INVESTIGATING OLIGOSACCHARYLTRANSFERASES OF *N*-LINKED GLYCOSYLATION USING *ESCHERICHIA COLI*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Cassandra M B Guarino

January 2013

INVESTIGATING OLIGOSACCHARYLTRANSFERASES OF *N*-LINKED GLYCOSYLATION USING *ESCHERICHIA COLI*

Cassandra M B Guarino, Ph. D.

Cornell University 2013

Escherichia coli is a powerful tool for elucidating many of the basic principles of biology. As a protein production host, *E. coli* can produce exogenous protein to upwards of 20% of the total cellular protein content. While *E. coli* is used to produce a wide variety of proteins for research and therapeutic purposes, this organism is limited in its capability to perform various post-translational modifications required for the proper function of most mammalian proteins. One such modification is *N*-linked glycosylation, the transfer of an oligosaccharide onto an asparagine residue within an acceptor peptide sequence. *N*-linked glycosylation can alter the solubility of a protein, enhance its effector function, and increase the serum half-life of therapeutic proteins. In recent years, researchers have reconstituted the *N*-linked glycosylation pathway from a distantly related gram-negative pathogen, *Campylobacter jejuni*, in non-pathogenic *E. coli*. While engineered strains of *E. coli* are now capable of *N*-linked glycosylation, the type of glycan transferred, the acceptor protein sequence modified, and the efficiency of protein modification remain limiting factors in these recombinant systems. To address these limitations, this work focused on understanding and improving the oligosaccharyltransferase (OST), the central enzyme of the *N*-linked glycosylation pathway. First, selection tools were designed to isolate

E. coli with enhanced glycosylation capabilities. One of these tools, the glycophage display system, was utilized to select for OSTs with altered acceptor site specificity. Next, we performed a functional analysis of twenty-three different bacterial OSTs. Harnessing the functional diversity in naturally occurring bacterial OSTs combined with structure-guided mutations yielded several interesting OSTs with unique acceptor site specificities not previously reported. Finally, to further simplify the model framework of glycosylation, we designed an *in vitro* prokaryotic-based cell-free system to efficiently synthesize glycoproteins. *E. coli* has proven to be a useful tool for developing a more thorough understanding of *N*-linked glycosylation, and continued research in the field of bacterial glycosylation will undoubtedly lead to new and important discoveries with significant impact on the biopharmaceutical industry.

BIOGRAPHICAL SKETCH

Cassandra Guarino was born in Plainview, New York on October 19th, 1982 and grew-up in Brightwaters, NY. She graduated from the Academy of St. Joseph, a small all-girls high school in Brentwood, NY, in 2000. In August of 2000, she left Long Island, with her horse, to begin her career as a student at Cornell University, majoring in Animal Science.

At the beginning of her undergraduate career, she was required to take a math course and decided to challenge herself with Engineering Mathematics, which led to the discovery of her passion for Engineering. Numerous engineering courses later, she decided to add a second major in 'Biological Engineering Technology', which appears on the transcript as 'Agricultural Technology.' In 2004, she received her bachelor's degree, magna cum laude, from Cornell University.

Following graduation, she purchased a farm in Slaterville Springs with her family, acquired a young horse, and accepted a job working for the Animal Health Diagnostic Center at Cornell University as a Medical Technician, where she worked on serological virology for two years before beginning her PhD in Comparative Biomedical Sciences. Following several laboratory rotations, she found herself at home in the Chemical Engineering lab of Matthew P. DeLisa, aka. DLRG, where they were just beginning to study *N*-linked glycosylation in *Escherichia coli*.

In 2010 and again in 2011, she received the NSF GK-12 fellowship, funding two years of her research and re-igniting her passion for teaching.

Outside of lab, her favorite activities include travelling through the woods on horseback, growing her own food, playing music with friends, creating useful objects out of yarn, as well as spending many hours teaching and inspiring middle and high school students to develop an excitement for Science, Technology, Engineering and Mathematics.

This work is dedicated to my Mom,

whose enthusiasm for knowledge-seeking and continuous perseverance

rubbed-off, like wet paint.

ACKNOWLEDGMENTS

I would like to acknowledge the many individuals who have made this experience worthwhile. First, I would like to thank Dr. Thomas Mansell, who took the time to train me on the basic techniques used in the DLRG and always made time to discuss ideas and offer solutions. I would also like to thank all of the glycopals, past and present, for their collaborations and productive discussions, including Dr. Adam Fisher, Dr. Eda Çelik, Dr. Anne Ollis, Dr. Jenny Baker, Linxiao Chen, Christine Endicott-Warner, Sean O'brian, Alyse Portnoff, Taylor Stevenson, and Jason Book (honorary member).

I would also like to acknowledge the NSF GK-12 fellowship program (DGE 0841291) and all those who made that experience possible, including Dr. Shivaun Archer, Dr. Christopher Schaffer, Dr. Michael Shuler, Kevin Dilley, Nev Singhota, and particularly my teacher partners Ron Reed, and Michelle Kornreich.

Additionally, I would like to thank Dr. Chris Martin, who joined somewhat late in the game, but who has been an amazing asset to help get to the finish.

Finally, I would like to express my sincere appreciation for each of my committee members, including Dr. David Putnam, Dr. Susan Quirk, and Dr. Robert Oswald for their time and efforts, and especially Dr. Matthew P. DeLisa for his continued guidance, assistance and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

Amp – ampicillin

Cm – chloramphenicol

FACS – fluorescence activated cell sorting

GT – GlycTag

IgG – immunoglobulin G

Kan – kanamycin

LLO – lipid-linked oligosacchride

LPS – lipopolysaccharide

OMV – outer membrane vesicle

OST – oligosaccharyltransferase

PEG – polyethylene glycol

RBS – ribosome-binding site

SBA – soybean agglutinin

SBA-488 – SBA-AlexaFluor488

Sp - spectinomycin

Tat – Twin-arginine translocation

TEV – tobacco etch virus

UndPP – undecaprenyl pyrophosphate

wt – wild-type

CHAPTER 1
INTRODUCTION

Background

The term 'protein' was first coined in 1838 by Jöns Jacob Berzelius, one of the founders of modern chemistry, to describe one class of the constituents of living things and to express the fundamental significance of certain substances in nutrition[1,2]. Today, many laymen still think of 'protein' as simply a substance of nutrition, but proteins are so much more than merely something we have to eat to survive. Proteins are miraculous nano-machines that perform various tasks, from catalysis of chemical reactions to structural recognition. Certainly, protein is an essential component of the diets of all animals, but the primary reason for this necessary class of molecules is to obtain the amino acid building blocks required for creating novel proteins in the host's body.

Recombinant DNA technology has allowed for closer investigation into protein structure and function. Through the use of molecular cloning techniques, foreign DNA can be introduced into an organism, and with the appropriate genetic material to code for replication and transcription, non-native and engineered proteins can be produced in a myriad of organisms. The first publication describing the recombinant DNA process for creating plasmids which could be replicated in *Escherichia coli* was published in 1973[3]. That work, led by Herbert W. Boyer and Stanley N. Cohen, produced one of the first major patents in biotechnology[4]. Nearly a decade after the initial proof of concept, human insulin became the first commercial protein produced by recombinant DNA

technology to be approved for use as a therapeutic in humans[5]. This version of insulin, known as Humulin®, and also referred to as biosynthetic human insulin (BHI), is produced using a fermentation process in *E. coli*[6].

E. coli is an excellent host for recombinant protein production. As of 2009, approximately 30% of all proteins produced for the biopharmaceutical market were made in *E. coli* [7]. *E. coli* remains an exceptionally popular and predominantly utilized bacterium for the production of recombinant proteins due to: (i) thorough characterization, both genetic and physiologic, (ii) well-established fermentation methodologies, (iii) rapid growth on inexpensive substrates, (iv) ease of modification through a broad range of simple molecular techniques, and most notably, (v) the capacity to accumulate foreign proteins to greater than 20% of the total cellular protein content[8,9]. While *E. coli* continues to prove itself to be remarkably useful for recombinant protein production, there are several drawbacks to this particular production host. The primary drawbacks have been the inability to produce proteins composed of multiple subunits or requiring substantial post-translational modifications, necessitating the use of more complex hosts such as yeast or mammalian cells[10]. Recent advancements in *E. coli* protein expression techniques have led to the ability to produce multiple-subunit molecules in *E. coli*, such as full-length immunoglobulin G molecules (IgGs)[11,12]. For full effector function, however, IgGs also require the post-translational modification known as *N*-linked glycoslation[13].

The focus of this thesis is on utilizing *E. coli* to better understand the process of *N*-linked glycosylation, with a primary focus on the oligosaccharyltransferase (OST). The OST is an enzyme that catalyzes the transfer of the glycan from a lipid carrier in the

membrane to the asparagine residue of a nascent peptide containing an *N*-linked glycosylation consensus sequence. For decades, *E. coli* has been utilized to elucidate many of the basic principles of biology, including the genetic code[14], gene regulation[15], recombinant DNA technology[3], and protein production and secretion[16]. Because *E. coli* does not contain native *N*-linked glycosylation machinery and yet can be engineered to perform this post-translational modification[17], it is an excellent host for the study of *N*-linked glycosylation.

Protein glycosylation in bacteria

Glycoproteins are ubiquitous in nature, and their existence has been known for over a century[18]. Protein glycosylation is a post-translational modification that involves the covalent linkage of a sugar molecule to a protein. There are three main types of protein glycosylation events that are known to occur: (i) *N*-linked glycosylation of an aspargine residue, (ii) *O*-linked glycosylation of a serine or threonine residue, and (iii) glycosylphosphatidyl inositol derivitization of the carboxy-terminal carboxyl group[19]. It has been shown that *N*-linked glycosylation has profound effects on protein structure and function[19,20], and this type of glycosylation is the focus of this thesis.

N-linked glycosylation is necessary for the viability and development of eukaryotic organisms[21]. In eukaryotes, *N*-linked glycosylation involves the co-translational transfer of a pre-assembled tetradecasaccharide, $Glc_3Man_9GlcNAc_2$-, from a dolichol pyrophosphate lipid carrier to the asparagine side chain of a nascent protein

within the consensus sequence, N-X-S/T, where X can be any amino acid except proline[22]. Recently, it has been shown that the core of this glycan, $Man_3GlcNAc$-, can be produced in engineered *E. coli*[23]. In eukaryotes, the initial glycan transfer occurs in the endoplasmic reticulum (ER), where the protein-bound carbohydrate can be trimmed and shuttled to the Golgi apparatus, at which point a variety of modifications take place, yielding heterogeneous glycoproteins[19]. For many years, the process of *N*-linked glycosylation was thought to be restricted to eukaryotes, *never* occurring in bacteria; however, it is now clear that we should 'never say never again'[24], as systems for *N*-linked glycosylation have been discovered in all domains of life[25–27].

The first evidence for protein glycosylation in non-eukaryotic cells was documented in archaea. *Halobacterium salinarium* were found to have glycosylated envelope proteins (e.g. surface layer proteins) based on a periodate-Schiff reagent reaction with proteins in an SDS-PAGE gel, which is an indication of carbohydrate presence. Covalent linkage of the carbohydrates to surface layer proteins was confirmed via tryptic and pronase digestion, suggesting the existence of both *O*- and *N*-linked glyans[28].

It was first suggested that bacteria may also be capable of *N*-linked protein glycosylation when the surface antigens of *Borrelia burgdorferia*, a helical shaped bacterium known as a spirochete and recognized as the causative agent of Lymes disease[29], were found to include carbohydrates[30]. It was later shown that these carbohydrates were *N*-linked to the surface lipoproteins OspA and OspB based on deglysoylation by PNGase F[31]. PNGaseF is known to specifically hydrolyze *N*-linked glycans at the β–aspartylglycosylamine bond, yielding ammonia, aspartate, and an

oligosaccharide with di-*N*-acetyl-chitobiose at the reducing end[32]. However, this data was recently refuted, suggesting that the glycoproteins in those experiments were contaminants from the growth media components[33].

A general method for protein glycosylation in bacteria was discovered in 1999, when a genetic locus from *Campylobacter jejuni* 81-176 was shown to be involved in the glycosylation of multiple proteins[34]. A few years later, it was determined that the locus responsible for general protein glycosylation in *C. jejuni* contains a protein (PglB) that is homologous to STT3, an essential component of the OST complex responsible for catalyzing *N*-linked glycan transfer in eukaryotic cells. This group further showed that the *C. jejuni* glycosylation locus could be functionally transferred to *E. coli* cells, giving these latter bacteria the ability to perform *N*-linked glycosylation. This was significant as *E. coli* normally lack a native *N*-linked glycosylation pathway[17]. Together, these studies provided the first conclusive evidence for the existence of *N*-linked glycosylation in the bacterial domain.

While not necessary for cell survival, the *N*-linked glycosylation machinery found in *C. jejuni* is essential for full competence of this pathogen[35]. Specifically, in the absence of the general protein glycosylation machinery, *C. jejuni* 81-176 has a reduced ability to adhere to intestinal cells and to colonize the intestinal tracts of mice[36]. Additionally, mutation of the *pglH* gene, which is responsible for the addition of the third GlcNAc residue of the *C. jejuni* glycan[37], causes *C. jejuni* to become deficient in its ability to adhere to and subsequently invade human epithelial cells *in vitro*[38].

Since the discovery that *C. jejuni* is capable of performing *N*-linked glycosylation, similar systems have been identified in numerous other mucosal pathogens[39]. The

primary similarity between the eukaryotic and bacterial *N*-linked glycosylation systems is the formation of a β-glycosylamide linkage by en-block transfer of a glycan from a lipid carrier in the membrane onto an asparagine amino acid within a specific consensus sequence[22]. Figure 1.1 depicts the overall similarities between these two systems. While there are many similarities, the glycan that is transferred, the lipid carrier upon which the glycan is assembled, and the consensus sequence that is recognized by the OST are different. In eukaryotes, the tetradecasaccharide $Glc_3Man_9GlcNAc_2$- is preferentially recognized by the OST[40], while in bacteria, a variety of different oligosaccharides can be recognized and transferred by the OST[23,41]. Dolichol pyrophosphate and undecaprenyl pyrophosphate (UndPP) are the lipid carriers upon which the glycan is assembled in eukaryotes and bacteria, respectively. The eukaryotic consensus sequence is generally accepted to be N-X-S/T, and in bacteria, specifically in *C. jejuni,* this sequence is extended to D/E-X_1-N-X_2-S/T, where X can be any amino acid besides proline[42].

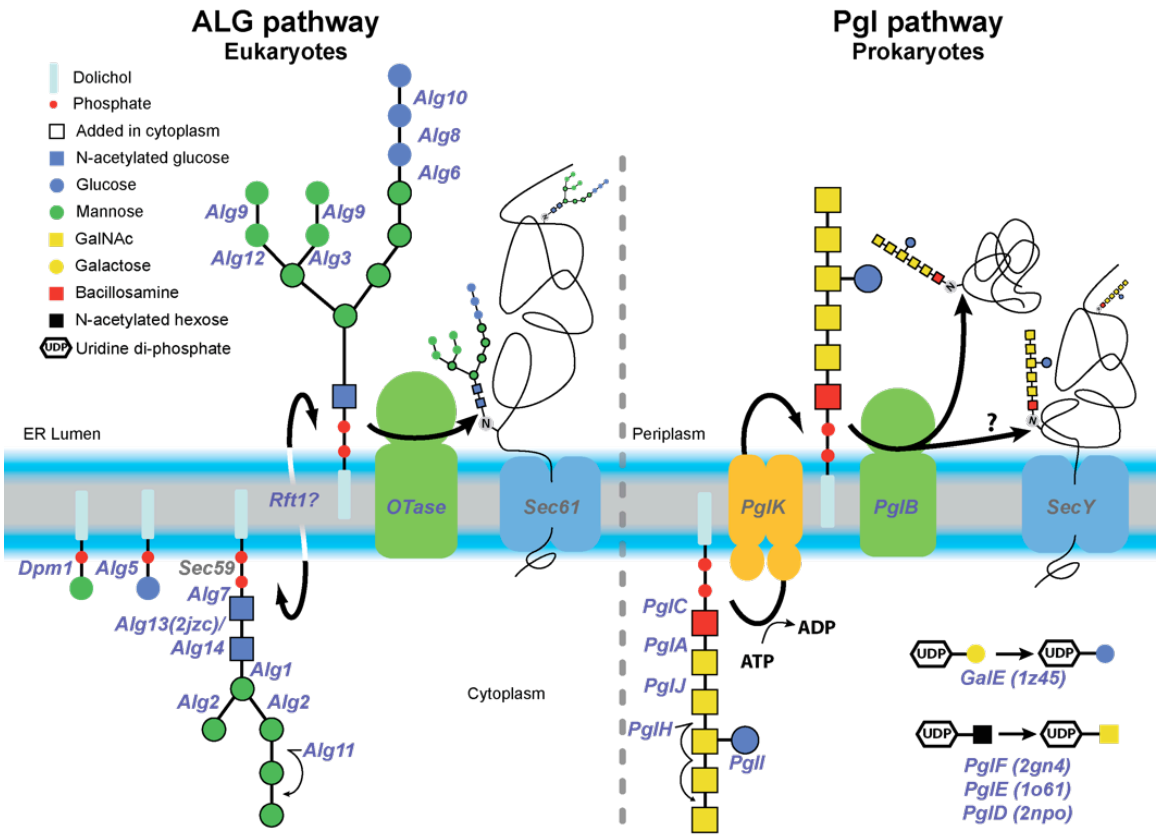


Figure 1.1. Eukaryotic vs. bacterial *N*-linked glycosylation. Comparison of *N*-linked glycosylation in prokaryotes (right) and eukaryotes (left). In both systems, several glycosyltransferases synthesize the glycan by sequential addition of nucleotide-activated sugars on a lipid carrier on the cytoplasmic face of the inner membrane. Once assembled, a flippase transfers the LLOs across the membrane where the OST catalyzes the transfer to Asn residues of periplasmic or ER substrate proteins. PglB is a single-subunit, integral membrane protein that is homologous to the catalytic subunit of the eukaryotic OST STT3 (note that PglB and STT3 complex are not drawn to scale). Whereas eukaryotes and archaea use an N-X-S/T acceptor sequence (where X is any amino acid but Pro), PglB requires an extended motif that includes an Asp or Glu residue in the -2 position (D/E-X-1-N- X+1-S/T, where X-1 and X+1 can be any amino acid except Pro). PglB can transfer sugars post-translationally to locally flexible structures in folded proteins. It remains to be determined whether PglB can transfer sugars co-translationally like eukaryotes. This figure was published[43] and is re-used here with permission.

The protein glycosylation operon from *C. jejuni* contains all the genes necessary to (i) build the *C. jejuni* heptasaccharide: GalNAc-1,4-GalNAc-1,4-(Glc1,3)-GalNAc-1,4-GalNAc-1,3-Bac-[44], where bacillosamine (Bac), 2,4-diacetamido-2,4,6-trideoxyglucopyranose, is a monosaccharide unique to bacteria[45], (ii) flip the glycan from

the cytoplasmic face to the periplasmic face of the inner membrane, and (iii) transfer the glycan onto an acceptor motif. *In vitro* assembly of the heptasaccharide using overexpressed, purified glycosyltransferases has led to the determination of the precise function of these enzymes[46,47]. Figure 1.2 shows a representation of the function of each of the enzymes in the *C. jejuni pgl* operon; not included in this figure are WlaA, a hypothetical protein of unknown function, WlaJ, a putative integral membrane protein of unknown function, GalE, an UDP-glucose-4-epimerase, which catalyzes the interconversion of UDP-galactose and UDP-glucose[48], and PglG, which does not play a role in glycan biosynthesis or transfer[37,49]**.** Native *E. coli* proteins can complement portions of the *C. jejuni N*-linked glycosylation pathway. For example, PglK, the *C. jejuni* flippase responsible for flipping the assembled glycan from the cytoplasmic face to the periplasmic face of the inner membrane, can be complemented by the native *E. coli* Wzx enzyme[50]. The *wzx* gene is thought to encode a flippase that participates in lipopolysaccharide (LPS) biosynthesis, specifically transferring UndPP-linked *O*-antigen units across the inner membrane[51]. Additionally, *O*-antigens, which compete for the UndPP lipid carrier, can be transferred in place of the *C. jejuni* glycan when the concentration of *O*-antigens on the inner membrane is increased[41,52]. Fortunately, for glycoprotein engineering purposes, *O*-antigen biosynthesis is not required for cell viability, as is demonstrated by the ability of cells to survive without *rfe* (also known as *wecA*), the gene responsible for the first step of *O*-antigen biosynthesis[53].

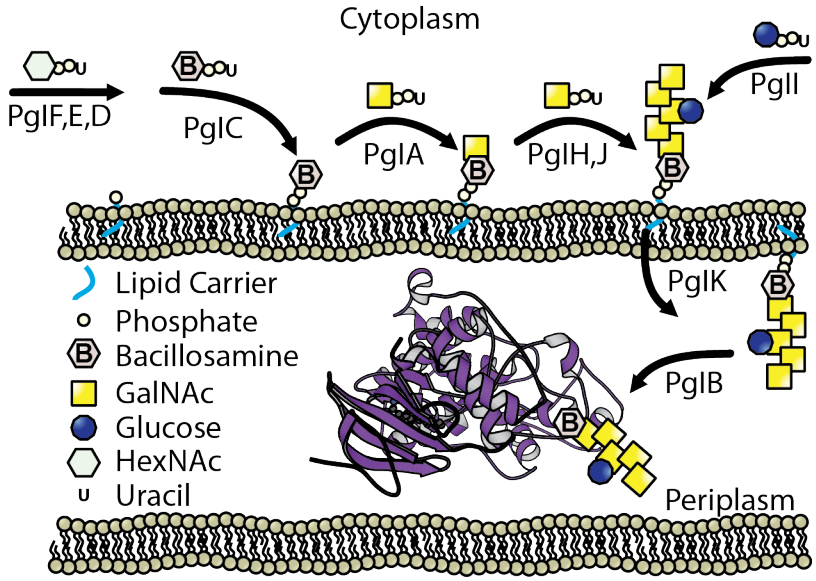


Figure 1.2. Enzymes in the *C. jejuni pgl* operon. This schematic shows the function of each of the enzymes in the *C. jejuni pgl* operon.

Oligosaccharyltransferases

The OST complex from *Saccharomyces cerevisiae* is composed of nine subunits: Wbp1, Swp1, Stt3, Ost1, Ost2, Ost3, Ost4, Ost5, and Ost6, where the first five are essential for the viability of cells and the last four are required for maximal OST activity[40]. Ost1, Swp1, Wbp1 and Ost2 are homologous to the mammalian OST subunits ribophorin I, ribophorin II, Ost48 and DAD1 respectively. Furthermore, homologues of Stt3 (Stt3A and Stt3B), and Ost3/Ost6 (N33 and IAP), and Ost4 are expressed in mammalian cells and assemble into multimeric complexes with the Ost1, Swp1, Wbp1 and Ost2 homologues to form OST complexes similar to those found in yeast[54]. Stt3 has been determined to be the catalytic subunit of the mammalian OST complex based on highly selective cross-linking to nascent polypeptide chains[55]. In particular, the five-

residue sequence, WWDYG, is believed to be the catalytic motif of the Stt3 subunit active site[56]. However, the Stt3 subunit alone is not capable of catalyzing glycan transfer[57].

A single subunit enzyme is ideal for bacterial expression and is the type of OST investigated in this thesis. Homologues of the yeast and mammalian OST have been identified that contain the WWDYG motif but do not form a multi-protein complex for function. For example, the OST from *C. jejuni,* PglB, is the only component necessary for glycan transfer[58]. Additionally, an Stt3 homolog from the archaean *Pyrococcus furiosus* is believed to function as a single subunit[59]. Recently, a handful of eukaryotic OSTs, derived from protozoans, have also been identified that can function independent of a multi-protein complex[40,60,61]. Although these protozoan Stt3 homologs are capable of replacing the entire OST complex in yeast, it has been proposed that in the absence of the complex, substrate specificity is diminished[60]. Diminished substrate specificity could lead to heterogeneous glycoprotein production, which is problematic when producing proteins for use as therapeutics and/or vaccines.

In both eukaryotes and bacteria, the peptide substrate for glycosylation must be translocated from the compartment in which it is translated (cytoplasm) to the compartment in which it is glycosylated (endoplasmic reticulum and periplasm, respectively). Several lines of evidence indicate that *N*-linked glycosylation is associated with the translocon [62,63]. It has been shown through chemical cross-linking studies and a split-ubiquitin yeast-two-hybrid system that the components of the OST complex of yeast interact with components of the Sec translocon[64]. The co-translational Sec translocon in yeast is a heterotrimeric complex known as the Sec61 complex, composed of Sec61p,

Sss1p, and Sbh1p, and a homologous translocation pathway exists in bacteria, known as SecYEG[65,66]. In addition to the SecYEG translocon, translocation of proteins to the bacterial periplasm can occur via the twin-arginine translocation pathway (Tat), which transfers fully folded proteins and protein complexes containing an appropriate signal peptide out of the cytoplasm[67,68]. While interaction between the SecYEG translocon and the bacterial OST have yet to be determined, it is clear that this interaction is not necessary for glycosylation, as fully folded proteins transported to the periplasm via the Tat translocation pathway are still able to be glycosylated[69,70].

Recently, the crystal structure of the *Campylobacter lari* OST in complex with a bound peptide substrate in the active site was solved[71]. This was the first full-length OST structure to be elucidated, including both the membrane-integrated domain and the membrane extrinsic domain. The previously solved structure of the C-terminal soluble domain of *C. jejuni* PglB[72] aligns well with the *C. lari* structure. However, the exclusion of the transmembrane helices in the *C. jejuni* structure caused a significant portion of the active site domain to be absent. With a three-dimensional structure now available to work from, structure-guided rational design and directed evolution can be implemented for the isolation of mutations that alter or improve bacterial OST function[73].

Significance

Glycoproteins in vaccine development

The surfaces of pathogenic bacteria are covered with a variety of unique carbohydrate structures. These carbohydrates are excellent candidates for vaccine development. It is well-established that carbohydrates alone will not produce protective immunity against a pathogen, however, conjugation of these glycan structures to proteins can offer varying degrees of immunity[74].

The earliest example of a carbohydrate-bound protein used for immunization is described by Avery and Goebel[75], who demonstrated the ability to produce glycan-specific antibodies in rabbits by vaccinating animals with specific carbohydrates covalently bound to proteins *in vitro*. Since that time, it has been clearly demonstrated that proteins conjugated to bacterial polysaccharide can offer immunity to infection. One noteworthy example is the *Haemophilus influenza* type b conjugate vaccine, which was first introduced in the late 1980s[76], with variations later approved[77]. *H. influenza* is now considered eradicated in some areas[78]. A comprehensive review on glycoconjugate vaccines was published in 2005 that provides further examples of glycoconjugate vaccines currently in use or in various stages of clinical trials[79].

Early glycoconjugate vaccines were produced by chemical linkage of the carbohydrate to a protein. The ability to attach specific carbohydrates to proteins *in vivo* opened a new pathway for glycoprotein vaccine production, which led to the formation of GlycoVaxyn, a privately held Swiss biopharmaceutical company founded in 2004. The

GlycoVaxyn technology is based on the observation that the *C. jejuni* OST is promiscuous. That is, in addition to transferring *N*-glycans, PglB is also capable of modifying target proteins with a variety of bacterial *O*-antigens when their concentration in the inner membrane is increased by preventing their export from the inner-membrane to the outer surface of *E. coli*[41]. The ability to produce glycoprotein conjugate vaccine candidates *in vivo* has since been demonstrated. For example, AcrA from *C. jejuni* or PAE from *Pseudomonas aeruginosa* can be glycosylated in *E. coli* cells co-expressing the *Shigella dysenteriae* type 1 O-antigen and PglB from *C. jejuni*[80]. More recently, it was found that by expressing a glycoprotein along with the *C. jejuni* OST directly in *Yersinia enterocolitica* O9 cells, a less pathogenic relative of the *Brucella abortus* pathogen, glycoproteins could be produced that elucidated immune reactivity to Brucella in mice, though protection was not achieved[81].

In addition to attachment of glycans to soluble proteins, outer membrane vesicles (OMVs) containing glycoproteins could be utilized as vaccines. OMVs are non-replicating membrane-encapsulated structures released from some gram-negative bacteria. They are composed of outer membrane lipids, outer membrane proteins, LPS, and contain periplasmic components in the lumen[82]. The sugars, proteins, and lipids found in OMVs are foreign to the mammalian system and are immunogenic[83], allowing them to be an effective adjuvant for vaccine delivery. Additionally, OMVs are internalized by mammalian cells[84], which could aid in vaccine delivery throughout the body. We have demonstrated that a variety of outer-membrane proteins can be glycosylated, as described in Chapter 2. Furthermore, our group has shown that it is possible for some outer membrane glycoproteins to be localized to OMVs[69].

There are several challenges associated with the production of glycoprotein and OMV vaccines produced from *E. coli*. One of these challenges is related to the OST, *C. jejuni* PglB, which has been found to only transfer oligosaccharides that contain an acetamido group at the C-2 carbon of the reducing end sugar[52]. Since many immunologically relevant bacterial carbohydrates do not contain the required acetamido group at the reducing end, efforts to discover and/or design OSTs with novel function is necessary[85]. The methods described in Chapter 2 of this thesis could be extended to select for OSTs that more efficiently transfer specific *O*-antigens, or other bacterial glycans, by replacing the *C. jejuni pgl* locus with a locus expressing the genes required to build the specific bacterial glycan of interest.

Therapeutic glycoproteins

The biopharmaceutical sector is a rapidly growing segment of the pharmaceutical industry. Biopharmaceuticals, which are predominantly recombinant protein therapeutics, but also encompass nucleic acid-based products and engineered cell or tissue-based products, were estimated to have a global market value of $70-80 billion in 2010[86]. However the sales of general biologics in that same year, which are primarily recombinant protein therapeutics and antibodies, are documented to have exceeded $100 billion, where therapeutic monoclonal antibodies accounted for approximately 48% of those sales [87]. Currently, approximately 70% of the human therapeutic proteins on the

market are *N*-linked glycoproteins[88]. As mentioned previously, monoclonal antibodies, specifically IgGs, require *N*-linked glycosylation for full effector function[13].

