCCAGTTTCAGCCGCTGGCAGGCGATATT
AGCCGTCAGAAAGCACTGACCACCAAAGCATTTATCAAAAACTATATCCGCGAC
GATTTCGATACCGAGAAAGGCGATAAAATCATTGTTACCAGCACCAGTTTTACAC
CGGATGAAGATATTGGTATTCTGCTGGGTGCACTGAAAATCTATGAAAACAGCTA
TGTGAAATTCGATAGCAGCCTGCCTAAAATCCTGTGTTTTATTACCGGCAAAGGT
CCGCTGAAAGAAAAATACATGAAACAGGTGGAAGAATATGACTGGAAACGTTGC
CAGATCGAATTTGTTTGGCTGAGCGCAGAAGATTATCCGAAACTGCTcCAGCTGT
GTGATTATGGTGTTAGCCTGCATACCAGCAGCAGCGGTCTGGATCTGCCGATGAA
AATTCTGGATATGTTTGGTAGCGGTCTGCCGGTTATTGCAATGAATTATCCGGTTC
TGGATGAACTGGTTCAGCATAATGTTAACGGTCTGAAATTTGTGGATCGTCGTGA
ACTGCATGAAAGCCTGATTTTTGCCATGAAAGATGCAGATCTGTATCAGAAACTG
AAAAAAAACGTTACCCAAGAAGCCGAAAATCGTTGGCAGAGCAATTGGGAACGT
ACCATGCGTGATCTGAAACTGATTCATATGGCAAGCATGACCGGTGGTCAGCAG
ATGGGTTAA

DNA sequence in bracket was omitted in ALG1dTMopt. Translation was similar to that of ALG1 and ALG1dTM.

TrxALG1opt

(ATGGGATCTGATAAAATTATTCATCTGACTGATGATTCTTTTGATACTGATGTAC
TTAAGGCAGATGGTGCAATCCTGGTTGATTTCTGGGCACACTGGTGCGGTCCGTG
CAAAATGATCGCTCCGATTCTGGATGAAATCGCTGACGAATATCAGGGCAAACT
GACCGTTGCAAAACTGAACATCGATCACAACCCGGGCACTGCGCCGAAATATGG
CATCCGTGGTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACC
AAAGTGGGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTG
GCCGGCTCTGGGTCCGGTGATGACGATGACAAG)ATGTTCCTGGAAATTCCGCGT
TGGCTGCTGGCACTGATTATTCTGTATCTGAGCATTCCGCTGGTTGTGTATTATGT
TATTCCGTACCTGTTCTACGGCAACAAAAGCACCAAAAAACGCATCATTATCTTC
GTTCTGGGTGATGTTGGTCATAGTCCGCGTATTTGTTATCATGCAATCAGCTTTAG
CAAACTGGGTTGGCAGGTTGAACTGTGTGGTTATGTTGAAGATACCCTGCCGAAA
ATTATCAGCAGCGATCCGAATATTACCGTGCATCACATGAGCAATCTGAAACGTA
AAGGTGGTGGCACCAGCGTGATTTTTATGGTTAAAAAAGTTCTGTTCCAGGTGCT
GAGCATCTTTAAACTGCTGTGGGAACTGCGTGGTAGCGATTATATTCTGGTTCAG
AATCCGCCTAGCATTCCGATTCTGCCGATTGCAGTTCTGTATAAACTGACCGGTT
GTAAACTGATCATCGATTGGCATAATCTGGCCTATAGCATTCTcCAGCTGAAATT
CAAAGGCAACTTTTATCATCCGCTGGTGCTGATTAGCTATATGGTGGAAATGATC
TTCAGCAAATTCGCCGATTATAACCTGACCGTTACCGAAGCAATGCGCAAATATC
TGATTCAGAGCTTTCATCTGAATCCGAAACGTTGTGCCGTTCTGTATGATCGTCC
GGCAAGCCAGTTTCAGCCGCTGGCAGGCGATATTAGCCGTCAGAAAGCACTGAC
CACCAAAGCATTTATCAAAAACTATATCCGCGACGATTTCGATACCGAGAAAGG
CGATAAAATCATTGTTACCAGCACCAGTTTTACACCGGATGAAGATATTGGTATT
CTGCTGGGTGCACTGAAAATCTATGAAAACAGCTATGTGAAATTCGATAGCAGC
CTGCCTAAAATCCTGTGTTTTATTACCGGCAAAGGTCCGCTGAAAGAAAAATACA
TGAAACAGGTGGAAGAATATGACTGGAAACGTTGCCAGATCGAATTTGTTTGGC
TGAGCGCAGAAGATTATCCGAAACTGCTcCAGCTGTGTGATTATGGTGTTAGCCT
GCATACCAGCAGCAGCGGTCTGGATCTGCCGATGAAAATTCTGGATATGTTTGGT
AGCGGTCTGCCGGTTATTGCAATGAATTATCCGGTTCTGGATGAACTGGTTCAGC
ATAATGTTAACGGTCTGAAATTTGTGGATCGTCGTGAACTGCATGAAAGCCTGAT
TTTTGCCATGAAAGATGCAGATCTGTATCAGAAACTGAAAAAAAACGTTACCCA
AGAAGCCGAAAATCGTTGGCAGAGCAATTGGGAACGTACCATGCGTGATCTGAA
ACTGATTCATATGGCTAGCATGACTGGTGGACAGCAAATGGGTTGA

