ENGINEERING WATER-SOLUBLE VARIANTS OF THE SINGLE-SUBUNIT OLIGOSACCHARYLTRANSFERASE

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Yong Hyun (Daniel) Kwon

August 2021

Abstracts

Oligosaccharyltransferase (OST) is a key enzyme in the asparagine-linked (*N*-linked) protein glycosylation pathway. OSTs exist in all domains of life and are capable of transferring a preassembled glycan from lipid carrier to an acceptor peptide. Bacterial OSTs are an single-subunit enzyme that are amenable to recombinant expression in model organism including *Escherichia coli*. As a result, bacterial OSTs have been used as models to explore the mechanism of the *N*-linked glycosylation process in nature. These developments, notwithstanding, recombinant expression and purification of the OST enzymes remain significant challenges. Bacterial OSTs are multi-pass transmembrane protein that requires intricate balance between protein synthesis rate and a pace of membrane insertion. Further, membrane protein purification often necessitates the use of ultracentrifugation and detergent, both of which limit process scalability and compatibility. To address these challenges, we proposed a protein engineering strategy called SIMPLEx or solubilization of integral membrane proteins with high levels of expression to generate water-soluble variants of the bacterial OST. Specifically, we designed several OST chimeras where the *N*-terminus of the OST is fused with the amphipathic protein including engineered human apolipoprotein A-I. Using *E. coli* culture as an expression platform, several SIMPLEx-OSTs could be expressed within the cytoplasmic fraction of the *E. coli*. Importantly, our engineered OSTs retain their biological activity and are able to *N*-glycosylate several acceptor proteins including therapeutic human growth hormone. Collectively, our OST-engineering strategy is anticipated to generate a new subclass of water-soluble *N*-OST enzymes with applications in bioproduction of the glycotherapeutics and glycovaccines.

BIOGRAPHICAL SKETCH

Yong Hyun (Daniel) Kwon received his B.S in Chemical Engineering from University of Minnesota-Twin Cities in 2017. Upon receiving Bachelor's degree, Yong Hyun returned to South Korea to serve in the army for two years. In 2019 he joined Professor Matthew DeLisa's research group at Cornell University. He will be completing his M.S degree after spending two years in DeLisa's research group under the mentorship of Thapakorn Jaroentomeechai. Yong Hyun will continue his studies in DeLisa group as a PhD student after pursuing his M.S degree.

ACKNOWLEDGMENTS

I would like to thank Professor Matthew DeLisa for his support and insightful guidance throughout my MS program. I would also like to thank Thapakorn Jaroentomeechai for his mentorship and providing me the wonderful opportunity to work on this project.

I would also like to thank entire DeLisa group for their friendship, advice, and support.

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction and Background

1.1 Glycobiology

Glycobiology is the study of sugar or glycan, an essential biomolecule presence in all domain of life. (1) Often times glycan moieties, as mono-, oligo-, or polysaccharides, are found covalently attached to the polypeptides to form glycoconjugates. (2) The protein-glycan linkage is catalyzed by an enzymatic reaction called protein glycosylation, which is one of the most common protein post-translational modification in nature. (3) (4) Most secreted proteins in eukaryotic cell are modified by such process, and the attached glycans significantly impact the characteristic of the recipient proteins including protein folding, stability, and biological activity. (5) Glycans are also found prevalently on the cell surface and have significant roles as a molecular signal to mediate several critical biological process including embryogenesis and cell proliferation. (6) While there are numerous protein-glycan linkages in nature, glycan attached to the nitrogen atom of the asparagine side chain (*N*-linked) or to the oxygen atom of the hydroxyl group within serine or threonine amino acids (*O*-linked) are the most common. As relevant to this thesis, the content below will focus primarily on the *N*-linked protein glycosylation process. (7)

1.1.1 *N*-linked protein glycosylation system

Many secreted proteins in eukaryotic cells undergo *N*-linked protein glycosylation, a process that is initiated in the lumen of the endoplasmic reticulum (ER) organelle. (8) Within the ER membrane, several glycosyltransferases in the Alg family work in succession to assemble GlcNAc2Man9Glc3 glycan on the dolichol-pyrophosphate, a polyisoprene-based phospholipid. This lipid-linked

oligosaccharides or LLOs can be recognized by key enzyme oligosaccharyltransferase (OST) that will transfer the preassembled glycan onto polypeptide acceptor and initiate the first step in *N*-linked protein glycosylation.

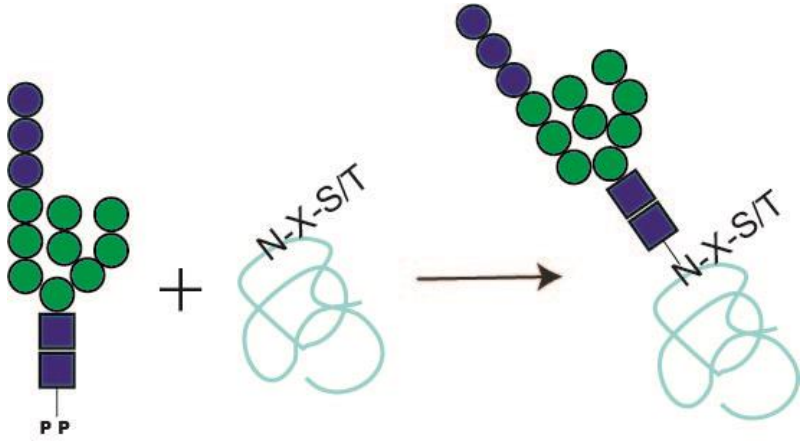

Figure 1. Schematic illustration of the key step in the *N*-linked protein glycosylation pathway

OST links glycan to the nitrogen atom of an asparagine within the specific amino acid sequence of Asn-X-Ser/Thr (where X denotes any amino acids except for proline) as shown in Figure 1. (13, 14) The nascent polypeptide carrying initial *N*-glycan is then transported to the Golgi apparatus for further glycan processing including trimming, elaborating, and editing. The final *N*-glycan on glycoprotein is structurally-diverse and can containing several unique monosaccharides including fucose and sialic acid cap. (15). (16)

In human, there are two catalytic subunits within the OST complexes, namely OST-A and OST-B as shown in Figure 2. OST-A is primary responsible for co-translational modification and OST-B is more responsible for post-translational modification. (17) OSTs are membrane-complex protein with eight different protein sub-units spanning over the cell membrane. While each subunit has different

functions during the *N*-linked protein glycosylation, a study has revealed that STT3, indicated in green color in Figure 2, are the catalytically active sub-units. (18)

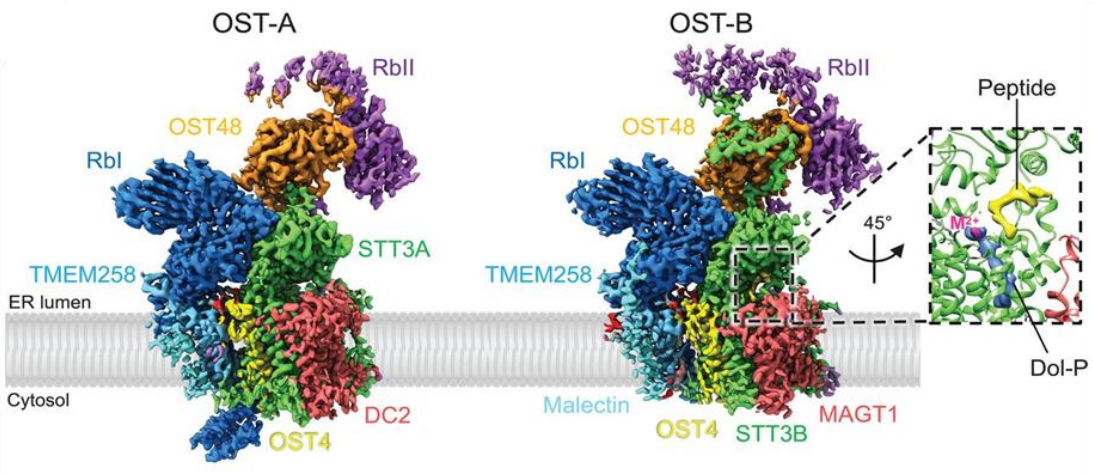

Figure 2. X-ray structures of OST-A and OST-B in human(19)

As shown from the above figure, Dol-P from the glycan cleavage (blue) and acceptor peptides (yellow) are located in STT3B region which indicates STT3 are the main active sites. STT3 are also present in other eukaryotes including yeast and insects (19, 20)

1.1.2 Application of protein N-glycosylation

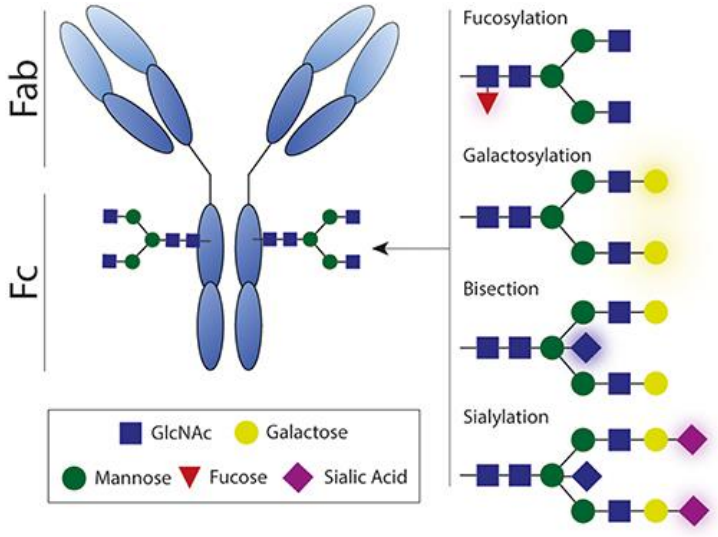

Figure 3. Human IgG with different combinations of glycans (21)

N-linked glycosylation system and human immune system are tightly related. Human immunoglobulin G (IgG) has a highly conserved *N*-glycosylation site at the Fc region as shown in Figure 3. (21, 22, 23) The asparagine located at position 297 (N297) can harbor one of the several glycan structures and each glycan structure can significantly impact the effector function of the antibody. (24) For example, studies have shown than sialylation on IgG-Fc glycan can work as a biomarker of diseases such as influenza H1 virus. (25, 26) Another recent studies have shown that afucosylated IgG are often found in critically ill COVID-19 patients and infusing fucosylated IgG could potentially treat the patients. (27, 28, 29)

1.2 Protein glycosylation in prokaryotes

1.2.1 Discovery of prokaryotic *N*-linked glycosylation

While the existence of prokaryotic glycosylation is now firmly established, the *N*-linked glycosylation was long believed to be restricted to eukaryotes for a long time. In 1999, Szymanski group discovered that *Campylobacter jejuni* (*C. jejuni*), a gram-negative bacteria, has the ability to carry out its own *N*-linked protein glycosylation system. (31, 32, 33)

