THESIS

THE DEVELOPMENT OF A MOLECULAR TOOLBOX TO EXAMINE THE ROLE OF

PROTEIN *O*-MANNOSYLATION IN *ACTINOBACTERIA*

Submitted by

Nakita Buenbrazo

to the

Department of Chemistry and Biology

In partial fulfillment of the requirements

For the Degree of Master of Science

Ryerson University

Toronto, Ontario

Winter 2016

Master's Committee:

    Advisor: Warren Wakarchuk

    Michael Suits

    Deborah Foster

AUTHOR'S DECLARATION FOR THE ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

1. Abstract

Protein glycosylation is the most abundant and diverse protein modification that occurs in all domains of life. It is defined as the covalent attachment of a carbohydrate moiety to a specific amino acid on a target protein. The functional role of this attachment is implicated in and spans various cell processes from cell signaling, cell defense, and pathogenesis – to name a few. A specific type of protein glycosylation, called protein *O*-mannosylation (POM) is a process found to be conserved from bacteria to man. In humans, POM is required for healthy cell function, and the absence of POM can cause fatal diseases. Certain prokaryotic species possess a related POM system, but it is poorly understood. It is our hypothesis that the analysis of the POM system in simpler organisms can aid in the characterization of this process and the functional role of the mannosylated proteins that are produced. However, the protocols to prove this theory do not yet exist. This thesis establishes a collection of developed protocols that can be used to characterize the POM systems from gram-positive species *Corynebacterium glutamicum* and *Cellulomonas fimi*. In addition the first ever evidence of a *C. fimi* glycoprotein being glycosylated by the endogenous *C. glutamicum* POM system is provided.

Acknowledgements

During the time in this degree I thankfully received support and help from many people.

In particular, I would like to express my gratitude to my supervisor Dr. Warren Wakarchuk for the guidance, expertise, and endless patience. Dr. Wakarchuk tailors his direction to the specific needs and personalities of his students, supporting everyone individually but allowing all of us to feel like a team. I am grateful for the chance to work in this laboratory and the mentorship of Dr. Wakarchuk to help me become an accomplished scientist.

I am grateful to Ryerson University and The Department of Chemistry and Biology for the opportunity to study in the molecular science graduate program and learn from the accomplished faculty. Thank-you to my committee for supporting me and for taking the time to be present and provide invaluable insight to this project

And finally, but not least, thank-you to my entire family, and especially my husband, Joseph Crampton, who has supported me in all endeavors and especially this one, the one that stuck, for the pursuit of happiness.

To all of my lab colleagues these past two years, we must endure together through all of the ConA blots of the future.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

6HIs	6 Histidine repeat ([CATCAC] X3)
AA	amino acid
ABC	ATP-binding cassette transporter
Ala	alanine
aph	aminoglycoside 3-N-acetyltransferase
Anti-6His-P	antibody 6histidine peroxidase
ASB	amidosulfobetaine
Asp	asparagine
BCA	Bicinchoninic acid assay
bp	base pair
BSA	bovine serum albumine
CDGs	congenital disorders of glycosylation
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CLb	CelLytic^{TM} reagent
ConA	Concanavalin A
Dol-P-Man	Dolichol phosphate β-D-mannose
ECL	enhanced chemiluminescent substrate
eGFP	enhanced green fluorescent protein
GH	glycosylhydrolase
Gly	glycine
GT	glycosyltransferase
GT39	glycosyltransferase family 39
hLP	hypothetical lipoprotein
IPTG	isopropyl β -D-thiogalactopyranose
IMAC	Immobilized mesh affinity chromatography
Kan	kanamycin
kDa	kilodalton
MCS	multiple cloning site
NEB	New England Biolabs
OD	optical density
PCR	polymerase chain reaction
POM	protein *O*-mannosylation
PMT	protein mannosyltransferase

PPM	polyprenol monophosphomannose
Pro	proline
PTM	post-translational modification
RBS	ribosome binding site
Ser	serine
Thr	threonine
UPOM	Unfolded protein *O*-mannosylation
UTR	untranslated region
wt	wild-type (*C .glutamicum* strain ATCC 13032)

2. Introduction

2.1. Prokaryotic Protein Glycosylation

Protein glycosylation is a quintessential and exceptionally complex event taking place in cells from all three domains of life: archaea, bacteria, and eukaryotes. It is referred to as the most abundant post-translational modification (PTM) in nature and is defined as the covalent attachment of a mono- or oligo- saccharide, a glycan, to a specific target site on a peptide molecule (Figure 1) [1, 2, 3]. This process is so complex, that for almost half of a century it was thought to be unique to eukaryotes. During this time it was believed that the simpler organisms, bacteria and archaea, would not undertake such a diverse and energetic feat like glycosylation, due to their short life span and lack of the cellular compartments that are required for glycosylation in eukaryotes [4, 5, 6]. This concept was dismissed after the discovery of the first glycoprotein, a surface layer (S-layer protein) with an oligosaccharide attachment, in an archaeon *Halobacterium salinarium* [5, 7, 8]. In fact, this discovery marks a transformative point in the field of Glycobiology; it was now likely that prokaryotes held the key to the evolution of this process from simple to complex organisms, a cellular investment that occupies 1-2% of the genome [3]. In humans, problems with glycosylation can lead to serious, even fatal disease [3, 9]. Most well studied are the congenital disorders of glycosylation (CDGs) causing developmental and neurological deficiencies in children and are directly linked to problems in glycan formation [10, 11, 12]. Understanding glycosylation is vital to understanding and curing CDGs, an objective that could be attainable within the studies of prokaryotic glycosylation.
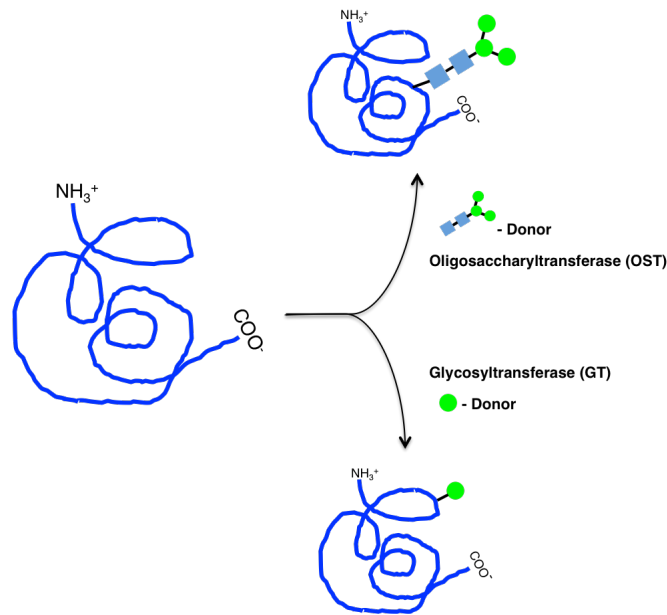
Figure 1: Glycosylation simple model. Two types of transferase enzymes are involved in the covalent attachment of a sugar or sugar chain to a target amino acid (AA) on a protein molecule. These transferase enzymes require a specific lipid linked donor molecule to provide the carbohydrate(s). Oligosaccharyltransferase (OST) transfers a pre-built sugar chain to a protein, while a glycosyltransferase (GT) transfers one sugar to either a protein or developing sugar chain.

The relative ease of studying and manipulating glycosylation pathways in simpler organisms has led to many significant discoveries [13], a majority of these confirm two important observations 1) bacterial glycans are, remarkably, more diverse than those found in eukaryotic cells [14], and 2) the inhibition of glycoprotein synthesis in pathogenic bacteria is not fatal, but rather affects the organisms ability to proliferate within a host [13, 15, 16]. It was clear that bacterial glycoproteins were not only evolutionarily significant but now an effective target for therapeutics against medically important human bacterial pathogens [6] (*Staphylococcus aureus* [1, 2], *Escherichia coli* [17], *Campylobacter* spp. [18], *Helicobacter pylori* [19], *Pseudomonas aeruginosa* [20], *Mycobacterium tuberculosis* [15, 21], and many others [6, 21]) especially in this new era of antibiotic resistant strains. Glycans exhibited on the surface of the pathogenic bacteria can evade the host immune system by binding to the host innate immune-system receptors [21, 22, 23]. Efforts have been focused on targeting the endogenous glycosylation pathways that allow

the production of these glycoconjugates, which would in turn affect their surface glycans [21]. The recently exposed extent and variety of prokaryotic glycans fuels the bright future of this field.

Although a lot of useful and compelling work has taken place, many fundamental questions remain unanswered, such as - what are the functional roles of carbohydrate modification, what are the mechanisms and target sites of protein glycosylation, how is this process regulated and able to rapidly adjust to environmental factors - to name a few. Our understanding of this process is limited, and in order to fill in the blanks efforts must be focused on the development of the tools and protocols required to efficiently study these biosynthetic pathways from genome to glycome in prokaryotic species.