Translation was similar to that of TrxALG1. Trx coding portion was bracketed.

TrxALG2

(ATGGGATCTGATAAAATTATTCATCTGACTGATGATTCTTTTGATACTGATGTAC
TTAAGGCAGATGGTGCAATCCTGGTTGATTTCTGGGCACACTGGTGCGGTCCGTG
CAAAATGATCGCTCCGATTCTGGATGAAATCGCTGACGAATATCAGGGCAAACT
GACCGTTGCAAAACTGAACATCGATCACAACCCGGGCACTGCGCCGAAATATGG
CATCCGTGGTATCCCGACTCTGCTGCTGTTCAAAAACGGTGAAGTGGCGGCAACC
AAAGTGGGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTG
GCCGGCTCTGGATCCGGTGATGACGATGACAAGCTCGCCCTT)ATGATTGAAAAG
GATAAAAGAACGATTGCTTTTATTCATCCAGACCTAGGTATTGGGGGCGCTGAAA

GGTTAGTCGTCGATGCAGCATTAGGTCTACAGCAACAAGGACATAGTGTAATCA
TCTATACTAGTCACTGTGATAAATCACATTGTTTCGAAGAAGTTAAAAACGGCCA
ATTAAAAGTCGAAGTTTATGGTGATTTTTTACCGACAAACTTTTTGGGTCGTTTTT
TTATTGTTTTCGCAACAATTAGACAGCTTTATTTAGTTATTCAATTGATCCTACAG
AAAAAAGTGAATGCGTACCAATTAATTATCATTGATCAACTGTCTACATGTATTC
CGCTTCTGCATATCTTTAGTTCTGCCACTTTGATGTTTTATTGTCATTTCCCCGACC
AATTATTGGCTCAAAGAGCTGGGCTATTGAAGAAAATATACAGACTACCATTTGA
CTTAATAGAACAGTTTTCCGTGAGTGCTGCCGATACTGTTGTGGTAAATTCAAAT
TTCACTAAGAATACGTTCCACCAAACGTTCAAGTATTTATCCAATGATCCAGACG
TCATTTATCCATGCGTGGATTTATCAACAATCGAAATTGAAGATATTGACAAGAA
ATTTTTCAAAACAGTGTTTAACGAAGGCGATAGATTTTACCTAAGTATAAATCGT
TTTGAGAAAAAAAAGGATGTTGCGCTGGCTATAAAGGCTTTTGCGTTATCTGAAG
ATCAAATCAATGACAACGTTAAGTTAGTTATTTGCGGTGGTTATGACGAGAGGGT
TGCAGAAAATGTGGAGTACTTGAAGGAACTACAGTCTCTGGCCGATGAATACGA
ATTATCCCATACAACCATATACTACCAAGAAATAAAGCGCGTCTCCGATTTAGAG
TCATTCAAAACCAATAATAGTAAAATTATATTTTTAACTTCCATTTCATCATCTCT
GAAAGAATTACTGCTCGAAAGAACCGAAATGTTATTGTATACACCAGCATATGA
GCACTTTGGTATTGTTCCTTTAGAAGCCATGAAATTAGGTAAGCCTGTACTAGCA
GTAAACAATGGAGGTCCTTTGGAGACTATCAAATCTTACGTTGCTGGTGAAAATG
AAAGTTCTGCCACTGGGTGGCTAAAACCTGCCGTCCCTATTCAATGGGCTACTGC
AATTGATGAAAGCAGAAAGATCTTGCAGAACGGTTCTGTGAACTTTGAGAGGAA
TGGCCCGCTAAGAGTCAAGAAATACTTTTCTAGGGAAGCAATGACTCAGTCATTT
GAAGAAAACGTCGAGAAAGTCATATGGAAAGAAAAAAAGTATTATCCTTGGGAA
ATATTCGGTATTTCATTCTCTAATTTTATTTTGCATATGGCATTTATAAAAATTCT
ACCCAATAATCCATGGCCCTTCCTATTTATGGCCACTTTTATGGTATTATATTTTA
AGAACTACTTATGGGGAATTTACTGGGCATTTGTATTCGCTCTCTCCTACCCTTAT
GAAGAAATAAAGGGCGAGCTTGAAGGTAAGCCTATCCCTAACCCTCTCCTCGGT
CTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTGA