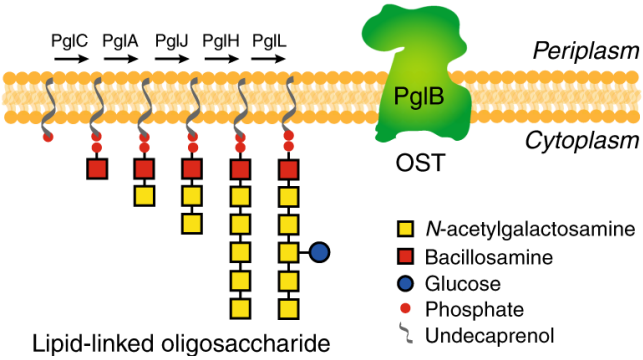

Figure 4. *N*-linked protein glycosylation pathway in *Camplyobacter jejuni* (67)

As shown in Figure 4, several glycosyltransferases within the pgl gene cluster build the lipid-linked oligosaccharide in *C. Jejuni*. (34) The mechanism for

the *N*-glycosylation system in prokaryotic is unique because the bacteria build their oligosaccharides on undecaprenyl pyrophosphate (Und-PP) at the inner membrane of the cells, while eukaryote uses Dol-PP at the ER membrane. (36, 37) In addition, most eukaryotic produce almost the same glycan structure ($Glc_3Man_9GlcNac_2$) at the early stage, while each bacteria display its own unique *N*-linked glycan structure. (38, 39)

1.2.2 Bacterial oligosaccharyltransferases (OST)

The bacterial OSTs are unique because they are single subunit OSTs that can carry out *N*-linked glycosylation process similar to the OST complex in eukaryotes. (40) The most well-studied bacterial OST is called PglB from *C. jejuni* and it shares high structural similarity with STT3 in eukaryotic OST complex. (41)

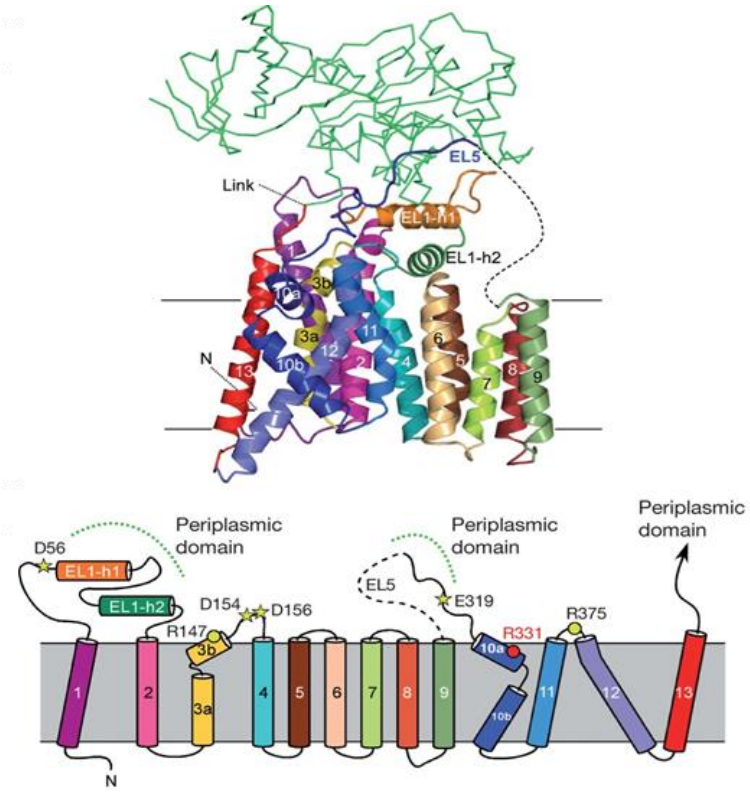

Figure 5. X-ray structure and topological schematic of bacterial OST (42)

However, the glycosylation sequon of the bacteria OST is more restricted and

requires amino acid sequence of D/E-X-N-X-S/T where X can be any amino acid except for proline. To point out, the optimal glycosylation sequon for *C. Jejuni* is DQNAT. (38, 39) This is because PglB requires the interaction with the negatively charged amino acid present at the -2 position for efficient catalytic activity. (42, 43, 44, 45)

The Figure 5 shows the structure of PglB, which consist of 13 alpha-helix domain spanning over the transmembrane domain which makes protein insoluble in nature. (46) The periplasmic domain of the PglB has the main active site and those sites are indicated in yellow stars in the figure. The red sphere located in R331 is responsible for acceptor peptide recognition. (40, 47) The molecular mass of PglB is measured at around 85 kDa and shares structural similarity with other bacterial OSTs such as *Campylobacter lari*. (47, 48)

The most remarkable feature of PglB is that it became a bridge to introduce the N-link glycosylation system to *Escherichia coli*. (50, 51) *E. coli* is absence with endogenous N-glycosylation machinery. *E. coli*, however, is one of the most well-studied model organisms, thus it offers an attractive platform to study *N*-linked glycosylation mechanism. (52, 53, 54) The *E. coli*-based *N*-link glycosylation system opened a new era in the field of synthetic glycobiology with application of cell-free system. (55, 56)

1.3 Cell-free biology

1.3.1 Brief overview of cell-free system

Cell-free system is based on *in vitro* and has become a powerful platform for studying synthetic biology applications. Traditionally, *in vivo* cell systems were heavily used for studying synthetic biology but have assorted constraints including

liability and viability of cells. (56, 57) The *E. coli*-based cell-free system can bypass those limitations and introduce exciting opportunities to exploit the natural biology with rational designs. (58)

Frequently, cell-free system refers to cell-free protein synthesis (CFPS). The CFPS provides direct access and manipulation of the cell environment to support co- or post- translational modifications. CFPS offer major advantages over the conventional *in vivo* system. (59) First, timeline for process that takes weeks to prepare *in vivo* can be significantly reduced down to a few days with CFPS. (60) Second, direct influence on the system provides easier access to add or synthesize new components and to optimize the reaction more precisely. In contrast, *in vivo* system cannot be modified until the end of the cell lysis. (61, 62) Lastly, CFPS can comprehend toxic proteins that living cells cannot maintain during the synthesis. (63) Taken these advantages together, CFPS has served as a key to accelerate the synthetic biology field. (64)

1.3.2 Cell-free synthetic glycobiology

Cell-free glycoprotein synthesis (CFGpS) was developed to fulfill the glycosylation machinery that CFPS lacks. (65) *E. Coli* lysates-based cell-free systems, which are not equipped to synthesize glycoprotein because *E. Coli* is deficient in glycosylation mechanism. (66) Some eukaryotic based system prepared from insect, trypanosomes, or mammalian cells could glycosylate protein. However, these techniques can only perform using their host-specific glycosylation with little to no control of glycosylation components. (67) Furthermore, eukaryotic platforms need technically demanding preparation and is inefficient for glycosylation studies due to low glycoprotein yields. (68)

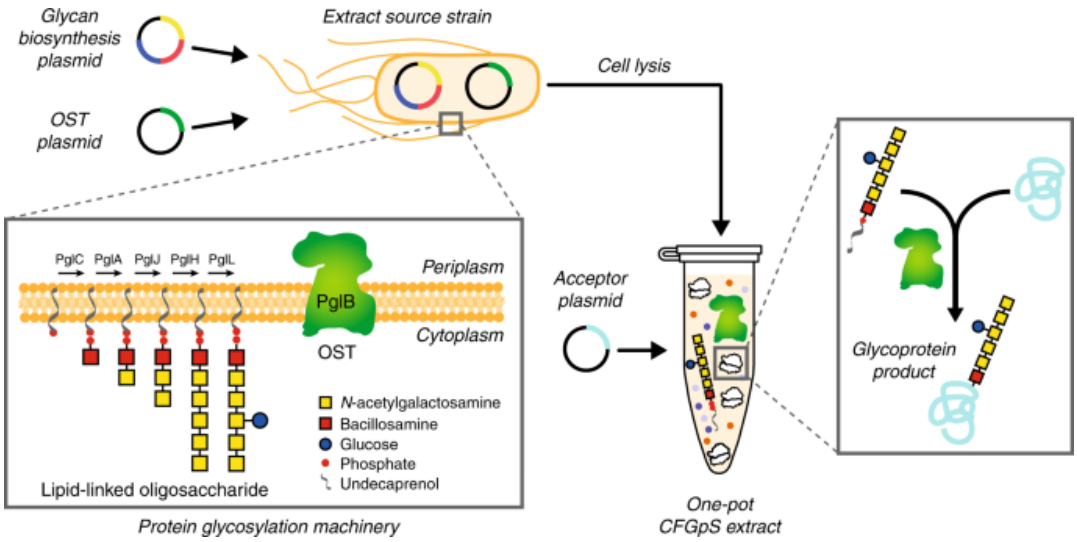

Figure 6. Schematic of one-pot cell-free glycoprotein synthesis (CFGpS) (67)

To address this challenge, *E. coli*-based lysate is augmented with bacterial N-linked glycosylation pathway including the one from *C. jejuni*. By individually preparing the necessary ingredients such as OST, LLOs, and acceptor proteins, the *in vitro* glycoprotein synthesis was achieved. However, preparation of each component including purified OST and extracted LLO are time-consuming. CFGpS can circumvent these laborious works by compressing all the process in a single-pot reaction as shown in Figure 6. As a result, CFGpS provides a modular platform for customizing different glycosylation pathways. (67)

1.4 Protein engineering strategy for recombinant membrane protein production

1.4.1 Brief overview of integral membrane protein (IMP)

Even with the promising cell-free technique, the field of glycobiology has a critical drawback that still holds various potential achievements. The integral membrane proteins (IMPs) plays essential roles in functioning cellular processes

such as transferring essential molecules in and out of the cells. Membrane proteins are often involved in drug designing that more than half of the developed drugs target IMPs for stimulating therapeutic machinery. (69) However, the inherit hydrophobic characteristic of the proteins makes full structural and functional characterization challenging. (70) Detergent purification is often used to solubilize the membrane proteins but finding optimal detergents and buffer conditions is time consuming and cost demanding. Furthermore, detergent solubilized proteins often loss their functions making hard to assess their innate biological activity. (71, 72) These challenges also apply to the bacterial OST and the hydrophobic nature of PglB has remain one of the biggest hurdles for fundamental understanding and applications of this class of protein in synthetic glycobiology.