2.2. Types of Bacterial Glycosylation

Though glycosylation is most commonly associated with the modification of a protein by a sugar molecule, other types of molecular targets exist, such as glycerophosphatides and sphingolipids (in eukaryotes) [11, 24], and lipids (in bacteria) [25, 26]. Known bacterial glycoconjugates include peptidoglycan, lipopolysaccharide, capsule, and teichoic acids [27]. Naming conventions for broad sub-categories of glycosylation denote the molecular linker between the sugar and the target molecule. Today, four of the six unique types of glycosylation have been discovered in bacteria: N-linked [27], O-linked [29], C-linked [23, 30], and most recently S-linked [31]. S-linked was only just reported by Stepper et al. (2011) to be a sulfur bridge between a Cysteine (Cys) residue on a protein and a glycan [23, 30]. C-linked is a rarity in bacterial cells but refers to the glycan attachment to a carbon molecule on a tryptophan side chain. N- and O- linked glycosylation are the most common types and found in both bacteria and eukaryotic cells (Figure 2) [1], to date only N-linked has been reported in archaea [7].
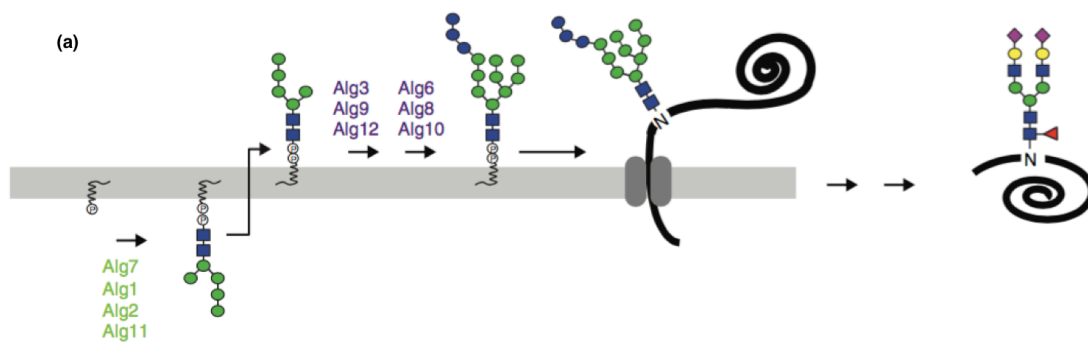
Figure 2. Model representations of N- and O-linked glycosylation. (a) N-linked glycosylation showing the specific GTs shown to have been functional in a step of the process. GTs involved in transfer to lipid donor are cytosolic and highly conserved (green), GTs involved in transfer to the protein are ER resident (eukaryotic) and less conserved as bacterial paralogs are periplasmic. (b) O-linked glycosylation process in the ER (eukaryotic) would be emulated in the plasma membrane of bacteria. The Dol-P-Man is the mannosyl donor for all processes involving mannose sugar moieties. This shows the assembly of core oligosaccharides usually by N-linked glycosylation (1), transfer on Ser/Thr residue via O-linked glycosylation (2), and formation of glycerophosphatide anchors (3).

N-linked glycosylation (Figure 2a) is the better known and studied type of glycosylation, and refers to the *en bloc* transfer of a glycan chain from a donor substrate to a target sequon on a peptide. The enzyme that catalyzes this reaction, oligosaccharyltransferase, recognizes the sequon Asn-Xaa-Ser/Thr (A-X-S/T), and attaches the glycan onto the invariant asparagine residue creating a nitrogen linker between the sugar and the amino acid (AA) [3, 5, 7, 31]. O-linked glycosylation (Figure 2b) is lesser studied, but is gaining momentum in the glycosphere as it appears to be a frequent PTM in bacterial cells. O-linked glycosylation differs from N-linked in three key ways: 1) the sugar and peptide are joined by an **oxygen** linker, 2) the attachment point is a hydroxyl group on a hydroxy AA residue of the protein; the most common being **serine (Ser) or threonine (Thr)** but have been reported on tyrosine, hydroxyproline, or hydroxylysine, and 3) the glycan is often constructed consecutively **after** the synthesis of the first sugar-AA bond [2, 32].

More distinct classification regarding types of glycosylation reactions follow logically, this can be done in a few different ways that include using the general type of molecular target, type/name of enzyme used to catalyze the reaction, and/or type of attached monosaccharide. The type of glycosylation central to this body of work is protein-*O*-mannosylation (POM) an oligosaccharide-independent glycosylation reaction, where by a mannose sugar is covalently attached to a Ser/Thr residue on a specific polypeptide.

2.3. Protein-*O*-mannosylation

Protein-*O*-mannosylation (POM) is a highly conserved type of glycosylation. Evidence of this type of glycosylation is found in fungi, bacteria, and eukaryotes [2,33]. Most of the work done thus far with regards to POM has been done in yeast cells [13] and has contributed to our knowledge of this process in human cells [34].

In humans, POM is essential to normal cell function, prominent CDGs and associated brain and ocular abnormalities are directly related to a dysfunctional O-mannosylation process [11, 35]. So far, cells that express mannosylated proteins have been isolated from brain tissue, nerve and muscular tissue, and on neuron-specific receptors [35, 36, 37]. An example of a notable O-mannosylated protein is α–dystroglycan, isolated from animal skeletal tissue [30]. The hypoglycosylation of this protein and the associated flawed mannosylation pathway is indicated in many significant CDGs such as Walker-Warburg syndrome (fatal), muscle-eye-brain disease, and Fukuyama congenital muscular dystrophy to name a few [35, 38]. As such, continued discovery and dissection of the O-mannosylation pathway and the molecular components involved have great medical significance. Consequently, the discovery of mannosylated protein in bacteria suggests that POM has evolutionary significance as well.

A majority of the mannosylated proteins discovered in any cell to this date are secreted or cell membrane proteins [33, 34, 39, 40]. In fact, in yeast and fungi most secreted and cell wall proteins are *O*-mannosylated, which helps to narrow in on a functional category for this PTM [13]. In order to produce these proteins cells must have the molecular machinery consisting of four mandatory components [33]: an activated mannose residue, a mannose-carrying lipid donor, a mannosyl-transferase enzyme, and the target protein containing the recognized sequon. The specific molecular make-up of some of these components can differ from cell to cell and is especially different between prokaryotes and eukaryotes [33], yet the steps of this reaction are conserved. Figure 3 demonstrates the POM reaction in eukaryotes (Figure 3a) and prokaryotes (Figure 3b). There are four basic steps of POM. First, the mannosylated lipid donor is synthesized through the action of a specific glycosyltransferase, which transfers the mannose from the activated donor GDP-α-D-Mannose to the phosphorylated lipid [41 – 44]. Second, the mannosylated lipid is flipped across the membrane where it will interact with the principal enzyme, the protein *O*-mannosyltransferase (*PMT*) [43]. Third, the mannose is then transferred

6

from the lipid to the *PMT* glycosyltransferase enzyme. Lastly, as the target protein is translated from the ribosome it is translocated through the membrane, the *PMT* recognizes the sequon required for mannosylation and, through an inverting mechanism, transfers the mannose to the hydroxyl AA (most likely a Ser or Thr residue) [45, 46]. Afterward, the primed glycoprotein will undergo further transferase reactions with distinctive enzymes to complete the glycan chain and produce the final glycoprotein product [47].
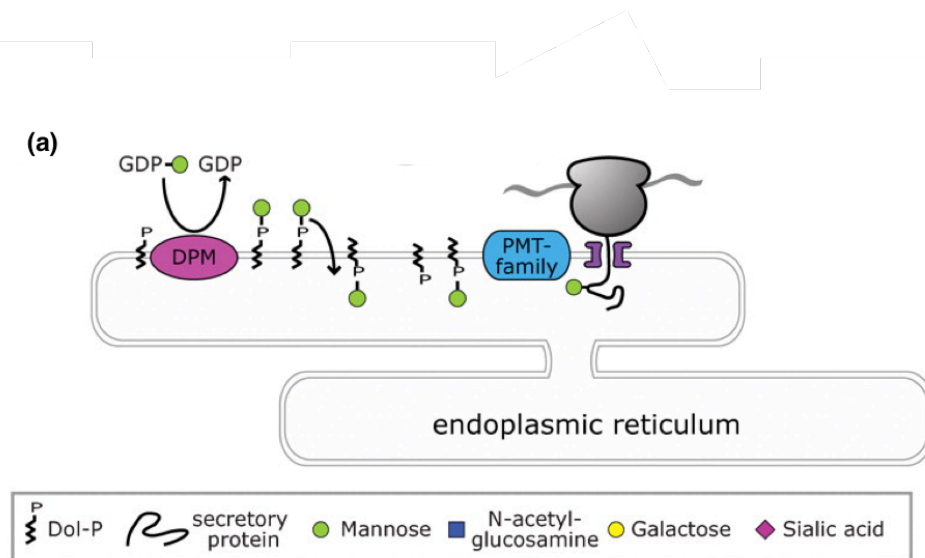
Image obtained from Lommel, M., & Strahl, S. (2009). Protein O-mannosylation: Conserved from bacteria to humans. *Glycobiology, 19*(8), 816–828. http://doi.org/10.1093/glycob/cwp066