Translation:

(MGSDKIIHLTDDSFDTDVLKADGAILVDFWAHWCGPCKMIAPILDEIADEYQGKLT
VAKLNIDHNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSG
SGDDDDKLAL)MIEKDKRTIAFIHPDLGIGGAERLVVDAALGLQQQGHSVIIYTSHCD
KSHCFEEVKNGQLKVEVYGDFLPTNFLGRFFIVFATIRQLYLVIQLILQKKVNAYQLII
IDQLSTCIPLLHIFSSATLMFYCHFPDQLLAQRAGLLKKIYRLPFDLIEQFSVSAADTV
VVNSNFTKNTFHQTFKYLSNDPDVIYPCVDLSTIEIEDIDKKFFKTVFNEGDRFYLSIN
RFEKKKDVALAIKAFALSEDQINDNVKLVICGGYDERVAENVEYLKELQSLADEYEL
SHTTIYYQEIKRVSDLESFKTNNSKIIFLTSISSSLKELLLERTEMLLYTPAYEHFGIVPL
EAMKLGKPVLAVNNGGPLETIKSYVAGENESSATGWLKPAVPIQWATAIDESRKILQ
NGSVNFERNGPLRVKKYFSREAMTQSFEENVEKVIWKEKKYYPWEIFGISFSNFILH
MAFIKILPNNPWPFLFMATFMVLYFKNYLWGIYWAFVFALSYPYEEIKGELEGKPIPN
PLLGLDSTRTG**HHHHHH***