1.4.2 Solubilization of IMPs with high level expression (SIMPLEx)

To address the challenge, a protein engineering strategy called SIMPLEx (solubilization of IMPs with high levels of expression) was introduced. The idea of SIMPLEx is to fuse truncated apolipoprotein A-I (ApoAI) to serve as a protein "shield". (73)

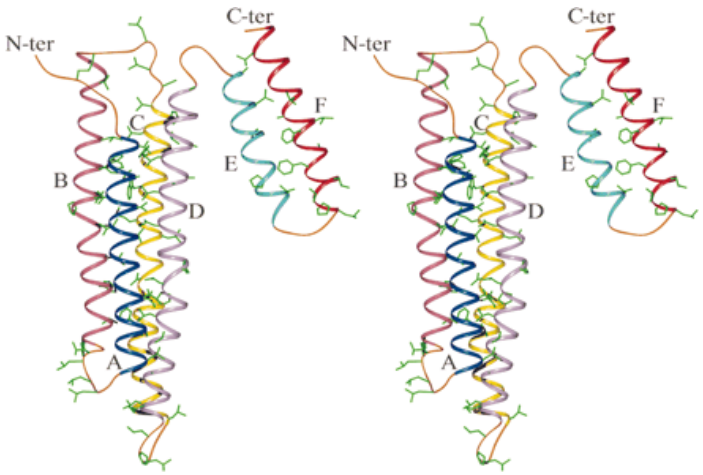

Figure 7. Crystal structure of human apolipoprotein A-I (73)

ApoAI from human plasma is an amphipathic protein that has both hydrophobic and hydrophilic surfaces. It is the major protein component in high density lipoprotein that binds to phospholipid molecules and assemble them into soluble disc or bilayer structures. (74) With this technique, direct *in vivo* expression of solubilized membrane proteins can be achieved. As a result, a structural conformation of the protein can be maintained in water without the detergent while retain the full protein activity.

A previous study has shown that fusing the lipid-binding domain of ApoAI (residues 44-243) into a C-terminal of the target while inserting protein decoy on the N-terminal exhibit remarkable solubility level of a target protein. Protein decoy is important because it is a tactic to make bacteria to synthesize full-length protein in the cytoplasm just like other natural soluble proteins. The SIMPLEx provides the greatest opportunity to study proteins at molecular level and reveal the full machinery that can accelerate the field of synthetic biology. (75, 76)

1.5 Research Objectives

1.5.1 Engineering water-soluble oligosaccharyltransferases using SIMPLEx

In this project, SIMPLEx-PglB chimeras were prepared with two major features: protein decoys and amphipathic protein on N-terminus of the PglB. Traditionally, the SIMPLEx-proteins have ApoAI shield on the C-terminus of the target protein as shown in Figure 8a. However, the same strategy generated low solubility yield for PglB so a few modifications have made to address this problem.

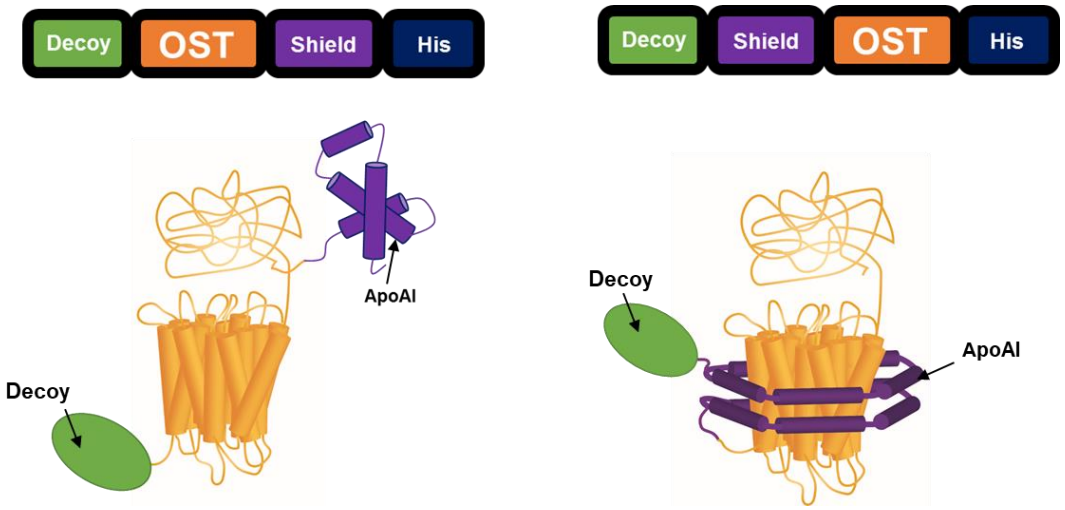

Figure 8. Schematic of engineering water-soluble bacterial OSTs (left: traditional approach, right: new approach)

First, the location of the ApoAI shield was shifted to the N-terminus of the PglB to properly shield the transmembrane domain, a hydrophobic region, as shown in Figure 8b. Second, the length of ApoAI was expanded to properly shield the massive molecular size of PglB. As a proof of principle, the new designed SIMPLEx-PglB variants will be analyzed in this project to answer the two fundamental protein engineering questions: (1) solubility in water and (2) functional activity.

2. Experimental Methods

2.1 Overview

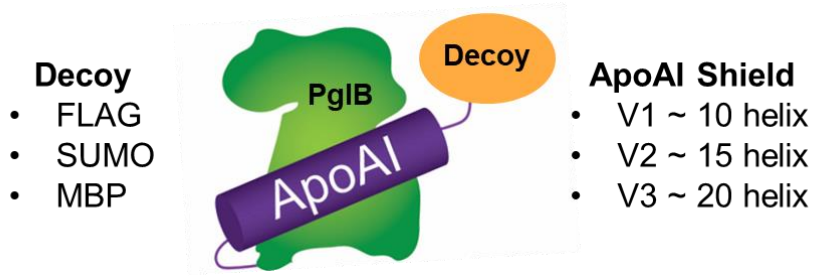

Figure 9. Illustration of engineered SIMPLEx-PglB variants

For this experiment, three protein decoys (MBP, SUMO, and FLAG) were selected to enhance the solubility of SIMPLEx-PglB. Each protein decoy has three different length of ApoAI proteins as shown in the Figure 9. Therefore, total of nine different variants of soluble-OST were engineered. Each variant was synthesized *in vivo* in same *E. coli* strains and was analyzed for their stability and activities. The constructs were named by the following rule [Decoy.ApoAI.V#.PglB]. As an example, PglB fused with FLAG decoy with shortest ApoAI shield is called FLAG.ApoAI.V1.PglB.

2.2 Plasmid constructions

The plasmid used for constructing SIMPLEx-PglB is pBAD expression plasmid encoded the *C. jejuni* PglB with N-terminal FLAG epitope tag, decoy protein, and ApoAI sequentially and with a C-terminal decahistidine affinity tag.

To prepare the lipid-linked oligosaccharides (LLOs), pMW07 based plasmid was used. It was encoded with *C. jejuni* protein glycosylation locus (pgl) with specific deletion of *Cj*PglB. (84)

The primary acceptor protein used for this project is scFv13-R4^{DQNAT}, which is an engineered single chain antibody fragment variant that can be solubly expressed

without disulfide bonds. pET28a(+) plasmid was used with encoding scFv13-R4 modified with a C-terminal DQNAT glycosylation sequence and two mutations at N34 and N77L to remove internal glycosylation site.

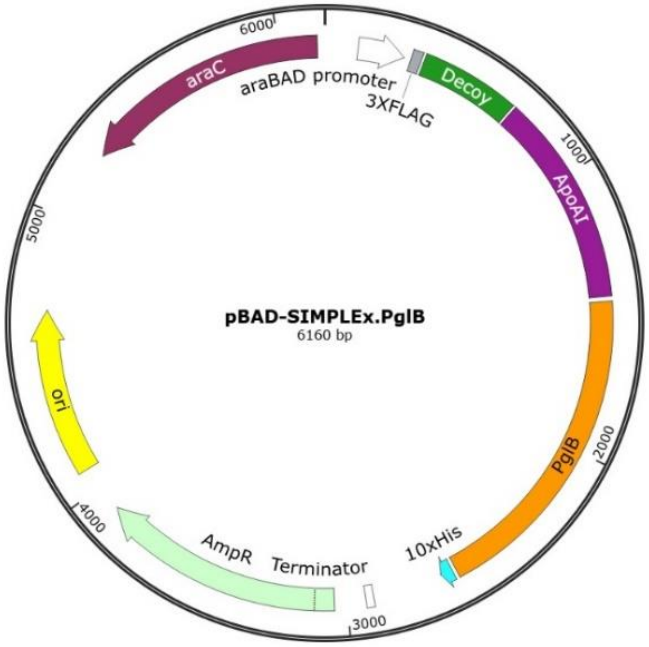

Figure 10. Plasmid map for SIMPLEx-PglB design

2.3 Protein expression and purification

2.3.1 Bacterial strains used in this study

BL21(DE3) and CLM24 *E. Coli* strains were used for the project. For expressing the SIMPLEx-PglB, BL21(DE3) was selected since it demonstrated highest level of expression. This strain was also used for expressing scFv13-R4^{DQNAT} acceptor protein. CLM24 is a derivative of W3110 *E.Coli* strains with deletion of Waal ligase. Therefore, the strain allows the accumulation of oligosaccharides on Und-PP, which is a suitable substrate of PglB.

2.3.2 Protein expression

The SIMPLEx-PglB plasmid encoded cells were overnight in 10mL of Luria-Bertani (LB; 10 g/L of tryptone, 5 g/L of yeast extract, 5 g/L of NaCl, pH 7.2) media supplemented with ampicillin (Amp) in 1:1,000 ratio at 37˚C. Cells were subcultured into 1L terrific broth (TB; 12 g/L of tryptone, 24 g/L of yeast extract, 0.4% of glycerol (v/v%), 10% of 0.17M KH_2PO_4/0.72M K_2HPO_4 phosphate buffer (v/v%)) with ampicillin at 37˚C shaking incubator until the OD_{600} level reaches between 0.7~0.8. The cells were then induced with the final concentration of 0.2% L-arabinose at 16˚C until the OD_{600} reaches 1.2~1.5. The samples were spin down at 8K g for 25 minutes. The cell pellets were resuspended with Resuspension Buffer A (50 mM HEPES, 250 mM NaCl, 1 mM EDTA, pH 7.5) that about 10mL of buffer was used per 1g of cell. The cells were lysis by using the Avestin Emulsifulx-C5 homogenizer. After the lysate, the cells were spin down at 15K g for 30 minutes at 4˚C to remove cell debris. The supernatants are collected for protein purification.

The same method was applied to express scFv13-R4^{DQNAT} except kanamycin replaced ampicillin and isopropyl β-D-1-thiogalactopyranoside (IPTG) was used as new inducer. The IPTG was added to a final concentration 0.1 mM. Upon inducing, the sample was incubated at 20ºC for 18 hours.