(b)

Cell Wall

Pmt 1 ⓟ

SecYEG

SecA

Cytosol

ⓟ = Decaprenyl phosphate

⬡ = Proximal mannose
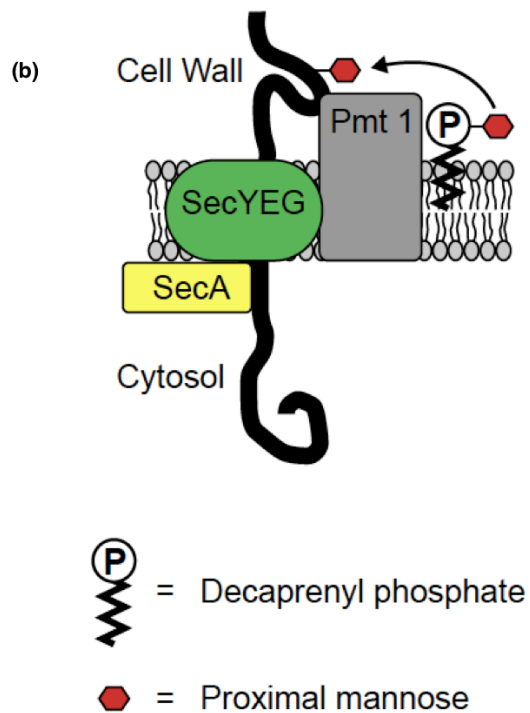
Figure 3. Model of protein *O*-mannosylation from eukaryote and prokaryote. (a) Dol-P-Man is synthesized in the ER membrane on the cytosolic side and flips to the luminal side by an unknown mechanism. The mannose is transferred from Dol-P-Man to a Ser/Thr residue on the protein co-translationally by PMT integral membrane protein. (b) In bacteria as the target protein is being translocated across the membrane into the periplasmic space, the PMT, also an integral membrane protein but within the plasma membrane transfers the mannose from the *PPM* to the target Ser/Thr residue.

The most evident POM difference between eukaryotes and prokaryotes is the location. Bacteria lack the cellular compartments of eukarya and thus POM occurs across the plasma membrane instead of the endoplasmic reticulum [1]. Elongation of the glycan chain is completed in the Golgi apparatus for eukaryotes and (most likely) in the periplasmic space in prokaryotes although this mechanism is poorly understood [1, 33]. Other notable differences follow. In prokaryotes, the initial reaction to synthesize the mannosylated lipid donor uses a polyprenol monophosphate as the initial lipid molecule and a polyprenyl monophosphomannose

synthase (a glycosyltransferase in GT-2) [48] to carry out the subsequent reaction creating a polyprenol monophosphomannose (*PPM*) as the lipid donor; in eukaryotes the lipid donor is a dolichol monophosphate (Figure S1) [49], although this donor shows significant homology to the polyprenyl monophosphate (Figure S1), the eukaryotic synthase and mannose product are specific to the dolichol molecule. The *PMT* enzyme catalyzes the most crucial part of the POM reaction [50]. Eukaryotic cells carry multiple homologs of this enzyme, supporting the fact that this PTM is necessary for cell viability [13]. However, multiple whole genome analysis studies have only ever found one *PMT* gene in any bacterial species [2, 39]. Remarkably, all *PMT* enzymes found to this point show enough homology to be grouped into one family, glycosyltranferase family 39 (GT39), indicating that the chief part of this process has been conserved throughout time, and throughout evolution [33]. To date, studies have focused on these enzymes in eukaryotic species, and little is known about the bacterial counterpart. It is evident, based on the prevalent and frequent nature of POM from bacteria to humans that further exploration of the bacterial *PMT* enzyme can be of great importance to the scientific community.

2.4. Glycosyltransferase Family GT39

In 1998 a database known as the Carbohydrate-Active Enzyme database, was created to collect and organize information on all of the enzymes that are involved in glycan reactions [51, 52]. Enzymes that build or breakdown complex carbohydrates and glycoconjugates are grouped into families based on sequence similarity and function [51]. The overwhelming majority of the enzymes are either a glycosylhydrolase enzyme (GH) or a glycosyltransferase enzyme (GT). There are approximately 120 000 GT sequences recorded belonging to 97 different glycosyltranferase families [52]. The glycosyltransferase family 39 exclusively contains all of the *PMT* enzymes discovered to date.

All *PMT* enzymes discovered are integral membrane proteins containing several transmembane domains and significant homology [2, 10]. The organism in which the *O*-mannosylation process has been studied the most is the Baker's yeast *Saccharomyces cerevisiae*, which contains seven *PMT* genes subdivided into the three *PMT* subfamilies (*PMT*1, *PMT*2, and *PMT*4) (10, 13, 21, 37). Figure 4 shows the predicted architecture of the *S. cerevisiae* PMT1.
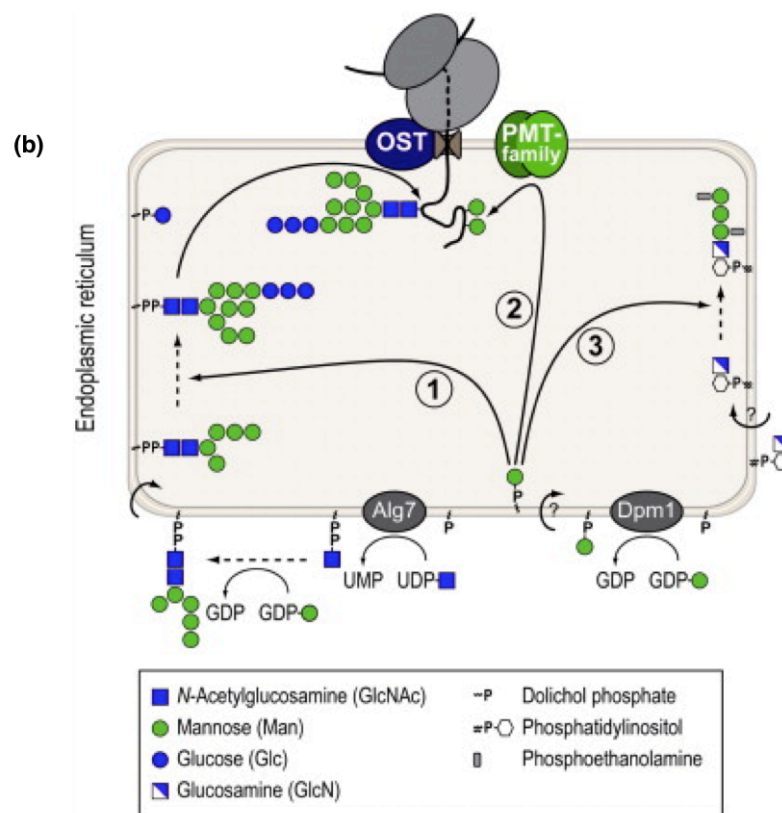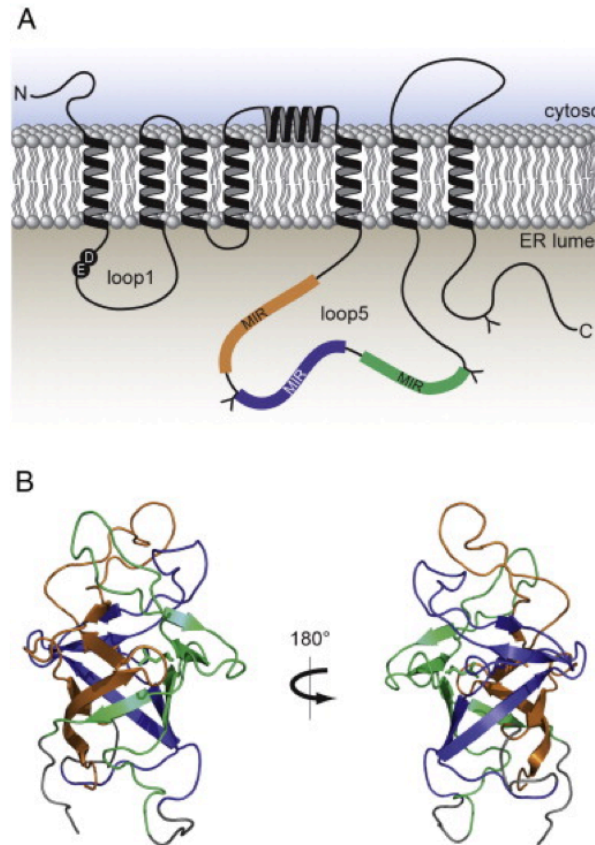

Image obtained from Loibl, M., & Strahl, S. (2013). Protein O-mannosylation: What we have learned from baker's yeast. *BBA Molecular Cell Research*, *1833*(11), 2438–2446

Figure 4. The predicted architecture of the *S. cerevisiae PMT*1. (A) Topology of PMT showing conserved D_{55} E_{56} (Asp-Glu) in the first luminal hydrophilic loop. (Y) N-glycosylation sites along the primary structure, and (MIR) mannosyl transferase inositol triphosphate- and ryanodine receptor. MIR is essential for enzymatic activity and is unique to eukaryotes and possible an ER specific domain. (B) Ribbon representation of *PMT*4 by SWISS MODEL server.