In addition to antibodies, a variety of other therapeutic proteins require *N*-linked glycosylation in order to achieve optimal therapeutic activity. One therapeutic glycoprotein of interest is erythropoietin, which functions to regulate red blood cell production and is used to treat anemia caused by chronic kidney disease or chemotherapy. Erythropoietin has also been shown to have nonhaematological effects including the prevention of ischaemia-induced tissue damage[89]. The complete absence of sugar residues on erythropoietin, as produced from non-engineered *E. coli*, results in a 1000-fold decrease in *in-vivo* activity[90]. Engineering the protein with two addition *N*-linked glycosylation sites, for a total of four, yields the derivative Darbepoetin alfa which exhibits a three-fold increase in serum half-life and increased *in-vivo* activity[91]. Another therapeutic glycoprotein engineered for improved efficacy is tissue plasminogen activator (t-PA), a serine protease used in the treatment of acute myocardial infarction[92]. A third generation t-PA, Tenecteplase (TNKase), was engineered such that the glycan was moved from residue 117 to residue 103, leading to eight-fold slower clearance from serum and 200-fold greater resistance to plasminogen-activator inhibition while maintaining function[93].

The significance of *N*-linked glycosylation in therapeutic proteins is clear, and the functional transfer of the *N*-linked glycosylation machinery from *C. jejuni* to *E. coli* has opened the door to engineering the cellular machinery of *E. coli* to produce therapeutic glycoproteins. Significant strides have been made in the engineering of *E. coli* towards this goal, including the biosynthesis of the core $Man_3GlcNAc_2$- structure of the

eukaryotic glycan in *E. coli*, which was transferred by the *C. jejuni* OST, albeit with low efficiency[23]. While the engineered $Man_3GlcNAc_2$- glycan was transferred to target proteins by the *C. jejuni* OST, this required the more specific bacterial acceptor sequence (D/E-X_1-N-X_2-S/T). Hence, any acceptor sequences in eukaryotic glycoproteins that lack a negatively charged residue two amino acids upstream of the glycosylation consensus sequence are not currently recognized in the engineered bacterial system.

Chapters 2 and 3 of this thesis describe efforts that have been made to attempt to overcome the limitations of the bacterial OST, including the development of a variety of selection tools to select for improved function, and a functional analysis of OST homologues from other bacterial species. Chapter 4 presents a novel method for characterizing OSTs *in vitro,* and Chapter 5 details proposed future work that will allow us to draw closer to the goal of creating 'humanized' and other engineered glycoproteins in *E. coli.*

CHAPTER 2
DEVELOPMENT OF A SELECTION TOOL FOR
N-LINKED GLYCOSYLATION IN *ESCHERICHIA COLI*

Introduction

The addition of an oligosaccharide to a protein sequence can have significant and dramatic impacts on the properties of the protein, including, but not limited to, improved stability, altered susceptibility to proteolysis, enhanced folding and solubility, varied pharmacokinetic properties, and enhanced effector function[13,94–97]. The introduction of *N*-linked protein glycosylation machinery into *E. coli* has opened the door to engineering therapeutic proteins and novel vaccine technology in this prolific protein production host. While it has been clearly demonstrated that *E. coli* can successfully glycosylate proteins with a variety of different oligosaccharides, both in our lab and by others[23,41,69], the efficiency of glycosylation, the specific site that is glycosylated, and the specificity of the glycan that is transferred are all areas that need to be improved in order to take full advantage of *E. coli* as a glycoprotein production host.

In order to make improvements to the glycosylation machinery in *E. coli*, we set-forth to develop a selection tool which could be utilized to isolate variant glyco-engineered *E. coli* displaying enhanced properties of glycosylation. To this end, it was necessary to develop a tool that allows for the maintenance of a link between the genotype and the phenotype of a specific protein sequence glycosylated with a specific sugar moiety. Our primary goal here was to design a tool that links genetic alteration of the glycosylation machinery, specifically, the OST, to the effects on glycosylation.

In our initial attempts, a surface display system was developed in which glycosylated proteins are displayed on the surface of *E. coli*. It was discovered that *E. coli* expressing the *C. jejuni* glycosylation machinery non-specifically displayed the *C. jejuni* oligosaccharide on their surface, and this challenge was overcome through genomic alterations. It was shown that glycoproteins could be displayed and detected on the surface of *E. coli,* however initial attempts to isolate glycosylation competent cells from those not capable of glycosylation were of limited success. Alterations to this system are currently being investigated, and the next steps underway to successfully utilize this surface display system are presented.

An additional attempt to design a glycosylation selection tool in *E. coli* involved the design of a glycosylation site located adjacent to a proteolytic cleavage site. Our hypothesis was that close proximity of a glycan to the cleavage site could interfere with the enzyme recognition and inhibit proteolytic cleavage. While the specific construct designed and presented here did not support our hypothesis, additional constructs that could prove successful are elucidated.

Finally, we established a phage display system, with which we could select for glycoproteins containing the previously established acceptor sequence from a pool of proteins with mutated acceptor sites[98]. We went on to utilize the phage display system to select for functional alterations of the OST. The results of this selection are presented.

Results and Discussion

Outer membrane glycoprotein display

Our hypothesis was that if we could display glycosylated proteins on the surface of *E. coli*, we would be able to separate *E. coli* that are capable of glycosylation from those that were not glycosylation-competent. Furthermore, we anticipated that cells carrying genetic alterations leading to more efficient glycosylation would display more glycan on their surface and could potentially be selected from a pool of cells with less efficient glycosylation.

The *C. jejuni N*-linked heptasaccharide binds to soybean agglutinin (SBA), a lectin that recognizes terminal GalNAc residues[17]. SBA is a large tetrameric protein, ~120kDa, that should not be capable of penetrating the *E. coli* outer membrane. Fluorescently labeled SBA-AlexaFluor 488 (SBA-488) was used to probe cells for the presence of cell surface *C. jejuni* glycans. Initial experiments led to the discovery that *E. coli* expressing the *C. jejuni pgl* locus displayed SBA-binding glycan on their surface in the presence or absence of glycosylated protein. This led to the hypothesis that the *C. jejuni* glycan was hijacking the native LPS biogenesis machinery of *E. coli*. To test this hypothesis, *E. coli* K-12 BW23113 with either the *rfaL* gene or the *rfaC* gene deleted were chosen as candidate hosts[99]. The *rfaL* gene, or *waaL*, encodes the *O*-antigen ligase that catalyzes the transfer of an *O*-antigen to the LPS core[100]. The *rfaC* gene, or *waaC*, encodes the LPS heptosyl transferase I that is responsible for the transfer of the first heptose residue onto the inner core of the LPS[101]. Both *rfaC* and *rfaL* knock-out cell

lines showed a dramatic decrease in fluorescence labeling compared to the parental strain, with the *rfaC* knock-out having the lowest level of outer membrane display of the *C. jejuni* glycan (Figure 2.1). This data supports our hypothesis that the *C. jejuni* glycan can hijack the LPS machinery of *E. coli.* Following Western blot verification of glycosylation of a native *C jejuni* substrate (data not shown), BW23113Δ*rfaC* was chosen as the cell line for surface display.

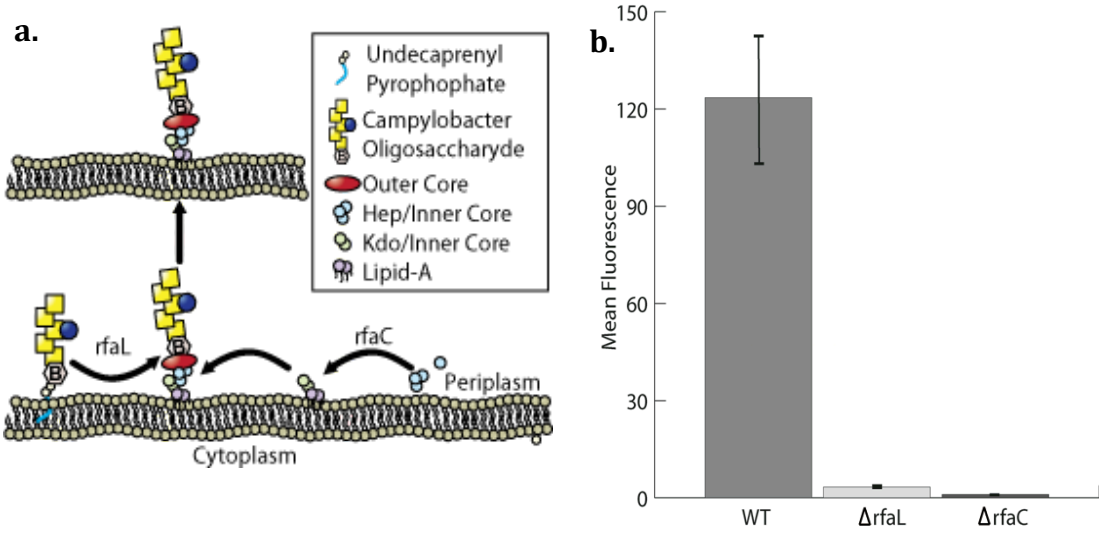


Figure 2.1. Selection of cell line for surface display. (a.) Schematic of the function of *rfaC* and *rfaL*. (b.) Mean fluorescence of cells expressing the *C. jejuni pgl* locus following labeling with SBA-488. The results are the average of four independent samples.

Prior to this work, only soluble periplasmic proteins had been glycosylated in *E. coli* cells carrying the *C. jejuni pgl* locus. Therefore, it was necessary to first investigate whether outer membrane proteins could be *N*-glycosylated. The first protein we investigated was *C. jejuni* OmpH1 (CjaA), a native *C. jejuni* surface lipoprotein, containing one glycosylation consensus sequence, $(_{137}\text{DSNIS}_{141})^{102,103}$. When produced in cells co-expressing the *C. jejuni pgl* locus, glycosylated CjaA could be detected by

Western blot, and in cells where the OST was mutated to prevent glycosylation (pgl_{mut}), no glycosylated CjaA was detected (Figure 2.2a). A significant increase in fluorescence was detected for BW23113Δ*rfaC* cells co-expressing the functional *C. jejuni pgl* locus and CjaA compared to those lacking a functional OST (Figure 2.2b). Introducing a D137A point mutation, which removes the negatively charged residue at the -2 position of the glycosylation consensus motif, making it not recognized by the OST, caused a loss of fluorescence signal in glycosylation-competent BW23113Δ*rfaC* cells, as expected; additionally, a C20A point mutation downstream of the signal peptide required for lipoprotein processing and localization[104] led to a similar loss of fluorescence signal (Figure 2.2b).

Following verification of the presence of glycosylated CjaA on the surface of cells, attempts were made to isolate SBA-488-labeled bacteria. For this experiment, we generated artificial libraries, consisting of 1:100 mixtures of cells expressing glycosylated CjaA to cells expressing aglycosylated CjaA. Following two rounds of screening by fluorescence activated cell sorting (FACS), no enrichment of glycosylation competent cells was achieved, leading to the hypothesis that the number of glycans on the surface of BW23113Δ*rfaC* cells was insufficient for efficiently separating cells.

In an attempt to increase the number of SBA-488-accessible glycans on the surface, a strategy previously utilized in the DeLisa lab research group to put multiple glycans onto maltose binding protein (MBP) was employed[69]. An engineered GlycTag (GT) acceptor sequence was inserted into the second extracellular loop of the β-barrel outer membrane protein OmpX from *E. coli,* as described previously for outer membrane display of peptide sequences[105]. The resulting OmpX-GT construct contained four

potential glycosylation sites in tandem. Glycosylated OmpX could be detected by Western blot only in cells co-expressing a functional *C.jejuni pgl* locus (Figure 2.2a). As seen above with CjaA, significant increase in fluorescence was detected in BW23113Δ*rfaC* cells co-expressing OmpX-GT and the functional *C. jejuni pgl* locus compared to those lacking a functional OST (Figure 2.2b). Outer membrane localization of glycosylated OmpX-GT in BW23113Δ*rfaC* cells expressing the *C. jejuni pgl* locus was further verified with fluorescence microscopy (Figure 2.2c).

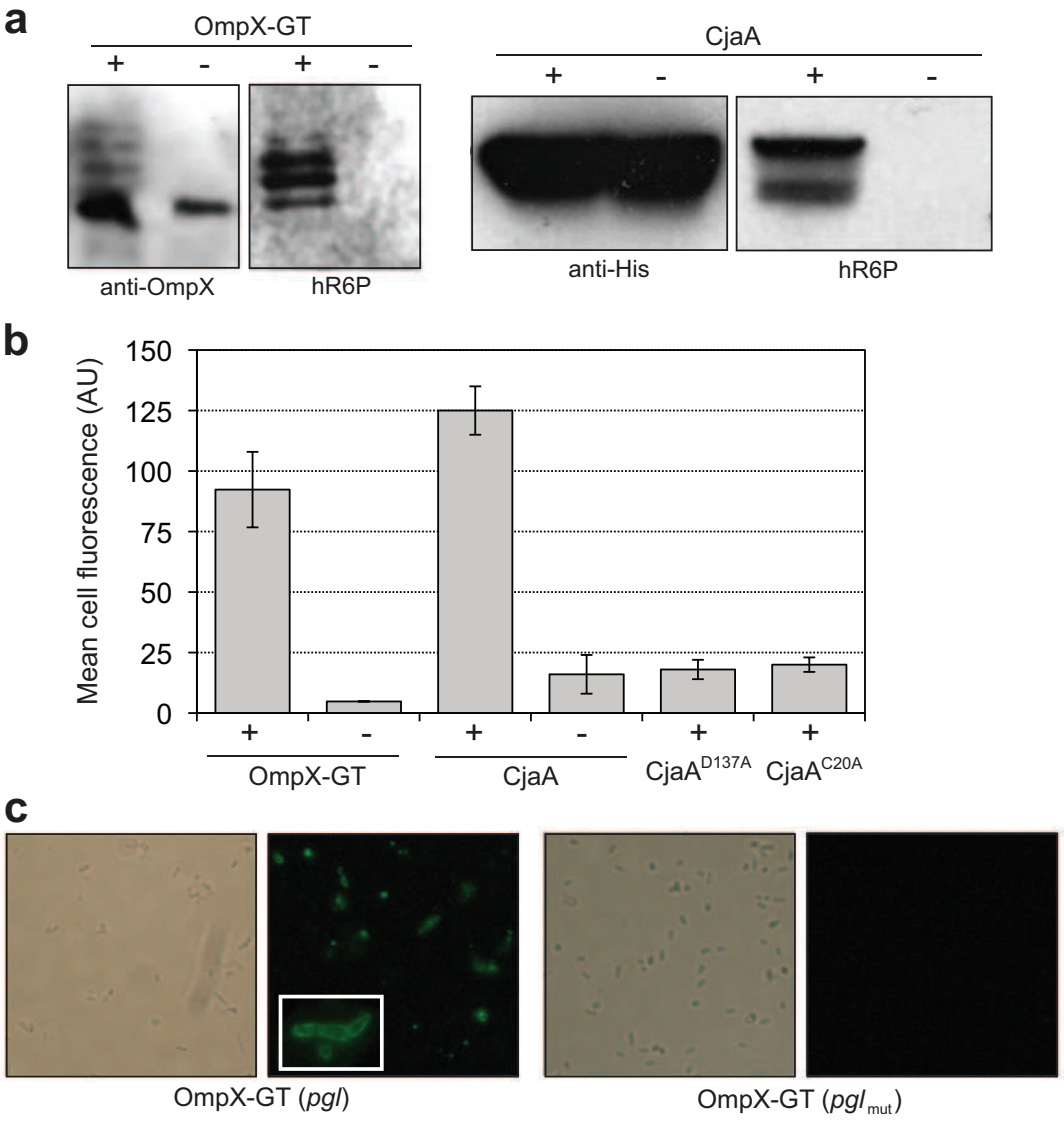


Figure 2.2 Expression of outer membrane glycoproteins. (a.) Western blot analysis of OmpX-GT or CjaA expression in cells co-expressing pACYC*pgl* (+) or pACYCpgl_{mut} (-). (b.) Flow cytometric analysis of BW23113ΔrfaC cells expressing OmpX-GT, $CjaA^{D137A}$, $CjaA^{C20A}$ or CjaA and co-expressing pACYC*pgl* (+) or pACYCpgl_{mut} (-). Cells were labeled with SBA-488. The data presented is the average of three independent replicates, with error bars representing the standard error of the mean. (c.) Light and fluorescence microscopy of BW23113ΔrfaC cells expressing OmpX-GT and co-expressing pACYC*pgl* (+) or pACYCpgl_{mut} (-). This figure was published[69] and is re-used here with permission.

Similar to the test sorts of CjaA, OmpX-GT was produced from both glycosylation competent cells and from cells lacking a functional OST, and a 1:100 mixture was used as a 'false library' for determining the sortability of this system. This time, instead of labeling with SBA-488 and sorting on the FACS Calibur, the cells were mixed with SBA-agarose beads. The beads were then washed, and the bound cells were recovered and tested for the presence of funtional OST. To our surpise, there was again no significant improvement in the ratio of glycosylation competent to non-glycosylation competent cells.

Currently, this selection tool is being re-visited by a post-doc in the DeLisa lab research group, Dr. Anne Ollis. In an effort to further increase the number of accessable glycans on proteins on the surface of *E. coli*, the FimA protein has been targeted. The *fimA* gene encodes the subunit of type 1 fimbriae in *E. coli*[106]. Fimbriae are long fibrous structures which protrude from the surface of bacteria and aid in the attachment to external surfaces for colonization. On the order of 500 fimbriae can be found on the surface of one *E. coli* cell, each containing approximately 1000 copies of the FimA subunit[107]. FimA fusions have been previously been utilized to succesfully display heterologous peptides on the surface of *E.coli*[108]. The *fimA* gene was cloned from the genome of *E. coli* W3110, and a single DQNAT glycosylation motif was inserted between residues S27 and V28 of mature FimA, yielding FimA-GT. This site was chosen based on succesful peptide display positions identified in the previously published study. Cells co-expressing FimA-GT along with the *C. jejuni pgl* locus showed enhanced binding to hR6, an antibody raised against the *C. jejuni* glycan[109], in comparison to cells lacking a functional OST (Figure 2.3b). Mixtures of glycosylation-competent cells

expressing FimA-GT with cells expressing FimA-GT but with an inactive glycosyaltion locus were prepared and selected on a plate coated with the hR6 antibody. Following recovery, cells were detected from a dilution as low as 1:1000, where, based on OD_{600}, approximately 100,000 cells were incubated in each well. Verification of glycosylation competency of the recovered cells was inconclusive, as the pgl locus could not be detected following cell re-growth. Studies to determine the level of FimA glycoprotein display and its utility in library screening are ongoing.

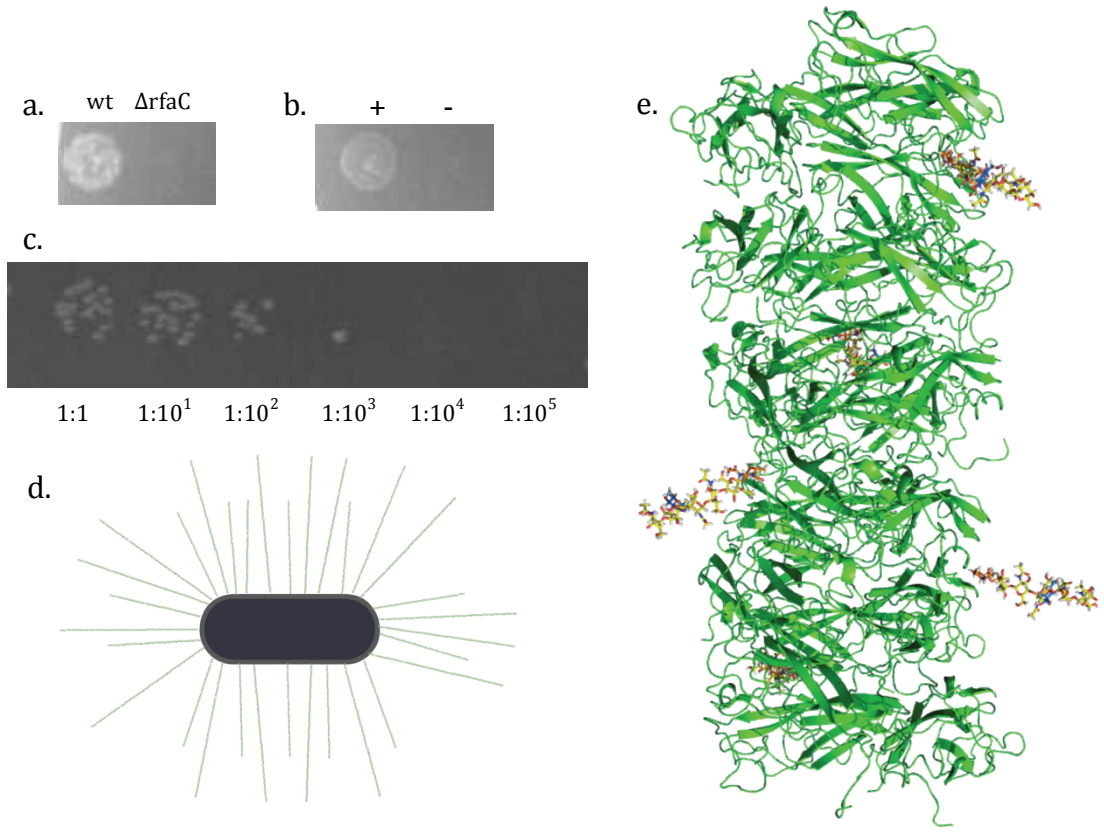


Figure 2.3. *E. coli* expressing glycosylated FimA bind to glycan-specific hR6 antibody. (a. - c.) Spot plate results from cells recovered following binding to hR6. (a.) BW23113 pgl+ vs BW23113ΔrfaC pgl+. (b.) CLM24 pgl+ pFimA-GT vs CLM24 pgl_{mut}.pFimA-GT (c.) Dilutions of CLM24 pgl+ pFimA-GT:CLM24 pgl_{mut} pFimA-GT; dilutions were based on normalized OD. (d.) Artistic rendition of fimbriaeted *E. coli*. (e.) Model of glycosylated FimA

The ability to display and detect glycoproteins on the surface of bacteria has been confirmed, but more research is necessary to improve the ability to isolate glycosylation competent cells from a background of cells that are not capable of glycosylation in order to utilize this system as a selection tool for *N*-linked glycosylation in *E. coli.* Improving the number of glycans displayed on the surface of *E. coli* by selecting a protein that is

displayed in significantly higher quantity on the surface may prove to be the solution that allows this system to be successfully utilized.

Glycan-protected protease cleavage

In general, the addition of an oligosaccharide confers a level of protease resistance to a protein; this can be due to the (i) increased rigidity of a protein caused by the orientation of the hydrophilic carbohydrates on the protein surface, or (ii) steric protection of susceptible protease cleavage sites by the large oligosaccharide structure[110]. Indeed, it has been shown that alterations to glycosylation can impact observed proteolytic cleavage patterns[111–113].

Since it is clear that glycosylation of a protein sequence can decrease proteolysis, we hypothesized that a glycosylation site located within or near a specific protease cleavage site could inhibit the cleavage site recognition and subsequent proteolytic degradation, as depicted in Figure 2.4. Therefore, we setout to create a fusion protein construct that was protected from proteolytic cleavage when glycosylated, and fully hydrolyzed when aglycosylated. In order to maintain the genotype-phenotype linkage necessary to develop this technology into a selection tool for *N*-linked glycosylation in *E. coli,* the fusion protein was designed as an integral membrane protein.

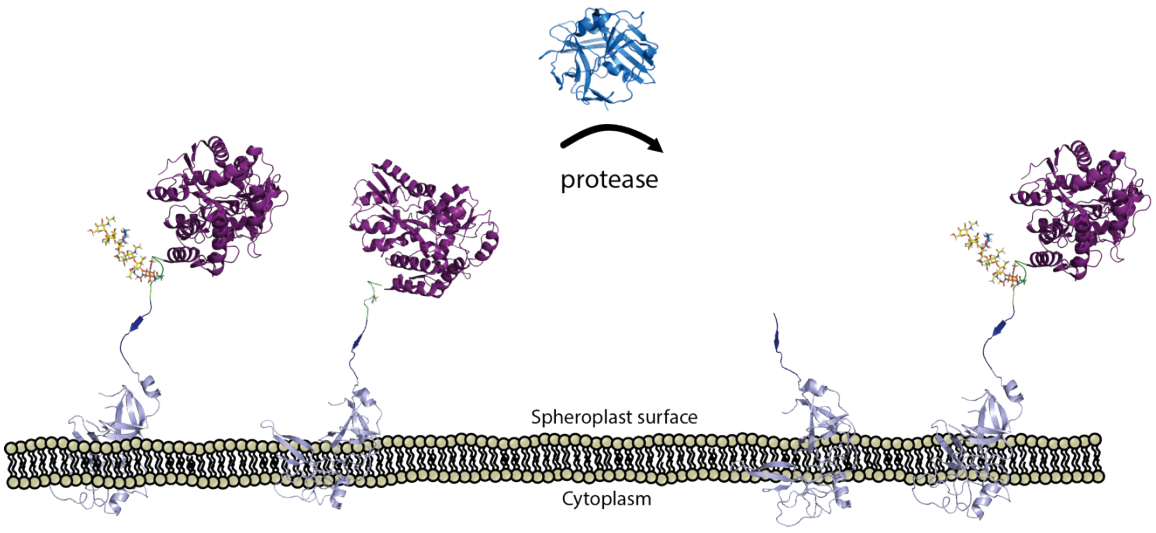


Figure 2.4 Schematic of glycan-protected protease cleavage. Our hypothesis is that, in the presence of a glycan, a protease cleavage site can be protected from degradation, as depicted here.

Many proteases cleave non-specifically or have broad substrate specificity. There are, however, a number of commercially available proteases with strict substrate specificities. The protease we chose to test first was tobacco etch virus (TEV) protease (Invitrogen; Ac-TEV). TEV protease recognizes the peptide sequence ENLYFQ(G/S) and cleaves the protein following the glutamine residue[114]. To test our hypothesis that a glycan could protect this site, a hybrid TEV protease cleavage site/bacterial *N*-linked glycosylation site, DQNATENLYFQG, was engineered into a target protein. The protease cleavage site served as a linker between MBP and g3p of M13 bacteriophage. The g3p protein is transiently inserted into the inner membrane prior to incorporation into phage[115], allowing for display of proteins on the inner membrane of *E. coli*. Displayed proteins can be selected after spheroplasting of the bacterial cells, a process which results in disruption of the outer membrane and hydrolysis of the peptidoglycan layers, leaving an intact cytoplasmic membrane enclosing the cytoplasm. Indeed, a technology termed

APEx has been described in which combinatorial antibody libraries fused to g3p were selected for enhanced ligand binding by isolating spheroplasts with high affinity for the ligand[116]. MBP was chosen as the fusion partner because it is a thoroughly characterized soluble periplasmic protein[117], and it can be easily selected for based on binding to an amylose resin (NEB).

The first step towards our goal of using glycan-protected protease cleavage as a selection tool for *N*-linked glycosylation was to verify that glycosylation of the site prevents proteolytic degradation. To investigate this, whole cell lysates were prepared from an equivalent number of *E. coli* cells co-expressing the MBP-GT-TEV-g3p protein and the *C. jejuni pgl* locus with or without a functional OST. Lysates were treated with Ac-TEV protease at 30°C for 30min, and the reactions, including controls lacking protease, were resolved on SDS-PAGE, and proteins were visualized by Western blot (Figure 2.5). Unfortunately, results with this construct indicated that the fusion protein was hydrolyzed by the protease regardless of the glycosylation state.

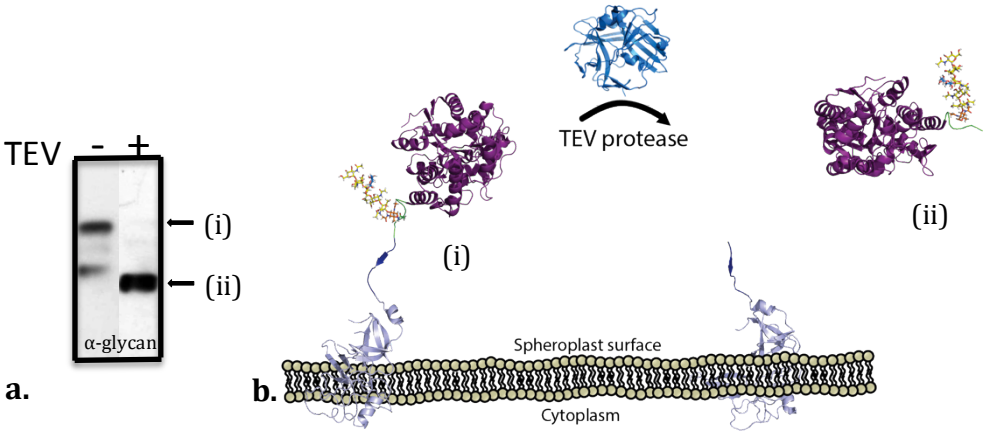


Figure 2.5. TEV protease cleavage site is not protected by adjacent glycan. (a.)Western blot detection of whole cell lysates from glycosylation competent *E. coli* expressing MBP-GT-TEV-g3p. Whole cell lysate with and without TEV protease digestion are depicted. The blot is probed with hR6 for detection of glycan. (b.) Schematic of protease degradation. (i) MBP-GT-TEV-g3p. (ii) MBP-GT.

It is clear from Figure 2.5 that the glycosylated fusion protein was degraded in the presence of TEV protease, indicating that this cleavage site was not protected by the presence of the *C. jejuni* glycan. While the specific construct described above did not demonstrate glycan-protected protease cleavage as hoped, our hypothesis that a glycosylation site located within or near a specific protease cleavage site could inhibit recognition and prevent degradation should not be rejected, as there are a variety of other constructs that hold promise and should be investigated. Three additional possibilities for construct design are described here.

Bovine enterokinase (Sigma), a highly specific serine protease, can be used to cleave an N-terminal FLAG epitope-tag from fusion proteins; the specific recognition site of bovine enterokinase is XDDDDKX$_{-P}$X, where X is any amino acid and X$_{-P}$ is any amino acid besides proline, and cleavage occurs following the lysine residue[118]. Based on

the structural sensitivity, as noted by the inability of the protease to recognize the sequence when the residue following the cleavage site is a proline, this protease could potentially be inhibited by a glycan positioned on an asparagine in the X_{-P} position, as in the sequence XDDDDKNAT.