For WT PglB, Resuspension Buffer A with additional 0.01% (w/v%) DDM was used. The same preparation protocol was followed except the supernatants after the first lysate spin down at 15K were ultracentrifuged at 28K for 2 hours at 4˚C. In this way, majority of the membrane debris can be eliminated and secure pure product.

2.3.3 Protein purification

Protein purification was performed immediately after the supernatants were

collected. The fresh Nickel resins was washed in 2mL of 10 mM imidazole buffer and centrifuged at 700g at 4˚C for five to six times. The Nickel resins were then added to the supernatants and incubated at 4˚C for one hour with SIMPLEx variants and scFv13-R4^{DQNAT}. WT PglB was incubated with the resin at 4˚C for 24 hours.

2.3.3.1 Nickel-based immobilized metal affinity chromatography

Once the incubation was finished, the sample was poured into empty gravity flow columns. Once the sample completely ran through the column, it was run once more and collected as the flow through sample. The column was washed with 4mL of 10 mM imidazole and collected as the wash fraction. Finally, the proteins were collected by using 2.5mL of 250 mM imidazole.

2.3.3.2 Size exclusion chromatography

Upon collecting the protein from Nickel column, the samples were then went through PD-10 desalting column protocol to remove imidazole ions. The column was washed with equilibrate buffer (20 mM HEPES, 150 mM NaCl, 1mM EDTA, pH 7.4) for four to five times and samples were ran through the column. The final samples were than collected with the same equilibrate buffer to store the final protein.

2.4 Lipid-linked oligosaccharide recombinant expression and extractions

E. coli strain CLM24 were used for carrying the plasmids for LLOs from *C. jejuni*. CLM24 is an engineered E. coli strains with deletion of WaaL chromosomal gene which is responsible for flipping the LLO from cytoplasm to periplasm. The same protocol was used to culture the cells as we used for the large-scale culture

except it was induced with 0.2% L-arabinose at 30˚C. After 16 hours, cells were resuspended with methanol and dried for overnight. The dried pellets were cut into small pieces and collected in 50mL Falcon tube with 12mL CM. The sample was water bath sonicated for five minutes. Then, the pellet samples were centrifuged at 3000g for 10 minutes. The liquid fraction is decanted and the whole process is repeated with decanting liquid, sonication, centrifuge, and decant. The 20mL of H_2O is added to the pellet sample and vortexed to quickly break the pellet. The same sonication, centrifuge, and decant protocol was proceeded. Then 18mL of CMW was added to the pellet, then vortexed and sonicated. 8mL of methanol was added to the pellet sample and centrifuged again at same condition. The supernatants were collected to a fresh 50mL tube. 8mL of chloroform and 2mL of H_2O is added to the sample and centrifuged at same condition. The aqueous layer, top fraction, is discarded while the organic layer, bottom fraction, is collected into glass dish and dried overnight at room temperature. The dried pellets are than resuspended in cell-free glycosylation buffer (10mM HEPES, 10mM $MnCl_2$, 0.1% (w/v%) DDM, and pH 7.5), allocated, and stored in -20^{o}C fridge.

2.5 Cell-free protein glycosylation

To perform a functional assay, *in vitro* glycosylation (IVG) imitating the natural glycosylation environment in one-pot system was used. The main ingredients for N-linked protein glycosylation such as extracted LLOs, purified acceptor protein with DQNAT tag, and purified PglB were mixed all together. Then supplement ingredients such as buffer for mimic cell environment and $MnCl_2$ were also added to enhance the glycosylation activity.

2.5.1 Crude lysate based cell-free protein glycosylation

The plasmid encoded cells were initially overnight in 5mL LB media with ampicillin at 1:1000 dilution. Cells were subculture at 50mL LB with same ampicillin ratio and left at 37˚C shaking incubator until the OD_{600} level reaches 0.7~0.8 per mL. The cells were induced with 0.2% L-arabinose at 16˚C until the final OD_{600} reaches 1.2~1.5. Each culture were transferred into fresh 50 mL tubes and centrifuged at 3,500 rpm for 12 minutes. Approximately 0.3g of cell pellets were harvested and 300 μL of resuspension buffer A were added. For WT PglB, 1% DDM buffer was used. The cells were lysated by the sonication method for 6 minutes at 40% amplitude for 10 seconds on/off. The lysates were than centrifuged at 15K for 10 minutes at 4˚C with microcentrifuge. The supernatants were than taken from the samples and used for the in vitro glycosylation reaction.

2.5.2 Purified OST based cell-free protein glycosylation

The individually purified scFv13-R4^{DQNAT} and PglB constructs were mixed with extracted LLO were prepared as above protocol. The reaction mixture is than placed on water bath at 30˚C for 16 hours. If necessary, the samples can be placed on water bath up to 20 hours. The samples were than taken out from water bath and mixed with beta-mercapoethanol to stop the glycosylation activity by denaturing the protein and for analyzing the protein by immunoblot analysis.

2.6 Glycoprotein analysis

SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) is part of electrophoresis method where the proteins can be separated in terms of their mass, which is crucial for analyzing the glycosylation activity. The mass of

glycoprotein and aglycoprotein is differ by the mass of glycan, meaning that glycoprotein have higher molecular mass. Therefore, SDS-PAGE separation was a key assessment to differentiate the two.

For preparing the samples, the sample buffer, which is the mixture of a reducing agent, beta-mercapoethanol, and SDS-page loading dye in 1:9 ratio is added to the sample. About 1/3 of sample volume of mixture is added. The sample is heated at either 95ºC for 10 minutes or 70ºC for 15 minutes. The reason for heating the sample is to prevent the formation of secondary and tertiary structures of the protein by disrupting hydrogen bonds and stretching the molecules. Upon heating, the samples were cooled down to the room temperature.

The samples are then loaded in the polyacrylamide gel in electrophoresis buffer. When the voltage is applied, the proteins will move along the gel in the direction of the positively charged anodes. Smaller size proteins can be transported easily than the larger proteins, which will migrate into gels more slowly and therefore causing proteins to be separated by their molecular size.

Western Blot is a technique where we transfer the proteins to a solid surface and visualize the target proteins by using a primary and secondary antibody. Upon gel electrophoresis, the target proteins get transferred to a membrane, typically polyvinylidene difluoride (PVDF), using electroblotting method. This method uses a transfer buffer solution and an electric current to drive negatively charged proteins in the gel onto positively charged PVDF membrane. In this way, the proteins can be transferred to a membrane while retaining their orientations from the gel.

Once the transfer is completed, the membrane goes through a blocking process to prevent the non-specific binding of antibody. Typically, 5% (w/v) bovine serum albumin (BSA) or non-fat dry milk are used for blocking the membrane

typically 30 minutes in slow agitation or overnight at 4ºC. The proteins from the solution will fill the empty sides of the membrane where target proteins have not been attached, resulting in providing clear background by eliminating unspecific binding of antibodies.

The membrane will then be washed with tris-buffered saline with Tween 20 (TBST) for five times for 30 minutes. The primary antibody, diluted in TBST buffer in 1:10,000 ratio most of the time, is introduced to the membrane and will bind to specific target proteins. The membrane can be incubated in the room temperature for one hour or overnight at 4ºC. Upon the end of incubation, the membrane is washed with TBST for several times to remove any unbounded antibody and clearer background can be achieved. In most cases, the membrane was washed every 10 minutes for six times.

The membrane is than incubated with the secondary antibody that can binds to the specific portion of the primary antibody. The secondary antibody is with a reporter enzyme such as horseradish peroxidase (HRP). HRP can enhance the protein signal by producing luminescence from cleaving the chemiluminescent substrate. This luminescence are the indications of the presence of the target protein and can be detected from the Chemi-Doc machine. Upon incubating the secondary antibody for 30 minutes, the membrane was washed for six times for every five minutes. The sample was than analyzed with BioRad Chemi-Doc.

3. Results

3.1 Protein engineering strategy renders OST enzymes water-soluble

The first fundamental question regarding the stability/solubility of the SIMPLEx-PglB was answered by performing immunoblot analysis with the crude lysates prepared from the *E. coli* cells expressing one of the SIMPLEx-PglB variants – FLAG.ApoAI.V1.PglB. Lysate from cells expressing WT PglB served as positive control and lysate from cells carrying empty plasmid served as negative control. The S/D/I denote for soluble fraction (supernatants collected from the first ultracentrifuge run), detergent soluble fraction (supernatants collected from the second ultracentrifuge run), and insoluble fraction (resuspended pellets from the second ultracentrifuge run) respectively.

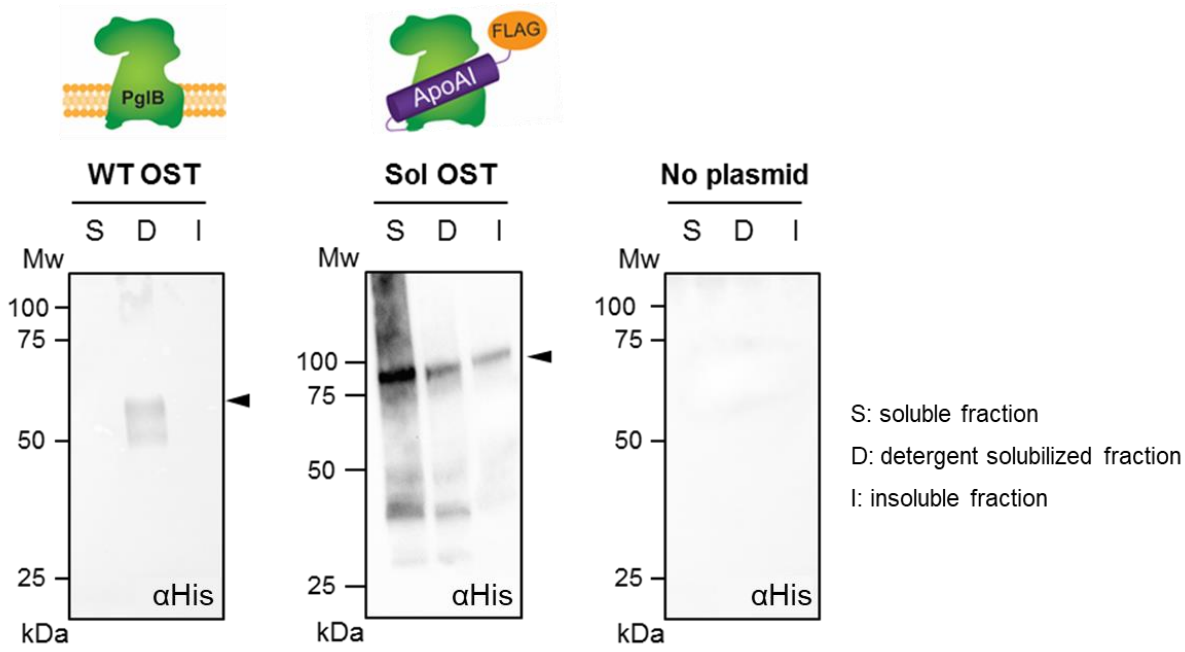

Figure 11. Immunoblot analysis of crude lysates for testing solubility of PglB variant.