Initially, multiple isoforms of *PMT* in the eukaryotes were thought to be solely due to functional redundancy in response to the necessity of the POM process in cells [21]. This theory was revised when Girrbach et al. (2003) reported that the members from the *PMT*1 family form heterodimers with *PMT*2 members, and members of the *PMT*4 family form homodimers [37]. Other types of fungi have been shown to have at least one *PMT* homologue from each subfamily [21]. Humans contain one *PMT* from each subfamily *PMT*2 and *PMT*4 [21]. Within an organism, all *PMT* genes seem to have between 50%-60% sequence similarities, but cross species *PMT*s have variable sequence differences [21]. For example, the only *PMT*s experimentally verified in prokaryotic species are from *M. tuberculosis* and the taxonomically related *C. glutamicum* [53, 54]. Vandervan et al. (2005) reported on the *M. tuberculosis* PMT having a 22-24% sequence similarity to the *PMT*1 from *S. cerevisiae* [53], and Mahne et al. (2006) reported the *C. glutamicum PMT* has a 15-18% to the same yeast enzyme [54]. However, these two prokaryotic *PMTs* have a 40% sequence similarity to each other [54]. Using bioinformatics, putative *PMT* genes from other bacteria have been uncovered including *Cellulomonas fimi* (this work), *Cellulomonas flavigena* (this work), *Streptomyces coelicor*, *Propionibacterium acnes*, *Mycobacterium leprae*, *Mycobacterium bovis*, and *Corynebacterium diphtheriae* [54].

Although three-dimensional structures of any of the *PMT* proteins have yet to be verified, topology-sensitive reporters, as well as hydropathy profiles of several *PMT* enzymes have been published [53 – 56]. *PMTs* across domains seem to maintain three characteristics 1) the N-terminal end is within the cytosol of the cell, where as the C-terminal end is on the other side of the membrane (endoplasmic reticulum or plasma membrane), 2) the structure contains multiple hydrophilic loops on either side of the membrane (three loops for eukarya [12, 13], and one or two loops for bacteria [53, 54]), and 3) the first periplasmic/luminal loop (outside the plasma membrane/endoplasmic reticulum) forms part of the catalytic site of the enzyme carrying the conserved *PMT* active site residues (D^{55} and E^{56}) [45]. Figure 5 shows the *M. tuberculosis* model of the *PMT* enzyme. Most research regarding the

structure-function relationship of the *PMT* enzyme has been on the *S. cerevisiae* *PMT*1 [13], although bacterial *PMT* sequences studied to date maintain the same characteristics and possibly a similar catalyzing mechanism [53, 54]. Like other GT enzymes the *PMT*s most likely use a direct displacement S_N2-like mechanism where an acidic AA, glutamate or aspartate, within the active site acts as an enzymatic base catalyst [4, 13]. Interestingly, it is unlikely that the conserved D_{55} and E_{56} in the first cytosolic loop (and suggested enzyme active site) act in this capacity due to the fact that the exchange of either one of these does not eliminate *PMT* activity [12]. Instead, it has been theorized that these highly conserved AAs are involved in electrochemical stabilization of the active site by way of the coordination and positioning of the required of divalent cations [57].
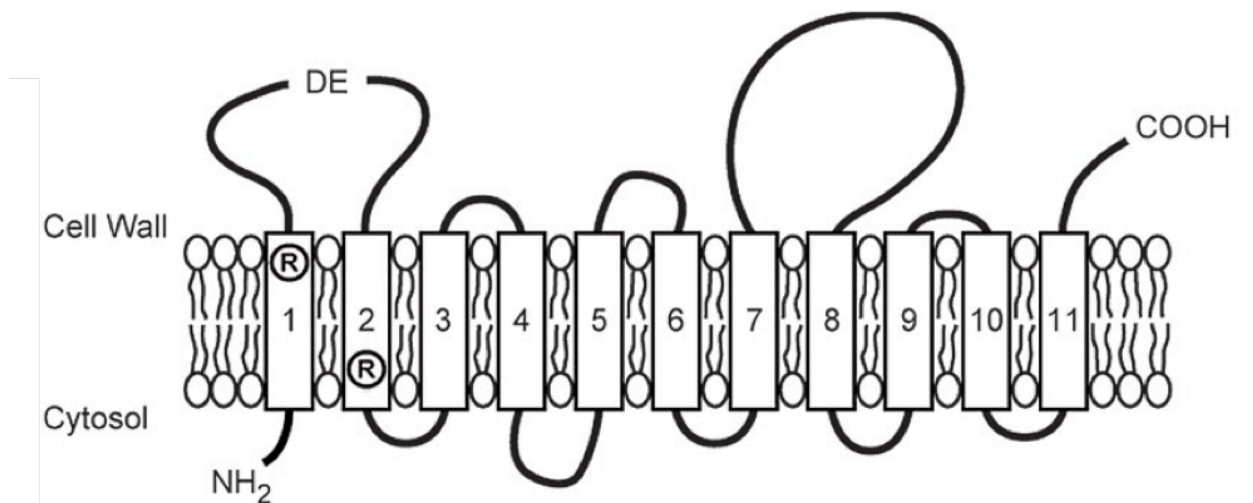
Image obtained from VanderVen, B. C., Harder, J. D., Crick, D. C., & Belisle, J. T. (2005). Export-Mediated Assembly of Mycobacterial Glycproteins Parallels Eukaryotic Pathways. *Science, 309*(5736), 941–943.

Figure 5: Model of *M. tuberculosis* PMT enzyme based on hydropathy profile. *M. tuberculosis* PMT contains two hydrophilic loops and a conserved D_{55} E_{56} (Asp-Glu) in the first periplasmic loop.

Besides the conserved D_{55} E_{56} active site AAs, the other conserved functional premise of PMTs is that *O*-mannosylation is initiated as the protein is being translocated across the membrane in its unfolded state [34, 58]. In bacteria, this constitutes a link between protein export, specifically Sec-mediated translocation, and glycosylation [53]. Tat-

mediated translocated proteins have also been found to be glycosylated thus suggesting a link, but these proteins have tertiary structure before export [59]. Notably, recent discoveries show that unfolded protein moving freely within the ER can also be mannosylated with the same *PMT*, this process is a branch of POM known as unfolded protein *O*-mannosylation (UPOM) and functions to signal for the termination of futile protein folding cycles [60]. Still, how the catalytic sites and target protein sequons of UPOM differ from POM, in addition to many other questions regarding the structure and function of this enzyme remain a mystery. In addition, the prokaryotic counterpart to UPOM has not been studied or discovered to date. The structural analysis of the *PMT* is necessary to infer the molecular mechanisms that shape the POM process. *PMTs* not only hold evolutionary significance but are also an attractive target for novel therapeutic drugs [15, 23, 27]; in the age of antimicrobial resistance this is an important element [21]. There is value to the continued dissection of the POM process and specifically the mechanism of the *PMT* that may provide insight into POM related disease processes in both humans and pathogens.

2.4.1. The PMT mannosyl-lipid donor molecule

In vitro studies in *Mycobacterium smegmatis* have showed that combining the cell membrane, GDP-[^{14}C]mannose, and synthetic peptides derived from endogenous mannosylated proteins, resulted in two unique molecules: polyprenol monophosphate [^{14}C]mannose and a [^{14}C]mannose-peptide [61, 62]. This experiment performed by Copper et al. (2002) is known as the cell-free peptide glycosylation assay and is reliable way to determine the glycosylation capabilities of a bacterium [62]. Although the mannolipid donor had been discovered prior to this assay, this experiment led to full characterization of the *PMT* lipid mannosyl donor and the enzymes that synthesize them in this and other organisms.