Another protease worth investigating is Factor Xa protease (NEB), which cleaves following the arginine in the specific sequence I(D/E)GRX, where X can be an amino acid besides proline and arginine[119]. We have noted that, in some cases, it is possible for bacterial OSTs to glycosylate an asparagine that has a negatively charged residue in the -3 position[120]. Again, the structural specificity of this protease is revealed by the inability to tolerate a proline following the cleavage site, indicating that the presence of a glycan at that position could impact protease recognition of the site. A site such as IDGRNAT could hypothetically be glycosylated by a bacterial OST, potentially inhibiting cleavage by Factor Xa protease.

Finally, and possibly the most promising protease, thrombin (GE Healthcare), is a serine protease that recognizes the peptide sequence LVPRGS and cleaves between the arginine and glycine[121]. It has been shown that this protease can tolerate a S/T in place of the valine in its recognition sequence[122]. Additionally, this protease does not tolerate hydrophilic residues near the N-terminus of the recognition sequence[123,124]. Therefore, it is likely that a hydrophilic glycan preceding this cleavage site could interrupt recognition by this protease. The extended sequence LTPRGVRL is an efficiently recognized sequence for thrombin[125]. A sequence such as DQNLTPRGVRL would likely be recognized in the absence of glycosylation, whereas glycosylation of the asparagine could potentially protect this site from proteolysis.

Phage display of glycoproteins

Phage display of proteins and peptides has proven to be a powerful tool for isolating protein subunits with increased binding affinity for a particular substrate[126]. The technique for displaying phage particles with associated fusion peptides was first described in 1985[127]. Since that time, phage display has been primarily utilized for the selection of improved binding of proteins to a ligand[128–131], and has also been used to select for increased protein stability[132,133], and increased enzymatic function[131,133]. Additionally, phage display has been utilized to elucidate the substrate specificity for a variety of proteases[125,134,135]. While significantly less common, phage display can also be utilized to investigate complex enzymatic pathways. One specific example is the use of M13 phage to provide the genotype-phenotype linkage necessary to elucidate mutations in DsbC, a native *E. coli* disulfide bond isomerase, that led to enhanced substrate folding in the periplasm of *E. coli*[136].

Here, we hypothesized that phage expressing proteins glycosylated with the *C. jejuni* glycan should be selected based on their increased affinity to a glycan-specific lectin. As previously mentioned, the *C. jejuni* glycan is known to have an affinity for SBA. Working in conjunction with Dr. Eda Çelik, a post-doc in the DeLisa lab research group, we developed a genetic screen for glycosylation in *E. coli* based on the display of *N*-linked glycoproteins on the tail-region of M13 phage particles[98]. The basis of this system is a genetic fusion between g3p, the minor coat protein of M13 phage, and a target protein containing an optimized glycosylation consensus sequence tag for *N*-linked glycosylation[69]. The fusion protein contains a signal peptide which directs its secretion to

the periplasm through the SecYEG translocon[137], where glycosylation can take-place if a functional *pgl* locus is co-expressed. Recovery of phage particles containing glycosylated protein on their surface was demonstrated following lectin affinity chromatography. Recovered phage are used to re-infect cells in order to establish the necessary genotype-phenotype linkage through replication of the phagemid. In initial studies, the phagemid carried the *N*-linked glycoprotein, but was later re-designed to also carry the OST, and could be further engineered to carry any of the components of the glycosylation machinery.

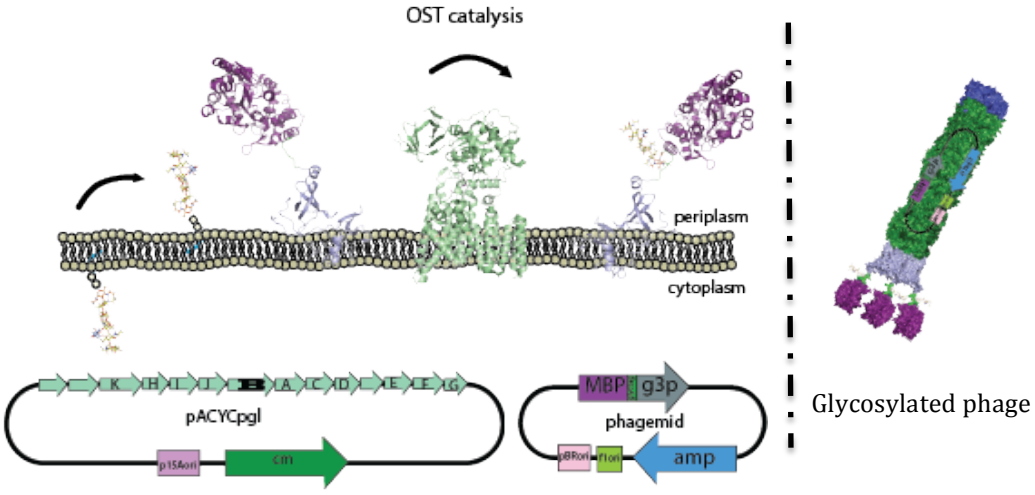


Figure 2.6. Schematic of glycophage production. Glycosylated phage particles are produced from TG1ΔwaaL cells co-expressing the *C. jejuni pgl* locus and a phagemid encoding the glycoprotein, MBP-GT-g3p. Production of phage particles is induced by infection with helper phage.

In proof-of-concept studies[98], we demonstrated the ability of the phage display system to be utilized in the selection of glycosylation consensus sequences that can be recognized by the OST. The optimal acceptor sequence for the OST of *C. jejuni* is DQNAT[138]. Utilizing this information, a C-terminal fusion of the DQNAT glycosylation tag was made to MBP. This construct was then cloned in-frame with g3p into pBAD24, a plasmid which carries the M13 phage origin of replication and thus serves as the phagemid for these experiments[139]. Western blot analysis verified that this construct was produced in *E. coli* and was glycosylated only when (i) the functional *C. jejuni* locus was present, and (ii) the DQNAT acceptor site was present[98], as verified by immunoblot with anti-MBP and hR6.

In order to demonstrate that this system was useful for the selection of glycosylated phage, TG1 *E. coli* cells expressing the *C. jejuni pgl* locus and a phagemid containing either MBP-GT-g3p or MBP-g3p were infected with VCSM13 helper phage

to induce the production of modified phage particles. Phage titers were quantified by infection of fresh TG1 cells with the phage preparations, followed by selection on ampicillin (Amp), the antibiotic resistance conferred by the phagemid, to determine colony forming units (CFU). Western blots of isolated phage particles confirmed the presence of glycosylated protein only in the case where the DQNAT site and functional *C. jejuni pgl* locus were present[98]. A mixing experiment was used to determine whether glycosylated phage particles could be isolated by binding to the SBA lectin. A 1:1 mixture of glycosylated to aglycosylated phage particles were incubated with the agarose-bound SBA, washed several times with PBS and PBS containing 30mM galactose, and finally eluted with PBS supplemented with 300mM galactose, which competitively binds SBA[140], allowing for the release of the bound phage particles. Utilizing a PCR-based assay, it was verified that 22/28 of the phagemids recovered encoded for the MBP-GT-g3p construct[98].

While the ratio of DQNAT-containing phagemid particles was significantly greater than the starting 1:1 mixture, it should be noted that the overall number of phage eluted following this initial sort was significantly below the theoretical maximum recovery expected if the actual starting mixture were 1:1. This low recovery can be explained by the fact that (i) glycosylation in *E. coli* has been shown to be relatively inefficient[141], and (ii) lectins exhibit weak binding for their carbohydrate ligands, generally requiring multivalency to achieve high avidity[142]. Furthermore, it was later found that we could select up to 5/22 phagemid constructs encoding MBP-GT-g3p from a $1:10^6$ mixture following just a single sort[98]. These data led us to the conclusion that this

system could be utilized for the selection of libraries of enzymes in the *N*-linked glycosylation pathway expressed in *E. coli*.

To demonstrate the ability of this selection tool to select libraries of enzymes, we first investigated whether a library of the acceptor sequences could be sorted. It has been previously shown that the bacterial OST, specifically *C. jejuni* PglB, requires a negatively charged residue two base-pairs upstream of the asparagine that is glycosylated[143]. Therefore, a library was created by random mutagenesis of the aspartic acid residue in the DQNAT acceptor site, leading to the production of phagemids expressing an XQNAT glycosylation site, of which only DQNAT and EQNAT sites should be glycosylated. A library of approximately $1x10^9$ phage was sorted as described above, and following a single round of sorting, 30/47 of the sequenced clones contained a negatively charge aspartic acid residue at the -2 position. While no clones were selected containing a glutamic acid residue, we have previously observed that the glycosylation of acceptor sequences is significantly greater in the presence of an aspartic acid residue as compared to a glutamic acid residue (data not shown).

In conjunction with our publication[98], another report describing the use of phage display to select for improved glycosylation phenotypes in *E. coli* was published[144]. This further supported our hypothesis that glycophage could be useful for selection of improved function of enzymes in the *C. jejuni pgl* locus, or any other *N*-linked glycosylation locus that could be expressed in *E. coli.*

Sorting an OST library with the phage display system

Following verification that the phage display system could be utilized to select phage particles displaying glycosylated proteins on their surface[98], we proceeded to sort a library of *C. jejuni* PglB in an attempt to select a variant that no longer required a negatively charged residue in the -2 position. Based on the recent crystal structure of *C. lari* PglB, a homolog of *C. jejuni* PglB, it was determined that the R331 residue of the OST potentially forms a salt bridge with the negatively charged residue in the -2 position of the acceptor peptide to be glycosylated[71]. In *C. jejuni* PglB, aligned to *C. lari* PglB, this residue is adjacent to an additional positively charged residue. Therefore, a library was created using NNK random primers, where $_{327}$RR$_{328}$ of *C. jejuni* PglB was randomly mutated. The resulting library was called the R^2 library.

Prior to sorting of this library, several modifications were made to the glycophage selection system. First, based on the results of Dürr et al.[144], it was determined that utilizing an antibody against the glycan, specifically hR6, could more efficiently select for glycosylated phage due to the increased binding affinity of antibodies as compared to lectins. Secondly, it was noted that a truncated version of g3p (tg3p) could lead to more efficient protein display on phage[145]. Finally, we were interested in selecting for glycosylation sequences that did not contain a negatively charged residue at the -2 position of the asparagine that gets glycosylated. Therefore, a phagemid was designed containing MBP-2xAQNAT-tg3p along with the *C. jejuni* PglB R^2 library. The MBP-2xAQNAT-tg3p and PglB library members were expressed bicistronically, each with their own ribosome-binding site (RBS).

Following two rounds of sorting over hR6, bound through biotin to streptavidin beads, ten clones were selected at random and characterized. It was found that one of these ten constructs produced phage particles with significantly increased glycosylation, however the size of the glycosylated protein was found to be significantly smaller than full-length MBP-2xAQNAT-tg3p fusion protein (Figure 2.7a). The mutation in this selected OST is $_{327}ER_{328}$. We next determined whether this selected OST had increased recognition for glycosylation sites lacking a negatively charged residue at the -2 position. This was tested by sub-cloning the gene encoding $PglB_{327ER328}$ into a plasmid with AcrA'0x', a version of the AcrA glycoprotein containing four glycosylation consensus sequences in which each is mutated to have an alanine at the -2 position. Unfortunately, no improvement in glycosylation over wild-type (wt) *C. jejuni* PglB was observed (Figure 2.7c).

The observed glycosylated product from the selected phagemid was smaller than the expected fusion protein, so we wanted to determine whether this altered phenotype was dependent on the selected OST. To test this, wt *C. jejuni* PglB was sub-cloned into the selected phagemid in place of *C. jejuni* $PglB_{327ER328}$. It was found that wt *C. jejuni* PglB was also capable of glycosylating the glycoprotein expressed from this construct (Figure 2.7d). These results, however, are not fully conclusive, as it is possible that the original phagemid containing *C. jejuni* $PglB_{327ER328}$ was not eliminated during the cloning process.

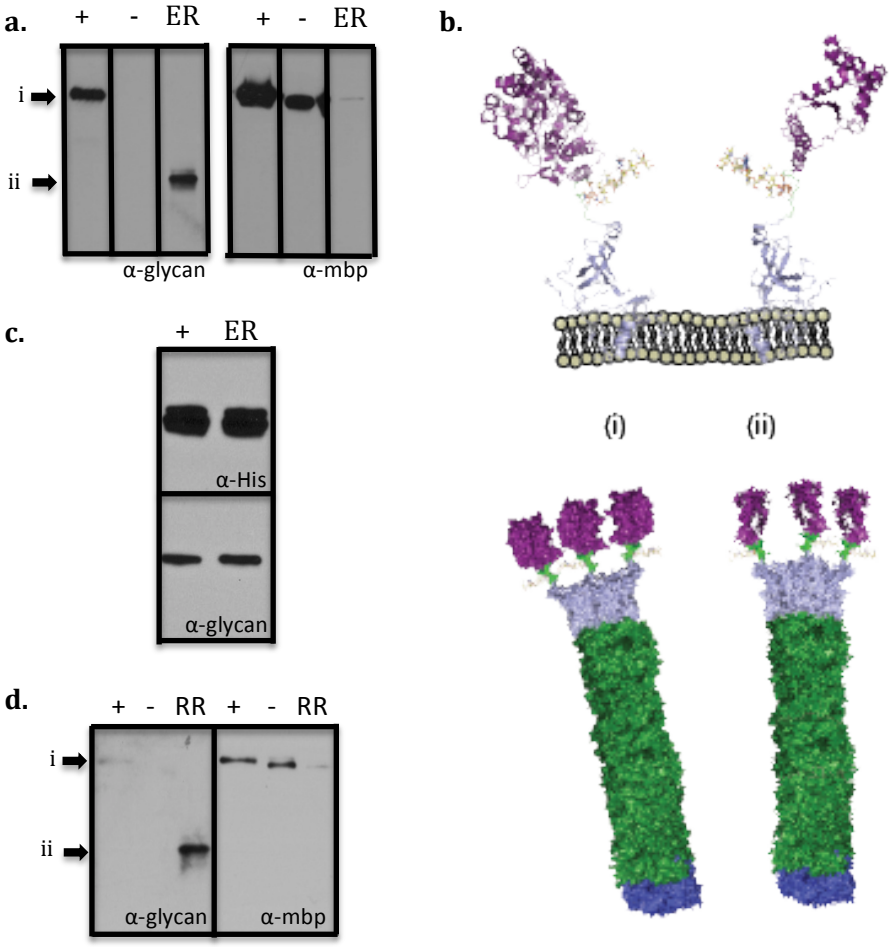


Figure 2.7. Investigating OST selection. (a.) Western blot of phage particles produced from glycosylation competent *E. coli* co-expressing a phagemid with (+)MBP-GT-tg3p and wt *C. jejuni* PglB, or (ER) MBP-2xAQNAT-tg3p bicistronic with the R^2 library member, *C. jejuni* PglB $_{327}ER_{328}$. (b.) Schematic of C-terminally truncated MBP-2xAQNAT-g3P in the membrane and integrated into phage particles. (c.) Western blot of periplasmic fraction from glycosylation competent CLM24 cells expressing AcrA'0x' and wt *C. jejuni* PglB (RR) or *C. jejuni* PglB $_{327}ER_{328}$. (d.) Western blot of phage particles produced from glycosylation competent *E. coli* expressing wt *C. jejuni* PglB bicistronic with (+) MBP-GT-tg3p and, (-) MBP-2xAQNAT-tg3p and (RR) MBP-2xAQNAT-tg3p from the phagemid containing *C. jejuni* PglB $_{327}ER_{328}$. (i) MBP-2xAQNAT-tg3p. (ii) truncated product.

Upon closer analysis, it was found that five random point mutations were present in the selected phagemid, four of these point mutations resided in the coding sequence for MBP. It is likely that the original phagemid construction contained a mixture of plasmids, some of which harbored these mutations. Interestingly, NCBI orf finder (http://www.ncbi.nlm.nih.gov/projects/gorf/) predicts that the mutated gene encodes a C-terminal truncation mutant, tMBP-2xAQNAT-tg3p. The size of this predicted truncation mutant is approximately 25kDa smaller than the full-length fusion, which is the size of the glycosylated product we observed (Figure 2.7, ii). As mentioned, the DQNAT sequence is the optimal acceptor sequence, and the mutation of the first residue in the sequence to an alanine may not be sufficient to fully disrupt the interaction of this peptide in the active site of the OST. In addition, the residue preceding the first AQNAT sequence in this construct is a negatively charged glutamic acid, which has been seen to relax the specificity of the OST when present in this location[120]. To confirm this theory, the novel coding region could be cloned freshly into a new vector and glycosylation by wt *C. jejuni* PglB and *C. jejuni* $PglB_{327ER328}$ could be analyzed.

It should be noted that the library sorted here was small, with only 32^2 possible unique OSTs, and that there is not necessarily a *C. jejuni* PglB R^2 library member that is capable of efficiently glycosylating the AQNAT glycosylation sites tested here. However, the specific mutation isolated from these library sorts correlates with the observed structure-guided mutations described in Chapter 3, where it was shown that the R331E mutation of *C. lari* PglB leads to enhanced glycosylation of scFVR4-AQNAT over that of wt *C. lari* PglB. Finally, while the results presented in Figure 2.7d indicate that wt *C. jejuni* PglB can also glycosylate the mutated acceptor sequence selected from

these library sorts, these results are not fully conclusive. Further analysis is necessary to determine whether *C. jejuni* $PglB_{327}ER_{328}$ offers improved glycosylation of the AQNAT acceptor site over wt *C. jejuni* PglB.

Materials and Methods

Bacterial Strains and Plasmids

DH5α cells were used for plasmid cloning, replication and storage; CLM24 cells were used for general glycoprotein production; BW23113ΔrfaC cells were used for surface display of glycoproteins. TG1 *E. coli* were utilized for phage production and grown in M9 medium for selection of the presence of the F' plasmid, followed by growth and induction in 2xTY medium; all other *E. coli* cells were grown in Luria-Bertani medium. Cells were grown at 37°C and induced at 30°C. Induction of glycoprotein and phage production proceeded overnight. Culture medium was supplemented with 50mM glucose during the growth phase and with 30mM arabinose during the induction phase. Antibiotics were supplemented into the culture medium where appropriate at the following concentrations: 100 μg/mL Amp, 20 μg/mL chloramphenicol (Cm), and 50 μg/mL kanamycin (Kan).

Table 2.1. Strains and Plasmids used in these studies

Strain or plasmid	Relevant genotype or description	Reference or source
E. coli		
DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	Laboratory stock
CLM24	W3110 (IN(*rrnD-rrnE*)1 rph-1) Δ*waaL*	Feldman et al.
TG1Δ*waaL*	*supE hsdΔ5 thi* Δ(*lac-proAB*) F' [*traD36 proAB lacI lacZ*ΔM15] Δ*waaL*	Laboratory stock
BW23113ΔrfaC	K-12 BW25113 Δ*waaC* (*rfaC)*	Laboratory stock[99]
Plasmids		
pBAD24	Cloning vector, aribinose-inducible, Amp^r	Guzman et al.
pPglΔB	pMW07 containing the *C. jejuni pgl* locus without pglB	Glycobia
pACYCpgl	pACYC184-based plasmid encoding the *C. jejuni* protein glycosylation locus cluster	Linton et al.
pACYCpgl *mut*	pACYC184-based plasmid encoding the *C. jejuni* protein glycosylation locus cluster, were pglB has been inactivated	Linton et al.

Biopanning of cells

FACS sorting was performed on a Becton Dickinson FACSCalibur. Following induction of outer membrane glycoprotein, cell dilutions were prepared as described and were probed with SBA-488 (Invitrogen) in PBS for 45 minutes in the dark, and run on the FACSCalibur. Cells with relative fluorescence over 10^2 were collected in PBS and concentrated by filtration; the filter paper was placed on a LB-agar plate containing the appropriate antibiotics. The glycosylation phenotype was confirmed by sequencing of pglB. In most cases, a PCR product for pglB was not obtained.

Panning of cells over hR6 was performed by binding of biotinylated hR6 antibody to a streptavidin plate (Pierce), following the manufacturers instructions. Cells were incubated on the plate for 1hr at 25°C. The wells were then washed 12 times with 200μL of PBS. Cells bound to the plate were recovered following a 1hr incubation at 37°C in

SOC supplemented with Amp and Cm. Recovered cells were spotted onto LB-agar plates supplemented with Amp and Cm.

Phage purification

Phage particles were produced from *E. coli* TG1 Δ*waaL* carrying a phagemid expressing a g3p fusion protein and pACYC*pgl*, *p*ACYC*pgl*mut or pPglΔB. 20mL cultures of cells were infected with VCSM13 helper phage at OD_{600} 0.5-0.6, and incubated at 37°C for 30min without shaking. Infected cells were then pelleted and resuspended in 100mL 2xTY medium supplemented with Cm, Amp, Kan and arabinose for 16hr at 30°C. Phage were purified using PEG/NaCl precipitation, and helper phage titers were determined as described elsewhere[146].

Biopanning of phage

SBA panning: Phage particles, 10^9 CFU in PBS, were incubated for 1hr at 25°C with 1mL of agarose-bound SBA (Vector Laboratories). Following incubation, agarose-bound SBA was pelleted at 100g for 5min and washed with 10mL of PBST (PBS with 0.1% Tween20) four times, followed by three washes with 10mL of 30mM galactose in PBST. Glycosylated phage particles were eluted in three steps using 1 mL of 300mM galactose in PBS. Eluted phage particles were then used to infect *E. coli* TG1 cells. DNA was recovered from infected cells for verification of genotype.

hR6 panning: The hR6 antibody was conjugated to biotin, allowing for the conjugation to streptavidin beads (Invitrogen), and panning of phage was carried-out as previously described[144].

Phagemid library construction

The *C. jejuni* PglB R^2 library was constructed by overlap extension PCR of the *C. jejuni* pglB gene using the internal NNK primer 5'-AGAAGATTATGTGGTAACTTGGNNKNNKTATGGTTATCCTGTGCGTTATTATAG-3' and its reverse complement. The forward primer for pglB contained an RBS at the 5' end, allowing the PCR product to be cloned into pBAD24 between XbaI and SbfI, following the MBP-2xAQNAT-g3p sequence.

Western blot analysis

Expression and glycosylation of g3p fusion constructs, AcrA, and outer membrane proteins were analyzed by immunoblot following SDS–PAGE of phage particles, periplasmic fractions and whole cell lysates, as indicated. Immunodetection was performed with monoclonal anti-His antibody (Abcam, ab1187), polyclonal anti-OmpX antibody[105], monoclonal anti-MBP antibody (NEB, E8038), and polyclonal anti-glycan serum hR6 (S. Amber and M. Aebi, personal communication), as indicated.

Conclusions

The surface display system is a promising tool for selection of glycosylation-competent *E. coli*. Research is currently underway to enhance the number of glycosylated proteins on the surface of *E. coli* in order to be able to more efficiently separate glycosylation competent cells from those that are not capable of or less efficient in the process of glycosylation. We discovered that the *C. jejuni* glycan can be displayed on the surface of wt *E. coli* in the absence of glycosylation, and we were able to prevent export of glycan to the surface by genomic removal of either one of two enzymes in the LPS pathway.

The *C. jejuni* glycan located upstream of the protease recognition sequence was not sufficient to prevent digestion with TEV protease. Although this specific construct did not produce the desired results, this does not rule out the possibility of creating a selection tool using this strategy. Currently, several other constructs are in development in an attempt to create a protease cleavage site that can be protected from protease recognition by the presence of the *C. jejuni* glycan.

Phage display of glycoproteins has been shown to be a promising method for selecting improvements to the glycosylation phenotype[98,144]. Extending this selection tool for the selection of variant OSTs brought to light some of the challenges of high-throughput selection technology – 'you get what you select for'. We discovered that, while we were able to select for improved glycosylation of proteins displayed on phage, these improvements were not necessarily associated only with alterations to the OST. Re-designing this system to stabilize the glycosylation machinery and acceptor protein,

through genome integration of the glycosylation machinery and/or the MBP-GT-g3p fusion, for example, and including only the target enzyme on the phagemid, could lead to significant improvements of this selection tool.

Overall, this work has provided insight into the process of *N*-linked glycosylation in *E. coli* and has laid the foundation for future selections where improvements to the glycosylation machinery are isolated using glycoengineered *E. coli.*

CHAPTER 3
FUNCTIONAL ANALYSIS OF BACTERIAL *N*-LINKED OLIGOSACCHARYLTRANSFERASES

Introduction

N-linked glycosylation is a common post-translational modification found in all eukaryotic cells, which is required for cell viability[147]. In the last decade, it has been clearly established that this modification is present in all domains of life[26,27,34]. While it is clear that *N*-linked glycosylation is an important aspect of most living things, there is still a significant amount of information left to be uncovered in order to successfully employ this process for therapeutic protein and vaccine development. Bacterial *N*-linked glycosylation is not required for cell viability, and *E. coli* does not naturally have the ability to produce *N*-linked glycoproteins. Therefore, the *E. coli* host is an excellent candidate for the study of *N*-linked glycosylation. The central enzyme of all *N*-linked glycosylation pathways is the OST, which catalyzes the transfer of an oligosaccharide onto a protein.

OSTs responsible for *N*-linked glycosylation are membrane-spanning enzymes that catalyze the transfer of an oligosaccharide from a lipid carrier in the membrane onto a protein within a target consensus sequence. In eukaryotic cells, this sequence is typically N-X-S/T, where X can be any amino acid besides proline. In bacteria, this sequence is not sufficient for glycosylation[148]. Glycosylation of native *C. jejuni* proteins occurs only when a negatively charged residue is located two residues upstream of the asparagine that gets glycosylated. Hence, the bacterial consensus sequence has been defined as D/E-X_1-N-X_2-S/T, where X_1 and X_2 can again be any amino acid beside

proline[143]. It was later discovered that the optimal acceptor sequence for the *C. jejuni* OST is DQNAT[138]. Currently, the most thoroughly characterized bacterial OSTs are from *C. jejuni* and its close relative *C. lari*.

PglB from *C. jejuni* is a single subunit protein containing 713 amino acids with 11 predicted N-terminal transmembrane helices, followed by a soluble C-terminal domain oriented towards the periplasm. This later domain, comprises approximately one third of the protein[149]. The *C. jejuni* OST is responsible for the glycosylation of more than 65 periplasmic and membrane proteins with various functions[26,39]. A crystal structure of the soluble domain of *C. jejuni* PglB was recently solved[72], which revealed a new structural motif within the predicted active site. Shortly thereafter, the crystal structure of full-length *C. lari* PglB was solved[71]. This structure contained a divalent metal ion cofactor, which had previously been suggested to be necessary for OST function[150], and an acceptor peptide containing the optimized glycosylation consensus sequence bound in the proposed active site. The structure also confirmed that *C. lari* PglB has 13 transmembrane helices, where the predicted periplasmic region between the second and third helix[149], crosses the membrane. This additional membrane spanning region places the predicted lipid carrier recognition domain on the cytoplasmic face of the membrane[71]. The inclusion of the acceptor peptide in the crystal structure of *C. lari* PglB revealed previously undiscovered interactions between the enzyme and its protein substrate. The five-residue sequence, WWDYG, was previously proposed to be the catalytic motif of the OST active site[56,151], and indeed these residues were found to be in close proximity to the acceptor peptide in the *C. lari* OST. All functional OSTs investigated to date contain a similar sequence, however some OST homologues contain slight variations in this

sequence, with W/N/F replacing Y[72]. The glycan substrate was not included in the *C. lari* PglB structure, however a non-structured region of external loop 5 was indicated as the likely interaction domain for glycan recognition[71].

It is evident that the glycan recognition regions of bacterial OSTs have relaxed specificity, as many of the bacterial OSTs studied to date demonstrate promiscuity in glycan recognition. The ability of the *C. jejuni* PglB to transfer O-antigens onto proteins at an appropriately situated asparagine residue has been previously demonstrated[41,73]. The OST from *Desulfovibrio desulfuricans* can similarly transfer O-antigens onto protein, albeit at a lower efficiency than *C. jejuni* PglB. Besides O-antigen promiscuity, the *D. desulfuricans* OST and one of the two *Helicobacter pullorum* OSTs can efficiently transfer the *C. jejuni* glycan[152,153]. Likewise, the *C. jejuni* OST can transfer the pentasaccharide *N*-glycan from *H. pullorum*[153]. Therefore, we hypothesized that other bacterial OSTs would likely exhibit similar promiscuity and be capable of transferring the *C. jejuni* glycan. Further, we hypothesized that some of the bacterial OSTs studied might have altered acceptor site specificity

In these studies, we analyze the ability of numerous bacterial OSTs (listed in Figure 3.1) to transfer the *C. jejuni* glycan onto proteins containing both native and modified glycosylation consensus sequences. Not surprisingly, we found that the OSTs with the greatest homology to *C. jejuni* PglB efficiently transferred the *C. jejuni* glycan onto native as well as engineered *C. jejuni* glycoproteins. A preliminary investigation into the unexpected variations in site specificity of the *W. succinogenes* OST is presented, as well as an unexpected relaxed acceptor site specificity observed for *C. jejuni* PglB. Several of the bacterial OSTs were observed to transfer the *C. jejuni* glycan to an altered

acceptor site that lacked the negatively charged residue at the -2 position. This result suggested some bacteria, such as *Desulfovibrio* strains, have evolved distinct acceptor site specificity. Finally, we performed structure-guided mutagenesis to engineer an active site variant of *C. lari* PglB exhibiting altered acceptor site specificity.