Having no protein band on the soluble fraction of the WT PglB confirms that the PglB is not water-soluble in nature. On the other hand, the protein band located in the soluble fraction of the SIMPLEx construct provides an evidence that

this engineered protein can be expressed as water-soluble domain in the detergent-free condition. The SIMPLEx-PglB show stronger band intensity than WT PglB meaning that overall expression level for SIMPLEx version was improvedThe remaining constructs were prepared in a same manner and categorized by the types of decoy proteins. The result can be found in Figure 12 and 13.

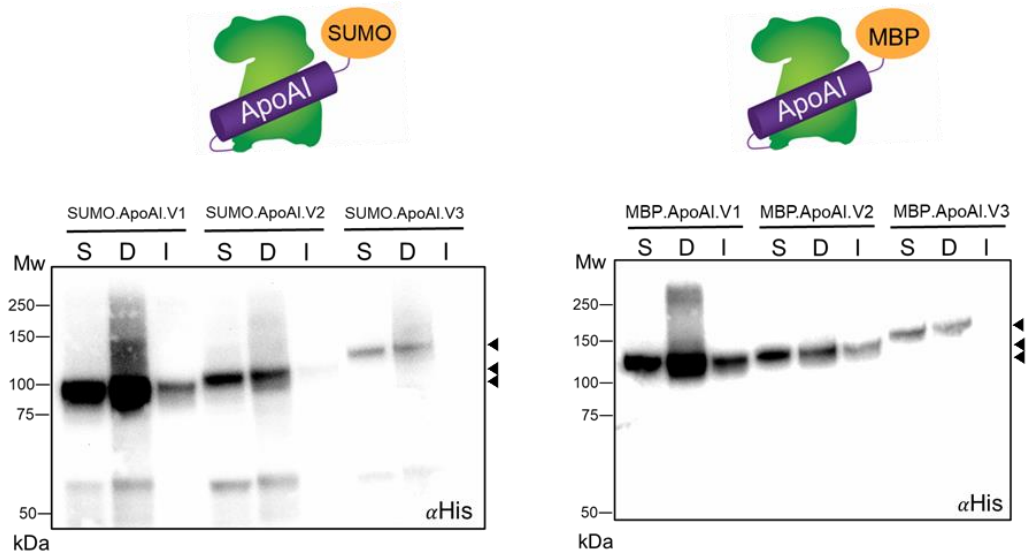

Figure 12. SIMPLEx-PglB with SUMO decoy proteins and MBP decoy proteins were expressed.

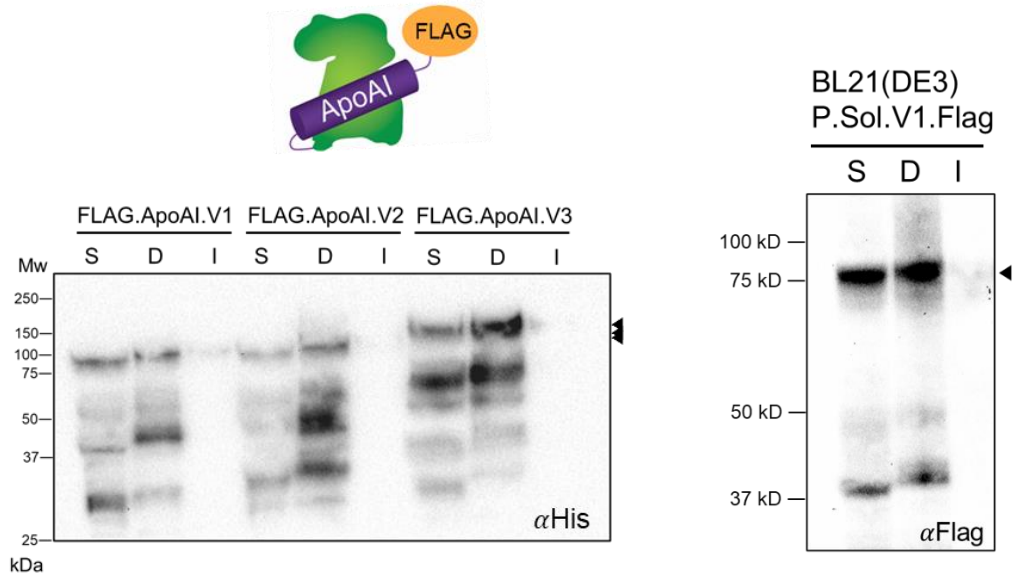

Figure 13. SIMPLEx-PglB with FLAG decoy proteins were expressed and western blot from both His-tag antibody and FLAG-tag antibody for comparison.

Figure 12 and 13 show that all the designed constructs are soluble in water. The data show that MBP decoy protein showed the highest expression level followed by SUMO decoy and FLAG decoy. SUMO also works as a solubility enhancer but at a lesser extent than the MBP decoy. FLAG decoy showed higher level of degraded product, therefore the anti-FLAG immunoblot was ran in parallel to confirm the expression of the full-length protein fusion

3.2 Engineered SIMPLEx-OST is functionally active

3.2.1 Positive control establishment with WT OST

Prior to test the activities of the constructs, the WT PglB was first tested to ensure the positive control and negative control data. This data was a useful guideline throughout the whole project. As shown in the Figure 14, the + signs indicate the component was added while the – signs mean the specific ingredient has been omitted.

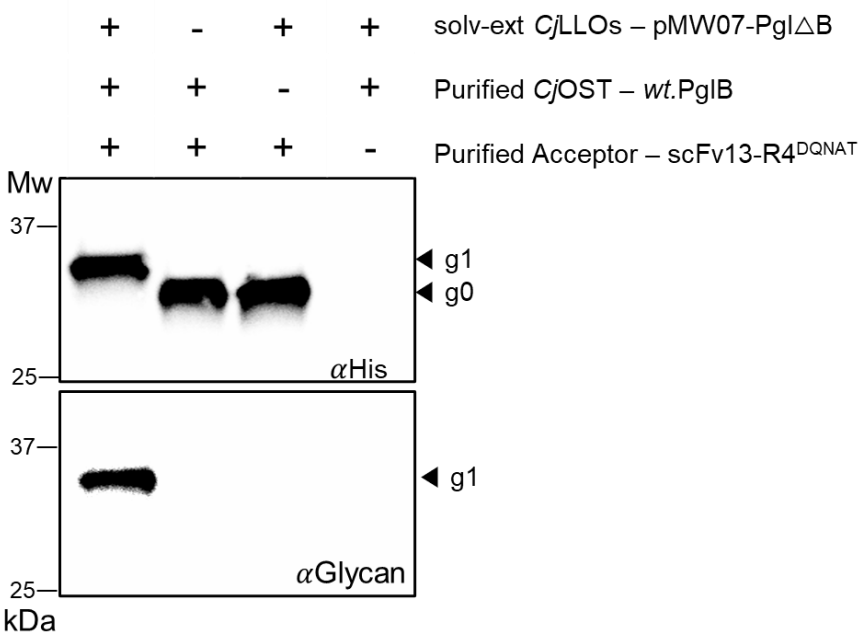

Figure 14. IVG blot established positive control data and negative control data. To be specific, the first lane with all the ingredients was the positive control data

whereas all the remaining lanes were considered as negative control. The sign g0 indicates aglycoprotein and g1 indicates glycoprotein. αGlycan blot is generated from house-made antibody that specifically recognizes the *Cj*LLOs. The presence of protein band in αGlycan blot confirms the formation of glycoprotein.

3.2.2 Cell-free glycosylation with crude lysates with solubilized OST

Running a reaction from crude lysates from the constructs was a rapid method to assess activity of the engineered constructs. WT indicates the purified WT PglB in the buffer containing DDM detergent and Lysate WT indicates WT PglB from the crude cell lysates. The amount of lysates in each reaction were normalized in terms of volume that 5 μL of lysates were added.

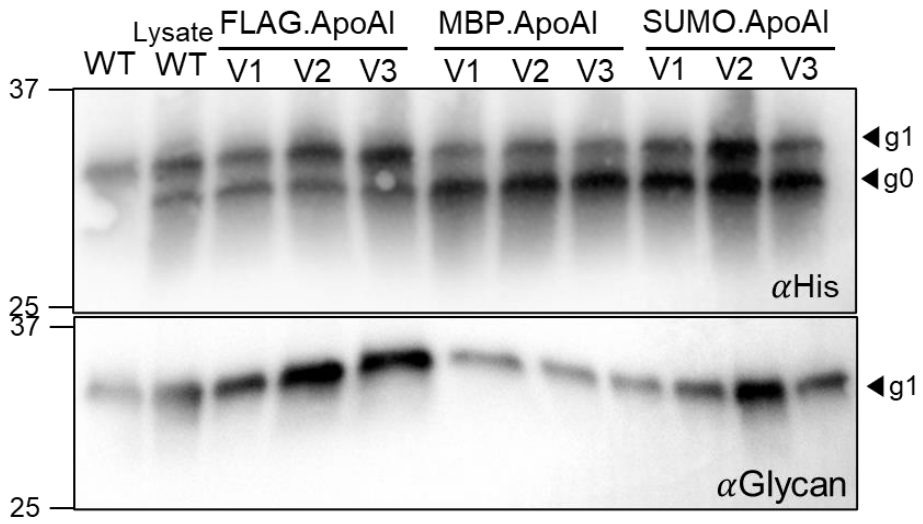

Figure 15. IVG blot with SIMPLEX-PglB prepared from crude lysates

All the SIMPLEx variants showed glycosylation activities according to the Figure 15. FLAG.ApoAI.V2.PglB, FLAG.ApoAI.V3.PglB, and SUMO.ApoAI.V2.PglB constructs showed relatively stronger protein band indicating more glycoproteins were synthesized. The MBP-fused PglB in general showed low glycosylation activity.

3.2.3 Cell-free glycosylation with purified SIMPLEx-OST

Ni-NTA purification protocol was able to capture and isolate the SIMPLEx-PglB construct. The Coomassie staining gel and western blot shows the presence of isolated construct. The FLAG.ApoAI.V2.PglB is used for demonstrating the Ni-NTA protocol method. The nanodrop was used to measure the concentration of the purified PglB and Coomassie staining gel was used to calculate the total yield of the PglB. The αHis blot was used to confirm that the isolated protein was SIMPLEx-PglB.