There is one specific mannosyl donor in all eukaryotes, and one similar specific

mannosyl donor in all *O*-mannosylating bacteria; they are Dolichol phosphate β-D-mannose (Dol-P-Man) and *PPM*, respectively (Figure S1). The interaction of the *PMT* with the mannosyl donor depends on the presence of an activated anchored mannose moiety on the periplasmic side of the membrane [63, 64], or luminal side in the ER for eukaryotes [44, 65]. Therefore, prior to the POM reaction a separate synthase enzyme anchored into the membrane acts to prep these lipid molecules by attaching an activated mannose on the cytoplasmic extension of the lipid donor [43]. The newly synthesized ppm or Dol-P-Man then flips to the other side of the membrane, by an unknown mechanism [13, 43, 66, 67], in order to interact with the *PMT* and/or other mannosyltransferase enzymes.

2.5. *Actinobacteria*

Actinobacteria are a gram-positive class of organisms that are the focus of Wakarchuk laboratory. These organisms are often recognized based on three characteristics: a high G + C content, ability to produce secondary metabolites, and ability to breakdown various biomass related substrates [16, 68]. Several *Actinobacteria* species have already been mentioned in the previous sections. Due to frequent *O*-glycosylation related discoveries in *Actinobacteria* organisms [16] they have the potential to become a model for studying POM processes in prokaryotes.

Irrespective of the interest in *O*-glycosylation, increasing the scope of knowledge regarding certain *Actinobacteria* would be a global benefit. The three relevant examples for this thesis are: *M. tuberculosis*, *Cellulomonas*, and *C. glutamicum*. *M. tuberculosis* is highly pathogenic species that is a major public health threat accumulating over three deaths per minute and continually evolving drug-resistant strains [15]. Inhibitory molecules that target any of the enzymes in the POM process may help to suppress the proliferation of these deadly bacteria in humans, particularly the *PMT* enzyme, which has recently been discovered to be crucial for

the virulence of this organism [15]. *Cellulomonas* and *C. glutamicum* are both a part of the *Coryneform* group and are the more specific focal point of this project. *Cellulomonas* is most commonly known for its unique ability to degrade the most abundant form of fixed carbon on earth, cellulose [69]. They produce rare enzymes called cellulases, which generate glucose as a by-product. Glucose can then be a convenient feedstock for bioethanol fermentation [69, 70]. Work in our laboratory has shown that new cellulases are still being discovered in this organism (Hirak Saxena MSc thesis work in progress). Currently, ethanol is our most popular "clean" alternative for petroleum, yet quick and cheap production is the bottleneck [70]. Thus, *Cellulomonas* is a global species of interest in our crucial efforts to acquire a clean energy source. *PMT* protein targets in *Cellulomonas* and related cellulose degraders have been shown to be metabolite based (this work). Thus, they may contribute to the proliferation of the species and be involved in the secretion and stability of secreted enzymes. *C. glutamicum* is the gram-positive model organism and for decades has been useful for industrial fermentative production of certain economically important AAs such as monosodium L-glutamate (more commonly MSG) and L-lysine hydrochloride [71]. Applications of these AAs include flavour enhancers in food and pharmaceutical products. Over 2 million tons of AAs are consumed worldwide; most of these are generated by *Coryneform* bacteria that have been optimized for efficient AA production [72].

Even though all of these species have been of interest to labs around the world for decades, there are many molecular characteristics and mechanisms that are still unknown. Novel discoveries are being made within *Actinobacteria* that aid in our multifaceted efforts to eradicate fatal disease, combat drug-resistant bacteria, obtain a source of clean energy, and even piece together the puzzle of cellular evolution. The feasibility of conquering these world problems is directly proportional to the growth and evolution of our application of molecular biology. As new systems and technology are being developed, becoming efficient and even affordable, and more communication portals open up for information transfer between scientists it is our

hope that we will contribute to the tools to enable breakthrough research. The development of the tools, such as reliable and versatile protocols, is necessary to study the process of POM in prokaryotic species.

2.6. Structure and Function of Mannosylated Proteins

Substrates of the *PMT* enzyme in *Actinobacteria* were first discovered in *C. fimi* when secreted cellulases were shown to bind to Concanavalin A (ConA) [73]. Later, glycoproteins in *M. tuberculosis* confirmed the presence of the glycosyltransferase in *Actinobacteria* by their reactivity with the ConA-peroxidase conjugate. ConA is a plant derived lectin that binds specifically to non-reducing terminal α-D-mannosyl and α-D-glucosyl groups and is the primary tool used for discovering mannosylated proteins in cells [75-77]. Simple ConA affinity assays led to the discoveries of many other mannosylated proteins in *M. tuberculosis*, among other soil bacteria. Further investigation lead to some accepted structural and functional concepts surrounding this topic but still the complete biological significance of this modification is poorly understood.

2.6.1. Structure of the *Actinobacteria* mannosylated proteins

Identifying *PMT* mannosylated proteins in a bacterial genome is challenging because one specific *PMT* recognized sequon has not been defined [23, 39, 78], as it has for N-glycosylation. Instead, multiple different sites of mannosylation have been discovered that all contain at the center, the mannosylation target Ser or Thr residue. The difference between modified and non-modified Ser/Thr seems to be the presence of proline (Pro) residues at positions -1 and/or +3 of the mannosylated AA with some exceptions [47, 79]. In general, a local environment of Pro, glycine (Gly) and/or alanine (Ala) residues is observed and theorized to be important for exposing the hydroxyl group of the Ser/Thr for

pending mannosylation [80, 81]. Other work has hypothesized that it may be the tertiary structure not primary that is important for the *PMT* recognition but questions are raised about the likelihood of this since mannosylation is coupled with unfolded protein translocation [58].

A software tool, NetOglyc, was developed and populated with known eukaryotic mannosylated proteins, which predicts mannosylation sites from gene strings that are inputted [82]. Herrman et al. (2000) proved the software was accurate to predict mannosylated proteins from the *M. tuberculosis* genome indicating that structural similarities exist between eukaryotic and prokaryotic glycoproteins [80]. Analysis of the genome found that approximately one third of the total lipoprotein population in the cell are mannosylated and have homology to other (possibly) unmodified proteins found in different organisms [80]. Further bacterial glycomic analysis has revealed that lipoproteins and mannosylation are frequently paired [16, 67, 74, 83]. This is predicted by a highly conserved consensus motif called a lipobox located near the N-terminal end of the protein sequence [67]. The lipobox sequence is observed in organisms from all domains, as a four AA sequence [Lys-Xaa-Ala/Gly-Cys], where the cysteine in the fourth position is lipidated following protein translocation and the cleavage of the upstream AAs [67]. Genes found to contain both the lipobox and at least one of the *PMT* recognized sequons for mannosylation are labeled as putative glycolipoproteins [16, 77, 84, this work]. Different types of experiments can be completed to confirm the presence, location, and chain-length of the mannosylation such as protein trypsin digest with mass spectrometry and/or ConA chromatography or blotting. Using these methods, glycolipoproteins have been confirmed in *M. tuberculosis* (10) [64, 75-77, 79-81, 84-87], *C. glutamicum* (4) [54], *S. coelicolor* [54, 80, 84] (5), and *C. fimi* (14) [this work]; structure analysis shows some of these proteins have relevant sequence similarity to each other [16]. Mannosylated proteins without the lipobox have been confirmed as well in *M. tuberculosis* (1) [64], *S. coelicolor* (1) [26], and *C. fimi* (6) [29, this work]. Proteins that have been further

characterized are shown to have a range of *O*-mannose sites along the gene, as little as 2 and up to 24 with an average of 8 – 10 [80]. Target sites are modified with $\alpha 1-2$ linked mannose, mannobiose, or mannotriose chain [80].

One of the few characterized *Actinobacteria O*-mannose protein is from *M. tuberculosis,* called the Apa protein [64]. Apa is an antigen that is the only non-lipidated glycoprotein discovered in *M. tuberculosis* so far, possibly because lipoproteins are difficult to accurately characterize due to their extreme hydrophobicity. The work by Dobos et al. (1996) revealed the 45 kDa Apa protein has four o-mannose modified sites [64]. All sites are unique, Pro rich and modified on a central Thr residue. Two sites contain a mannobiose, and one each mannose and mannotriose [64]. Other non-modified Ser/Thr residues were observed, all lacking a Pro rich vicinity, and possible thus sterically hindered or hidden from the *PMT* catalytic site.