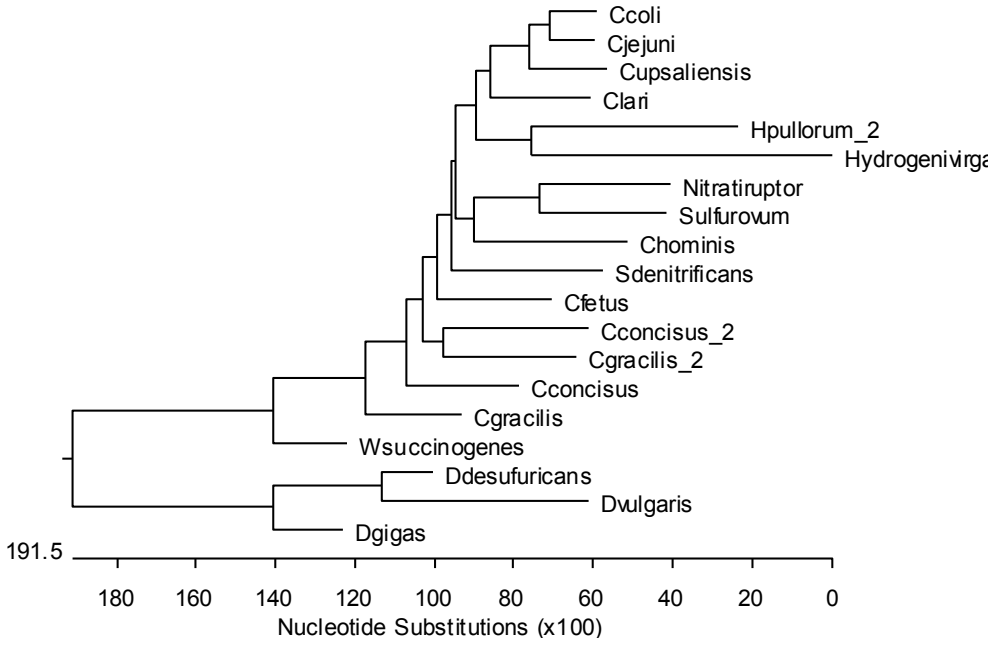


Figure 3.1 Phylogenic tree of OSTs investigated in this study. The tree was created following alignment of OST sequences by the ClustalW method using MegAlign from DNASTAR Lasergene 8.

Results and Discussion

Native C. jejuni glycoproteins

Twenty-three different PglB homologues from seventeen unique bacterial species and two closely related subspecies were tested for their ability to transfer the native *C. jejuni* glycan onto two model glycoproteins from *C. jejuni* (Figure 3.2). The first glycoprotein analyzed was AcrA, a native *C. jejuni* glycoprotein with two native glycosylation sites. Here, we used a version of AcrA called AcrA4x that was engineered to contain two additional glycosylation sites by replacing the residues at the -2 positions of two naturally occurring NXT sites with aspartic acid: F115D, and T145D [143]. The second protein analyzed was Cj0114, a native *C. jejuni* glycoprotein with four native glycosylation sites. Each glycoprotein was expressed bicistronically from a plasmid that also expresses the OST. A C-terminal 6x-His tag was also added to each glycoprotein for ease of detection. The order of the OSTs in Figure 3.2 is based on predicted phylogenetic distance of the OST from *C. jejuni* PglB. *C. coli* encodes the most closely related OST, having 80.8% amino acid sequence identity. Phylogenetic distances are difficult to determine with significant accuracy for the more distantly related OSTs, such as those from *Desulfovibrio* bacteria, which have only 13-15% amino acid identity to *C. jejuni* PglB.

To estimate the efficiency with which the different OSTs transferred the *C. jejuni* *N*-glycan onto the two *C. jejuni* substrates tested here, we performed Western blot analysis. Efficiency can be estimated from a blot using anti-His antibodies, where the number of bands corresponds to the different glycoforms (eg. a-, mono-, tri-, and tetra-

glycosylated) and the intensity of the bands corresponds to the relative amount of protein at that size. The *C. jejuni* oligosaccharide is approximately 1kDa, which allows for distinct separation of a-, mono-, di-, tri-, and tetra-glycosylated proteins.

The OSTs that were most efficient at glycosylating AcrA4x were the OSTs from *C. jejuni, C. coli, and C. concisus* based on the ability of these enzymes to generate predominantly fully glycosylated glycoprotein targets (Figure 3.2). Comparatively, the *C. lari* OST, which has been previously shown to be capable of glycosylating a non-canonical site, $_{257}$DANSGT$_{262}$, for a total of five possible sites in AcrA4x[154], proved to be less efficient at overall glycosylation under these conditions. The efficiency of glycosylation at this non-canonical site is significantly lower than that of the native sites, therefore it is not surprising that is was not detected.

Previous studies indicate that *C. jejuni* PglB should be capable of efficiently glycosylating all four of the sites in Cj0114[153]. In multiple replicates, however, only two of those sites were efficiently glycosylated by *C. jejuni* PglB, with some indication of low-level glycosylation of a third site. The *C. coli* and *C. concisus* OSTs were the only other OSTs investigated here that show evidence of glycosylating more than one of the four sites in Cj0114. Sequencing results confirmed the presence of all four glycosylation sites in this construct, hence the lack of tetra-glycosylated Cj0114 was not the result of a cloning error. It should be noted that while these results are at odds with previously published data[153], they are consistent with our earlier observation that the *C. jejuni* OST is more efficient at glycosylating acceptor sites having an aspartic acid at the -2 position versus a glutamic acid at the -2 position (data not shown). These results are also

supported by our glycophage selection results, where the majority of acceptor site sequences selected by this system contained an aspartic acid in the -2 position[98].

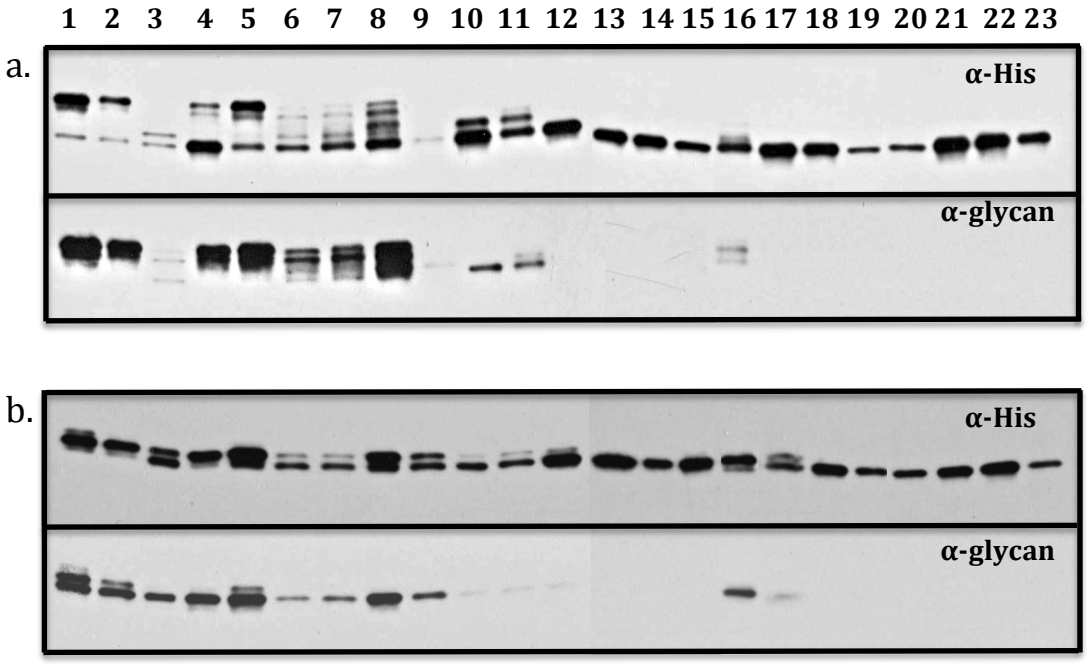


Figure 3.2 Glycosylation of a native and a modified *C. jejuni* substrate by bacterial OSTs. Western blot of periplasmic fraction from CLM24 cells co-expressing pACYCpgl B::Kan with the indicate OST and either (a.) AcrA4x or (b) Cj0114. The bacterial OSTs are denoted as follows: 1 *C. jejuni*; 2 *C. coli*; 3 *C. upsaliensis*; 4 *C. lari*; 5 *C. concisus* ; 6 *C. fetus*; 7 *C. fetus*; 8 *C. gracilis*; 9 *C. hominis*; *10 Wolinella succinogenes*; 11 *Sulfurimonas denitrificans*; *12 Sulfurovum sp* NBC37-1; 13 *Nitratiruptor sp* SB155-2; 14 *C. gracilis** ; 15 *C. concisus**; 16 *H. pullorum* MIT; 17 *H. pullorum*; 18 *H. pullorum* MIT*; 19 *H. pullorum**; 20 *D. gigas*; 21 *D. desulfuricans*; 22 *D. vulgaris*; *23 Hydrogenivirga sp* 128-5-R1-1. An asterisk indicates a second OST from the same organism)

The overall expression level of AcrA4x was low when co-expressed with the OSTs from both *C. upsaliensis* and *C. hominis*, however, the anti-glycan blot verifies that these enzymes were still capable of glycosylating this protein. A possible explanation for this observed low expression level is that the expression of these particular OSTs caused

increased stress on the cells, leading to lower overall expression and/or increased proteolysis of the AcrA-4x target protein.

OSTs from two *C. fetus* subspecies both exhibited the same glycosylation patterns. Specifically, mono- and tri-glycosylated AcrA4x were the prominent glycoforms generated by the *C. fetus* OSTs, indicating that these OSTs may have preferential specificity for particular sites in this protein. Along similar lines, both *C. fetus* OSTs glycosylated only one of the four sites in Cj0114, to approximately the same efficiency, further indicating a particular sequence preference.

C. gracilis, C. concisus and *H. pullorum* all contain a second OST in their genomes, OST*, which is more distantly related to *C. jejuni* PglB, and none of these OSTs were capable of glycosylating the native *C. jejuni* substrates tested here. In agreement with these data, the *H. pullorum* OST* was previously shown to be incapable of glycosylating Cj0114, however in those studies, two of the potential sites in Cj0114 were glycosylated by the *H. pullorum* OST[153]. In these studies, OSTs from both *H. pullorum* subspecies glycosylated only one site in Cj0114, and only the *H. pullorum* MIT subspecies was capable of glycosylating AcrA4x.

The OST from *Sulfurovum sp.* did not show evidence of glycosylating AcrA, but did glycosylate one site in Cj0114, and the OSTs from *Hydrogenivirga sp., Nitratiruptor sp,* and the *Desulfovibrio* bacteria did not prove capable of glycosylating either AcrA4x or Cj0114 with the *C. jejuni* oligosaccharide. These data indicate that these OSTs are (i) not well expressed in *E. coli,* (ii) cannot recognize the *C. jejuni* glycan, or (iii) have variation in their glycosylation site specificities.

The OST from *D. desulfuricans,* which is more homologous to eukaryotic and archaeal OSTs than the *Campylobacter* OSTs are, was recently reported to glycosylate one of the two native sites in AcrA when recombinantly expressed in *E. coli*[155]. Specifically, this group found that the site, $_{271}$DNNNST$_{276}$, was glycosylated at N274. In our *in vivo* experiments, there is no evidence of this site being glycosylated. The likely explanation for this apparent discrepancy is that different expression levels of the OST and glycoprotein could have an impact on the enzyme function. The same cell line and glycosylation machinery were used in the *D. desulfuricans* experiments as were used here, but the glycoprotein expression vector and the OST expression vector were different, where the glycoprotein was produced from a high-expression pET vector, and the OST was expressed from a separate vector, pMLBAD, originally designed for optimal expression in *Burkhoderia*[156].

Collectively, these data support our hypothesis that numerous bacterial OSTs are capable of recognizing the *C. jejuni* glycan and transferring it onto a target protein. One important point is that the glycoproteins tested here contained the generally accepted bacterial consensus sequence[143]. Our goal of identifying bacterial OSTs with variations in acceptor site specificity is descried below.

Variations in acceptor site-specificity of PglB homologues

A particularly interesting result from the glycosylation of native *C. jejuni* substrates was that of the *Wolinella succinogenes* OST, where only mono-glycosylated AcrA4x was observed. In order to determine which site was being glycosylated, four different AcrA'3x' variants were created. These mutants each contained one point mutation that removed the negatively charged residue from the -2 position of one of the acceptor site sequence, D115A (Δ1), D121A (Δ2), D145A (Δ3), or D273A (Δ4). The *W. succinogenes* OST was cloned bicistronically with each of these AcrA'3x' variants. The ability of this OST to glycosylate the different AcrA'3x' proteins was investigated. These experiments revealed that removing the negatively charged residue preceding N117 prevented the *W. succinogenes* OST from glycosylating AcrA'3x' (Figure 3.3). Based on these results, it can be concluded that the *W. succinogenes* OST has altered site specificity. While this OST can glycosylate the asparagine of $_{115}$DENAS$_{119}$, the negatively charged residue is required. The inability to glycosylate the other three acceptor sites of AcrA4x indicates that the D/E-X_1-N-X_2-S/T site alone is not sufficient to promote glycosylation of asparagine by this OST.

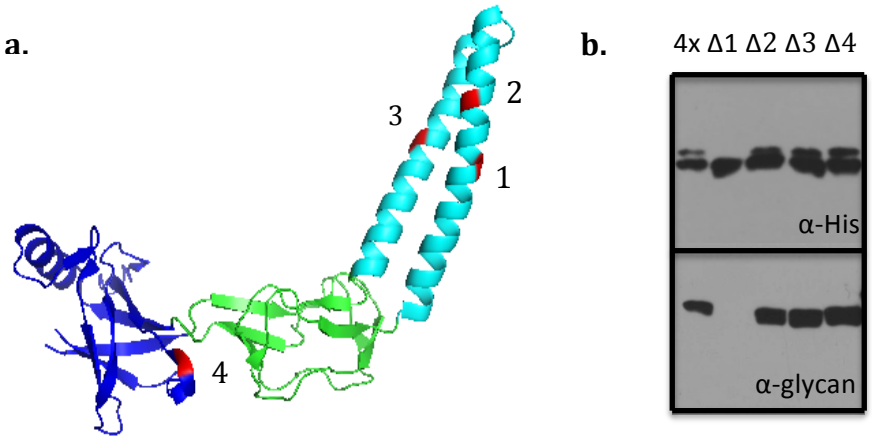


Figure 3.3 Analysis of the glycosylation of AcrA by the *W. succinogenes* OST (a.) Hypothetical structure of AcrA with the four glycosylation sites labeled in red, adapted from the structure of MexA[20]. (b.)Western blot of periplasmic fraction from CLM24 cells expressing *W. succinogenes* OST along with the *C. jejuni* glycosylation operon sans pglB and one of five engineered AcrA constucts, as indicated

A subset of the bacterial OSTs were tested for their ability to transfer the *C. jejuni* glycan onto a mutated version of AcrA4x, called AcrA'0x', for which each of the four glycosylation sites were mutated to contain an alanine at the -2 position of the glycosylation consensus sequence (Figure 3.4a). AcrA'0x' should not be glycosylated by *C. jejuni* PglB given its known acceptor site specificity. However, we were less certain about the *C. lari* PglB homolog because it was shown that this OST has relaxed site specificity compared to the *C. jejuni* OST. Specifically, it was shown that the *C. lari* OST glycosylated the $_{271}$DNNNST$_{276}$ site when D271 was mutated to alanine (D271A) or when N273 was mutated to glutamate of lysine[154]. In our expression system, the *C. lari* OST was not capable of glycosylating AcrA'0x'. Remarkably, however, *C. jejuni* PglB was capable of generating a mono-glycosylated form of AcrA'0x'.

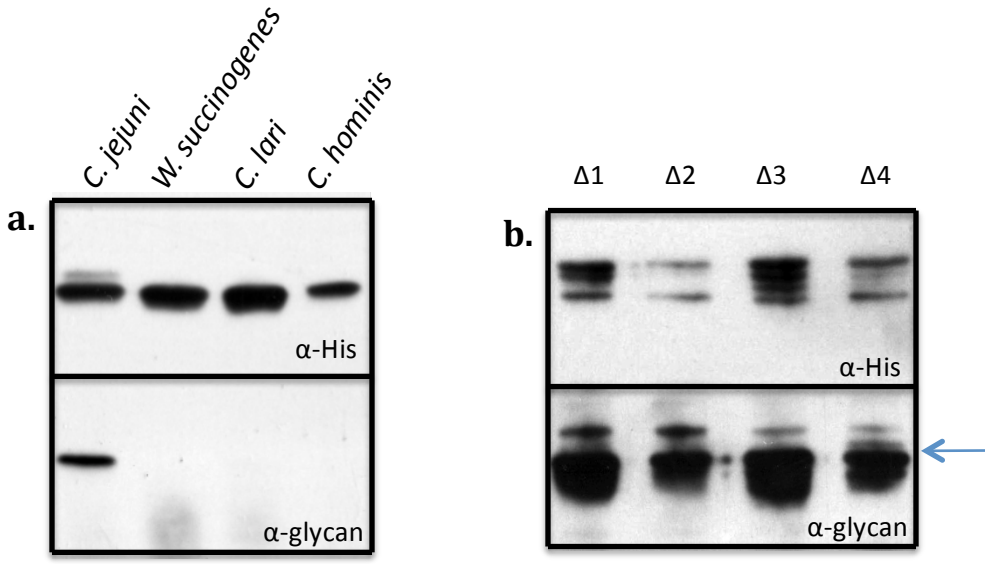


Figure 3.4 Analysis of glycosylation of altered acceptor site by *C. jejuni* PglB. Western blot of periplasmic fraction from CLM24 cells co-expressing the *C. jejuni pgl* locus without PglB and (a.) AcrA'0x' and the OST indicated, or (b.) AcrA'3x' with alanine in the -2 position at the indicated site and *C. jejuni* PglB.

To determine which of the four $AX_1NX_2(S/T)$ sites of AcrA'0x' was glycosylated, *C. jejuni* PglB was tested for its ability to glycosylate the four different AcrA'3x' constructs described above (Figure 3.4b). It was found that $_{271}$ANNNST$_{276}$ site is glycosylated by *C. jejuni* PglB, albeit at lower efficiency than for the other three sites, as indicated by the lack of signal for the fourth band in the anti-His blot. The arrow in the anti-glycan blot of Figure 3.4b denotes the fourth band, where the band above in all four lanes is a background band. This allows us to conclude that this fourth site of AcrA is indeed glycosylated by *C. jejuni* PglB in the absence of a negatively charged residue at the -2 position. This is the first report, to our knowledge, that rejects the hypothesis that a negatively charged, acidic, residue is strictly required for catalysis of *C. jejuni* PglB.

Our original hypothesis was that some bacterial OSTs would contain natural variations in site specificity. To determine if any of the other bacterial OSTs we have investigated have relaxed specificity, each OST was tested for its ability to transfer a glycan onto scFvR4-AQNAT. The scFvR4 protein was selected as the glycoprotein target because of its highly soluble nature[157] and its ability to be glycosylated when appended with a GT at its C-terminus[43]. The scFvR4-AQNAT construct was cloned into pBAD24, and each of the OSTs was cloned into a separate vector, pSF, which introduced a FLAG epitope tag to the C-terminus for detection via Western blot. Our results indicated that the *C. coli* and *C. upsaliensis* OSTs are capable of transferring a glycan onto scFvR4-AQNAT (shown in Figure 3.5 for the *C. coli* OST). Comparing the sequences of the *C. coli* and *C. upsaliensis* OSTs to the *C. lari* OST, several amino acid sequence variations were found in common. Of particular interest is the P325 residue of *C. lari*; just downstream of this residue is the R331 residue believed to help stabilize the acceptor peptide in the active site of the OST (Figure 3.6a). In both the *C. coli* and *C. upsaliensis* OSTs there is an arginine aligned to the R331 residue of the *C. lari* OST, and the residue aligned to P325 of the *C. lari* OST is a leucine. The rigid nature of proline could force the R331 residue into the active site, and a leucine in this position could alter the structure enough to explain the relaxed specificity observed here.

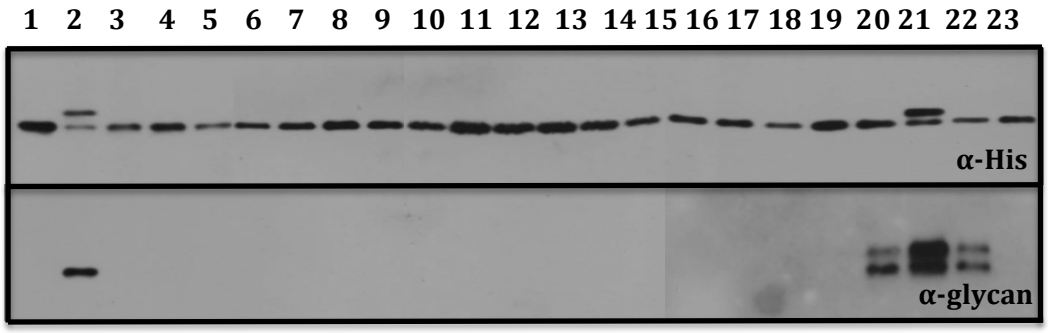


Figure 3.5. Glycosylation of a modified acceptor site by bacterial OSTs. Western blot of periplasmic fraction from CLM24 cells expressing pACYCpgl B::Kan, scFVR4-AQNAT from pBAD24, and pBOST (noted OST). The bacterial OSTs are denoted as follows: 1 *C. jejuni*; 2 *C. coli*; 3 *C. upsaliensis*; 4 *C. lari*; 5 *C. concisus* ; 6 *C. fetus fetus*; 7 *C. fetus venerealis*; 8 *C. gracilis*; 9 *C. hominis*; *10 W. succinogenes*; 11 *S. denitrificans*; *12 Sulfurovum sp* NBC37-1; 13 *Nitratiruptor sp* SB155-2; 14 *C. gracilis** ; 15 *C. concisus**; 16 *H. pullorum* MIT; 17 *H. pullorum*; 18 *H. pullorum* MIT*; 19 *H. pullorum**; 20 *D. gigas*; 21 *D. desulfuricans*; 22 *D. vulgaris*; *23 Hydrogenivirga sp* 128-5-R1-1. (* Second OST from organism)

The most remarkable results presented in Figure 3.5 are those of the OSTs from the three *Desulfovibrio* bacteria. While none of these OSTs were capable of glycosylating native *C. jejuni* substrates, all three *Desulfovibrio* OSTs tested glycosylate not one, but two sites in scFvR4-AQNAT. There is only one AQNAT site engineered in the construct, so these OSTs are clearly glycosylating some other site. Upon closer investigation of the scFvR4 sequence, it is evident that there are a variety of possible sites that could be getting glycosylated. The two asparagine residues with the highest likelihood to be glycosylated by the *Desulfovibrio* OSTs are $_{32}$FSNYS$_{36}$ and $_{75}$RDNAT$_{79}$, both of which contain the standard eukaryotic NX(S/T) motif but lack a negatively charged residue at the -2 position. Site directed mutagenesis combined with mass spectrometry is necessary to confirm which of the sites is glycosylated. As mentioned, the *Desulfovibrio* OSTs are distantly related to the *Campylobacter* OSTs, and more closely related to archaeal and eukaryotic OSTs than are the *Campylobacter* OSTs.

Therefore, it seems logical that the *Desulfovibrio* OSTs may recognize glycosylation consensus sequences that more closely resemble those of the eukaryotes.

These data further support our hypotheses that (i) most bacterial OSTs exhibit broad specificity for the glycan, allowing them to transfer the non-native *C. jejuni* glycan, and (ii) some bacterial OSTs exhibit relaxed acceptor site specificity compared to *C. jejuni* PglB.

Structure-guided alteration to site-specificity

The crystal structure of the *C. lari* OST[71] indicated the likely formation of a salt bridge between R331 of the OST and the aspartic acid residue at the -2 position of the glycosylation acceptor site, with a distance of <4Å between the two residues (Figure 3.6a). To determine whether this residue, and thus the putative salt bridge, is required for glycosylation, we mutated this position in the *C. lari* OST to a neutral residue (e.g., R331G) or a negatively charged residue (R331E). Not surprisingly, mutation of this residue to negatively charged residue abolished glycosylation of the four standard bacterial acceptor sites in AcrA4x containing a negative residue at the -2 position (Figure 3.6b). This result is likely a product of the charge clash introduced by this mutation. However, the *C. lari* PglB mutant carrying a neutral R331G mutation was able to glycosylate AcrA4x as efficiently as the wt *C. lari* OST (Figure 3.6b).

We next tested whether swapping the charge at the -2 position from a negatively charged residue to a positively charge arginine would restore glycosylation in the context

of the *C. lari* R331E OST. For this experiment, an scFvR4-RQNAT construct was created and the *C. lari* R331E OST mutant was tested for its ability to transfer a glycan to this construct (Figure 3.6c). Unfortunately, the *C. lari* R331E OST was unable to glycosylate scFvR3-RQNAT, indicating that simple swapping of the charged amino acids was insufficient to restore glycosylation. Interestingly, the scFVR4-AQNAT construct, which was not glycosylated by the wt *C. lari* OST was successfully glycosylated by the *C. lari* R331E OST (Figure 3.6c). Thus, the R331E mutation appears to have altered the acceptor site specificity of *C. lari* PglB. This mutant enzyme has lost the ability to glycosylate sites with negatively charged residues at the -2 position (e.g., DFNRS), but it has acquired the ability to glycosylate acceptor sites which lack a charged residue in the -2 position of the acceptor site (e.g., AQNAT)

Harnessing the functional diversity in naturally occurring bacterial OSTs and combining this with structure-guided mutations has yielded several interesting OSTs with unique acceptor site specificities not yet reported.

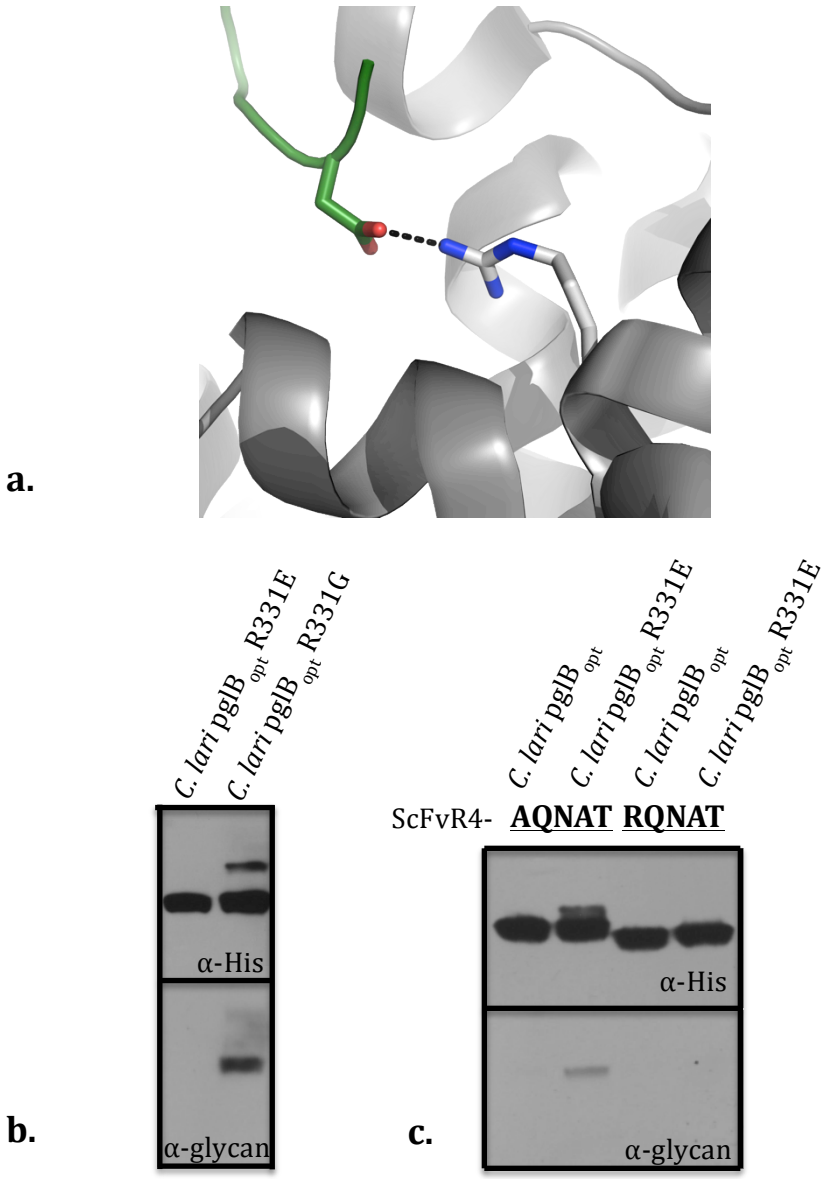


Figure 3.6 Mutational analysis of *C. lari* PglB. (a.) Close-up of the interaction between R331 of *C. lari* PglB (grey) and the aspartic acid in the -2 position of the glycosylation acceptor site (green); the dashed line corresponds to a distance of 2.4Å (from the 3RCE structure in the protein data bank[71]). (b.) Western blot of periplasmic fractions from CLM24 cells co-expressing pACYCpgl B::Kan, the noted *C. lari* OST and either (b.) 4xAcrA or (c.) scFvR4-AQNAT or -RQNAT.

Materials and Methods

Bacterial Strains and Plasmids

The bacterial strains and plasmids used in the study are listed in Table I. *E. coli* were grown at 37°C overnight in Luria-Bertani broth and protein inductions were carried-out at 30°C. Amp (100μg/mL), Cm (20μg/mL), and spectinomycin (Sp, 75μg/mL) were added to the medium as required for plasmid maintenance. Template DNA for the OSTs was obtained from ATCC where available, or provided as a generous gift from the laboratory of Bil Clemons of the California Institute of Technology.

OSTs were cloned using directional cloning between XbaI and SbfI of pBad24, downstream of the acceptor protein sequence in the bicistronic constructs, or following the RBS in the pSF plasmid. The pSF plasmid was constructed by insertion of a truncated MBP sequence, preceded by an XbaI cut site and followed by the coding sequence for a FLAG-tag, with an SbfI cut site in-between into pSN18[43] between NcoI and SalI. The pBS plasmid was constructed for expression of the acceptor protein on a third plasmid by combining the spectinomycin resistance cassette and pSC101 ori from the pZ expression vectors[158], digested with AvrII and AatII with the expression region of pBad24, copied with primers pBadAvrIIfor (5'-AACATACCTAGGATCGATGCATAATGTGCCTGTC-3'), and pBadAatIIrev (5'-AAGATTGACGTCGATGCCTGGCAGTTTATGG-3').