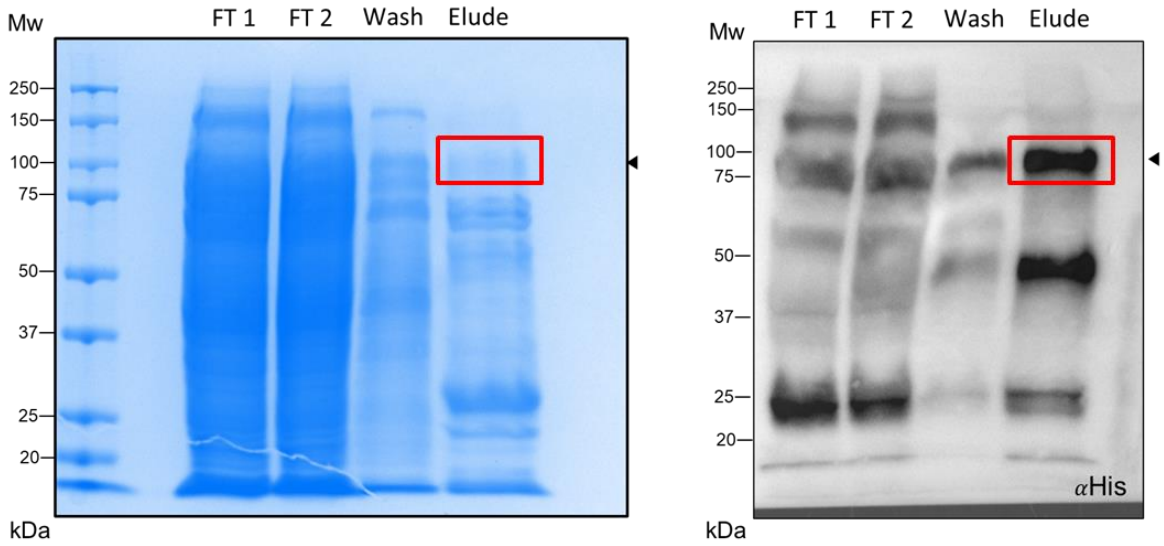

Figure 16. Ni-NTA purification protocol isolated FLAG.ApoAI.V2.PglB

Once the proteins were purified, they were normalized in terms of mass for functional assays.. For the Figure 17 and Figure 18, approximately 10 and 30 µg of PglB were added to the reaction respectively.

In Figure 17, the result is in similar agreement with the lysate blots that the blots with high activities were also observed. After adding the enzyme up to ~30 µg in the reaction, FLAG.ApoAI.V2.PglB could resemble about ~80% of activities compared to the WT PglB. In addition, FLAG.ApoAI.V1.PglB, FLAG.ApoAI.V3.PglB and

SUMO.ApoAI.V2.PglB showed strong activities in Figure 18 as well. While the MBP fused SIMPLEx-PglB showed low performance in general, the MBP.ApoAI.V3.PglB worked best among the other MBP constructs.

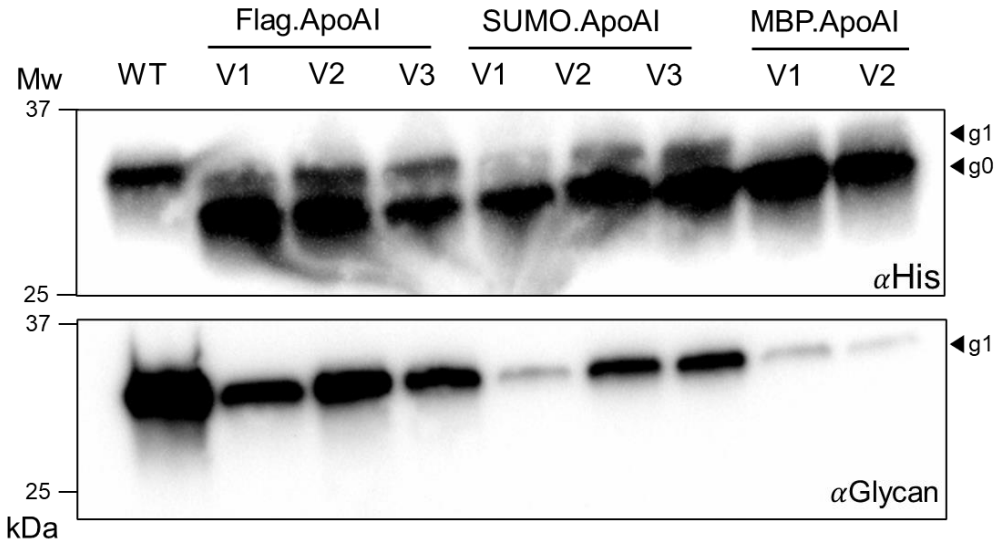

Figure 17. IVG blot with purified OSTs with 10μg of each.

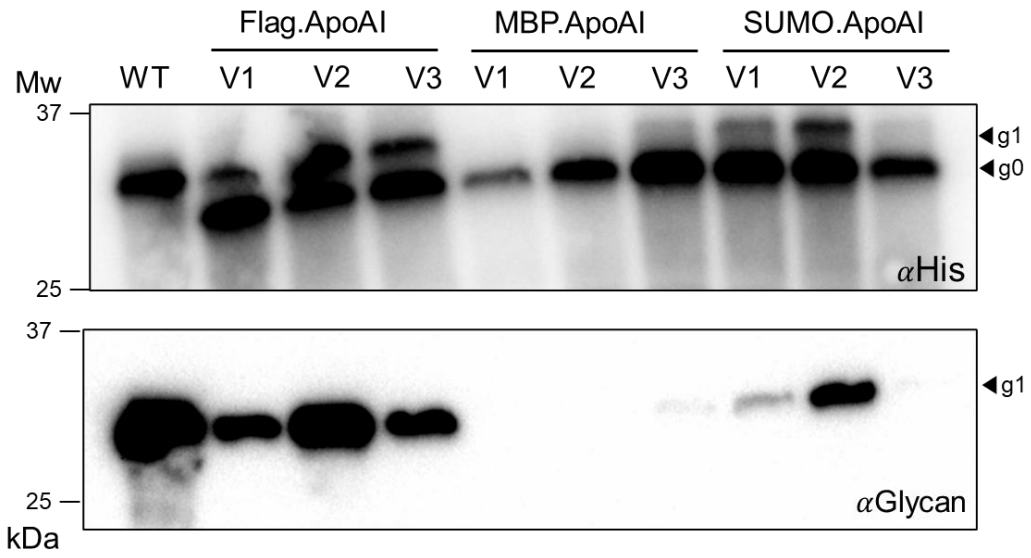

Figure 18. IVG blot with purified OSTs with 30μg of each.

3.2.4 Optimization of cell-free glycosylation reaction

In order to optimize the glycosylation efficiency, all the reaction

components such as LLOs glycan donor and acceptor proteins were freshly prepared. The four constructs used in this blot were FLAG.ApoAI.V2, FLAG.ApoAI.V3, SUMO.ApoAI.V2, and MBP.ApoAI.V3.

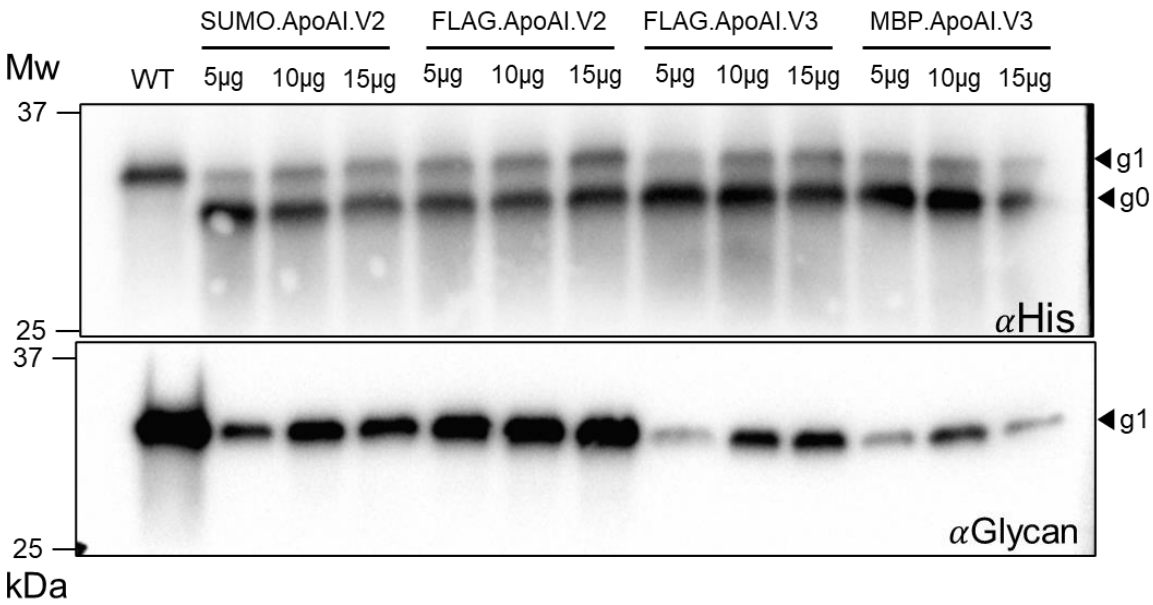

Figure 19. Optimized IVG blot with fresh ingredients

The result is in consensus with previous data that FLAG.ApoAI.V2 could resemble the highest glycosylation activity. Therefore, the FLAG.ApoAI.V2 was selected as a best construct for further functional analysis.

3.3 SIMPLEx-OST modifies human protein

The Figure 20 shows a result for IVG with MBP-fused human growth hormone (hGh) as an acceptor protein. MBP-fused hGh protein has molecular mass ~64 kDa and has DQNAT tag at an internal site within the hGh amino acid sequence. Unlike scFv13-R4^{DQNAT} which contains artificial glycosylation tag at the C-terminus, the glycosylation site of the hGh is derived from a natural sequon within the hGh sequence. The amino acid sequence around this glycosylation site has been

minimally mutated to the acceptable sequeon for PglB.

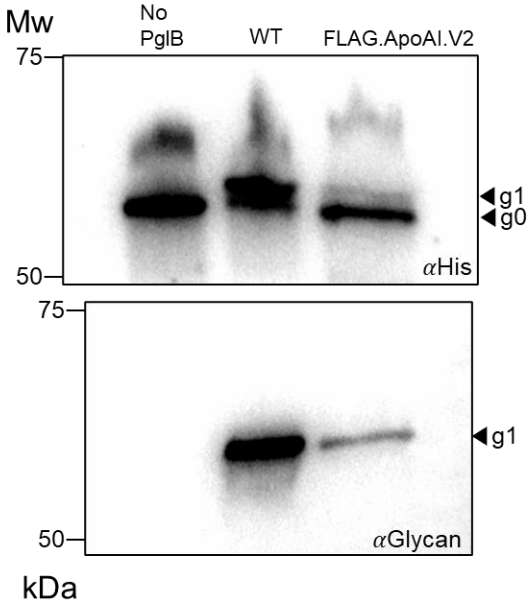

Figure 20. IVG blot with glycosylation on human growth hormone

Since the protein has MBP decoy fused, the mass of the product is located in a higher MW region which makes it harder to resolve the glycoprotein band from the aglycosylated one, but the anti-Glycan blot confirms the presence of glycan on hGh acceptor protein.