pglB

ATGGCTAGCATGACTGGTGGACAGCAAATGGGTTTGAAAAAAGAGTATTTAAAA
AACCCTTATTTAGTTTTGTTTGCGATGATTGTATTAGCTTATGTTTTTAGTGTATTT
TGCAGGTTTTATTGGGTTTGGTGGGCAAGTGAGTTTAACGAGTATTTTTTCAATA
ATCAATTAATGATCATTTCAAACGATGGCTATGCTTTTGCTGAGGGCGCAAGAGA
TATGATAGCAGGTTTTCATCAGCCTAATGATTTGAGTTACTATGGATCTTCTTTAT
CCGCGCTTACTTATTGGCTTTATAAAGTCACACCTTTTTCTTTTGAAAGTATCATT
TTATATATGAGTACTTTTTTATCTTCTTTGGTGGTGATTCCTATTATTTTACTAGCT
AATGAATACAAACGTCCTTTAATGGGCTTTGTAGCTGCTCTTTTAGCAAGTATAG
CAAACAGTTATTATAATCGCACTATGAGTGGGTATTATGATACGGATATGCTGGT
AATTGTTTTACCTATGTTTATTTTATTTTTTATGGTAAGAATGATTTTAAAAAAAG
ACTTTTTTTCATTGATTGCCTTGCCGTTATTTATAGGAATTTATCTTTGGTGGTATC
CTTCAAGTTATACTTTAAATGTAGCTTTAATTGGACTTTTTTTAATTTATACACTT
ATTTTTCATAGAAAAGAAAAGATTTTTTATATAGCTGTGATTTTGTCTTCTCTTAC
TCTTTCAAATATAGCATGGTTTTATCAAAGTGCCATTATAGTAATACTTTTTGCTT
TATTCGCCTTAGAGCAAAAACGCTTAAATTTTATGATTATAGGAATTTTAGGTAG
TGCAACTTTGATATTTTTGATTTTAAGTGGTGGGGTTGATCCTATACTTTATCAGC
TTAAATTTTATATTTTTAGAAGTGATGAAAGTGCGAATTTAACACAGGGCTTTAT
GTATTTTAATGTTAATCAAACCATACAAGAAGTTGAAAATGTAGATTTTAGCGAA
TTTATGCGAAGAATTAGTGGTAGTGAAATTGTTTTCTTGTTTTCTTTGTTTGGTTTT
GTATGGCTTTTGAGAAAACATAAAAGTATGATTATGGCTTTACCTATATTGGTAC
TTGGGTTTTTAGCCTTAAAAGGGGGGCTTAGATTTACCATTTATTCTGTACCTGTA
ATGGCTTTAGGATTTGGTTTTTTATTGAGCGAGTTTAAGGCTATATTGGTTAAAAA
ATATAGCCAATTAACTTCAAATGTTTGTATTGTTTTTGCAACTATTTTGACTTTAG
CTCCAGTATTTATCCATATTTACAACTATAAAGCGCCAACAGTTTTTTCTCAAAAT
GAAGCATCATTATTAAATCAATTAAAAAATATAGCCAATAGAGAAGATTATGTG
GTAACTTGGTGGGATTATGGTTATCCTGTGCGTTATTATAGCGATGTGAAAACTT
TAGTAGATGGTGGAAAGCATTTAGGTAAGGATAATTTTTTCCCTTCTTTTGCTTTA
AGCAAAGATGAACAAGCTGCGGCTAATATGGCAAGACTTAGTGTAGAATATACA
GAAAAAAGCTTTTATGCTCCGCAAAATGATATTTTAAAATCAGACATTTTACAAG
CCATGATGAAAGATTATAATCAAAGCAATGTGGATTTATTTCTAGCTTCATTATC
AAAACCTGATTTTAAAATCGATACACCAAAAACTCGTGATATTTATCTTTATATG
CCCGCTAGAATGTCTTTGATTTTTTCTACGGTGGCTAGTTTTTCTTTTATTAATTTA
GATACAGGAGTTTTGGATAAACCTTTTACCTTTAGCACAGCTTATCCACTTGATGT
TAAAAATGGAGAAATTTATCTTAGCAACGGAGTGGTTTTAAGCGATGATTTTAGA
AGTTTTAAAATAGGTGATAATGTGGTTTCTGTAAATAGTATCGTAGAGATTAATT
CTATTAAACAAGGTGAATACAAAATCACTCCAATCGATGATAAGGCTCAGTTTTA
TATTTTTTATTTAAAGGATAGTGCTATTCCTTACGCACAATTTATTTTAATGGATA
AAACCATGTTTAATAGTGCTTATGTGCAAATGTTTTTTTTGGGAAATTATGATAA
GAATTTATTTGACTTGGTGATTAATTCTAGAGATGCTAAAGTTTTTAAACTTAAA
ATTTAA

Translation:

MASMTGGQQMGLKKEYLKNPYLVLFAMIVLAYVFSVFCRFYWVWWASEFNEYFF
NNQLMIISNDGYAFAEGARDMIAGFHQPNDLSYYGSSLSALTYWLYKVTPFSFESIIL
YMSTFLSSLVVIPIILLANEYKRPLMGFVAALLASIANSYYNRTMSGYYDTDMLVIVL
PMFILFFMVRMILKKDFFSLIALPLFIGIYLWWYPSSYTLNVALIGLFLIYTLIFHRKEKI
FYIAVILSSLTLSNIAWFYQSAIIVILFALFALEQKRLNFMIIGILGSATLIFLILSGGVDPI
LYQLKFYIFRSDESANLTQGFMYFNVNQTIQEVENVDFSEFMRRISGSEIVFLFSLFGF
VWLLRKHKSMIMALPILVLGFLALKGGLRFTIYSVPVMALGFGFLLSEFKAILVKKYS
QLTSNVCIVFATILTLAPVFIHIYNYKAPTVFSQNEASLLNQLKNIANREDYVVTWWD
YGYPVRYYSDVKTLVDGGKHLGKDNFFPSFALSKDEQAAANMARLSVEYTEKSFY
APQNDILKSDILQAMMKDYNQSNVDLFLASLSKPDFKIDTPKTRDIYLYMPARMSLIF
STVASFSFINLDTGVLDKPFTFSTAYPLDVKNGEIYLSNGVVLSDDFRSFKIGDNVVS
VNSIVEINSIKQGEYKITPIDDKAQFYIFYLKDSAIPYAQFILMDKTMFNSAYVQMFFL
GNYDKNLFDLVINSRDAKVFKLKI*