Table 3.1. Strains and plasmids used in this study

Strain or plasmid	Relevant genotype or description	Reference or source
E. coli		
DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	Laboratory stock
CLM24	*E. coli* W3110 (IN(rrnD-rrnE)1 rph-1) *waaL*	Feldman et al.[41]
Plasmids		
pACYCpgl B::kan	pACYC184-based plasmid encoding the *C. jejuni* protein glycosylation locus cluster, where pglB has been replaced with kan^r, Cm^r	Linton et al.[37]
pBAD24	Cloning vector, aribinose-inducible, Amp^r	Guzman et al.[139]
pBS	Cloning vector, arabinose-inducible, Sp^r	This study
pSF	Cloning vector, arabinose-inducible, Amp^r	This study

Introduction of point mutations to AcrA

Point mutations were introduced into AcrA using the QuikChange II site directed mutagenesis kit (Agilent Technologies Inc., Santa Clara, CA). Complementary primers were designed to mutate the aspartic acid residue to an alanine for each of the four glycosylation consensus sequences in AcrA using the primer design software provided through the Agilent website.

Preparation of periplasmic fractions

An overnight culture, grown in the appropriate antibiotics and 0.2% glucose, was sub-cultured 1:100 into media containing the appropriate antibiotics and 0.02% glucose. Following growth at 37°C, shaking, to OD_{600} ~0.5, cultures were induced with 0.4% arabinose and grown overnight at 30°C. Induced cultures were normalized by OD_{600} and

resuspended in ice-cold Buffer F (20% sucrose, 30 mM Tris pH 8.5, 1 mM EDTA, 1 g/L lysozyme) to a final concentration of 10 OD_{600}/mL and incubate at 4°C with gentle agitation for 2hr. Following centrifugation at 10,000xg for 10 minutes, the supernatant was collected as the periplasmic fraction.

Western blot analysis

Expression and glycosylation of AcrA, Cj0114 and scFv13-R4-GT was analyzed by immunoblot following SDS–PAGE of the periplasmic fraction. Immunodetection was performed with monoclonal anti-His antibody (Abcam, Cambridge, MA), and polyclonal anti-glycan serum hR6 (S. Amber and M. Aebi, personal communication).

Conclusions

We have demonstrated the ability of *E. coli* to express functional OSTs from a variety of bacterial species, including nine *Campylobacter* species, *W. succinogens, S. denitrificans, Sulfurovum sp, H. pullorum* and three *Desulfovibrio* species. As far as we are aware, this is the first report of the functional expression in *E. coli* for many of these OSTs, including the OSTs from *C. upsaliensis*, *C. concisus*, *C. fetus*, *C. gracilis*, *C. hominis*, *W. succinogens, S. denitrificans, Sulfurovum sp*, *D. gigas*, and *D. vulgaris.* Interestingly, OSTs from the more distantly related *H. pullorum*, particularly the MIT subspecies, exhibited efficient glycosylation of the *C. jejuni* substrates.

OSTs from the *Desulfovibrio* species did not prove capable of glycosylating the native *C. jejuni* substrates, contrary to published data[155], likely due to differences in expression conditions. However, the OSTs from each of the three *Desulfovibrio* species were capable of glycosylation of the scFvR4-AQNAT construct at two sites. While the specific identity of the individual sites that were glycosylated by these OSTs were not elucidated in these studies, these data provide evidence that the *Desulfovibrio* OSTs recognize a glycosylation consensus sequence that is different from that of the well-characterized *C. jejuni* OST.

A variety of different factors influence the function of OSTs, from expression levels of the various components of the system, to the specific induction conditions implemented, making engineering *E. coli* to reliably perform *N*-linked glycosylation a nontrivial task. Because *E. coli* are well understood and thoroughly characterized, overcoming these challenges will prove possible. One way to improve the reliability of

this system is to integrate the major components of the glycosylation machinery into the genome of an appropriate *E. coli* strain, increasing the stability of the glycosylation machinery by decreasing plasmid stress on cells.

The data presented here is a preliminary analysis of the functional characterization of bacterial OSTs in *E. coli.* Suggestions for the future directions of this work are presented in Chapter 5 of this thesis.

CHAPTER 4
A PROKARYOTE-BASED CELL-FREE TRANSLATION SYSTEM THAT EFFICIENTLY SYNTHESIZES GLYCOPROTEINS

Introduction

[1]Cell-free protein-synthesizing systems are emerging as an attractive alternative to conventional expression systems that rely on living cells (for a review, see [159]). This is because, over the past decade, cell-free protein synthesis reactions: (i) can be completed in less than a day; (ii) use reagents whose costs are down; (iii) fold complex proteins by routinely forming disulfide bonds; and (iv) can be scaled to 100 L. Two main approaches have been used for *in vitro* transcription/translation: one is based on cell-free extracts (CFEs), often derived from *Escherichia coli*, rabbit reticulocytes or wheat germ and the second is based on reconstituted protein synthesis from purified components[160].

Because of their ability to co-activate multiple biochemical networks in a single integrated platform[161], cell-free systems are increasingly used in many important bio-technology and synthetic biology applications [162–164].

The ability to accurately and efficiently glycosylate proteins in a cell-free system would have advantages for many areas of basic and applied research, especially given the importance of *N*-linked glycosylation in protein folding, quality control, sorting, degradation, secretion and activity[165]. Unfortunately, the best characterized and most

[1] Adapted with permission from Guarino, C. & DeLisa, M. P. A Prokaryote-Based Cell-Free Translation System That Efficiently Synthesizes Glycoproteins. *Glycobiology* **22**, 596–601 (2012).

widely used cell-free translation systems based on *E. coli* are incapable of making glycoproteins because *E. coli* lack glycosylation machinery. Likewise, rabbit reticulocyte and wheat germ CFE systems cannot perform this post-translational modification because they lack microsomes[166]. This can be overcome by supplementing eukaryotic CFEs with microsomal vesicles (e.g. canine pancreas microsomes[167,168]), but the resulting systems do not always faithfully process the target protein due to poor compatibility between some CFEs and microsomal vesicles[168,169]. An alternative strategy for creating a cell-free translation system that can execute *N*-linked glycosylation is to prepare CFEs from specialized cells such as hybridomas[170], trypanosomes[169], insect cells[166] or mammalian cells[171]. However, these systems are technically difficult to prepare and typically result in inefficient glycosylation and low product yields. Moreover, in all the above systems, the glycosylation process is effectively a "black-box" and thus difficult to control.

To address these issues, two novel cell-free translation/glycosylation systems—termed "glycoCFE" and "glycoPURE"— were created here by combining existing *in vitro* translation systems with a reconstituted *N*-linked glycosylation pathway. Purified glycosylation components were derived from the protein glycosylation locus (pgl) present in the genome of the Gram-negative bacterium *Campylobacter jejuni*. This gene cluster encodes an *N*-linked glycosylation system that is functionally similar to that of eukaryotes and archaea, involving an OST that catalyzes the en bloc transfer of preassembled oligosaccharides from lipid carriers onto asparagine residues in a conserved motif [N-X+1-S/T in eukaryotes and D/E-X−1-N-X+1-S/T in bacteria[42], where X−1 and X+1 are any residues except proline] within polypeptides (Figure 1.1). Several observations suggested that the *C. jejuni* glycosylation machinery was ideally suited for

use in a cell-free translation/glycosylation system. First, *E. coli* transformed with the entire pgl gene cluster can perform *N*-linked protein glycosylation[17], thereby providing a convenient host for producing the necessary components in a pure and active form. Since *E. coli* lacks native glycosylation machinery, the potential for contamination from background *N*- or *O*-linked systems is eliminated. Second, *C. jejuni* OST, named PglB (*Cj*PglB), is a single-subunit enzyme that is active when solubilized in detergent[71] and, unlike the homologous STT3 subunit of higher organisms, does not require any accessory components for its activity. Third, *Cj*PglB can transfer sugars post-translationally to locally flexible structures in folded proteins[70], suggesting that protein glycosylation can be achieved without supplementing a functional membrane system (e.g. microsomes).

Results

Preparation of N-linked glycosylation components

To begin, functional reconstitution of bacterial *N*-linked glycosylation *in vitro* was attempted. Minimally, this required three components: an OST, a lipid-linked oligosaccharide (LLO) and an acceptor protein carrying the D/E-X−1-N-X+1-S/ T motif. For the OST, CjPglB was expressed in the membrane fraction of *E. coli* cells, solubilized with 1% N-dodecyl-β-D-maltopyranoside (DDM) and purified to near homogeneity by nickel affinity chromatography followed by gel filtration (Supplementary data, Figure S4.1). Separately, *E. coli* cells carrying the *C. jejuni* pgl locus were used for producing the oligosaccharide donor. This gene cluster encodes enzymes that carry out the biosynthesis of a GlcGalNAc5Bac heptasaccharide (where Bac is bacillosamine) and its

transfer from membrane-anchored UndPP to asparagine residues. Here, a modified version of this cluster that carried an inactivated pglB gene[17] was transferred to *E. coli* SCM6 cells and used to prepare LLOs. SCM6 cells were chosen for several reasons. First, these cells lack the WaaL enzyme that naturally transfers oligosaccharides (e.g. O-antigens, glycans) from the lipid carrier UndPP onto lipid A, which in turn shuttles the oligosaccharides to the outer leaflet of the outer membrane[41]. Thus, in the absence of WaaL, the desired lipid-linked glycans accumulate in the inner membrane. Second, the LPS and enterobacterial common antigen initiating GlcNAc transferase, WecA, is removed. Thus, this strain should only produce LLOs with GlcGalNAc5Bac at the reducing end. In support of this notion, previous mass spectrometry analysis of LLOs extracted from an *E. coli* strain similar to the one used here (i.e. ΔwaaL ΔwecA) revealed that only LLOs containing GlcGalNAc5Bac heptasaccharide were detected[172]. For the oligosaccharide acceptor, the model glycoprotein AcrA from *C. jejuni*[148] was purified from the periplasm. AcrA presents two consensus D/E-X1-N-X2-S/T sites that are glycosylated by CjPglB[42]. Alternatively, a glycoengineered single-chain variable fragment (scFv) called scFv13-R4-GT, which carried a C-terminal glycosylation tag (GT) consisting of four consecutive DQNAT motifs separated from one another by consecutive glycine residues[42], was similarly purified.

Functional reconstitution in vitro of the C. jejuni protein glycosylation pathway

To evaluate the reconstituted glycosylation pathway, CjPglB OST was combined with LLOs extracted from *E. coli* cells and purified AcrA. This reaction resulted in

efficient glycosylation of both AcrA sites as evidenced by the mobility shift of nearly all of the AcrA from the unmodified (g0) to the fully glycosylated (g2) form (Figure 4.1A). This activity was dependent on PglB and LLOs. Doubling the LLO concentration resulted in the appearance of the g0 and g1 forms of AcrA, in addition to g2, suggesting slightly less efficient glycosylation. Importantly, glycosylation activity was lost when lipid extracts from cells lacking the pgl cluster or an inactive CjPglB mutant was used (Figure 4.1A). These results were corroborated by detecting glycosylated AcrA with serum specific for the *C. jejuni N*-glycan (Figure 4.1A). Nearly identical results were observed when the glycoengineered scFv13-R4-GT protein was used as the oligosaccharide acceptor (Figure 4.1A). It should be noted that g2, g3 and g4 were the predominant glycoforms detected here, with barely detectable levels of g1. To demonstrate that other OSTs could be used in this system, *in vitro* glycosylation of AcrA was also performed using Campylobacter lari PglB (ClPglB), which is 56% identical to that of *C. jejuni*[154]. This resulted in nearly equal amounts of the g0, g1 and g2 forms of AcrA under the conditions tested (Figure 4.1B). To be useful for translation/glycosylation reactions, the purified glycosylation components must tolerate long-term storage and freeze-thaw cycles. To test this, the components were stored separately at −20°C for 3 months. No changes were made to the storage buffers except that the final concentration of glycerol in the PglB samples was increased to 10%. Each of the components was thawed and refrozen 5–10 times during this period, after which an *in vitro* reaction with ClPglB was performed. This reaction yielded the glycosylation of AcrA that appeared to be only slightly less efficient than the glycosylation observed with freshly purified components (compare Figure 4.1B and C).

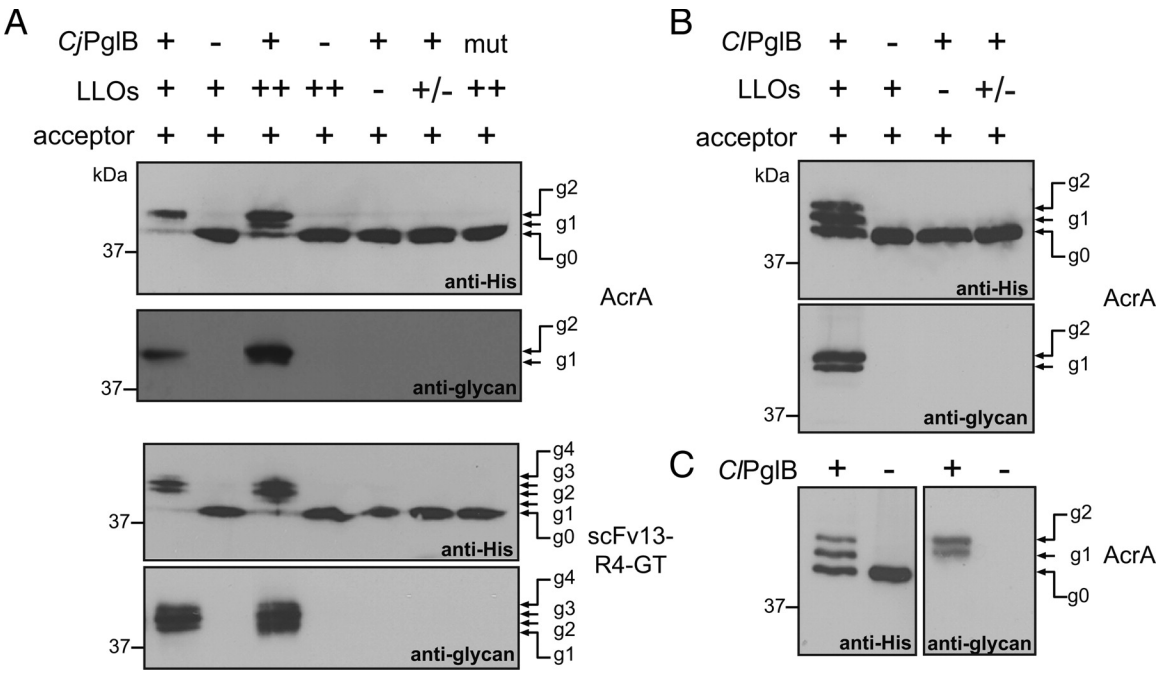


Figure 4.1. Reconstituted glycosylation with defined components. (**A**) *In vitro* glycosylation assay using purified OST, extracted LLOs and purified acceptor proteins produced in *E. coli*. Immunoblots detecting acceptor protein AcrA and scFv13-R4-GT (both anti-His) or glycans (anti-glycan). Reactions included 3 μg wild-type *Cj*PglB, 5 (+) or 10 (++) μL of LLOs and 5 μg of acceptor protein. Controls included the omission of different components (−), inactivated PglB (mut) and LLOs from SCM6 cells with empty pACYC (+/−). Glycosylation yields a mobility shift from the unmodified (g0) to the glycosylated forms (g1 and g2). (**B**) Same as in (A) but with purified *Cl*PglB. (**C**) Immunoblot detecting AcrA following *in vitro* glycosylation using 3-month-old freeze-thawed components.

Cell-free translation of protein targets

To determine whether existing cell-free translation systems could synthesize protein targets of interest, both an *E. coli* CFE-based protein synthesis system and the PURE (protein synthesis using recombinant elements) system that uses purified translation components and T7 RNA polymerase[160] were evaluated. This involved

priming the CFE and PURE systems with three different AcrA DNA sequences cloned in a T7 promoter-driven pET vector. Using the CFE system, ~150–250 μg/mL of each AcrA variant was produced as a full-length polypeptide in 1 h (Figure 4.2A). AcrA carrying its native signal peptide accumulated to the highest level but also experienced the greatest amount of degradation. In contrast, AcrA carrying a PelB signal peptide in place of the native signal and AcrA lacking a signal peptide each accumulated to a slightly lower concentration but experienced no visible degradation. The PURE system similarly produced all three AcrA variants as full-length polypeptides albeit at a slightly lower level (~100 μg/mL/h of each) than the CFE-based system (Figure 4.2A). Both systems were also able to generate appreciable amounts of scFv13-R4-GT (Supplementary data, Figure S4.2a). It should be noted that this scFv was previously optimized for expression under non- oxidizing conditions (i.e. in the absence of disulfide bonds)[173] and thus did not require special transcription/translation conditions.

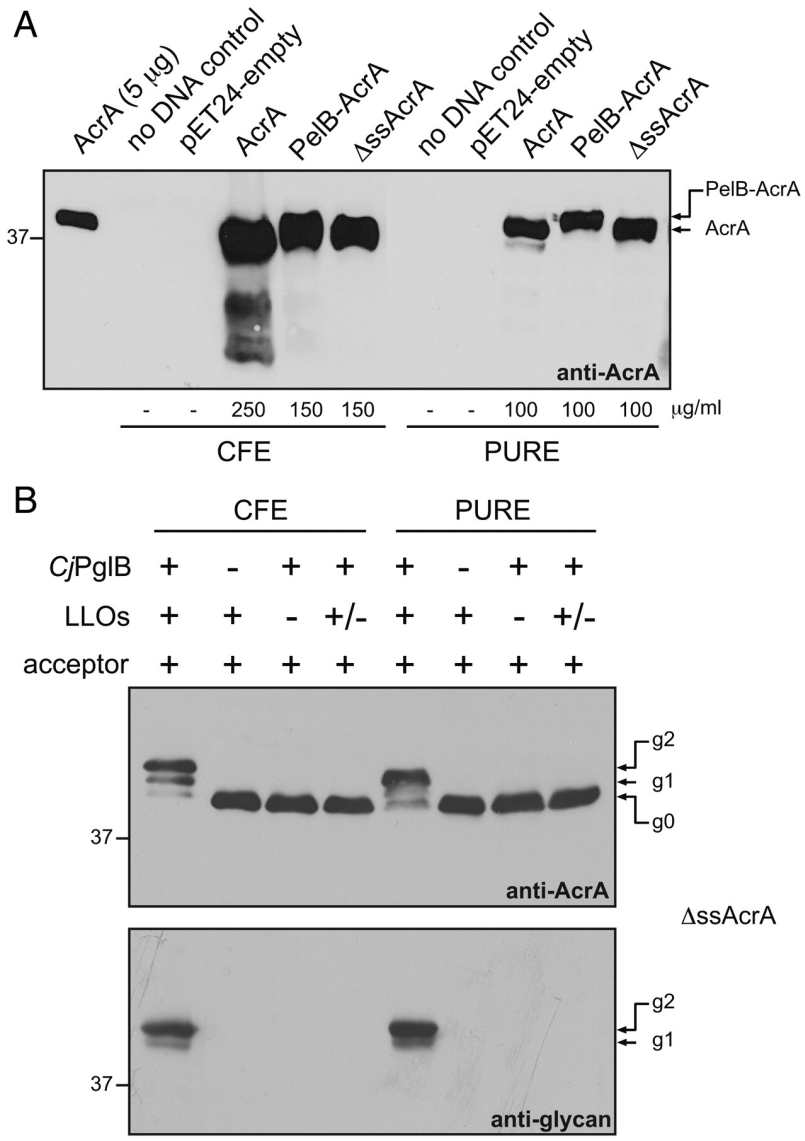


Figure 4.2. Cell-free translation/glycosylation of AcrA. (**A**) Immunoblot detecting different AcrA constructs (anti-AcrA) produced by *in vitro* translation using either *E. coli* CFEs or purified translation components (PURE). AcrA concentration was estimated by comparing band intensities to that of purified AcrA loaded in lane 1. (**B**) Immunoblot detecting ΔssAcrA expression (anti-AcrA) and glycosylation (anti-glycan). ΔssAcrA was produced by cell-free translation/glycosylation using either the CFE or the PURE systems that were primed with pET24(AcrA-cyt). Controls included the omission of different components (−) or LLOs from SCM6 cells with empty pACYC (+/−).

Cell-free translation and glycosylation of target glycoproteins

Encouraged by these results, the glycoCFE and glycoPURE translation/glycosylation systems were constructed by combining the purified glycosylation components (minus the acceptor protein) with one of the cell-free translation systems. The plasmid pET24 (AcrA-cyt) that encodes AcrA without an N-terminal signal peptide was chosen to evaluate these systems because it gave rise to significant amounts of target protein in both translation systems with no detectable degradation. When either the CFE or the PURE system were primed with this plasmid along with CjPglB and LLOs, AcrA was produced primarily as the doubly glycosylated g2 glycoform with lesser amounts of g1 and virtually no detectable unmodified AcrA (Figure 4.2B). It was estimated that ~100–150 μg of glycosylated AcrA was produced in a 1 mL reaction volume after 12 h. Likewise, scFv13-R4-GT was efficiently produced by both the glycoCFE and glycoPURE systems, with ~50% of the protein in the fully glycosylated g4 form and 50% in the g3 form (Supplementary data, Figure S4.2b). Both systems produced ~50–100 μg/mL of glycosylated scFv13-R4-GT in 12 h. Thus, the glycoCFE and glycoPURE systems contain all the components essential for efficiently translating *N*-linked glycoproteins.

Discussion

A major advantage of the open prokaryote-based translation/ glycosylation systems developed here is that the supply of purified glycosylation components as well as their substrates and cofactors[71] can be provided at precise ratios. Likewise, the concentration of inhibitory substances such as proteases and glycosidases that catalyze the hydrolysis of glycosidic linkages can be reduced or eliminated entirely. Additionally, the *in vitro* systems permit the introduction of components that may be incompatible with *in vivo* systems such as certain LLOs that cannot be produced or flipped in vivo. This level of controllability is unavailable in any previous translation/glycosylation system and is significant for several reasons.

First, it helps to avoid glycoprotein heterogeneity, which is particularly bothersome in fundamental studies to assess the contribution of specific glycan structures or in pharmaceutical glycoprotein production. Along these lines, the glycoCFE and glycoPURE systems should allow the examination of factors that interact with or stimulate the glycosylation machinery and promote increased acceptor site occupancy. While the glycosylation efficiency observed here with CjPglB exceeded the level typically observed in vivo[42,69,70] it should be pointed out that further study of the reaction conditions should lead to increases in productivity and glycosylation efficiency. Second, it facilitates the integration/co-activation of multiple complex metabolic systems and pathways in vitro including transcription, translation, protein folding and glycosylation. Therefore, the glycoCFE and glycoPURE systems should provide a unique opportunity for studying the interplay of these important mechanisms under conditions where system complexity is reduced and structural barriers are removed. For instance, since the

bacterial OST can glycosylate locally flexible structures in folded proteins[70] and also structured domains of some proteins[174], these systems should help to decipher the influence of protein structure on glycosylation efficiency.

Also, since bacterial and eukaryotic glycosylation mechanisms display significant similarities, these bacterial systems could provide a simplified model framework for understanding the more complex eukaryotic process. Third, it allows for further customization of the system by re- constituting additional or alternative steps (both natural and unnatural) in the glycosylation pathway. For instance, the sequential activities of the glycosyltransferases in the pgl pathway have been reconstituted in vitro[46] and could easily be integrated with the translation/glycosylation reactions into a single integrated platform. While glycoengineered *E. coli* have the potential to provide a wide array of UndPP-linked glycans[41,175], the ability to extend beyond bacterial glycans could be achieved by supplementation with specific glycosyltransferases and the requisite activated sugars. This approach could be used for making eukaryotic glycan mimetics[176] and would allow finer control over the diversity of glycoforms that can be used for modifying target proteins in vitro. Since CjPglB has relaxed specificity toward the glycan structure[41], all of these UndPP-linked glycans are likely to be suitable substrates. Even if CjPglB should prove insufficient, the demonstration here that two different OSTs could be used interchangeably suggests that virtually any single-subunit OST including those from other bacteria, archaea and even some eukaryotes[61] could be used in these systems. In support of this notion, the Leishmania major and Pyrococcus furiosus single-subunit OSTs can be functionally expressed in *E. coli* membranes[177,178].

Finally, because one is not limited to natural glycans, the glycoCFE and

glycoPURE systems should permit synthesis of hybrid natural/unnatural or even completely artificial glycans. For example, the addition of synthetic sugar-nucleotide donor substrates and/or mutant glycosyltransferases and OSTs having new specificities should enable the construction of a glycosylation system founded on a noncanonical glycan code. For all of these reasons, we anticipate that the glycoCFE and glycoPURE systems will be useful additions to the cell-free translation and glycobiology tookits alike.

Materials and methods

Protein purification

For the purification of CjPglB, *E. coli* strain C43(DE3) (Lucigen, Middleton, WI) was freshly transformed with plasmid pSN18[70], a modified pBAD expression plasmid encoding *C. jejuni* pglB with a C-terminal decahistidine affinity tag. Cells were grown in 1.5 L of terrific Broth supplemented with 100 μg/mL of Amp at 37°C. When the optical density (A600) of the culture reached ~1.0, cells were induced by the addition of 0.02% arabinose (w/v) for 4.5 h at 30°C. All following steps were performed at 4°C unless specified differently. Cells were harvested by centrifugation, resuspended in 25 mM Tris, pH 8.0, and 250 mM NaCl and lysed by three passages through a French press (SLM-Aminco; 10,000 PSI, SLM Instruments, Inc., Urbana, IL). Following the removal of cell debris by centrifugation, the membrane fraction was isolated by ultracentrifugation at 100,000 × g for 1 h. Membranes containing PglB were resuspended in 25 mM Tris–HCl, pH 8.0, 250 mM NaCl, 10% glycerol (v/v) and 1% DDM (w/v) (DDM, Anatrace, Affymetrix, Inc., Santa Clara, CA) and incubated for 2 h. The insoluble fraction was

removed by ultracentrifugation at 100,000 × g for 1 h. All subsequent buffers contained DDM as the detergent. The solubilized membranes were supplemented with 10 mM imidazole, loaded onto a Ni-NTA superflow affinity column (Qiagen, Valencia, CA) and washed with 60 mM imidazole before PglB was eluted with 200mM imidazole. The purified protein was then injected onto a Superdex 200 gel filtration column using AKTA-FPLC (GE Healthcare, Waukesha, WI). Eluate fractions were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS–PAGE) and stained with Coomassie blue to identify the fractions containing PglB (Supplementary data, Figure S4.1). The protein was desalted with a PD10 desalting column (GE Healthcare) into 20 mM Tris, pH 7.5, 100 mM NaCl, 5% glycerol (w/v) and 0.05% DDM (w/v) and concentrated to 5–10 mg/mL in an Amicon centricon with a molecular mass cutoff of 100kDa. Expression and purification of the inactive CjPglB mutant was performed identically in C43(DE3) cells except carrying plasmid pSN18.1, which encodes an inactive copy of pglB subcloned from pACYCpglmut (see below), were used. ClPglB was purified from BL2-Gold(DE3) cells (Stratagene, La Jolla, CA) carrying plasmid pSF2 as described elsewhere[71]. For long-term storage at −20°C, the glycerol content in PglB samples was increased to 10% (w/v).

Purification of AcrA and scFv13-R4-GT was from periplasmic fractions isolated from BL21(DE3) cells carrying plasmid pET24(AcrA-per)[148] or pET24-ssDsbA- scFv13-R4-GT (see below). Periplasmic extracts were prepared as described previously[154], supplemented with imidazole to reach a final concentration of 10 mM, sterile- filtered (0.22 μm) and purified by nickel affinity chromatography using Ni-NTA superflow affinity column (Qiagen, Valencia, CA).

Isolation of LLOs *Escherichia coli* SCM6 cells transformed with pACYCpglmut[17], which codes for the biosynthesis of the *C. jejuni* LLO and an inactivated *C. jejuni* pglB gene (W458A and D459A), were grown in 1 L of Luria-Bertani supplemented with 25 μg/mL of Cm at 37°C. When the A600 reached ~1.0, cells were harvested by centrifugation and the pellet was lyophilized to dryness for 20 h at −80°C and 0.04 mbar. All subsequent steps were performed using glass tubes and glass pipettes. Homogenized pellets were extracted in 25 mL of 10:20:3 CHCl3:MeOH:H2O followed by centrifugation at 3000×g for 30min. The supernatants were evaporated using a rotary evaporator (Büchi, Flawil, Sankt Gallen, Switzerland), after which the resulting pellet was resuspended in 1 mL of 10:20:3 CHCl3:MeOH:H2O and sonicated until homogenous. The sample was dried under nitrogen gas at 37° C, dissolved in 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.5, 1 mM MnCl2 and 0.1% DDM (w/v) and stored at −20°C. An identical procedure was followed to extract lipids from SCM6 cells carrying empty pACYC.