4. Discussion

4.1 Stability of SIMPLEx-PglB in detergent-free environment

Traditionally, SIMPLEx fused proteins have ApoAI shield located on the C-terminus of the target protein. The previous efforts from our group have shown that having ApoAI shield at the C-terminus of PglB is not sufficient to solubilize this membrane protein. This is because the transmembrane domain, of PglB is in the N-terminus followed by a large periplasmic, water soluble domain at the C-terminus. We hypothesize this periplasmic domain of the PglB will obstruct the interaction between ApoAI and PglB's transmembrane portion, which is necessary for rendering PglB water-soluble. We then redesigned the construct by fusing ApoAI at the N-terminus and found that such design significantly improved water solubility of the PglB (Figure 11). (77) In addition, the extended ApoAI chain can cover the massive size of PglB structure which also enhance the stability of SIMPLEx-PglB.

The exact folding mechanism of MBP is currently inexplicable, but a study showed that MBP is one of the most effective solubilizing agents. (78) MBP protein is known for preventing the aggregation pathway in protein folding and this could explain the highest solubility level achieved from MBP decoy. (Figure 12) The SUMO pathway is more conserved in eukaryotes and almost absent in prokaryotes. (79) Fusing SUMO into recombinant proteins in *E. coli* dramatically improve the protein folding and solubility for some polypeptides, SUMO decoy could outperform than MBP decoy. (80) However, in this study, MBP worked as better solubilized enhancer. FLAG decoy is short hydrophilic eight amino-acid peptides which rarely affect the solubility of the fused protein. (Figure 13) This explain the lowest solubility level achieved with FLAG fused SIMPLEx variants. Taken these together, MBP was the best enhancer followed by SUMO and FLAG.

4.2 Functional assays of the SIMPLEx-OST

The SIMPLEx-PglB in the crude lysates showed activities by synthesizing glycoproteins in an *in vitro* glycosylation assay. Since the contamination of membrane vesicles within cell lysate is common, the activity from the lysate generated from *E. coli* overexpressing membrane-bound, wild-type PglB was also observed. (Figure 15) The blot nevertheless was a useful guidance to find the constructs with high glycosylation activity. The IVG blots show consensus that FLAG.ApoAI.V2.PglB was the best construct. The overall results from purified OST IVG (Figure 16 and 17) show that protein activities and solubility are not directly correlated and hence high solubility does not always define high glycosylation efficiency.

The possible rational for lacking activities from MBP fused PglB could be explained by the massive molecular mass of the variants. All three MBP fused SIMPLEx-PglB are found between 125 kDa and 150 kDa, which could hinder the catalytic activities of the enzyme by sterically hindering enzyme-substrate interaction. Cleaving MBP decoy could perhaps increase the performance. Additionally, different reaction environments could improve the activities of PglBs. We could design a *in vivo* glycosylation pathway where all the reactions happen in the cytoplasm of *E. coli*. This provides native environments for PglB to carry out glycosylation reaction. Other improvements can be made with using outer membrane vesicle bilayer platform, which mimics the environments of gram-negative bacteria.

The length of ApoAI chain seem to affect the PglB activity that ApoAI.V2 showed higher activity for FLAG and SUMO fused-PglB. This did not apply to MBP fused-PglB as the ApoAI.V3 showed highest glycosylation activity. The reason that

ApoAI.V2 outperformed ApoAI.V3 could be that the longer ApoAI shield in V3 may interfere with the catalytic active site of the PglB, hindering the glycosylation reaction. However, this was not true for MBP.ApoAI.V3.PglB that further structural studies are required. To mathematically interpret the glycosylation efficiency, the kinetic assays could also be performed. While the current kinetic assays may not reveal the full data, preliminary numerical study can add the credibility of the data. One potential approach is the use of *in vitro* glycosylation assay with fluorescently labelled DQNAT peptides. By measuring the concentrations of substrates and products as reaction progresses , kinetic rates that describe the activity of PglB could be found and used as preliminary evidence for finding best constructs.

4.3 Glycosylation of human proteins using soluble OST

The acceptor protein scFv13-R4^{DQNAT} is engineered protein with C-terminus DQNAT. While this protein was useful for analyzing the glycosylation abilities of the constructs, the native N-glycosylated acceptor protein will be more attractive when it comes to therapeutic purposes or vaccine developments. Human growth hormone, used for children with growth hormone deficiency, contains native glycosylation site within the internal region, making glycosylation more complicated. (81) Therefore, our finding that SIMPLEx-PglB can glycosylate human growth hormone was a significant achievement. While the glycan used for this project is from *C. jejuni*, eukaryotic glycans will be introduced to improve the stability or half-life of the protein and this will be the next part of the experiment.

5. Future perspective

5.1 Expand the repertoire of water-soluble OSTs from diverse organism

While *C jejuni* PglB is the most studied bacterial OST, several studies have identified other bacterial OST orthologs. These OST orthologs display distinct glycosylation sequon specificity, which can be exploited to modify therapeutic proteins at their native sequon beyond DQNAT. Furthermore, several single cell eukaryotes such as *Trypanosoma* and *Leishmania* species contain one or more single subunit OSTs in their genome, and these represent another interesting group of OST for engineering. As an example, a preliminary data showed evidence that *Desulfovibrio marinus* (*D. marinus*) OST can mimic the glycosylation activity of *C jejuni* PglB. The other OSTs will be fused with SIMPLEx for further molecular level studies.

5.2 Investigating glycan structures that can be recognized by SIMPLEx-OSTs

The biochemical mechanism for bacterial OST is still unclear that reason for the different glycan specificities are hard to explain. Therefore discovering the glycan structure for different OSTs will be a useful guidance for antibody or vaccine developments. Eventually, transferring complex eukaryotic glycan structures such as sialic acid or fucose attached glycans could be the ultimate goal for generating a glycan library.

5.3 Perform biochemical and biophysical characterization of engineered OST enzymes

Solubilization of a large, multipass transmembrane protein while retaining its biological activity is challenging. It is intriguing to uncover the roles that each component in the SIMPLEx fusion plays in order to achieve this feat. To this end, I

plan to measure kinetic profiles of the SIMPLEx-OST in compared to its WT counterpart. Substrates including several glycosylation sequons and LLOs will be tested in an *in vitro* assay to identify any changes in substrate specificity. Finally, various biophysical measurements including dynamic light scattering (DLS) and small-angle X-ray scattering (SAXS) will be used to measure particle size and envelop of the molecule, respectively.

5.4 Structural elucidation of SIMPLEx-OST

Cryo-EM structural elucidation will be a useful tool for studying the structure of SIMPLEx-OST. This will confirm the hypothesis that ApoAI forms disk shape around the transmembrane domain of OSTs and retain the structural conformation.

6. Conclusion

In this study, a novel protein engineering strategy to solubilize bacterial OST has been investigated. The SIMPLEx technique allows *in vivo* expression of bacterial OSTs while maintaining their structural conformation in water without the need of detergent. The N-terminal truncated ApoAI enhances the solubility of variants by properly shield the hydrophobic transmembrane domain. The length of ApoAI proteins affects the expression level of variants that shorter version of shields tends to show higher solubility. Among the three protein decoys, MBP is the most effective in term of solubilization of the PglB constructs followed by SUMO and FLAG epitopes. While the exact mechanism of how those protein impact the solubility is still under investigation, MBP and SUMO worked as great solubility enhancer. The *in vitro* functional assay showed FLAG.ApoAI.V2.PglB with highest N-glycosylation activity. MBP-fused SIMPLEx-PglBs in contrast showed generally low glycosylation ability. Taken the best working construct, FLAG.ApoAI.V2.PglB demonstrated its N-glycan transfer ability to human growth hormone, providing potential application in biomanufacturing of glycoprotein therapeutics.

For the next part of the project, the FLAG.ApoAI.V2.PglB will be further analyzed for its glycan and acceptor peptide specificities. Further a collection of water-soluble OSTs using single subunit OST from diverse organism, including *Campylobacter* and *Desulfovibrio* species as well as protists, will be created. Lastly, the structure elucidation of SIMPLEx-OST with Cryo-EM technique is anticipated to reveal the exact mechanism of how SIMPLEx fusion, specifically ApoAI, renders OST water-soluble, providing opportunities for further optimization of the next generation engineering. To sum up, the OST-engineering strategy will shed light on the mechanism of diverse glycosylation pathways and enable new applications in

glycomedicines and glycomaterials.

7. References

1. Neelamegham, S., et al. (2019). "Updates to the Symbol Nomenclature for Glycans guidelines." Glycobiology **29**(9): 620-624.

2. Calo, D., et al. (2010). "Protein glycosylation in Archaea: Sweet and extreme." Glycobiology **20**: 1065-1076.

3. Rademacher, T. W., et al. (1988). "GLYCOBIOLOGY." Annual Review of Biochemistry **57**: 785-838.

4. Maverakis E, Kim K, Shimoda M, Gershwin M, Patel F, Wilken R, Raychaudhuri S, Ruhaak LR, Lebrilla CB (2015). "Glycans in the immune system and The Altered Glycan Theory of Autoimmunity" J Autoimmun. **57**: 1–13.

5. Helenius, A. and Aebi, M. (2001). "Intracellular functions of N-linked glycans". Science **291**: 2364–2369.

6. Sinclair, A. M. and Elliott, S. (2005). "Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins". J. Pharm. Sci. **94**: 1626-1635.

7. Imperiali, B. & O'Connor, S. E. (1999). "Effect of N-linked glycosylation on glycopeptide and glycoprotein structure". Curr. Opin. Chem. Biol. **3**: 643–649.

8. Matsumoto, S. et al. (2013) "Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation". Proc. Natl. Acad. Sci. USA **110**): 17868–73.

9. Cherepanova, N., Shrimal, S. & Gilmore, R. (2016). "N-linked glycosylation and homeostasis of the endoplasmic reticulum". Curr. Opin. Cell Biol. **41**: 57–65.

10. Helenius, A. & Aebi, M. (2004). "Roles of N-linked glycans in the endoplasmic reticulum". Annu. Rev. Biochem. **73**: 1019–49.

11. F. Schwarz, M. Aebi. (2011). "Mechanisms and principles of N-linked protein glycosylation**".** Curr. Opin. Struct. Biol. **21**: 576-582.

12. M. Aebi, R. Bernasconi, S. Clerc, M. Molinari. (2010). "N-glycan structures: recognition and processing in the ER"**.** Trends Biochem. Sci. **35**: 74-82.