pglBopt

ATGGCAAGCATGACCGGTGGTCAGCAGATGGGTCTGAAAAAAGAATATCTGAAA
AACCCGTATCTGGTGCTGTTTGCAATGATTGTTCTGGCCTATGTTTTTAGCGTGTT
TTGCCGTTTTTATTGGGTTTGGTGGGCAAGCGAATTTAACGAGTATTTTTTCAACA
ACCAGCTGATGATTATCAGCAATGATGGTTATGCATTTGCCGAAGGTGCACGTGA
TATGATTGCAGGTTTTCATCAGCCGAATGATCTGAGCTATTATGGTAGCAGCCTG
AGCGCACTGACCTATTGGCTGTATAAAGTTACCCCGTTTAGCTTTGAAAGCATCA
TCCTGTATATGAGCACCTTTCTGAGCAGCCTGGTTGTTATTCCGATTATTCTGCTG
GCCAACGAATATAAACGTCCGCTGATGGGTTTTGTTGCAGCACTGCTGGCGAGCA
TTGCAAATAGCTACTATAATCGTACCATGAGCGGCTATTATGATACCGATATGCT
GGTTATTGTGCTGCCGATGTTTATCCTGTTTTTTATGGTTCGCATGATCCTGAAAA
AAGATTTCTTTAGCCTGATTGCACTGCCGCTGTTCATTGGTATTTATCTGTGGTGG
TATCCGAGCAGCTATACCCTGAATGTTGCACTGATTGGTCTGTTTCTGATTTACAC
CCTGATTTTTCATCGTAAAGAGAAAATCTTTTATATCGCCGTTATCCTGAGCAGTC
TGACCCTGAGTAATATTGCCTGGTTTTATCAGAGCGCCATTATTGTTATTCTGTTT
GCACTGTTTGCCCTGGAACAGAAACGTCTGAACTTTATGATTATTGGCATTCTGG
GTAGCGCCACACTGATCTTTCTGATTCTGAGCGGTGGTGTTGATCCGATTCTGTAT
CAGCTGAAATTCTATATCTTTCGCAGTGATGAAAGCGCAAATCTGACCCAGGGTT
TTATGTATTTTAACGTGAACCAGACCATCCAAGAGGTGGAAAATGTTGATTTCAG
CGAATTTATGCGTCGTATTAGCGGTAGCGAAATCGTTTTTCTGTTTAGCCTGTTTG
GTTTTGTTTGGCTGCTGCGTAAACATAAAAGCATGATTATGGCACTGCCGATTCT
GGTTCTGGGTTTTCTGGCACTGAAAGGTGGTCTGCGTTTTACCATTTATAGCGTTC
CGGTTATGGCCCTGGGTTTTGGTTTTCTGCTGAGTGAATTTAAAGCCATCCTGGTG
AAAAAATACAGCCAGCTGACCAGCAATGTTTGCATTGTTTTTGCAACCATTCTGA
CCCTGGCACCGGTTTTTATTCACATCTATAACTATAAAGCACCGACCGTGTTTAG
CCAGAATGAAGCAAGCCTGCTGAATCAGCTGAAAAACATTGCCAATCGTGAAGA
TTATGTTGTGACCTGGTGGGATTATGGTTATCCGGTTCGTTATTATTCCGATGTTA
AAACCCTGGTGGATGGTGGTAAACATCTGGGTAAAGATAACTTTTTTCCGAGCTT
TGCACTGAGCAAAGATGAACAGGCAGCAGCAAATATGGCACGTCTGAGCGTTGA
ATATACCGAGAAAAGTTTTTATGCACCGCAGAACGATATTCTGAAAAGCGATATT