Cell-free translation and glycosylation

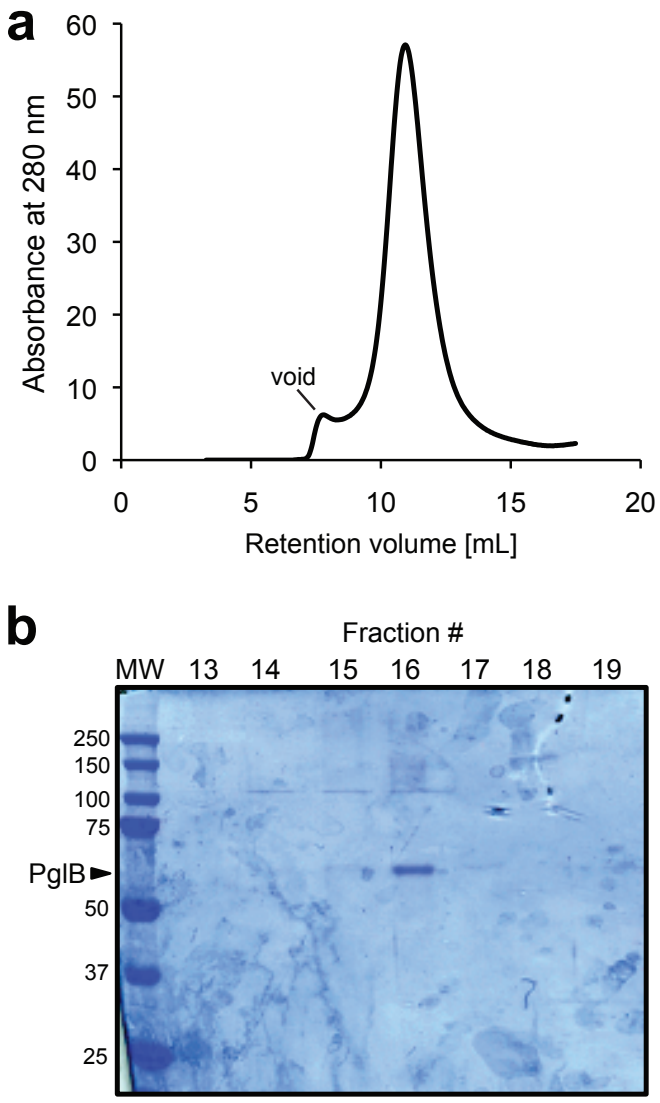
For in vitro glycosylation of purified acceptor proteins, a 50 μL solution containing 3 μg of purified PglB, 5–10 μL of extracted LLOs and 5 μg of purified AcrA or scFv13-R4-GT in 10 mM HEPES, pH 7.5, 1 mM MnCl2 and 0.1% DDM (w/v) was incubated for 12 h at 30°C. For in vitro translation of AcrA and scFv13-R4-GT in the absence of glycosylation, a 50 μL reaction was prepared using the S30 T7 High-Yield Expression System (Promega, Fitchburg, WI) or PURExpress (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions. A total of 1 μg of the following plasmids were added to each reaction: pET24b (Novagen, Madison, WI);

pET24-AcrA encoding full-length *C. jejuni* AcrA with a C-terminal hexahistidine tag[148]; pET24(AcrA-per) encoding a version of AcrA with an N-terminal PelB signal peptide in place of its native export signal[148]; pET24(AcrA-cyt) encoding a version of AcrA without an N-terminal export signal (ΔssAcrA)[148] and pET24- ssDsbA-scFv13-R4-GT encoding the expression-optimized scFv13-R4 intrabody gene[173] with an N-terminal signal peptide from *E. coli* DsbA for secretion and a C-terminal GT[69] followed by a FLAG and a hexahistidine epitope tag. For in vitro translation/glycosylation reactions, 50 μL of translation reactions was supplemented with 3 μg purified PglB, 5 μL extracted LLOs, 1 μg purified plasmid DNA, 1 mM MnCl2 and 0.1% DDM (w/v) and incubated for 12 h at 30°C. DDM was chosen for in vitro translation/glycosylation because it was previously observed to be well tolerated in an *E. coli*-derived CFE system[179].

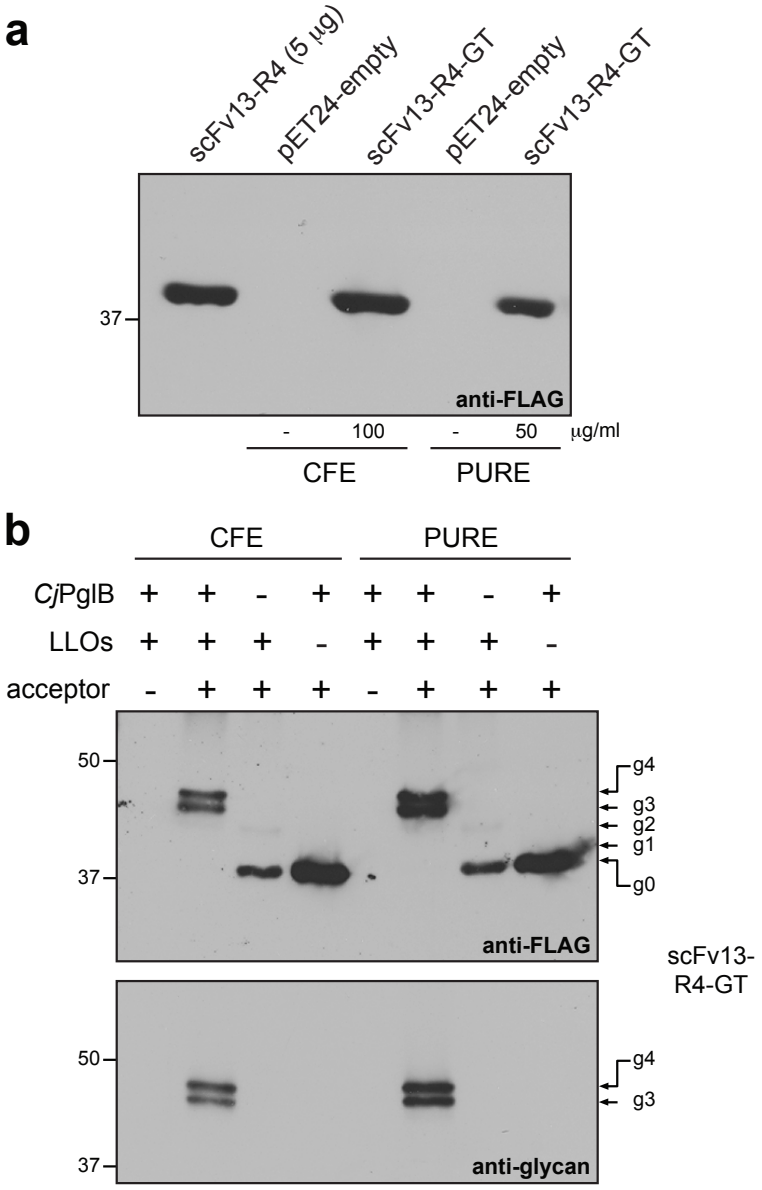
Western blot analysis

Expression and glycosylation of AcrA and scFv13-R4-GT was analyzed by immunoblot following SDS–PAGE. Immunodetection was performed with monoclonal anti-His antibody (Qiagen, Valencia, CA), monoclonal anti-FLAG antibody (Abcam, Cambridge, MA), polyclonal anti-AcrA serum[17] and polyclonal anti-glycan serum hR6[177]. All in vitro translation samples were treated with RNase A (Roche Diagnostics GmbH, Mannheim, Germany) prior to SDS–PAGE to reduce the irregularity of gel electrophoresis due to excess RNA. All experiments were performed at least in triplicate, and representative samples are shown.

Supplementary data



Supplemental Figure S4.1. Purification of bacterial OST. *CjPglB* was expressed in *E. coli* C43(DE3) cells and purified to near homogeneity. Elution fractions (as indicated) from gel filtration columns were examined by SDS-PAGE, and the Coomassie Blue-stained gel images **(b)** are shown together with the elution profiles **(a).** MW, molecular weight standard.



Supplemental Figure S4.2. Cell-free translation/glycosylation of scFv13-R4-GT. (a) Immunoblot detecting different scFv13-R4-GT (anti-FLAG) produced by *in vitro* translation using either *E. coli* cell-free extracts (CFE) or purified translation components (PURE). Estimates of the scFv13-R4-GT concentration were determined by comparison of band intensities to that of the purified scFv13-R4-GT sample loaded in lane 1. **(b)** Immunoblot detecting scFv13-R4-GT expression (anti-FLAG) and glycosylation (anti-glycan). The scFv13-R4-GT protein was produced by cell-free translation/glycosylation using either the CFE or PURE systems that were primed with pET24-ssDsbA- scFv13-R4-GT. Controls included omission of different components (-).

CHAPTER 5
CONCLUSIONS AND FUTURE DIRECTIONS

Significant progress has been made in our understanding of the process of *N*-linked glycosylation since the advent of the works described in this thesis. It is now clear that a wide variety of bacterial species have the ability to *N*-glycosylate proteins[26], and we have shown that many of the OSTs from these organisms function in *E. coli*. Further, it is now possible to produce the core $Man_3GlcNAc_2$- structure of the eukaryotic glycan in *E. coli*[23], and we have developed methodologies that could be extended to select for functional variants of OSTs that more efficiently transfer that glycan onto protein. Finally, *in vitro* assays for OST activity have been developed[150], and our *in vitro* system further simplifies the model framework for glycosylation[43]. Here, I present several ideas for future work in this field, followed by a summary of the works presented thus far.

Improving therapeutic protein production in *Escherichia coli*

Unlike mammalian cells and yeast, *E. coli* do not require glycosylation in order to survive. This along with their ease of culture makes them the ideal candidate for engineering glycosylation machinery to produce both specific and unique glycoproteins. Examples of modifications that could be engineered into *E. coli* to improve their therapeutic protein production potential are elucidated here.

Two of the main challenges associated with the production of therapeutic proteins are their lack of stability when stored over prolonged periods and rapid degradation and/or clearance from the body once administered. One solution that has begun to overcome these challenge is PEGylation. PEGylation is a process in which polyethylene glycol (PEG) chains are added to a protein, effectively increasing the protein's molecular weight and shielding it from degradation by proteolytic enzymes, leading to prolonged shelf-life and increase half-life *in vivo*[180]. The original PEGylated proteins were heterogeneous mixtures, with different numbers of PEG molecules and linkages, but regulations now require specific homogeneous conjugations to achieve approval[181]. Currently, PEG can be specifically added to the *N*-terminus, free cysteine residues[182], unnatural amino acids [183], disulfide bridges[184], the *C*-terminus[185], and glutamine residues[186]. However, it is possible that certain modifications, such as the addition of cysteine residues or modified amino acids, could negatively impact protein function. Another possibility for site-directed PEGylation of proteins is through attachment to a glycan. This process is known as glycoPEGylation and can be used to incorporated PEG at a specific location in a protein normally occupied by a glycan, and therefore less likely to interfere with the active site[181].

GlycoPEGylation has been demonstrated by enzymatic conjugation *in vitro* of a GalNAc monosaccharide to specific serine or threonine residues of non-glycosylated proteins produced from *E. coli*. The GalNAc monosaccharides are then conjugated to PEG by enzymatic transfer of a sialylated PEG molecule[187]. A similar, though more complex, procedure was developed for the glycoPEGylation of complex mammalian *N*-linked glycans attached to glycoproteins expressed from CHO cells[188]. Recently, a

protein modified using this technology has entered clinical trials[189]. Utilizing the glycosylation capabilities of engineered *E. coli,* it should be possible to create *N*-linked glycans than can be easily linked to PEG for increased stability of therapeutic proteins.

The native WecA protein from *E. coli* is capable of transferring a GlcNAc to the UndPP lipid carrier[23,190], and it has been demonstrated that this residue can replace bacillosimine as the aspargine-linked sugar residue of the *C. jejuni* oligosaccharide[191]. Therapeutic proteins should not contain bacillosimine, as this monosaccharide is not found in humans, and it would likely prove to be immunogenic. PglA is the enzyme in the *C. jejuni* glycosylation pathway that transfers the first GalNAc residue onto bacillosimine[192], and based on the successful construction of the *C. jejuni* oligosaccharide without bacillosimine, it is evident that it can also transfer GalNAc onto GlcNAc. Since the *C. jejuni* OST is capable of transferring the GalNac-Bac- disaccharide onto peptides *in vitro*[151], it should also transfer a GalNAc-GlcNAc- disaccharide onto proteins. Therefore, it should be possible to engineer *E. coli* to produce proteins with the *N*-linked GalNAc-GlcNAc- disaccharide by simply co-expressing PglA, PglB, and a therapeutic protein candidate with an *N*-linked glycosylation consensus sequence. PEG could be conjugated specifically to this glycoprotein using a single enzymatic step, simplifying the previously described method[187].

It may also be possible to create more stable therapeutic glycoproteins from bacteria without the addition of PEG. The primary benefits of PEGylation are the lack of immune response to PEG and the increase in molecular mass of the target protein. Instead of PEG, large polysaccharide chains could be added to specific sites in proteins, eliminating the *in vitro* conjugation step. O-antigens provide a viable template for such

large polysaccharides, however most O-antigens also have a high likelihood of initiating an immune response in the host. Clever engineering of these sugars will be necessary for this type of modification to have a positive impact on protein therapeutics. First, a core non-immunogenic structure would have to be engineered, mimicking standard structures found in the human body. Next, O-antigen polymerases such as Wzy[193], would have to be engineered to recognize and polymerize the non-immunogenic subunit. Finally, to maintain homogeneity of chain length, regulators of O-antigen chain length[194] would have to be engineered to recognize the novel substrate.

Uncovering glycosylation site specificity in bacterial OSTs

In order to exploit *E. coli* as a host for production of therapeutic proteins, it would be ideal to have a larger set of enzymes to work with an not be limited to OSTs that require extended sequence modifications for efficient glycan transfer. In our comparative analysis of bacterial OSTs, we discovered OSTs with altered glycosylation acceptor site specificity. The most interesting of these OSTs are those from the *Desulfovibrio* genus. These OSTs were demonstrated to be capable of transferring the *C. jejuni* oligosaccharide onto a model protein that did not contain the typical bacterial *N*-linked glycosylation consensus sequence. In order to determine the specific sites glycosylated by these enzymes, a combination of site-directed mutagenesis and mass-spectrometry has been proposed. While these methods will allow us to determine the specific sites glycosylated by these enzymes in this particular protein, the information we obtain will only represent a subset of all of the possible sites that these enzymes can recognize.

Here, I propose a method for rapid determination of the glycosylation site specificity of any OST that can be expressed in *E. coli*, native or engineered. The method involves a novel implementation of the phage display technology described in Chapter 2 of this thesis. The significance of the proposed method is two-fold: (i) the specific consensus sequences recognized by many of the OSTs discussed in this thesis are currently unknown, and novel glycosylation sites could prove useful in glycoprotein engineering, furthermore (ii) OSTs that are engineered to recognize novel sequences could be inadvertently altered to recognize unexpected sequences, and the screening methods proposed here will be useful for more thorough, high-throughput characterization of engineered OSTs.

The concept behind the proposed method is the design of a library of peptides containing potential glycosylation sites. The size of the peptide library is limited by the transformation efficiency of *E. coli,* where the maximum expected library size is approximately 10^9. We could make the assumption that all OSTs recognize N-X-(S/T) sites, but this may not be true, therefore I propose a library of random peptides, each containing a central asparagine residue, flanked by three randomized amino acids on either side. For full coverage of this library, ten transformations would need to be combined.

It has been shown that phage expressing a GT-g3p fusion, without the soluble MBP protein, but including the signal sequence of MBP (ssMBP) for targeting the protein to the periplasm, can be glycosylated and displayed on phage particles[98]. Therefore, to simplify the system, this would be the ideal construct. The library could easily be produced through overlap extension PCR, followed by insertion into the

phagemid by the Gibson assemble method[195], which avoids the use of restriction enzymes that could inadvertently truncate library members. Two complementary primers containing the ssMBP sequence, followed by 5'-NNKNNKNNKAATNNKNNKNNK-3', followed by ~20bp of the 5' end of g3p would be utilized to insert the randomized GT into the phagemid construct.

Once the phagemid library was produced, it would be transformed into TG1 ΔwaaL co-expressing the *C. jejuni pgl* locus without PglB, ideally from the genome, and a plasmid encoding the OST to be investigated. Phage particles would then be produced from these cells and selected using the previously described method[144]. Following two to four rounds of sorting, individual clones would be analyzed for glycosylation. This could be done in a high-throughput method, where phage particles are produced in 96-well plates from individual clones, and glycosylated phage are detected in an ELISA format (data not shown). Clones capable of glycosylation would be sequenced, and glycosylation would be verified by Western blot. To confirm the specific glycosylation sites, site-directed mutagenesis and/or mass spectrometry would be performed.

This straightforward method could help to elucidate the glycosylation site specificities for a variety of OSTs that can be expressed in bacteria. The discovery of novel glycosylation sites will be useful in the future design of glycoproteins in the prolific *E. coli* protein production host.

Summary of the presented works

This thesis focused on utilizing *E. coli* as a host for the study of one of the important branches of the field of glycobiology, *N*-linked glycosylation. Specifically, the emphasis of this research was to further our understanding of and improve the function of the OST, the enzyme responsible for catalyzing the transfer of an oligosaccharide onto an asparagine residue of a glycoprotein. The discovery that *N*-linked glycosylation could be engineered into *E. coli*, an organism that does not require this modification for cell survival, opened a unique avenue for the study of this important post-translational modification. Furthermore, the advent of glycosylation-competent *E. coli* has provided a new platform for the production of therapeutic glycoproteins and vaccine candidates.

In order to create OSTs with improved function, it was first necessary to develop a selection tool that could be utilized to isolate variants with the desired glycosylation phenotype, while maintaining a genotype linkage. A variety of approaches have been taken towards this goal, three of which were described here. The most promising tool developed was the glycophage display system[98], which was implemented for the selection of bacterial OSTs with relaxed acceptor site specificity. Although the phage display system requires further development in order to stabilize the components to more effectively select for functional variants of OSTs, it should prove useful in the future. Additionally, a variation of this technology was presented for the efficient elucidation of bacterial OST acceptor site specificity.

The *C. jejuni* OST and the closely related *C. lari* OST are the most thoroughly characterized of the bacterial OSTs to date. As more bacterial genomes are sequenced, it

is expected that many other species will contain genes with significant homology to these *Campylobacter* enzymes. In an effort to discover OSTs with novel function, a variety of bacterial OSTs were expressed in *E. coli.* Each of these OSTs was tested for its ability to transfer the *C. jejuni* glycan onto different protein substrates. Several OSTs with altered acceptor site specificity where discovered, and methodologies for further characterizing these enzymes have been proposed.

In order to more thoroughly characterize OSTs, an *in vitro* system was developed that simplifies the model framework of glycosylation[43]. The reduced complexity of the *in vitro* system allows for closer investigation of each component of the glycosylation machinery. By combining cell-free translation systems with a plasmid encoding a glycoprotein and adding LLOs and purified OST, we efficiently produced glycosylated proteins. This method could be utilized to introduce components that are incompatible with the *in vivo* system, such as unnatural or even artificial glycans. The system was proven to be compatible with both the *C. jejuni* OST and the *C. lari* OST, so it follows that it should be useful for observing the function of a variety of other single-subunit OSTs.

E. coli has proven to be a useful tool for developing a more thorough understanding of *N*-linked glycosylation. Continued research in the field of bacterial glycosylation will undoubtedly lead to new and remarkable discoveries with significant impact on the biopharmaceutical industry.

CHAPTER 6
EXPANDING THE HORIZONS:
BIOLOGICAL ENGNEERING FOR THE NEXT GENERATION

Introduction

The National Science Foundation GK-12 program, collaborating with university institutions, provides financial support to graduate students enrolled in science, technology, engineering and mathematics (STEM) programs. Cornell's Learning Initiative in Medicine and Bioengineering GK-12 (CLIMB) program is such a program. The CLIMB program partners science teachers in middle and high schools in surrounding school districts with graduate fellows in the field of Biomedical Engineering. The Department of Biomedical Engineering at Cornell University, through the CLIMB program, provided funding for two years of the aforementioned research.

Designed to provide more than financial support to graduate student fellows during their research, the CLIMB program creates a unique platform in which the graduate student can enhance communication expertise, improve research design skills as well as develop advanced teaching skills with the ability to translate graduate research for a K-12 audience. Furthermore, the program was designed to enhance the scientific knowledge base and skill set of K-12 educators and benefit the K-12 students by placing a guest scientist in the classroom as an expert and a role model.

Ronald Reed, a 20-year veteran teacher of Biology and Living Environment at Cortland Enlarged City School District in Cortland, N.Y., was my teacher partner during my first year in the CLIMB program. Michelle Kornreich, a sixth grade General Science

teacher at Boynton Middle School in Ithaca, N.Y. was my teacher partner during my second year in the program. The educators were invited to the Cornell Biomedical laboratory at Olin Hall to participate in a six-week summer scientific immersion program. During their summer participation they each were expected to conduct a novel scientific research project. The teachers also attended BME5875: Frontiers in Biomedical Research for Teachers, a class designed to broaden their knowledge of current research endeavors. Weekly meetings attended by both educator and fellow addressed teaching methodologies, with a strong focus on inquiry-based teaching. At the end of the six-week program, the educators were required to present their research in a poster session.

The CLIMB 2010 summer research project, with Ronald Reed, focused on the synthesis and design of carbohydrate microarrays which could be used in both diagnostic and research applications. Current methods for production of carbohydrate microarrays require expensive techniques for in-vitro carbohydrate production. Here, we proposed a method for making a carbohydrate microarray by decorating bacteriophage *in vivo* with carbohydrates, specifically *O*-antigens originating from pathogenic bacteria, similar to the phage display methods described in Chapter 2 of this thesis. The glycophage are produced by helper phage infection of cells carrying an *O*-antigen operon, an OST, and the gene for a phage coat protein engineered to contain an *N*-glycosylation site. Specifically, we investigated the ability of phage to display the glycan in higher valiancy by fusing the acceptor peptide to the major coat protein, g8p, instead of the minor coat protein, g3p. The OST used in this study, *C. jejuni* pglB, can be functionally expressed in non-pathogenic *E. coli* and can transfer *O*-antigens onto asparagine residues within

specific *N*-glycosylation sites, as describe in the former chapters of this thesis. Phage expressing the g8p-GT fusion protein did not prove to be glycosylated (data not shown).

The CLIMB 2011 summer program with teacher partner Michelle Kornreich was a different experience. Mrs. Kornreich was trained as an elementary teacher and had recently been tasked to teach middle school General Science. Not extensively trained in any advanced science, she lacked the basic understanding of the central principles of molecular biology. Shortly into the summer immersion program, it became apparent that she did not have a sufficient knowledge base to complete the proposed project. She was experiencing frustration, bordering on defeat. This was not the intention of the CLIMB program. Therefore, it became necessary to alter the directions of the proposed project. We chose to place more focus on developing Michele's understanding of basic biology, and less focus on accomplishing a specific project. Michelle shadowed me in the laboratory, learning about the process of cloning a gene of interest and making mutations to DNA that led to variations in protein function. Specifically, we worked on a project that involved creating and testing point mutations of *C. jejuni* PglB for altered catalytic activity. She finished her summer having accomplished a great deal and entered her classroom in the Fall with a new passion and appreciation for Science which she passed on to her students.

As part of the intensive summer program, a curriculum centered on bioengineering topics was designed. The curriculum specifically connected the teacher's newly acquired scientific knowledge to tangible human health problems, with the objective of sparking student interest in the fields of Science and Engineering.

The following school year, an average of ten hours per week was spent preparing curricular material and working in the teacher partners' classrooms, developing additional teaching and laboratory modules, lecturing and interacting with students, and assisting with laboratory and instructional tasks. I also had the opportunity during my classroom time to observe each of my educator partners' teaching techniques to help me to craft better highly interactive lesson plans.

The Integration of the research for the GK-12 experience was challenging as most young students view bacteria as pathogens. The lessons began by introducing the fact that bacteria, specifically non-pathogenic *E. coli*, are commonly used as protein production hosts, with the capacity to survive while more than 20% of the total protein content is composed of foreign protein. Describing proteins such as insulin, which can be produced in *E. coli* and used for therapeutic purposes provided a translation to a concrete concept that the students could relate to. One goal of the GK-12 experience was to use this protein production host to teach students the central dogma of molecular biology: DNA→RNA→Protein in a way that allows them to visualize the process. In the high school classrooms, students were challenged to design DNA coding for a fluorescent protein. When the cells expressed the protein from their engineered DNA sequence, the cells glowed. This exercise engaged and fascinated the students.

What follows are the curriculums developed for a high school Biology class and a 6th grade General Science class, respectively.

Curriculum 1:
Cloning a Protein of Interest

Introduction

This teaching module was designed in collaboration with Ronald Reed, a science teacher at Cortland Enlarged City School District in Cortland, NY, for 9th and 10th grade students enrolled in Living Environment, aka. Biology. The New York State – Living Environment standards covered in this curriculum include: Std. 1. Analysis, inquiry and design, Std. 4.2. Genetic inheritance, Std. 4.3. Organisms change over time, and Std. 4.4. Continuity of life.

The scheduled time frame for this curriculum is 2 periods of 45 minutes each, and 2 periods of 90 minutes each.

Description

Students are exposed to recombinant DNA technology and the process of cloning a protein of interest. Starting with the DNA sequence for a gene coding for a fluorescent protein, students help design primers to make copies of the gene. Next, students learn about gel electrophoresis, which is used to purify the gene copies from a PCR reaction. A model ligation reaction activity is used to demonstrate how to construct an expression plasmid. Finally, the students get to work with cells expressing the fluorescent protein(s) they helped clone.

Objectives

- Students will develop a deeper understanding of the central dogma of molecular biology: DNA → RNA → Protein.
- Students will understand how electrophoresis can be used to separate molecules.
- Students will be able to describe how to make a protein of interest, and they will be able to provide examples of proteins that are of interest to scientists.
- Students will understand that while many bacteria can make you sick, they can also be very useful.

Materials

Hand-outs:

Part I, PCR; Part II, electrophoresis; Part III, model ligation reaction

Day 1: Scissors, Codon table

Day 2: NeoSCI gel electrophoresis kit [components of kit available through BioRad]:

Four dyes and two unknown mixtures, Agarose, Tris/Borate/EDTA buffer (TBE).

Electrophoresis chamber and power supply, Micropipette, p20, with tips

Day 3: Scissors, Tape

Day 4: LB agar plates with kanamycine and IPTG (kanamycine and IPTG available through Sigma), MC4100 *E. coli* cells containing fluorescent protein expression vectors (pET28) [similar expression vectors and comparable competent cells available through invitrogen], Micropipette tips and 1.7mL epi-tubes

Blue light box (for imaging plates from Day 4)

Science Content for the Teacher

Most people view bacteria as pathogens that make you sick, but not all bacteria are bad. In fact, bacteria, specifically non-pathogenic *Escherichia coli*, are commonly used as protein production hosts, with the capacity to produce more than 20% of the total protein content as foreign protein. Proteins such as insulin can be produced in *E. coli* and used for therapeutic purposes.

The primary goal of this curriculum is teach students the central dogma of molecular biology: DNA→RNA→Protein in a way that allows them to visualize the process. To clone a protein, on must first obtain a copy of the gene sequence for the protein. Copies are made from a template sequence through a process known as PCR. Once the PCR product is copied, it is digested with restriction enzymes and ligated into a plasmid that has also been digested with the same enzymes. A plasmid is a circular piece of DNA that contains (1) an origin or replication (to allow the plasmid to be replicated in the cells), (2) a gene that confers antibiotic resistance (to force the bacteria to maintain the plasmid), and (3) a multiple cloning site (a region flanked with a promoter and terminator with numerous restriction enzyme cut sites in between for ease of cloning).

Students will be given the DNA sequence for a fluorescent protein, and with the guidance of the teacher and fellow, they will be challenged to design primers to clone this sequence into an expression vector. Over several days, the students will be exposed to all the steps of the cloning process, including separating molecules using gel electrophoresis, and an interactive model ligation reaction. When the protein from the students' engineered DNA sequence is expressed in bacteria, the cells will 'glow,' in so much as they will convert UV light into green light, which we can see.

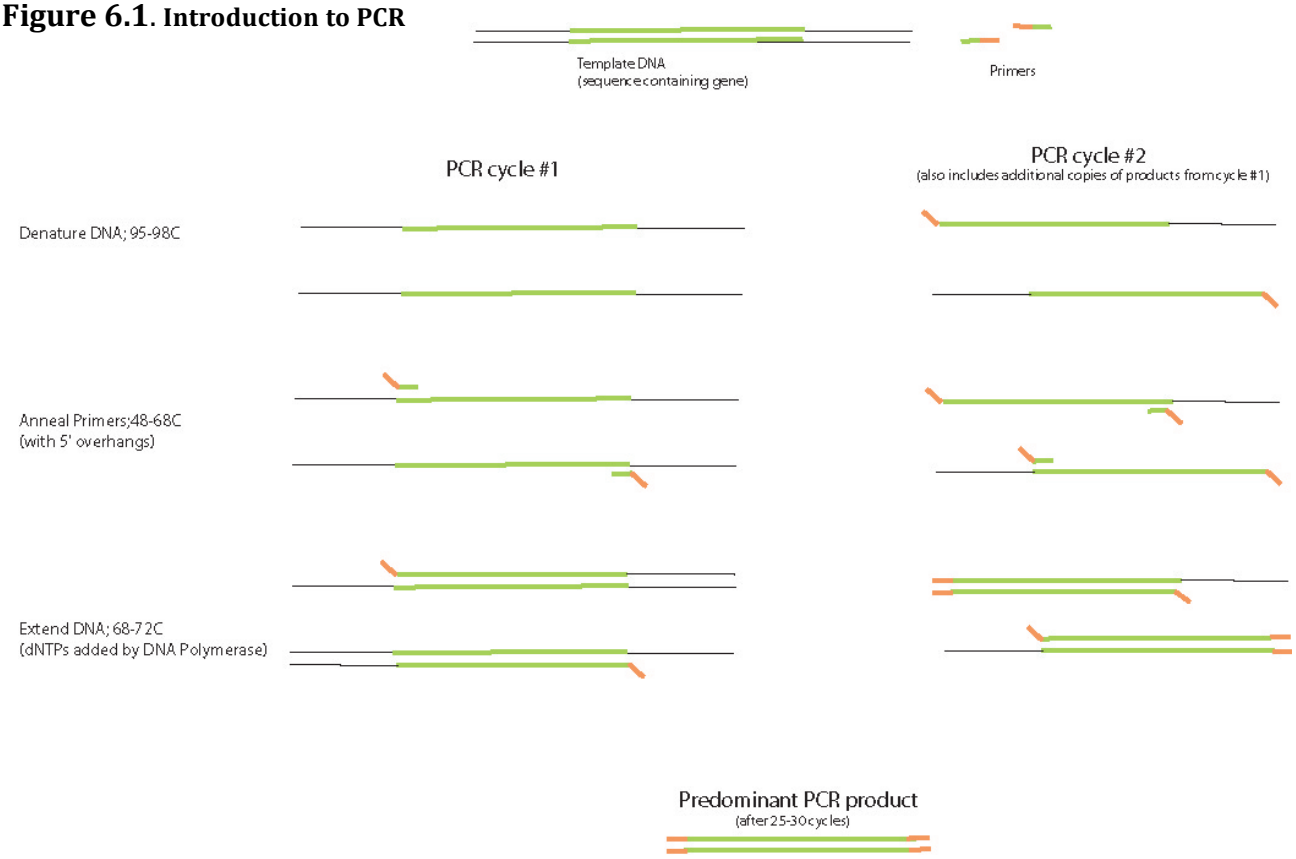
PCR:

Polymerase chain reaction is a methodology used to make many copies of a particular gene sequence. The reaction is composed of the template DNA, primers, nucleotide bases (dNTPs), polymerase and a buffer. By cycling the temperature from 98°C for 30s to melt the DNA template, to 50-56°C for 30s to anneal the primers, to 68-72°C for 1min/kb of product to extend, or copy, the template, and repeating this cycle 25-30 times, millions of copies of the DNA sequence can be made.