2.6.2. Functions of the *Actinobacteria* mannosylated proteins

A harder parameter to define is the function of the protein, which can only be completely elucidated when the structure is also defined. However, certain experiments have shown that many of the *M. tuberculosis* glycoproteins are immunologically active during the host-bacteria interaction [15, 21, 88-90]. This information fueled the on-going efforts to discover the mechanism of these reactions and their role in the virulence of the organisms. Not surprisingly, certain C-type lectins have been shown to interact with the mannosylated proteins from *M. tuberculosis* [6, 16]. C-type lectins are a type of cell receptors on human and animal macrophages and dendritic cells that specifically recognize glycoconjugates exposed on the bacteria cell wall [16]. More importantly, interaction of the bacteria with these lectins influences the extent of the infection. Certain lectins are specific to mannose moieties and bind to these exposed on

the bacterial cell wall [91, 92]. This interaction can promote different responses based on the type of mannosylated protein bound such as: producing cytokines that inhibit immune response, phagocytosis of the bacteria, apoptosis of the host macrophage cells, T-cell response and inhibition of antigen processing and presentation [90-93]. De-glycosylation of these proteins can decrease the potency of the immune response in the host suggesting that not just the protein, but the *PMT* enzyme is crucial to virulence of the *M. tuberculosis* species [15]. Although a lot of work has been done to link virulence and the mannosylated proteins, there is evidence to support that the functions are more diverse. Some groups have reported that the mannose chains on the some proteins from *M. tuberculosis* and *S. coelicolor* can decrease the potential for protease degradation [81], or/and bind metabolic substrates to enhance uptake into the cell and in *C. glutamicum* one glycoprotein has been discovered to be important for bacterial growth stimulation and intracellular communication [54].

The information gathered thus far on the structure and function of these *PMT* target proteins does not extend past what is mentioned above. It is clear that to further our understanding of the biological significance of this process in bacteria more foundational work is required.

2.7. Project Goals

In the broad scope of this project, which is to discover how protein *O*-mannosylation relates to the biology of an *Actinobacteria*, the objectives of this work are foundational. The aim was to provide the toolbox necessary to characterize the mannosylation of *C. fimi* proteins and the process of protein *O*-mannosylation in *Actinobacteria*. The molecular toolbox prepared is advantageous to a modest laboratory, using equipment and techniques that are standard for any biochemistry lab. For this reason, the protocols developed are more convenient and beneficial to

society. The model bacteria chosen for this work is *C. glutamicum* as it has the benefit of a more developed system of molecular biology protocols. *C. glutamicum* cells also contain the *PMT* enzyme and some confirmed mannosylated proteins and thus a suitable model to study the protein *O*-mannosylation system in *Actinobacteria*.

The protocols developed in this work include:

- The reliable genetic manipulation of the gram-positive strain *C. glutamicum*, and the subsequent inducible recombinant protein expression

- Assembly of the *C. glutamicum*/*E. coli* shuttle vector that is optimized for expression of the target proteins in both strains

- Ni-charged affinity chromatography of enriched recombinant 6xHis-tagged proteins

- Immuno- and lectin- blotting that allow for confirmation of the synthesized protein tags as well as post-translational mannose modification

- And the efficient solubilization and enrichment of gram-positive lipoproteins from a prepared cell lysate

The encompassing goal of the above toolbox is two-fold 1) to allow a user, equipped with the gene sequence of any *Actinobacterial* putative glycoprotein or glycolipoprotein, and efficiently, with little equipment, confirm the protein is post-translationally modified with a mannose moiety, and 2) provide the protocols necessary to begin studying the POM process in detail in a simple organism, which is the overall goal of this extensive work. As mentioned above, this process in prokaryotes is poorly understood, and understanding the mechanisms in a simple organism can have evolutionary and medical significance. We hypothesize that the

PMT enzymes across *Actinobacteria* are conserved in structure and function and will modify putative glycoproteins from other species.

In this work our interest is in the *C. fimi* putative glycoproteins and glycolipoproteins that were discovered to contain one or more of the *PMT* recognized sequons, and subsequently confirmed by peptide digest and mass spectrometry to contain mannose chains. Some of these proteins have sequence similarity to other characterized bacterial proteins, and some are unique. We set out to confirm that these proteins can be modified by a foreign but related POM system, and therefore can be further characterized using this system. It is almost impossible to complete reliable molecular work in *Cellulomonas* as the protocols have not been developed and the characteristics of this species make it difficult to work with. Most of the work related to POM in *Actinobacteria* has been done within *Mycobacteria* or *Streptomyces* and still very little is known about the mechanisms of this reaction. It is our goal to shed light on this process in other *Actinobacteria* that will help to further the understanding of POM across all domains.

2.7.1. Control Protein – *AmyE*

As mentioned above most of the putative glycoproteins contain the recognized signal sequence for lipidation. Lipoproteins are notoriously difficult to study due to their hydrophobicity. Since the end goal of this work is to enrich and characterize the putative mannosylated proteins it was beneficial to consider that most of these proteins would contain the lipid moiety and thus a reliable protocol needed to be developed to encompass that fact. In all protocol development, a bonafide and dependable positive control is necessary. This positive control is a *C. glutamicum* native lipoprotein called *AmyE*, that functions as a maltose binding protein embedded in the plasma membrane with proximity to an ABC transporter [67]. Recombinant *AmyE* is readily expressed in *E. coli* and *C. glutamicum* using

the acquired shuttle vector and was extremely useful as a positive control in a more diverse manner. *AmyE* was used to develop the lipoprotein solubilization protocols, as well as the recombinant 6His-tag protein enrichment, and many instances of confirmation of recombinant protein expression in both species. *AmyE* is a 48 kDa protein without the post-translational lipidation, and a 55 kDa protein in its mature form [67]. In recent work were *AmyE* was confirmed as a lipoprotein, it was discovered that the operon containing the genes for the lipid mannose donor, *PPM* is required for the proper acylation of *AmyE* [64]. Although *AmyE* is, most likely, not glycosylated, this finding suggests some type of functional link between *AmyE* and mannosylation and further demonstrates the multitudinous effects that the interruption of the POM process may have on the homeostasis of the *Actinobacteria* cell.

3. Background Data

The first ever indication of an *O*-mannosylated protein from Actinobacteria was found in *Cellulomonas fimi*, on two secreted cellulase enzymes [29]. Further work regarding the structural details of these glycoproteins has not been done but it appeared that the mannose residue provided a functional role of protease resistance on the linker between the two domains of the enzyme [29]. Current evidence demonstrates that from across all domains of life a similar POM system exists within the cells, thus the work of examining this process in simpler organisms can provide great value. The preliminary work done by former student Ms. Connie Li and the mass spectrometry facility at the National Research Council in Ottawa helped to provide a fruitful starting point to the work described in this thesis. The results of this work are briefly described below.

3.1. *C. fimi* O-mannose proteins

ConA lectin blots were performed on the cell lysates of the *Cellulomonas* cells grown in different carbon sources: carboxy-methyl cellulose, starch, xylan, and glucose. The results in Figure 6 show that the mannosylated proteins are differentially expressed based on the carbon available. The confirmation of this result is ongoing. This provides insight into the functional role of these proteins, perhaps for the response of environmental signals, and resource competition with other Actinobacteria in the vicinity. The proteins were further elucidated by peptide digestion and mass spectrometry. In total 20 mannosylated proteins were identified (Table S1) with varying glycosylation sites and varying amounts of total mannose moieties in each chain. Of these proteins 14 also contain the conserved lipobox sequence in relatively close proximity to the most N-terminal tryptic peptide (Table S2). In addition, some of the discovered putative *O*-mannosylated proteins have sequence homology to other characterized bacterial proteins, such as periplasmic

binding proteins, extracellular solute binding proteins, isomerases, and penicillin-binding transpeptidase. Some also had sequence homology to putative glycoproteins in other *PMT* containing Actinobacteria, while some proteins were completely unique.
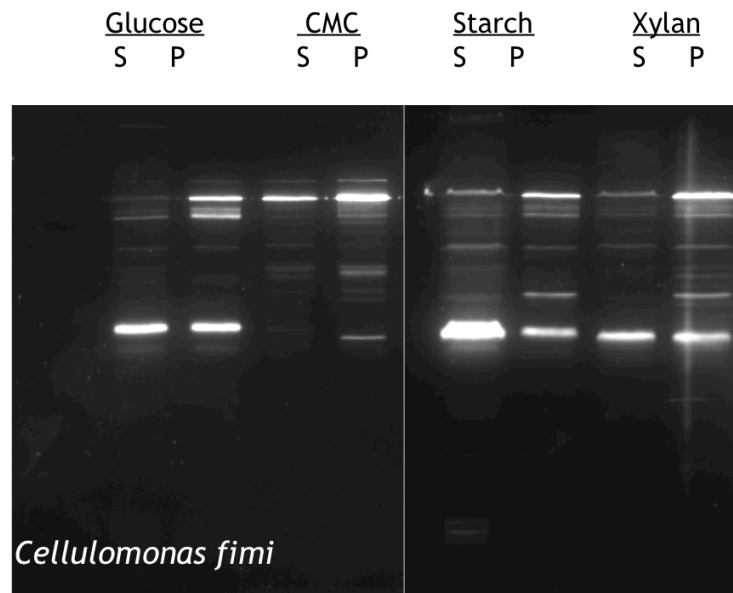
Figure 6: ConA lectin blot showing *C. fimi* mannosylated proteins differentially expressed under different carbon sources. S (supernatant fraction) and P (pellet fraction. CMC – Carboxy methyl cellulose. (Experiment performed by Denis Brochu NRC)

In order to begin the work described here, the hypothetical proteins that contain the sites of greatest mannosylation were chosen (between 7 and 11 hexose molecules). These proteins are expressed differentially based on the available carbon source, and so have the potential of providing a valuable layer of discovery. At a later date, due to difficulty of exposing the 6His-tag on the above chosen proteins, a fourth putative glycoprotein was chosen based on its sequence homology to a protein within the *C. glutamicum* genome. This is a predicted isomerase protein that is not a lipoprotein. Table 1 shows the four glycoproteins that were chosen for this project. It displays the following information with respect to each protein: suggested annotation (if available), the carbon source on which glycosylated form was discovered, the mass of the

protein (kDa), the number of mannose residues in the attached glycan chain, and the glycopeptide sequence evidently in close proximity, often next to, the lipobox.