13. M. Welti. (2012). "Regulation of dolichol-linked glycosylation". Glycoconj. J. **30**: 51-56.

14. C.B. Hirschberg, P.W. Robbins, C. Abeijon. (1998). "Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus". Annu. Rev. Biochem. **67**: 49-69.

15. M.D. Snider, O.C. Rogers. (1984). “Transmembrane movement of oligosaccharide-lipids during glycoprotein synthesis”. Cell. **36**: 753-761.

16. Kelleher, D. J. & Gilmore, R. (2006). “An evolving view of the eukaryotic oligosaccharyltransferase”. Glycobiology **16**: 47–62.

17. Wacker, M., et al. (2006). "Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems." Proceedings of the National Academy of Sciences **103**: 7088-7093.

18. Ramírez, A. S. *et al*. (2017). “Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs.” Glycobiology **27**: 525–535

19. Ramírez, A. S., et al. (2019). "Cryo–electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B." Science **366**: 1372-1375.

20. Matsumoto, S., et al. (2013). "Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation." Proceedings of the National Academy of Sciences **110**(44): 17868-17873.

21. van Erp, E. A., et al. (2019). "Fc-Mediated Antibody Effector Functions During Respiratory Syncytial Virus Infection and Disease." Frontiers in Immunology **10**(548).

22. Fang J., Richardson J., Du Z., Zhang Z. (2016). “Effect of Fc-glycan structure on the conformational stability of IgG revealed by hydrogen/deuterium exchange and limited proteolysis.” Biochemistry. **55**:860–868.

23. Kohler G., Milstein C. (1975). “Continuous cultures of fused cells secreting antibody of predefined specificity.” Nature. **256**: 495–497.

24. Li, W., et al. (2017). "Crystallizable Fragment Glycoengineering for Therapeutic Antibodies Development." Frontiers in Immunology **8**(1554).

25. Mimura Y., Katoh T., Saldova R., O’Flaherty R., Izumi T., Mimura-Kimura Y., Utsunomiya T., Mizukami Y., Yamamoto K., Matsumoto T., et al. (2018). “Glycosylation engineering of therapeutic IgG antibodies: Challenges for the safety, functionality and efficacy.” Protein Cell. **9**: 47–62.

25. Cymer F., Beck H., Rohde A., Reusch D. (2018). “Therapeutic monoclonal antibody N-glycosylation–structure, function and therapeutic potential.” Biologicals. **52**: 1–11.

26. Higel F., Seidl A., Sörgel F., Friess W. (2016). “N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal

antibodies and Fc fusion proteins." European Journal of Pharmaceutics and Biopharmaceutics. **100**: 94–100.

27. Larsen, M. D., et al. (2021). "Afucosylated IgG characterizes enveloped viral responses and correlates with COVID-19 severity." Science **371**(6532): eabc8378.

28. C. Huang, Y. Wang, et al, (2020). "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China." Lancet **395**: 497–506

29. C. Ceraolo, F. M. Giorgi, (2020). "Genomic variance of the 2019-nCoV coronavirus." Journal of Medical Virology. **92**: 522–528

30. Nothaft, H. & Szymanski, C. M. (2013). "Bacterial protein n-glycosylation: New perspectives and applications." J. Biol. Chem. **288:** 6912–6920.

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

32. Weerapana, E. & Imperiali, B. (2006). "Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems." Glycobiology **16**: 91–101.

33. Jervis, A. J. et al. (2012). "Characterization of the structurally diverse N-linked glycans of *Campylobacter* species." Journal of Bacteriology. **194**: 2355–2362.

34. Nothaft, H. and Szymanski, C. M. (2010). "Protein glycosylation in bacteria: sweeter than ever." Nature Reviews Microbiology. **8**: 765–778

35. Nothaft, H., et al. (2009). "Study of free oligosaccharides derived from the bacterial N-glycosylation pathway." Proceedings of the National Academy of Sciences of the United States of America **106**: 15019–15024

36. Valderrama-Rincon, J., Fisher, A., Merritt, J. *et al.* (2012). "An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*." Nature Chemical Biology **8:** 434–436.

37. Glasscock, C.J. et al (2018). "A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation". *Metabolic engineering,* **47**: 488-495 .

38. Larkin A, Imperiali B. (2011). "The expanding horizons of asparagine-linked glycosylation." Biochemistry. **50**: 4411-26.

39. Schäffer, C. and P. Messner (2017). "Emerging facets of prokaryotic glycosylation." FEMS Microbiology Reviews **41**(1): 49-91.

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

41.Guarino, C. and DeLisa, M. P. (2012). “A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins.” Glycobiology **22**: 596–601.

42. Lizak, C., Gerber, S., Numao, S. *et al.* (2011). *“* X-ray structure of a bacterial oligosaccharyltransferase.” Nature **474:** 350–355.

43. Ünligil, U. M. and Rini, J. M. (2000). “Glycosyltransferase structure and mechanism.” Current Opinion in Structural Biology. **10**: 510–517.

44. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P. and Walker, S. (2011). “Structure of human O-GlcNAc transferase and its complex with a peptide substrate.” Nature **469:** 564–567.

45. Silberstein, S., Gilmore, R. (1996) “Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase”. FASEB Journal. **10**: 849–858.

46. Zufferey, R. et al. (1995). “Stt3, a highly conserved protein required for yeast oligosaccharyl transferase activity *in vivo*.” EMBO Journal. **14**: 4949–4960.

47. Schulz, B. L. et al. (2009). “Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency.” Proceedings of the National Academy of Sciences of the United States of America. **106**: 11061–11066.

48. Barre, Y. *et al*. (2017). “A conserved DGGK motif is essential for the function of the PglB oligosaccharyltransferase from *Campylobacter jejuni*.” Glycobiology. **27**: 978–989.

49. Ihssen, J., et al. (2015). "Increased efficiency of *Campylobacter jejuni N*-oligosaccharyltransferase PglB by structure-guided engineering." Open Biology **5**(4): 140227.

50. Glover, K. J., Weerapana, E., Numao, S. & Imperiali, B. (2005). “Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*.” Cell Chemical Biology. **12**: 1311–1315

51. A. Dell, A. Galadari, F. Sastre, P. Hitchen. (2010). “Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes” Journal of Microbiology. 148178

52. Wacker, M., et al. (2002). "N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into *E. coli*.’ Science **298**(5599): 1790-1793.

53. Merritt, J. H., et al. (2013). "Glycans-by-design: Engineering bacteria for the biosynthesis of complex glycans and glycoconjugates." Biotechnology and Bioengineering **110**(6): 1550-1564.

54. Baker JL, Çelik E, DeLisa MP. (2013). “Expanding the glycoengineering toolbox:

the rise of bacterial N-linked protein glycosylation." Trends in Biotechnology. **31**(5):313-323.

55. Fisher, A. C., et al. (2011). "Production of Secretory and Extracellular N-Linked Glycoproteins in *Escherichia coli*." Applied and Environmental Microbiology **77**(3): 871-881.

56. Feldman, M. F., et al. (2005). "Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*." Proceedings of the National Academy of Sciences of the United States of America **102**(8): 3016-3021.

57. Pandhal, J., et al. (2012). "Systematic metabolic engineering for improvement of glycosylation efficiency in Escherichia coli." Biochemical and biophysical research communications **419**(3): 472-476.

58. Martin, R. W., et al. (2018). "Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids." Nature Communications **9**(1): 1203.

59. Swartz, J. (2006). "Developing cell-free biology for industrial applications." Journal of Industrial Microbiology and Biotechnology **33**(7): 476-485.

60. Carlson, E. D., et al. (2012). "Cell-free protein synthesis: Applications come of age." Biotechnology Advances **30**(5): 1185-1194.

61. Oza, J. P., et al. (2015). "Robust production of recombinant phosphoproteins using cell-free protein synthesis." Nature Communications **6**(1): 8168.

62. Martin, R. W., et al. (2018). "Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids." Nature Communications **9**(1): 1203.

63. Karim, A. S. and M. C. Jewett (2016). "A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery." Metabolic Engineering **36**: 116-126.

64. Jewett, M. C., et al. (2008). "An integrated cell-free metabolic platform for protein production and synthetic biology." Molecular Systems Biology **4**(1): 220.

65. Kaiser, L., et al. (2008). "Efficient cell-free production of olfactory receptors: Detergent optimization, structure, and ligand binding analyses." Proceedings of the National Academy of Sciences **105**(41): 15726-15731.

66. Dudley, Q. M., et al. (2016). "Cell-Free Mixing of *Escherichia coli* Crude Extracts to Prototype and Rationally Engineer High-Titer Mevalonate Synthesis." ACS Synthetic Biology **5**(12): 1578-1588.

67. Jaroentomeechai, T., et al. (2018). "Single-pot glycoprotein biosynthesis using

a cell-free transcription-translation system enriched with glycosylation machinery." Nature Communications **9**(1): 2686.

68. Burgenson, D., et al. (2018). "Rapid recombinant protein expression in cell-free extracts from human blood." Scientific Reports **8**(1): 9569.

69. Wallin, E. and G. V. Heijne (1998). "Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms." Protein Science **7**(4): 1029-1038.

70. Wagner, S., et al. (2006). "Rationalizing membrane protein overexpression." trends in biotechnology. **24**: 364–371.

71. Tate, C. G. (2001). "Overexpression of mammalian integral membrane proteins for structural studies." FEBS Letters **504**(3): 94-98.

72. Scott, D. J., et al. (2013). "Stabilizing membrane proteins through protein engineering." Current Opinion in Chemical Biology **17**(3): 427-435.

73. Zhang, Q., et al. (2011). "New amphiphiles for membrane protein structural biology." Methods **55**(4): 318-323.

74. Gursky, O. and D. Atkinson (1996). "Thermal unfolding of human high-density apolipoprotein A-1: implications for a lipid-free molten globular state." Proceedings of the National Academy of Sciences **93**(7): 2991-2995.

75. Mizrachi, D., et al. (2015). "Making water-soluble integral membrane proteins in vivo using an amphipathic protein fusion strategy." Nature Communications **6**(1): 6826.

76. Mizrachi, D., et al. (2017). "A water-soluble DsbB variant that catalyzes disulfide-bond formation in vivo." Nature Chemical Biology **13**(9): 1022-1028.

77. Ollis, A. A., et al. (2014). "Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity." Nature Chemical Biology **10**(10): 816-822.

78. Costa, S., et al. (2014). "Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system." Frontiers in Microbiology **5**(63).

79. Kapust, R. B. and D. S. Waugh (1999). "Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused." Protein Science **8**(8): 1668-1674.

80. Marblestone, J. G., et al. (2006). "Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO." Protein Science **15**(1): 182-189.

81. Flintegaard, T. V., et al. (2010). "N-Glycosylation Increases the Circulatory

Half-Life of Human Growth Hormone." Endocrinology **151**(11): 5326-5336.