Polymerase is an enzyme which creates new DNA from 5'to 3', as it reads a sequence 3' to 5'. The polymerase used in this reaction needs to be heat stable in order to survive the temperature cycling of the reaction. The discovery of a polymerase from thermophilic bacteria, which has evolved to withstand exceedingly high temperatures, made this possible. The buffer in the reaction contains magnesium, which is needed to stabilize and maintain the structure of the polymerase in order for it to function properly.

Primer design is a key element in any PCR. The primers indicate where the polymerase should begin copying, at either the beginning or end of the sequence. Therefore, only the region between the primers (a forward primer that is a direct copy of the sequence, and a reverse primer that is the reverse complement of the end of the sequence) gets copied. Additionally, 5' overhangs can be added to the primers to introduce the DNA sequence for restriction enzyme cut sites at the beginning and end of the sequence (Figure 6.1.).

Figure 6.1. Introduction to PCR



Gel Electrophoresis:

Gel electrophoresis is most commonly used for separation of biological macromolecules such as deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or protein; however, gel electrophoresis can also be used for separation of nanoparticles and small molecules.

Electrophoresis refers to the movement of a charged particle in an electrical field. Gels suppress the thermal convection caused by application of the electric field, and can also act as a sieving medium, retarding the passage of molecules. DNA Gel

electrophoresis is usually performed for analytical purposes, often after amplification of DNA via PCR, but it may also be used as a preparative technique.

Following PCR amplification of our DNA encoding the fluorescent proteins we are cloning, the DNA was run on a gel. The DNA fragments, which can be visualized on a blue (or UV) light box after staining with a SyberSafe® or Ethidium bromide, were cut from the gel to separate our product from the other components of the PCR reaction.

Ligation Reaction:

DNA ligase is an enzyme which will ligate, or join together, two pieces of DNA by catalyzing the formation of a phosphodiester bond in the backbone of the DNA.

One common way of cloning is to cut a plasmid with restriction enzymes, some of which leave overhanging 'sticky ends,' and ligate it to DNA coding for your protein of interest (POI) which has been cut by the same enzymes (the cut-sites are usually introduced to your POI gene sequence by adding 5' overhangs to the PCR primers). Restriction enzymes are enzymes that recognize specific sequences in DNA and cut in a specific way. These enzymes were originally discovered in bacteria and are used by the bacteria to digest and destroy foreign DNA.

Pre-requisites

This module is best taught near the end of the school year, after the students have been exposed to the majority of the course material, as there are many opportunities for 'making connections'. Specifically, the students should know where DNA is found in

cells, that DNA is double stranded and made-up of nucleotide base pairs, that DNA encodes genes that code for protein, that the DNA is transcribed into mRNA before being translated into protein, and that proteins are composed of amino acids. This module could be easily incorporated into an AP Biology class.

Classroom Procedure

Day 1. Cloning a protein of interest, Part I: PCR

(45-90 minute class)

Begin class by showing an image of an *Aequorea victoria* jellyfish, the animal in which green fluorescent protein (GFP) was originally discovered. Inform the students that the fluorescence of this animal is caused by the expression of the GFP protein, and that their task is to make bacteria fluoresce by making them produce this protein. Challenge the students to tell you how to make a protein.

Students should be able to come up with the idea that in order to make a protein, we need to start with the DNA sequence for the protein. Once it is established that we need the DNA, introduce the concept of polymerase chain reaction, which will be required to make copies of the gene. One or more videos that describe and animate the process can be used. This includes 'polymerase chain reaction Animated', available at: http://www.youtube.com/watch?v=HMC7c2T8fVk.

Once the students are clear on the basics of PCR, distribute the 'Cloning a protein following a basic introduction to polymerase chain reaction. Groups of 3-5 students will work together to determine which of the primers they have been given will work in a

PCR reaction to copy the DNA sequence for their protein of interest. DNA sequences for two proteins are provided. A range of primers that will bind the template, and some that will not, are provided. For ease of alignment, students will be asked to cut-out the primer sequences. Some of the primers have mutations, including lacking a start coding or a stop codon, and when the students discover these, they will be challenged to determine what effect using those primers would have on their final product – a codon table will be necessary for the identification of the start and stop codons. Two sets of primers will work to copy each template, but one set will add a 5' overhang at each end to introduce restriction enzyme cut sites. When the students discover these primers, this will lead into an introduction to one commonly used type of cloning, involving digestion with restriction enzymes.

SOLUTIONS:

1. Forward sequence, 2. Forward sequence with 5' overhang, 3. Forward sequence without start codon, 4. Reverse complement of forward sequence, 5. Reverse sequence, 6. Reverse sequence with 5' overhang, 7. Reverse sequence without stop codon, 8. Reverse complement of reverse sequence, 9.Random sequence.

Students may not have sufficient time in a 45 minute class period, following the introduction to PCR, to finish the activity. In this situation, the answers to the six questions should be discussed at the beginning of the next class. When presenting answers, students should be pushed to share their own answers before being given the correct answers.

<u>Day 2</u>. Cloning a protein of interest, Part II: Gel Electrophoresis

(45-90 minute class)

Preparation: Have at least one electrophoresis set-up in the classroom, and a gel casting set-up ready to go. Just before class begins, melt the 0.8-1% agarose (in TBE), by microwaving (approximately 1min/50mL, or until completely melted), and allow to cool for a few minutes. While students are making their way into class, elect a volunteer to pour the gel and insert the comb to create wells.

Initiate class with: 'While you weren't looking, I ran a PCR with the primers you helped select in the last class. Now, we need to verify that the reaction worked, and we need to separate our product from the rest of the components of the PCR reaction.' This is a good time to get the students to remind each other what the components of a PCR reaction are. Ask the students how they propose to separate these components (the student(s) who helped to pour the gel should have some idea).

Introduce the concept of electrophoresis. Draw a model gel on the board, with a band representing the PCR product, and a second band representing the primers, and ask which band is which. Students should be able to reason that the smaller molecules will travel faster. [Students will be running the PCR reaction on the gel, however, they will not see the DNA until after the gel is stained with an appropriate staining solution. To avoid using potentially hazardous chemicals, our class did not stain the DNA in the gel, but they were shown a picture of their PCR product run on a gel stained with ethidium bromide.]

By this point, the gel(s) should be solidified. If this is the student's first experience with using micropipettes, the general mechanism of the pipette should be described. Additionally, using the micropipettes is a good opportunity to get the students to relate 1uL to 1L (how many mL in a L, how many uL in a ml, how many uL in a L?). Each student should have an opportunity to pipette some dye with the micropipette, even if there are not enough lanes in the gel(s) for everyone to pipette into the gel.

Students should be divided into groups, and while one group is learning to pipette and loading the gel, the other groups should be working on the 'Cloning a Protein of Interest: Part II' hand-out on electrophoresis. The packet includes the size (MW) of the four dyes that students will be running on the gel, as well as the chemical structure, from which they may be able to determine the charge on the molecule. Students are asked to predict how far and in which direction each of the dyes will run.

The gel should be run for approximately 20 minutes in order to separate the dyes. Once the gel is finished running, the students should compare their expected results to the observed results and come-up with possible explanations for any differences. Additionally, based on the results, the students should be able to determine what dyes were included in each of the 'unknown' mixtures.

<u>Day 3</u>. Cloning a protein of interest, Part III: Model ligation reaction

(45 minute class)

Begin the class with an image of a restriction endonuclease, or restriction enzyme, attached to its recognition sequence within a strand of DNA. Describe what happens to the DNA when the restriction enzyme recognizes a specific sequence of DNA. Ask the students why an organism would develop such an enzyme – what purpose could it serve? Continue to explain that scientists now utilize these enzymes to cut DNA in specific patterns in order to assemble new DNA sequences.

Returning to that gene we copied with PCR and verified/cleaned-up with gel electrophoresis – what would happen if we put that DNA into bacteria? What would the bacteria do with it? [Hopefully, the students will surmise that it would get degraded.] How can we convince the bacteria to maintain our DNA? What has to happen to the plasmid in order for it to be passed on to the next generation of cells?

At this point, the 'Cloning a protein of interest: PartIII' handout should be distributed. Students should work in groups of at least four, with each student having the three parts of the plasmid to work with. Students should find the restriction enzyme cut sites and cut the 'DNA' accordingly, yielding three pieces with overhanging 'sticky ends.' Students should try to arrange their pieces in as many ways as possible. There is one preferred arrangement, but the students should discover that many other arrangements are possible, particularly when students combine their pieces to make larger plasmids. However, only plasmids that contain the antibiotic resistance marker and the origin of replication will be maintained and replicated by bacteria. And, only those

plasmids that have the protein of interest oriented in the correct direction in relation to the promoter region will be capable of expressing the protein.

<u>Day 4</u>. Cloning a protein of interest, Part IV: The Art of Science

(90 minute class)

Preparation: LB agar plates containing the antibiotic for which the plasmid confers resistance (kanamycin, in the case of pET28), and the inducer for the plasmid, if not constitutively active (IPTG, in the case of pET28), should be prepared in advance [Pre-poured sterile agar plates containing antibiotic can be purchased from Sigma-Aldrich]. Overnight cultures, grown at 37°C, should be inoculated the night before the activity. The morning of the activity, the cells should be aliquoted into small epi-tubes, <100uL each.

This activity is designed as the culminating experience for the module. During the first half of the class, students will be given plates and small aliquots of the bacteria containing an expression vector harboring the protein(s) of interest. Several different fluorescent proteins are available – in our class, we used five different proteins: GFP (green), CFP (cyan), mStrawberry (pink), mTangerine (orange), and mHoneydew (light green).

Prior to students obtaining their aliquots of bacteria, SAFETY must be emphasized. Although this bacteria is considered non-pathogenic, it should be treated

with caution. Students should wear gloves when handling the bacteria, and the bench tops should be cleaned with 70% EtOH after use.

Using a pipette tip as a 'paint brush', and the bacterial cultures as 'paint', students will draw images on the plates, an excellent opportunity for artistically inclined individuals to share their talents. The plates need to be incubated overnight at 30°C in order for the cells to grow.

Figure 6.2. Example of student 'bacto-art'. Students, working with several bacterial cultures expressing different fluorescent proteins as 'paint' and pipet tips as 'brushes,' created a variety of different bacto-art samples, which were later imaged under blue light.

Once the students have cleaned-up the lab area, they should be challenged to come up with more practical uses for genetic engineering and recombinant DNA technology through a class discussion. Some possible examples include: therapeutic protein production (insulin is the most common example), making large quantities of a protein to study its function, and creating biosensors that turn on/off based on environmental factors. At this point, it can also be emphasized that bacteria do not have the same machinery as eukaryotic cells, and that the processing of proteins in bacteria is

different – in other words, the same protein sequence made in bacteria versus mammalian cells, could have a completely different shape, and therefore different function. One example of a protein that requires eukaryotic cell culture is erythropoietin, a protein commonly used to treat anemia. When erythropoietin is made in bacteria, the protein product is 1000x less effective than the same amino acid sequence produced in a mammalian cell line – the primary reason for the difference, in this case, is the lack of specific types of glycosylation, the addition of sugar groups, to proteins in bacteria.

Extension activity

As an extension to this project, a lesson can be provided regarding fluorescence. Fluorescence is the emission of light by a substance that has absorbed light, or other electromagnetic radiation, of a different wavelength. Fluorescent proteins, therefore, don't 'glow in the dark', but they absorb light and emit a different wavelength of light with less energy.

Safety

The bacteria used for cloning and expressing protein is considered non-pathogenic, however it should be handled with care. Glove should be worn at all times, and changed if soiled. All surfaces should be cleaned with a 70% Ethanol (EtOH) solution.

Students should always wash hands before leaving the laboratory .

TAE – contains Tris Base, Acetic Acid and EDTA, all of which are skin irritants, thus gloves should be worn at all times.

Agarose –skin irritant, flush with water

xylene cyanol – skin irritant, flush with water

orange g – skin irritant, flush with water

methyl green – skin irritant, flush with water

bromophenol blue – skin irritant, flush with water

Waste Disposal Method - All chemical and biological waste disposal must be in accordance with current local, state, and federal regulations.

Assessment

Students were given a pre-test and a post-test in order to measure their previous knowledge and assess the level of comprehension gained through the activities. A short-answer questionnaire and/or a multiple choice problem set can be used. The post-test is the same test as the pre-test, in order to provide direct comparison for statistical analysis.

Curriculum 2:
Where in the Middle School can we find the most bacteria?

Introduction

This teaching module was designed in collaboration with Michelle Kornreich, a science teacher at Boynton Middle School in Ithaca, NY, for 6th grade students enrolled in General Science. The New York State – Intermediate Level Science standards covered in this curriculum include: Std. 1. Analysis, inquiry and design, Std. 4.4. Continuity of life, Std. 4.6. Interdependence of life, Std. 4.7. Humans impact their environment.

The schedule time frame for this curriculum is three 35 minute class periods.

Description

The primary objective of this activity is to make students aware of the microbial life in their environment and to help them better understand our complex ecosystem. Students will develop and test a hypothesis to address the question, 'where in the middle school can we find the most bacteria?'

Objectives

- Develop hypotheses and test by sampling the environment.
- Collect qualitative and quantitative data.
- Analyze the results and provide suggestions for improving the experiment.
- Develop new hypotheses based on the initial experimental results.

Materials

1-2 petri dishes for each pair of students, LB agar, q-tips, water, parafilm®, permanent markers, index cards

Science Content for the Teacher

Our environment is full of life that is smaller than the eye can see. These microorganisms play an essential role in the viability of our ecosystem – not only in the world around us, but also on and within each us. It is estimated that there are more than 10x more bacterial cells on the human body than there are human cells, and with an estimate of 10 trillion human cells composing the average human body, that means there are approximately 100 trillion bacteria covering us! But don't worry, by weight, we are definitely mostly human – human cells are, on average, 10x larger than most bacteria, and the human body is composed of a significant amount extracellular matrixes, compounds produced by our cells that help to hold us together.

Every surface of the human body that is exposed to the environment is covered in bacteria. The majority of the bacterial cells that cohabitate with the human body reside in the gut. These gut microorganisms are necessary for our survival, providing us with

many essential vitamins, digesting complex carbohydrates to usable forms, destroying cancer-inducing compounds[196], and much more.

The natural flora that exist on and within the human body act, in many cases, as a defense mechanism against potential disease-causing, aka pathogenic, bacteria. For example, *Staphylococcus epidermidis*, a natural member of the normal skin flora, produces an enzyme, serine protease Esp, which has been shown to inhibit the growth of the more significantly pathogenic *Staphylococcus aureus*[197].

The world around us is also covered in bacteria. Most locations, unless sterilized with a sanitizing agent like chlorine, have an entire bacterial ecosystem in place. Microorganisms inhabit even places where no one would think life could exist, like thermal vents. The unique microorganisms that exist in thermal vents have provided scientists with enzymes that have significantly impacted genetic research – specifically, these organisms code for heat-tolerant DNA polymerases that can be used for replicating DNA in a test tube in the laboratory, *in-vitro*.

Scientists have studied many of the different microorganisms in the environment, and new organisms are being discovered regularly. A variety of tools for understanding and manipulating biological systems have been discovered and/or designed in bacteria. Bacteria can be manipulated to produce proteins for use in humans – the first example to come to market was human Insulin, made in *Escherichia coli* from recombinant DNA, and marketed by Eli Lily and company in starting 1982. Prior to production in bacteria, insulin was purified from pigs, and many people were allergic to the pig protein.

The purpose of this activity is to help students to gain a better understanding of the world around them. Microorganisms play a major role in our everyday lives, even

though many of us don't even realize they are there. Many of the practices we undertake to maintain sterile environments and to help our body fight-off infection, such as antibiotics, are potentially harmful to the delicately balanced ecosystem that maintains us.

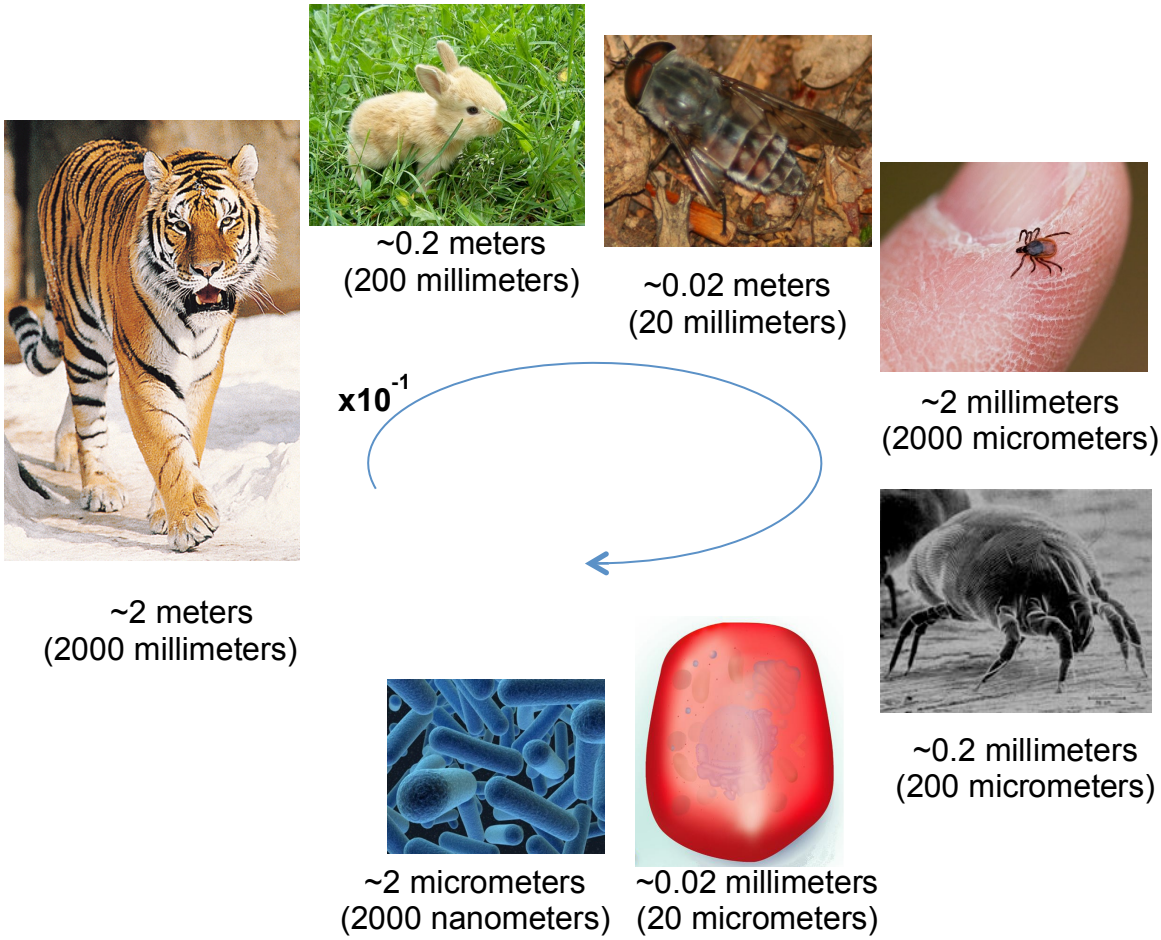


Figure 6.3. Understanding size. Bacteria are small, approximately 10x smaller than an animal cell. This series of images is designed to help students visualize size difference in orders of magnitude. Included in this figure are approximate metric measurements, providing the opportunity to introduce/re-emphasize the prefixes used to describe small things in metric units. Public stock photos used here for educational purposes.

Preparation

LB agar plates can be obtained pre-poured and ready-to-go from Sigma (L5542). Be sure not to purchase plates containing antibiotics, as this will prevent the growth of a larger portion of the sampled bacteria.

Q-tips, cut in half while attempting to maintain sterility, should be prepared ahead of time. In addition, aliquots of sterile water should be prepared for each team (tap water will suffice if sterile water is not available).

Classroom Procedure

Engage: A 4-minute video entitled 'Understanding Microbial Life' is used to initiate student interest in finding-out where in their school they can find the most bacteria.

Explore: Students are broken-up into teams and asked to come-up with a hypothesis as to where they believe the most bacteria live in their school. Teams then go out and sample the locations they agree to test. After three days incubating at room temperature, the results are analyzed by both qualitatively and quantitatively observing the plates.

Explain: The results from all of the classes are compiled and presented to each class. The specifics of this presentation depend somewhat on the results, and results from multiple classes can be compiled.

Expand: Based on the observed results, students are challenged to come-up with new hypotheses and improved methods to address the question.

Day 1: *Opening day –*

Students watch a 4-minute video entitled 'Understanding Microbial Life'. Students are then broken-up into teams and asked to come-up with and agree upon a hypothesis about where in their school they think they will find the most bacteria.

Following an initial period of discussion among groups, the students are given a demonstration on how to sample a location. A q-tip is dipped in water, wiped across the surface being sampled, and then streaked across the plate from left to right, down the length of the plate. Be certain to emphasize that the gelatin-like substance (LB agar) is what the bacteria feed on in order to grow (ie. If they streak their bacteria sample on the lid instead of on the agar, they will not find any bacteria after the plate is appropriately incubated.)

All plates should be labeled along the edge with the group name and sample location – labeling the plates on the lid or on the bottom will make it difficult to collect data.

Plates can be incubated for ~3 days at room temperature, or 1-2 days at warmer temperatures. Incubation temperature should not exceed body temperature, 37°C.

Longer incubation times will lead to more fungal/mold growth. Bacteria and fungus can be distinguished by their morphology – bacteria generally form round colonies, while fungus and mold have a variety of interesting morphologies. In some cases, the fungus on a plate can inhibit bacterial growth around it. This is how penicillin, the first antibiotic used to treat bacterial infections in humans, was discovered.

Day 2: *Data collection day –*

The plates are sealed with parafilm® before the students handle them in order to prevent aerosol of bacteria or fungal spores during the data collection.

Qualitative data collection: The plates from each team are displayed along the outer edge of the class, with a corresponding number written on an index card. Each student has a chance to observe each plate. Based on their observations, students decide up the three plates they believe have the most bacteria, taking into consideration both in variety and number.

Quantitative data collection: Following the qualitative observations, each group takes their plate(s) back to their workspace to count colonies. Some plates will have a significant number of colonies, and students will have to devise methods to

count - one possibility is to divide the plate into approximately equal sections and count only one of those section, then multiplying by the total number of sections.

Following the data collection, students should report their data. This can be done directly into a computer program, such as excel, or written on a collection sheet. Students should also record their data in their journal.

The teacher should compile the data from all of the classes in order to have meaningful results to report to the classes for Day 3. If computers are available to the students, an extra day of the activity can be implemented, where students play with the raw data, create graphs and try to come to meaningful conclusions.

Day 3: *Closing day –*

Students are given a 'pop quiz' at the beginning of class to help assess their understanding of what we have discussed and explored over the course of this activity. The questions in the hand-out correspond to topics covered in the video they watched on 'Day 1'. Students are given the opportunity to watch the video a second time as they fill-out the answers on their 'pop quiz'.

Students are asked to explain why we are able see the bacteria (because they multiplied and there are millions). Students are also challenged to come-up with explanations for strange results. For example, if the plates are incubated at room temperature, and one student samples bacteria from their body, the colonies will

be significantly smaller than those from bacteria sampled from a location that is normally at room temperature.

The number of trials will be a limiting factor in good data analysis, and the students should come to this conclusion. In some cases, interesting data will emerge. In our class, we found that the peanut-free cafeteria table had significantly less bacteria than on of the regular tables, but only two data points, it is difficult to be certain of this difference.

One question that should be asked is whether of not they cultured *all* of the bacteria in their location. The answer is no, as only some bacteria can rapidly divide at the temperature and on the feed stocks used in this experiment.

Extension Activities

Following the initial data collection, many students will likely have more questions. As an extension to this activity, students can design a new set of experiments to test their new hypotheses. Alterations to the experiment can include testing different growth conditions, altering the food source for the bacteria or changing the temperature and/or duration of growth.

Safety

Following growth of the microorganisms on the LB agar plates, the plates should be sealed with parafilm to avoid any bacterial or fungal spores from becoming air-born. The petri dishes with LB agar should not contain any significant pathogens and can be disposed of in the regular trash after being sealed with parafilm

Waste Disposal Method - All chemical and biological waste disposal must be in accordance with current local, state, and federal regulations.

Conclusions

The passion that brought graduate students to their chosen field of study is often forgotten in the pressure and hectic pace of experimentation, scientific papers, and graduate course work. While there were several challenges associated with participation as a fellow in the Cornell's Learning Initiative in Medicine and Bioengineering GK-12 (CLIMB) program, including curriculum preparation and at least 10 hours each week dedicate to CLIMB classroom activities outside of the university laboratory, sharing science with middle school and high school teachers and students was a valuable experience. The CLIMB program reacquainted and reconnected me with my own passion for science while igniting an appetite and excitement in encouraging future scientists.

There was an interesting synergy between me and the partner teachers as we were all both student and teacher. My partner teachers, Ron Reed and Michelle Kornreich, are exceptional educators. I was fortunate to observe each educator in their environment and watch as they empowered their students using their own unique teaching skill set. They both taught me techniques on how to translate complicated scientific concepts for the young middle school/high school minds. Since neither teacher had any prior experience with microbiology or molecular biology techniques, laboratory procedure proved a challenge. A significant amount of time was required to acclimate them to general laboratory procedures, but they were eager students. The knowledge exchange was a remarkable experience for each of us.

Students in both the 9th grade and 6th grade classrooms were excited to have their 'visiting scientist' teach them about cutting-edge scientific progress. In both the 9th grade and 6th grade classrooms, there was a notable change in attitude towards Science. Several

of the students in the 9th grade classroom requested the scientific papers that I addressed in some of the lectures, and a number of the 6th grade students expressed interest in perusing careers in science.

Participation as a fellow in this program has inspired me. While it is not my desire to run and manage a middle school or secondary classroom, I look forward finding future opportunities to be a visiting scientist to help inspire more young people to take a serious interest in scientific research.

Bibliography:

1. Jons Jacob Berzelius (Swedish chemist) -- Britannica Online Encyclopedia. at <http://www.britannica.com/EBchecked/topic/62958/Jons-Jacob-Berzelius>
2. Hartley, H. Origin of the Word 'Protein'. , *Published online: 11 August 1951; | doi:10.1038/168244a0* **168**, 244–244 (1951).
3. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proc Natl Acad Sci U S A* **70**, 3240–3244 (1973).
4. Hughes, S. S. Making dollars out of DNA. The first major patent in biotechnology and the commercialization of molecular biology, 1974-1980. *Isis* **92**, 541–575 (2001).
5. Johnson, I. S. Human insulin from recombinant DNA technology. *Science* **219**, 632–637 (1983).
6. Chance, R. E. & Frank, B. H. Research, development, production, and safety of biosynthetic human insulin. *Diabetes Care* **16 Suppl 3**, 133–142 (1993).
7. Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E. & Villaverde, A. Microbial factories for recombinant pharmaceuticals. *Microbial Cell Factories* **8**, 17 (2009).
8. Waegeman, H. & Soetaert, W. Increasing recombinant protein production in Escherichia coli through metabolic and genetic engineering. *Journal of Industrial Microbiology & Biotechnology* **38**, 1891–1910 (2011).
9. Schmidt, F. R. Recombinant expression systems in the pharmaceutical industry. *Applied Microbiology and Biotechnology* **65**, 363–372 (2004).
10. Jonasson, P., Liljeqvist, S., Nygren, P.-A. & Ståhl, S. Genetic design for facilitated production and recovery of recombinant proteins in Escherichia coli. *Biotechnol Appl Biochem* **35**, 91–105 (2002).
11. Mazor, Y., Van Blarcom, T., Iverson, B. L. & Georgiou, G. E-clonal antibodies: selection of full-length IgG antibodies using bacterial periplasmic display. *Nat. Protocols* **3**, 1766–1777 (2008).
12. Mazor, Y., Van Blarcom, T., Mabry, R., Iverson, B. L. & Georgiou, G. Isolation of engineered, full-length antibodies from libraries expressed in Escherichia coli. *Nat Biotechnol* **25**, 563–5 (2007).
13. Ferrara, C. *et al.* Modulation of therapeutic antibody effector functions by glycosylation engineering: Influence of Golgi enzyme localization domain and co-expression of heterologous beta1, 4-N-acetylglucosaminyltransferase III and Golgi alpha-mannosidase II. *Biotechnology and Bioengineering* **93**, 851–861 (2006).
14. Gardner, R. S. *et al.* Synthetic polynucleotides and the amino acid code, VII*. *Proc Natl Acad Sci U S A* **48**, 2087–2094 (1962).
15. Lin, S. & Riggs, A. D. The general affinity of lac repressor for E. coli DNA: Implications for gene regulation in procaryotes and eucaryotes. *Cell* **4**, 107–111 (1975).
16. Benson, S. A., Hall, M. N. & Silhavy, T. J. Genetic analysis of protein export in Escherichia coli K12. *Annu. Rev. Biochem.* **54**, 101–134 (1985).