Table 1: *C. fimi* putative glycoproteins chosen for examination in this thesis.

Gene	Annotation	Carbon Source	Mass (Da)	# Man Residues	**Lipobox** and Glycopeptide Sequence
Celf_1421	Hypothetical Protein	Xylan	13 290	8 to 11	MKLPATRTRLAVGA ATAALAATAF**LAGC**S DSGAEDEPTAVETT TESSTEDEAASGQV EVIDQTHLLLATAQD TLEAR
Celf_0777	Hypothetical Protein	CMC>Xylan	28 730	7	MSARRRGVHPGDF SVLRSRLFALPAAVL LTAT**LAAC**GGGSDS PSAEKTSASSEKST PTPTPEVAELTTEDF VAR
Celf_3336	Extracellular Solute Binding Family 1	CMC>Xylan	47 730	7	MRHRPALRGRTAVL RSAALLAVGALA**LTA C**AGSGEPAAEATSS GPPEPVEIR
Celf_2022	Peptide prolyl Isomerase	CMC>Xylan	26 700	7	

4. The Genetic Manipulation of *Corynebacterium glutamicum*

4.1. Experimental Procedures

C. glutamicum has a thick peptidoglycan layer interspersed with mycolic acids making up the outer layer of the cell, which makes genetic manipulation challenging. During this thesis, several electrotransformation methods were explored to uncover the most efficient protocol. The aim was to develop a method that would allow a degree of weakening of the cell wall that would enable the plasmid to enter the cell during electroporation, but also maintain the integrity of the cell wall thus maintaining cell homeostasis. The variables that are altered include: growth conditions, wash buffers, plasmid sequence, transformation protocol, and out-growth parameters.

4.1.1. Bacterial strains and plasmid preparation

Two types of *C. glutamicum* cells were used in this project. *C. glutamicum* R163 is a restriction-deficient strain donated by Dr. Chris Whitfield (University of Guelph, Guelph, Canada) and mostly used for the gram-positive protein expression. *C. glutamicum* ATCC 13032 strain was a gift from Dr. Lothar Eggeling (Research Centre Jülich, Jülich, Germany). This strain was treated as the *C. glutamicum* wild-type control strain and used for some foundational experiments regarding this species.

Two different plasmids (Figure 7) were used to troubleshoot the electroporation protocol for *C. glutamicum* cells, pCGE-04 and pTGR5. pCGE-04 is a plasmid generated at the inception of this project by Dr. Lisa Willis (University of Toronto, Toronto, Canada). It is a derivative of the pEKEx2 *E. coli/C. glutamicum* shuttle vector generated by Dr. Bernhard Eikmanns (Institut fur Biotechnologie, Jülich, Germany). This plasmid lacked its own ribosome binding site (RBS) and thus was

only a promoter fusion vector, changes were made to make it a translational fusion vector by inserting sequences to add an RBS and the restriction sites required to produce translational fusion. The pTGR5 plasmid is a completely synthetic *E. coli/C. glutamicum* shuttle vector generated and generously donated by Dr. Paulo Ravasi (Universidad Nacional de Rosario, Rosario, Argentina) [94]. The plasmid was optimized for gene expression in both *C. glutamicum and E. coli* containing only the optimized sequences for efficient protein expression. The pTGR5 expressible protein is enhanced green fluorescent protein (*eGFP*), which allows for rapid clone confirmation. Both plasmids contain TAC promoters and are inducible with isopropyl B-D-thiogalactopyranose (IPTG).
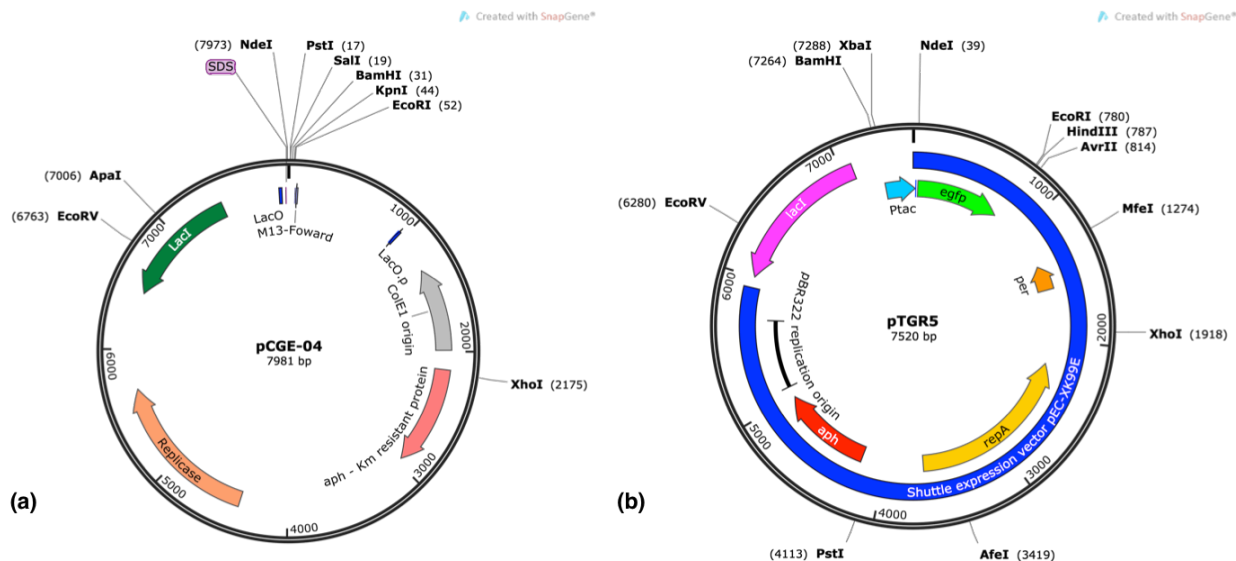
Figure 7: Plasmid maps of the pCGE-04 (a) and pTGR5 (b) plasmid. In pCGE-04, target genes were inserted between the NdeI and BamHI or NdeI and KpnI restriction enzyme sites. In pTGR5, target genes replaced the *egfp* gene, between NdeI and EcoRI restriction enzyme sites. Both plasmids contain a Lac operon, and kanamycin resistance.

Plasmids were introduced into ElectroMAX^{TM} DH10B^{TM} *E. coli* (Thermo Fisher Scientific Inc., Waltham, MA, USA) and retrieved using the Gene Aid Plasmid Mini Prep Kit ((FroggaBio, Toronto, Canada) (refer to section 5.1.4 for protocol) to create plasmid stocks for subsequent electroporation of *C. glutamicum.* Plasmid stocks were stored in TE Buffer (10mM Tris, 1 mM EDTA, pH 8.0) at

4°C. The DNA concentrations were measured by spectrophotometry. A Beckman Spectrophotometer was used to measure the absorbance of the solution at λ 260 nm (A_{260}).

4.1.2. Gene Pulser method (Gene Pulser electroprotocol) [95]

The wild-type *C. glutamicum* strain (ATCC 13032) was plated on LB Miller/agar media, consisting of (per liter) 10 g tryptone, 10 g NaCl, 5 g yeast extract, 15 g agar. Cells were incubated at 30°C overnight. 10 mL LB Miller with 1% (w/v) sterile glucose (LB Miller/Glc) was inoculated from the fresh plate. Cells were cultivated overnight (16 h) in a 125 mL baffled flask at 30°C with shaking at 180 rpm. 100 mL cell growth medium LB Miller/Glc in a 1-L baffled flask was inoculated with the overnight liquid culture to an optical density at λ 600 nm (OD_{600}) of 0.3. Cells were grown at 30°C with shaking at 180 rpm, and harvested 4-6 h after inoculation, or more specifically when the OD_{600} reached 1.0. Cells were harvested and washed in sterile and cold conditions. Cells were separated into 2 x 250 mL centrifuge bottles and centrifuged at 4000 *g* for 10 min, afterward the supernatant was discarded. Cells were washed 3 times in 25 mL of a 15% (v/v) sterile glycerol solution, then resuspended in 0.2 mL of the same solution. 40 µL aliquots of cells were kept on ice until transformation.