CTCCAGGCCATGATGAAAGATTATAACCAGAGCAATGTTGACCTGTTTCTGGCAA
GCCTGAGCAAACCGGATTTCAAAATTGATACCCCGAAAACCCGTGATATCTATCT
GTATATGCCTGCACGTATGAGCCTGATTTTTAGCACCGTTGCAAGCTTTAGCTTTA
TCAATCTGGATACCGGTGTTCTGGATAAACCGTTTACCTTTAGTACCGCATATCC
GCTGGATGTGAAAAACGGTGAAATTTATCTGAGCAATGGTGTGGTTCTGAGTGAT
GATTTTCGCAGCTTTAAAATCGGTGATAATGTGGTTAGCGTGAACAGCATTGTTG
AAATCAACAGCATTAAACAGGGCGAGTATAAAATCACCCCGATTGATGATAAAG
CCCAGTTCTATATCTTCTATCTGAAAGATAGCGCAATTCCGTATGCCCAGTTTATC
CTGATGGATAAAACCATGTTTAACAGCGCCTATGTGCAGATGTTTTTTCTGGGCA
ACTATGATAAAAACCTGTTCGATCTGGTGATCAATAGCCGTGATGCCAAAGTGTT
TAAACTGAAAATCTAA

Translation was similar to that of *pglB*.

Acknowledgements

My advisor, Professor Kazuhito Fujiyama, started his career earlier and began his professorship at Department of Biotechnology, Graduate School of Engineering, and International Center for Biotechnology (ICB) in February 2009. In October 2009, I joined his lab under Frontier Biotechnology, Master - Ph.D. program, as a first batch of international student. His enthusiasm, patience and understanding has helped keep me working through; from Master thesis defense, publication writings and this Dissertation. I would also like to thank my thesis committee, Professor Shigenori Kanaya, and Professor Takuya Nihira for their generosity and contributions. Besides, this work could not be completed without my co-advisor, Assistant Professor Takao Ohashi who critically read and made comments on my writings.

Actually I first began my student life at Osaka University in my final year of undergraduate study, by joining FrontierLab@OsakaU student exchange program in Professor Takuya Nihira's laboratory. I am also indebted to him for his encouragement for me to enroll in this course which I am completing.

Activities in Fujiyama laboratory were much facilitated by Associate Professor Ryo Misaki. Without him, I would not be able to obtain materials and chemicals. Luckily enough at Fujiyama laboratory, I have met both Japanese and international students whose diverse backgrounds gave me extensive view of Japanese and overseas cultures. My sincere thanks are also for the secretariat of ICB, for helping me with Japanese documents and guiding me through official processes.

Opportunely, I also had a chance to meet visiting Professors and visiting researchers, not only from Thailand but from around the globe. They have kept me updated on what was going on in research and academia, which sometimes was out of my sight. It was also felt

enlightening talked to them. Also, I have met vibrant students from Thailand who joined short-term short-visit program. I was enjoyed spending times with them as well.

At the International Center for Biotechnology, I must thank students whose shaped up enjoyable research community. I could not list all of names; some graduated and some are trying hard, without quitting nor giving up towards their dissertations. I hope that we can work in academic, somedays, to make a good circle of international exchange.

The financial support of the Ministry of Education, Culture, Sports, Science & Technology (MEXT) of Japan is gratefully appreciated and acknowledged.

I must gratefully acknowledge the *Escherichia coli*, and the variations I have created. This organism has been asked to withstand large plasmids and a total of four antibiotics, as well as the stress of expressing multiple proteins. It is definitely a simple yet clever organism, and I am sure that it can be further engineered for applications of great value to humans.

Finally, I would like to say thank you to my parents and family members, for allowing me to study abroad for an extended period of time. Thanks are also due to my friends, both in Japan and Thailand and beyond. The vagaries of fate have led our paths to cross, and I for one could not be more pleased. All of your kindnesses are very much appreciated.

Akkaraphol Srichaisupakit

September 2014

Osaka University