17. Wacker, M. *et al.* N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into E. coli. *Science* **298**, 1790 (2002).
18. Eichwald, E. Beiträge zur Chemie der gewebbildenden Substanzen und ihrer Abkömmlinge. I. Ueber das Mucin, besonders der Weinbergschnecke. *Ann. Chem. Pharm. Annalen der Chemie und Pharmacie* **134**, 177–211 (1865).
19. O'Connor, S. E. & Imperiali, B. Modulation of protein structure and function by asparagine-linked glycosylation. *Chem Biol* **3**, 803–12 (1996).
20. Imperiali, B. & O'Connor, S. E. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Current Opinion in Chemical Biology* **3**, 643–649 (1999).
21. Hauptmann, P. *et al.* Defects in N-glycosylation induce apoptosis in yeast. *Molecular Microbiology* **59**, 765–778 (2006).
22. Weerapana, E. & Imperiali, B. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16**, 91R–101 (2006).
23. Valderrama-Rincon, J. D. *et al.* An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat. Chem. Biol.* **8**, 434–436 (2012).
24. Benz, I. & Schmidt, M. A. Never say never again: protein glycosylation in pathogenic bacteria. *Mol Microbiol* **45**, 267–76 (2002).
25. Alex, M., Schmidt, er, Riley, L. W. & Benz, I. Sweet new world: glycoproteins in bacterial pathogens. *Trends in Microbiology* **11**, 554–561 (2003).
26. Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Micro* **8**, 765–778 (2010).
27. Yurist-Doutsch, S., Chaban, B., VanDyke, D. J., Jarrell, K. F. & Eichler, J. Sweet to the extreme: protein glycosylation in Archaea. *Molecular Microbiology* **68**, 1079–1084 (2008).
28. Mescher, M. F., Strominger, J. L. & Watson, S. W. Protein and Carbohydrate Composition of the Cell Envelope of Halobacterium salinarium1. *J Bacteriol* **120**, 945–954 (1974).
29. Burgdorfer, W. *et al.* Lyme disease-a tick-borne spirochetosis? *Science* **216**, 1317–1319 (1982).
30. Coleman, J. L. & Benach, J. L. Lectin Binding to Borrelia burgdorferia. *Annals of the New York Academy of Sciences* **539**, 372–375 (1988).
31. Sambri, V., Stefanelli, C. & Cevenini, R. Detection of glycoproteins in Borrelia burgdorferi. *Archives of Microbiology* **157**, 205–208 (1992).
32. Maley, F., Trimble, R., Tarentino, A. & Plummer, T. Characterization of Glycoproteins and Their Associated Oligosaccharides Through the Use of Endoglycosidases. *Anal. Biochem.* **180**, 195–204 (1989).
33. Štěrba, J. *et al.* Flagellin and Outer Surface Proteins from Borrelia burgdorferi Are Not Glycosylated. *J Bacteriol* **190**, 2619–2623 (2008).
34. Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. & Guerry, P. Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Molecular Microbiology* **32**, 1022–1030 (1999).
35. Larsen, J. C., Szymanski, C. & Guerry, P. N-Linked Protein Glycosylation Is Required for Full Competence in Campylobacter jejuni 81-176. *J Bacteriol* **186**, 6508–6514 (2004).

36. Szymanski, C. M., Burr, D. H. & Guerry, P. Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infect Immun.* **70**, 2242–2244 (2002).
37. Linton, D. *et al.* Functional analysis of theCampylobacter jejuniN-linked protein glycosylation pathway. *Molecular Microbiology* **55**, 1695–1703 (2005).
38. Karlyshev, A. V. *et al.* 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).
39. Szymanski, C. M. & Wren, B. W. Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Micro* **3**, 225–237 (2005).
40. Hese, K., Otto, C., Routier, F. H. & Lehle, L. The yeast oligosaccharyltransferase complex can be replaced by STT3 from Leishmania major. *Glycobiology* **19**, 160–171 (2009).
41. Feldman, M. F. *et al.* Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc Natl Acad Sci U S A.* **102**, 3016–3021 (2005).
42. Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO* **25**, 1957–1966 (2006).
43. Guarino, C. & DeLisa, M. P. A Prokaryote-Based Cell-Free Translation System That Efficiently Synthesizes Glycoproteins. *Glycobiology* **22**, 596–601 (2012).
44. Young, N. M. *et al.* Structure of the N-Linked Glycan Present on Multiple Glycoproteins in the Gram-negative Bacterium, Campylobacter jejuni. *J. Biol. Chem.* **277**, 42530–42539 (2002).
45. Amin, M. N., Ishiwata, A. & Ito, Y. Synthesis of asparagine-linked bacillosamine. *Carbohydrate Research* **341**, 1922–1929 (2006).
46. Glover, K. J., Weerapana, E. & Imperiali, B. In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc Natl Acad Sci U S A.* **102**, 14255–14259 (2005).
47. Olivier, N. B., Chen, M. M., Behr, J. R. & Imperiali, B. In vitro biosynthesis of UDP-N,N'-diacetylbacillosamine by enzymes of the Campylobacter jejuni general protein glycosylation system. *Biochem* **45**, 13659–13669 (2006).
48. Fry, B. N. *et al.* The galE Gene of Campylobacter jejuni Is Involved in Lipopolysaccharide Synthesis and Virulence. *Infect. Immun.* **68**, 2594–2601 (2000).
49. Kelly, J. *et al.* Biosynthesis of the N-Linked Glycan in Campylobacter jejuni and Addition onto Protein through Block Transfer. *J Bacteriol.* **188**, 2427–2434 (2006).
50. Alaimo, C. *et al.* Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25**, 967–976 (2006).
51. Liu, D., Cole, R. & Reeves, P. An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *J. Bacteriol.* **178**, 2102–2107 (1996).
52. Wacker, M. *et al.* Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *PNAS* **103**, 7088–7093 (2006).
53. Klena, J. D. & Schnaitman, C. A. Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1. *Molecular Microbiology* **9**, 393–402 (1993).

54. Nilsson, I. *et al.* Photocross-linking of nascent chains to the STT3 subunit of the oligosaccharyltransferase complex. *J. Cell Biol.* **161**, 715–725 (2003).
55. Karamyshev, A. L. *et al.* Mapping the Interaction of the STT3 Subunit of the Oligosaccharyl Transferase Complex with Nascent Polypeptide Chains. *J. Biol. Chem.* **280**, 40489–40493 (2005).
56. Abu-Qarn, M., Eichler, J. & Sharon, N. Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Current Opinion in Structural Biology* **18**, 544–550 (2008).
57. Heesen, S. te, Knauer, R., Lehle, L. & Aebi, M. Yeast Wbp1p and Swp1p form a protein complex essential for oligosaccharyl transferase activity. *EMBO J.* **12**, 279–284 (1993).
58. Szymanski, C. M., Logan, S. M., Linton, D. & Wren, B. W. Campylobacter - a tale of two protein glycosylation systems. *Trends in Microbiology* **11**, 233–238 (2003).
59. Igura, M. *et al.* Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO* **27**, 234–243 (2008).
60. Castro, O., Movsichoff, F. & Parodi, A. J. Preferential transfer of the complete glycan is determined by the oligosaccharyltransferase complex and not by the catalytic subunit. *Proc Natl Acad Sci U S A.* **103**, 14756–14760 (2006).
61. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J. & Aebi, M. All in One: Leishmania major STT3 Proteins Substitute for the Whole Oligosaccharyltransferase Complex in Saccharomyces cerevisiae. *Mol. Biol. Cell* **19**, 3758–3768 (2008).
62. Scheper, W., Thaminy, S., Kais, S., Stagljar, I. & Römisch, K. Coordination of N-glycosylation and protein translocation across the endoplasmic reticulum membrane by Sss1 protein. *J Biol Chem* **278**, 37998–8003 (2003).
63. Wang, L. & Dobberstein, B. Oligomeric complexes involved in translocation of proteins across the membrane of the endoplasmic reticulum. *FEBS Lett* **457**, 316–22 (1999).
64. Chavan, M., Yan, A. & Lennarz, W. J. Subunits of the Translocon Interact with Components of the Oligosaccharyl Transferase Complex. *J. Biol. Chem.* **280**, 22917–22924 (2005).
65. Rapoport, T. A. Protein transport across the endoplasmic reticulum membrane. *FEBS Journal* **275**, 4471–4478 (2008).
66. Tian, P. & Bernstein, H. D. Identification of a post-targeting step required for efficient cotranslational translocation of proteins across the E. coli inner membrane. *J Biol Chem* (2009).at <http://www.ncbi.nlm.nih.gov/pubmed/19211555>
67. Waraho, D. & DeLisa, M. P. Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3692–3697 (2009).
68. Choi, J. H. & Lee, S. Y. Secretory and extracellular production of recombinant proteins using Escherichia coli. *Applied Microbiology and Biotechnology* **64**, 625–635 (2004).
69. Fisher, A. C. *et al.* Production of Secretory and Extracellular N-Linked Glycoproteins in Escherichia coli. *Applied and Environmental Microbiology* **77**, 871–881 (2010).

70. Kowarik, M. *et al.* N-Linked Glycosylation of Folded Proteins by the Bacterial Oligosaccharyltransferase. *Science* **314**, 1148–1150 (2006).
71. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350–355 (2011).
72. Maita, N., Nyirenda, J., Igura, M., Kamishikiryo, J. & Kohda, D. Comparative Structural Biology of Eubacterial and Archaeal Oligosaccharyltransferases. *Journal of Biological Chemistry* **285**, 4941 –4950 (2010).
73. Ihssen, J. *et al.* Structural insights from random mutagenesis of Campylobacter jejuni oligosaccharyltransferase PglB. *BMC Biotechnology* **12**, 67 (2012).
74. Weller, S., Reynaud, C.-A. & Weill, J.-C. Vaccination against encapsulated bacteria in humans: paradoxes. *Trends in Immunology* **26**, 85–89 (2005).
75. Robbins, J. B., Schneerson, R., Szu, S. C. & Pozsgay, V. Bacterial polysaccharide–protein conjugate vaccines. *Pure and Applied Chemistry* **71**, 745–754 (1999).
76. Insel, R. & Anderson, P. Response to Oligosaccharide Protein Conjugate Vaccine Against Haemophilus-Influenzae-B in 2 Patients with Igg2 Deficiency Unresponsive. *N. Engl. J. Med.* **315**, 499–503 (1986).
77. Tai, J. *et al.* Haemophilus-Influenzae Type-B Polysaccharide Protein Conjugate Vaccine. *Proc. Soc. Exp. Biol. Med.* **184**, 154–161 (1987).
78. Vadheim cM, G. D. Eradication of haemophilus influenzae type b disease in southern california. *Arch Pediatr Adolesc Med* **148**, 51–56 (1994).
79. Jones, C. Vaccines based on the cell surface carbohydrates of pathogenic bacteria. *An. Acad. Bras. Cienc.* **77**, 293–324 (2005).
80. Ihssen, J. *et al.* Production of glycoprotein vaccines in Escherichia coli. *Microbial Cell Factories* **9**, 61 (2010).
81. Iwashkiw, J. A. *et al.* Exploiting the Campylobacter jejuni protein glycosylation system for glycoengineering vaccines and diagnostic tools directed against brucellosis. *Microb. Cell Fact.* **11**, 13 (2012).
82. Kesty, N. C. & Kuehn, M. J. Incorporation of Heterologous Outer Membrane and Periplasmic Proteins into Escherichia coli Outer Membrane Vesicles. *J. Biol. Chem.* **279**, 2069–2076 (2004).
83. Bauman, S. J. & Kuehn, M. J. Purification of outer membrane vesicles from Pseudomonas aeruginosa and their activation of an IL-8 response. *Microbes and Infection* **8**, 2400–2408 (2006).
84. Kesty, N. C., Mason, K. M., Reedy, M., Miller, S. E. & Kuehn, M. J. Enterotoxigenic Escherichia coli vesicles target toxin delivery into mammalian cells. *The EMBO Journal* **23**, 4538–4549 (2004).
85. Terra, V. S. *et al.* Recent developments in bacterial protein glycan coupling technology and glycoconjugate vaccine design. *Journal of Medical Microbiology* **61**, 919–926 (2012).
86. Walsh, G. Post-translational modifications of protein biopharmaceuticals. *Drug Discovery Today* **15**, 773–780 (2010).
87. Dimitrov, D. S.in *Therapeutic Proteins* (Voynov, V. & Caravella, J. A.) **899**, 1–26 (Humana Press, 2012).
88. Pandhal, J. & Wright, P. N-Linked glycoengineering for human therapeutic proteins in bacteria. *Biotechnology Letters* **32**, 1189–1198 (2010).

89. Bahlmann, F. H. & Fliser, D. Erythropoietin and renoprotection. *Curr Opin Nephrol Hypertens* **18**, 15–20 (2009).
90. Elliott, S. *et al.* Control of rHuEPO biological activity: The role of carbohydrate. *Experimental Hematology* **32**, 1146–1155 (2004).
91. Sinclair, A. M. & Elliott, S. Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *Journal of Pharmaceutical Sciences* **94**, 1626–1635 (2005).
92. Wilcox, R. G. *et al.* Trial of tissue plasminogen activator for mortality reduction in acute myocardial infarction: Anglo-Scandinavian study of early thrombolysis (ASSET). *The Lancet* **332**, 525–530 (1988).
93. Verstraete, M. Third-generation thrombolytic drugs. *The American Journal of Medicine* **109**, 52–58 (2000).
94. Raju, T. S., Briggs, J. B., Chamow, S. M., Winkler, M. E. & Jones, A. J. S. Glycoengineering of Therapeutic Glycoproteins: In Vitro Galactosylation and Sialylation of Glycoproteins with Terminal N-Acetylglucosamine and Galactose Residues. *Biochemistry* **40**, 8868–8876 (2001).
95. Liu, H., Wang, L., Brock, A., Wong, C.-H. & Schultz, P. G. A Method for the Generation of Glycoprotein Mimetics. *J. Am. Chem. Soc.* **125**, 1702–1703 (2003).
96. Sethuraman, N. & Stadheim, T. A. Challenges in therapeutic glycoprotein production. *Current Opinion in Biotechnology* **17**, 341–346 (2006).
97. Walsh, G. & Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat Biotech* **24**, 1241–1252 (2006).
98. Çelik, E., Fisher, A. C., Guarino, C., Mansell, T. J. & DeLisa, M. P. A filamentous phage display system for N-linked glycoproteins. *Protein Science* **9999**, NA (2010).
99. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008 (2006).
100. Pérez, J. M., McGarry, M. A., Marolda, C. L. & Valvano, M. A. Functional analysis of the large periplasmic loop of the Escherichia coli K-12 WaaL O-antigen ligase. *Mol Microbiol* **70**, 1424–40 (2008).
101. Brabetz, W., Muller-Loennies, S., Holst, O. & Brade, H. Deletion of the Heptosyltransferase Genes rfaC and rfaF in Escherichia Coli K-12 Results in an Re-Type Lipopolysaccharide with a High Degree of 2-Aminoethanol Phosphate Substitution. *European Journal of Biochemistry* **247**, 716–724 (1997).
102. Cordwell, S. J. *et al.* Identification of membrane-associated proteins from Campylobacter jejuni strains using complementary proteomics technologies. *Proteomics* **8**, 122–139 (2008).
103. Wyszyńska, A., Życka, J., Godlewska, R. & Jagusztyn-Krynicka, E. K. The Campylobacter jejuni / coli cjaA (cj0982c) Gene Encodes an N-Glycosylated Lipoprotein Localized in the Inner Membrane. *Curr Microbiol* **58**, 181–188 (2008).
104. Juncker, A. S. *et al.* Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **12**, 1652–1662 (2003).
105. Rocco, M. A. *et al.* Site-Specific Labeling of Surface Proteins on Living Cells Using Genetically Encoded Peptides that Bind Fluorescent Nanoparticle Probes. *Bioconjugate Chem.* **20**, 1482–1489 (2009).

106. Klemm, P. The fimA gene encoding the type-1 fimbrial subunit of Escherichia coli. *European Journal of Biochemistry* **143**, 395–399 (1984).
107. Brinton, C. C. The structure, function, synthesis and genetic control of bacterial pili and a molecular model for DNA and RNA transport in gram negative bacteria. *Transactions of the New York Academy of Sciences* **27**, 1003–1054 (1965).
108. Klemm, P. & Schembri, M. A. Fimbriae assisted bacterial surface display of heterologous peptides. *International Journal of Medical Microbiology* **290**, 215–221 (2000).
109. Aebi, M. Generous gift. *Swiss Federal Institue of Technology Zurich*
110. Opdenakker, G., Rudd, P. M., Ponting, C. P. & Dwek, R. A. Concepts and principles of glycobiology. *FASEB J* **7**, 1330–1337 (1993).
111. Knight, J., Beug, H., Marshall, J. & Hayman, M. J. Abnormal glycosylation of the env-sea oncogene product inhibits its proteolytic cleavage and blocks its transforming ability. *Oncogene* **2**, 317–326 (1988).
112. Herrmann, J. L., O'Gaora, P., Gallagher, A., Thole, J. E. & Young, D. B. Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from Mycobacterium tuberculosis. *EMBO J* **15**, 3547–3554 (1996).
113. Langsford, M. L. *et al.* Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. *FEBS Letters* **225**, 163–167 (1987).
114. Kapust, R. B., Tözsér, J., Copeland, T. D. & Waugh, D. S. The P1' specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **294**, 949–955 (2002).
115. Boeke, J. D. & Model, P. A prokaryotic membrane anchor sequence: carboxyl terminus of bacteriophage f1 gene III protein retains it in the membrane. *PNAS* **79**, 5200–5204 (1982).
116. Harvey, B. R. *et al.* Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries. *PNAS* **101**, 9193–9198 (2004).
117. Fikes, J. & Bassford, P. Export of Unprocessed Precursor Maltose-Binding Protein to the Periplasm. *J. Bacteriol.* **169**, 2352–2359 (1987).
118. He, H., Yi-Ru, G. & Yan, S. Cloning and fusion expression of bovine enterokinase light chain gene in Escherichia Coli. *Yichuan* **25**, 685–690 (2003).
119. Eaton, D., Rodriguez, H. & Vehar, G. A. Proteolytic processing of human factor VIII. Correlation of specific cleavages by thrombin, factor Xa, and activated protein C with activation and inactivation of factor VIII coagulant activity. *Biochemistry* **25**, 505–512 (1986).
120. Fisher, A. C. Unpublished data revealing gycosylation of scFVR4-EAQNAT by bacterial oligosaccharyltransferases. (2011).
121. Liu, D., Xu, R., Dutta, K. & Cowburn, D. N-terminal cysteinyl proteins can be prepared using thrombin cleavage. *FEBS Lett* **582**, 1163–1167 (2008).
122. Petrassi, H. M. *et al.* A strategy to profile prime and non-prime proteolytic substrate specificity. *Bioorganic & Medicinal Chemistry Letters* **15**, 3162–3166 (2005).
123. Lottenberg, R., Hall, J. A., Blinder, M., Binder, E. P. & Jackson, C. M. The action of thrombin on peptide p-nitroanilide substrates. Substrate selectivity and

examination of hydrolysis under different reaction conditions. *Biochim. Biophys. Acta* **742**, 539–557 (1983).
124. Chang, J.-Y. Thrombin specificity. *European Journal of Biochemistry* **151**, 217–224 (1985).
125. Gallwitz, M., Enoksson, M., Thorpe, M. & Hellman, L. The Extended Cleavage Specificity of Human Thrombin. *PLoS ONE* **7**, e31756 (2012).
126. Schmitz, U., Versmold, A., Kaufmann, P. & Frank, H. G. Phage display: a molecular tool for the generation of antibodies--a review. *Placenta* **21 Suppl A**, S106–112 (2000).
127. Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* **228**, 1315–1317 (1985).
128. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**, 433–455 (1994).
129. Rader, C. & Barbas, C. F., 3rd Phage display of combinatorial antibody libraries. *Curr. Opin. Biotechnol.* **8**, 503–508 (1997).
130. Smith, G. P. Surface presentation of protein epitopes using bacteriophage expression systems. *Curr. Opin. Biotechnol.* **2**, 668–673 (1991).
131. Viti, F., Nilsson, F., Demartis, S., Huber, A. & Neri, D. Design and use of phage display libraries for the selection of antibodies and enzymes. *Meth. Enzymol.* **326**, 480–505 (2000).
132. Watters, A. L. & Baker, D. Searching for folded proteins in vitro and in silico. *Eur. J. Biochem.* **271**, 1615–1622 (2004).
133. Forrer, P., Jung, S. & Plückthun, A. Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. *Curr. Opin. Struct. Biol.* **9**, 514–520 (1999).
134. Hills, R. *et al.* Identification of an ADAMTS-4 cleavage motif using phage display leads to the development of fluorogenic peptide substrates and reveals matrilin-3 as a novel substrate. *J. Biol. Chem.* **282**, 11101–11109 (2007).
135. Cloutier, S. M. *et al.* Substrate specificity of human kallikrein 2 (hK2) as determined by phage display technology. *Eur. J. Biochem.* **269**, 2747–2754 (2002).
136. Lafond, R., Zhan, X. & Georgiou, G. Screening and selection strategies for disulfide isomerase activity. *Methods Mol. Biol.* **230**, 239–257 (2003).
137. Gannon, P., Li, P. & Kumamoto, C. The Mature Portion of Escherichia-Coli Maltose-Binding Protein (mbp) Determines the Dependence of Mbp on Secb for Export. *J. Bacteriol.* **171**, 813–818 (1989).
138. Chen, M. M., Glover, K. J. & Imperiali, B. From Peptide to Protein: Comparative Analysis of the Substrate Specificity of N-Linked Glycosylation in C. jejuni. *Biochem* **46**, 5579–5585 (2007).
139. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol.* **177**, 4121–4130 (1995).
140. Swamy, M. J., Krishna Sastry, M. V., Khan, M. I. & Surolia, A. Thermodynamic and kinetic studies on saccharide binding to soya-bean agglutinin. *Biochem. J.* **234**, 515–522 (1986).

141. Pandhal, J. *et al.* Systematic metabolic engineering for improvement of glycosylation efficiency in Escherichia coli. *Biochem. Biophys. Res. Commun.* **419**, 472–476 (2012).
142. Monsigny, M., Mayer, R. & Roche, A. C. Sugar-lectin interactions: sugar clusters, lectin multivalency and avidity. *Carbohydr Lett* **4**, 35–52 (2000).
143. Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *The EMBO Journal* **25**, 1957–1966 (2006).
144. Dürr, C., Nothaft, H., Lizak, C., Glockshuber, R. & Aebi, M. The Escherichia coli Glycophage Display System. *Glycobiology* (2010).doi:10.1093/glycob/cwq102
145. Paschke, M. Phage display systems and their applications. *Appl. Microbiol. Biotechnol.* **70**, 2–11 (2006).
146. Hertveldt, K., Beliën, T. & Volckaert, G. General M13 phage display: M13 phage display in identification and characterization of protein-protein interactions. *Methods Mol. Biol.* **502**, 321–339 (2009).
147. Dwek, R. A., Butters, T. D., Platt, F. M. & Zitzmann, N. Targeting glycosylation as a therapeutic approach. *Nature Reviews Drug Discovery* **1**, 65–75 (2002).
148. Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S. & Aebi, M. The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology* **15**, 361–367 (2005).
149. Li, L. *et al.* Overexpression and topology of bacterial oligosaccharyltransferase PglB. *Biochem. Biophys. Res. Commun* **394**, 1069–1074 (2010).
150. Kohda, D., Yamada, M., Igura, M., Kamishikiryo, J. & Maenaka, K. New oligosaccharyltransferase assay method. *Glycobiology* **17**, 1175–1182 (2007).
151. Glover, K. J., Weerapana, E., Numao, S. & Imperiali, B. Chemo-enzymatic synthesis of glycopeptides using PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. *Chem Biol.* **12**, 1311–1315 (2005).
152. Ielmini, M. V. & Feldman, M. F. Desulfovibrio desulfuricans PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology* (2010).doi:10.1093/glycob/cwq192
153. Jervis, A. J. *et al.* Characterisation of N-linked protein glycosylation in Helicobacter pullorum. *J Bacteriol* (2010).doi:10.1128/JB.00211-10
154. Schwarz, F. *et al.* Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology* **21**, 45 –54 (2011).
155. Ielmini, M. V. & Feldman, M. F. Desulfovibrio desulfuricans PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology* **21**, 734–742 (2011).
156. Lefebre, M. D. & Valvano, M. A. Construction and Evaluation of Plasmid Vectors Optimized for Constitutive and Regulated Gene Expression in Burkholderia cepacia Complex Isolates. *Appl. Environ. Microbiol.* **68**, 5956–5964 (2002).
157. Philibert, P. & Martineau, P. Directed evolution of single-chain Fv for cytoplasmic expression using the β-galactosidase complementation assay results in proteins highly susceptible to protease degradation and aggregation. *Microbial Cell Factories* **3**, 16 (2004).

158. Lutz, R. & Bujard, H. Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucl. Acids Res.* **25**, 1203–1210 (1997).
159. Katzen, F., Chang, G. & Kudlicki, W. The past, present and future of cell-free protein synthesis. *Trends Biotechnol.* **23**, 150–156 (2005).
160. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
161. Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J. & Swartz, J. R. An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol. Syst. Biol.* **4**, 220 (2008).
162. Ryabova, L. A., Desplancq, D., Spirin, A. S. & Plückthun, A. Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones. *Nat. Biotechnol.* **15**, 79–84 (1997).
163. Noireaux, V., Bar-Ziv, R. & Libchaber, A. Principles of cell-free genetic circuit assembly. *Proc Natl Acad Sci U S A* **100**, 12672–12677 (2003).
164. Yang, J. *et al.* Rapid expression of vaccine proteins for B-cell lymphoma in a cell-free system. *Biotechnol. Bioeng.* **89**, 503–511 (2005).
165. Helenius, A. & Aebi, M. Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annual Review of Biochemistry* **73**, 1019–1049 (2004).
166. Tarui, H., Imanishi, S. & Hara, T. A novel cell-free translation/glycosylation system prepared from insect cells. *J. Biosci. Bioeng.* **90**, 508–514 (2000).
167. Lingappa, V. R., Lingappa, J. R., Prasad, R., Ebner, K. E. & Blobel, G. Coupled cell-free synthesis, segregation, and core glycosylation of a secretory protein. *Proc Natl Acad Sci U S A* **75**, 2338–2342 (1978).
168. Rothblatt, J. A. & Meyer, D. I. Secretion in yeast: reconstitution of the translocation and glycosylation of alpha-factor and invertase in a homologous cell-free system. *Cell* **44**, 619–628 (1986).
169. Moreno, S. N., Ip, H. S. & Cross, G. A. An mRNA-dependent in vitro translation system from Trypanosoma brucei. *Mol. Biochem. Parasitol.* **46**, 265–274 (1991).
170. Mikami, S., Kobayashi, T., Yokoyama, S. & Imataka, H. A hybridoma-based in vitro translation system that efficiently synthesizes glycoproteins. *J. Biotechnol.* **127**, 65–78 (2006).
171. Shibutani, M., Kim, E., Lazarovici, P., Oshima, M. & Guroff, G. Preparation of a cell-free translation system from PC12 cell. *Neurochem. Res.* **21**, 801–807 (1996).
172. Reid, C. W. *et al.* Affinity-capture tandem mass spectrometric characterization of polyprenyl-linked oligosaccharides: tool to study protein N-glycosylation pathways. *Anal. Chem.* **80**, 5468–5475 (2008).
173. Martineau, P., Jones, P. & Winter, G. Expression of an antibody fragment at high levels in the bacterial cytoplasm. *J. Mol. Biol.* **280**, 117–127 (1998).
174. Mansell, T. J. & DeLisa, M. P. Unpublished observations.
175. Yavuz, E., Maffioli, C., Ilg, K., Aebi, M. & Priem, B. Glycomimicry: display of fucosylation on the lipo-oligosaccharide of recombinant Escherichia coli K12. *Glycoconj. J.* **28**, 39–47 (2011).

176. Schwarz, F. *et al.* A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat Chem Biol* **6**, 264–266 (2010).
177. Aebi, M. Personal communication.
178. Igura, M. & Kohda, D. Selective Control of Oligosaccharide Transfer Efficiency for the N-Glycosylation Sequon by a Point Mutation in Oligosaccharyltransferase. *Journal of Biological Chemistry* **286**, 13255–13260 (2011).
179. Klammt, C. *et al.* Evaluation of detergents for the soluble expression of alpha-helical and beta-barrel-type integral membrane proteins by a preparative scale individual cell-free expression system. *FEBS J.* **272**, 6024–6038 (2005).
180. Harris, J. M. & Chess, R. B. Effect of pegylation on pharmaceuticals. *Nature Reviews Drug Discovery* **2**, 214–221 (2003).
181. Pasut, G. & Veronese, F. M. State of the art in PEGylation: The great versatility achieved after forty years of research. *Journal of Controlled Release* **161**, 461–472 (2012).
182. Marshall, S. A., Lazar, G. A., Chirino, A. J. & Desjarlais, J. R. Rational design and engineering of therapeutic proteins. *Drug Discovery Today* **8**, 212–221 (2003).
183. Cornish, V. W., Hahn, K. M. & Schultz, P. G. Site-Specific Protein Modification Using a Ketone Handle. *J. Am. Chem. Soc.* **118**, 8150–8151 (1996).
184. Shaunak, S. *et al.* Site-specific PEGylation of native disulfide bonds in therapeutic proteins. *Nat Chem Biol* **2**, 312–313 (2006).
185. Thom, J., Anderson, D., McGregor, J. & Cotton, G. Recombinant Protein Hydrazides: Application to Site-Specific Protein PEGylation. *Bioconjugate Chem.* **22**, 1017–1020 (2011).
186. Fontana, A., Spolaore, B., Mero, A. & Veronese, F. M. Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. *Advanced Drug Delivery Reviews* **60**, 13–28 (2008).
187. DeFrees, S. *et al.* GlycoPEGylation of recombinant therapeutic proteins produced in Escherichia coli. *Glycobiology* **16**, 833–843 (2006).
188. Østergaard, H. *et al.* Prolonged half-life and preserved enzymatic properties of factor IX selectively PEGylated on native N-glycans in the activation peptide. *Blood* **118**, 2333–2341 (2011).
189. Negrier, C., Knobe, K., Tiede, A., Giangrande, P. & Møss, J. Enhanced pharmacokinetic properties of a glycoPEGylated recombinant factor IX: a first human dose trial in patients with hemophilia B. *Blood* **118**, 2695–2701 (2011).
190. Price, N. P. & Momany, F. A. Modeling bacterial UDP-HexNAc: polyprenol-P HexNAc-1-P transferases. *Glycobiology* **15**, 29R–42R (2005).
191. Schwarz, F. *et al.* A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat Chem Biol* **6**, 264–266 (2010).
192. Weerapana, E., Glover, K. J., Chen, M. M. & Imperiali*, B. Investigating Bacterial N-Linked Glycosylation: Synthesis and Glycosyl Acceptor Activity of the Undecaprenyl Pyrophosphate-Linked Bacillosamine. *J Am Chem Soc.* **127**, 13766–13767 (2005).

193. Kim, T.-H. *et al.* Characterization of the O-antigen Polymerase (Wzy) of Francisella tularensis. *J. Biol. Chem.* **285**, 27839–27849 (2010).
194. Kalynych, S., Yao, D., Magee, J. & Cygler, M. Structural Characterization of Closely Related O-antigen Lipopolysaccharide (LPS) Chain Length Regulators. *J. Biol. Chem.* **287**, 15696–15705 (2012).
195. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009).
196. Hirayama, K. & Rafter, J. The role of probiotic bacteria in cancer prevention. *Microbes and Infection* **2**, 681–686 (2000).
197. Iwase, T. *et al.* Staphylococcus epidermidis Esp inhibits Staphylococcus aureus biofilm formation and nasal colonization. *Nature* **465**, 346–349