For electroporation, 200 – 300 ng pCGE-04 plasmid was added to one aliquot of electrocompetent *C. glutamicum* cells. 5 µL of sterile water was added to a separate aliquot of cells to be used as a negative control. Cells were transferred to a BIO-RAD Gene Pulse® Cuvette, 0.2 cm electrode gap. The dry cuvette was placed into the BIO-RAD MicroPulser^{TM}, which was set using the Ec2 pre-programmed parameters (12.5 kV/cm field strength), yielding a pulse duration ranging between 2-4 msec. Immediately after electroporation 1 mL of SOC media, consisting of (per liter) 10 g tryptone, 5 g yeast extract, 10 mM NaCl, 2.5

mM KCl, 10 mM $MgCl_2$ 10 mM $MgSO_4$, 20 mM glucose, was added to the cuvette. Suspension was transferred to a 12 mL culture tube and incubated with shaking for 2 h, 180 rpm, 30°C. 100 and 200 µL volumes of cells were plated on selective LB Miller/agar + 50 mg L^{-1} Kanamycin (Kan). Plates were incubated for 48 h at 30°C. Transformation efficiency was calculated as colony forming units (cfu) per µg plasmid DNA.

4.1.3. Handbook of *Corynebacterium glutamicum* electroporation method [96]

C. glutamicum ATCC 13032 electrocompetent cells was prepared in the same way as described in 4.1.2., substituting the LB Miller/Glc with commercial nutrient-rich Vegitone media containing 9% sorbitol (VegS). *Note: the sorbitol must be autoclaved separately from the media and combined at room temperature in sterile conditions.* The cells were harvested at an OD_{600} of 1.75 and centrifuged at 5000 g for 10 min. Cells were resuspended in 2 mL TG buffer containing 1 mM Tris HCl, pH 7.5; 10% (v/v) glycerol, then 20 mL of TG buffer was added, cells were centrifuged, and supernatant was discarded. This step was repeated. Final resuspension of cells was done in 2 mL 10% (v/v) glycerol solution. 150 µL aliquots were dispensed in cold 1.5 mL microcentrifuge tubes and kept on ice until transformation.

The electroporation protocol from 4.1.2. was followed, however all components were kept on or near ice in order to maintain ice-cold conditions. After cells were transferred to the cuvette, 0.8 mL cold, sterile 10% glycerol was added over the cell solution being careful not to disrupt the cells. Following electroporation, cell solution was transferred to 4 mL of prewarmed VegS media and heat shocked at 46°C for 6 min. Cells were incubated with shaking for 1 h, 180 rpm, 30°C. All cells were plated on selective VegS, 1.5% agar + 15 mg L^{-1} Kan and incubated at

30°C for 48 hrs.

4.1.4. The van der Rest et al. electroporation method [97]

This method was used with strains *C. glutamicum* ATCC 13032 and *C. glutamicum* R163. *C. glutamicum* 10 mL liquid overnight cultures were prepared in the same way as described in 4.1.2. substituting the LB Miller/Glc with LB Miller containing 2% (w/v) sterile glucose (LB Miller/Glc2). 100 mL of EPO medium in a 1-L baffled flask, prepared as described by van der Rest et al. (1999), made up of LB Miller, 400 mg isonicotinic acid hydrazide (isoniazid), 2.5 g glycine, and 0.1 mL Tween 80 was inoculated with the overnight culture to an OD_{600} of 0.3. Cells were allowed to grow 18°C for 28 h with shaking at 120 rpm. Ideally, the OD_{600} at this time is equal to 1.0. Cells were harvested and washed as described in 4.1.2. using a 10% (v/v) sterile glycerol solution. Before centrifugation, cells were chilled on ice for 10 min. Final suspension of cells was done in a 0.5 mL volume of the same glycerol solution and 100 μL aliquots were maintained on ice until transformation.

The electroporation protocol from 4.1.2. was followed while maintaining ice-cold conditions. Immediately after electroporation cells were resuspended in 1 mL of nutrient-rich Vegitone media and the suspension was transferred to a 1.5 mL microcentrifuge tube. The heat shock and out-growth steps from 4.1.3. were followed. Afterwards cells were plated on VegS, 1.5% agar + 25 mg L^{-1} Kan and incubated at 30°C for 48 hrs.

4.1.4.1. Adaptations made to the van der Rest et al. electroporation method

Trials of the *C. glutamicum* electroporation protocol were completed in the same way as described in 4.1.4 with the following changes: replace vegitone media

with 2YT media consisting of (per liter) 16 g tryptone, 10g yeast extract, 5 g NaCl, and sorbitol was removed from the out growth plates.

4.1.4.2.   Growth curve of *C. glutamicum* wt cells in the presence of different concentrations of isoniazid

A standard growth curve was prepared using *C. glutamicum* ATCC 13032 cells. A 20 mL overnight culture was prepared as in 4.1.1. and was used to inoculate 6 – 500 mL baffled flasks, which contained 50 mL LB Miller/Glc2 and varying concentrations of isoniazid: 0 mg L^{-1}, 1 mg L^{-1}, 2 mg L^{-1}, 4 mg L^{-1}, 8 mg L^{-1}, and 12 mg L^{-1}. Media was inoculated to a final OD_{600} of 0.1. Cells were grown at 30°C with shaking at 180 rpm. At *t* (time after inoculation) 2 h, all of the flasks were removed and the OD_{600} of a 1 in 10 dilution of the cells was recorded. This step was repeated for the following time points: *t* 4 h, *t* 6 h, *t* 8 h, *t* 10 h. The results were graphed using Microsoft Excel.

4.1.5. The Lessard electroporation method for *C. glutamicum* (adapted from the Follettie et al. protocol) [98]

C. glutamicum R163 cells were used for this method. 10 mL liquid overnight cultures were prepared the same way as described in 4.1.2. using 2YT media instead of LB Miller/Glc. The overnight culture was used to inoculate 200 mL of MB medium containing (per liter) 15 g tryptone, 5 g soytone, 5 g yeast extract, 5 g NaCl, 35 g glycine, in a 1-L flask to an OD_{600} of 0.1. Cells were grown at 30°C with shaking at 180 rpm until the OD_{600} reached 0.2 (~2.5 h). The flask was removed and ampicillin was added to a final concentration of 2 mg L^{-1}. The flask was returned to the incubator for another 1.5 h. Cells were harvested and washed as described in 4.1.2 using 30 mL of sterile wash buffer EPB1 containing 20 mM Hepes, 5% (v/v) glycerol, pH 7.2. Centrifugation was completed in a

swinging bucket rotor at 2000 g for 25 min at 4°C. The cells were washed 3 times and final resuspension was done with 1.5 mL sterile EPB2 buffer containing 5 mM Hepes, 15% (v/v) glycerol, pH 7.2. 150 µL aliquots of competent cells were prepared in 1.5 mL microcentrifuge tubes and stored at -80°F.

Two different plasmids were prepped for this experiment, pCGE-04 and pTGR5, trials were carried out separately for each plasmid. Cell aliquots were thawed on ice. 200 – 300 ng of the plasmid prep was incubated with the cells, on ice, for 5 min. Cells were electroporated as described in 4.1.2, but kept cold throughout the experiment. Immediately after electroporation 500 µL of sterile 2 YT was added to the cuvette, suspension was transferred to a culture tube Cells were grown at 30°C with shaking at 180 rpm for 4 h. Cells were plated on selective LB Miller/agar + 25 mg L^{-1} Kan. Plates were incubated for 48 h at 30°C.

4.1.6. Confirmation of *C. glutamicum* molecular clones

With low concentrations of the antibiotic Kan (< 50 mg L^{-1}) in the selective media plates the opportunity for contamination is higher. Therefore, a secondary measure of clone confirmation is required.

For the electroporation of cells using the pCGE-04 plasmid clones were chosen from the selective media plates, and grown overnight in 10 mL of LB Miller/Glc media containing 25 mg L^{-1} Kan at 30°C with shaking at 180 rpm. The plasmid was retrieved from the *C. glutamicum* cells using a protocol supplied by Qiagen (Hilden, Germany) and adapted to suit the requirements of this project. The adapted *C. glutamicum* mini-prep protocol is in the supplementary methods. The buffers from the GeneAid Plasmid Mini Prep Kit (FroggaBio, Toronto, Canada) were used. The eluted plasmid was stored in TE Buffer at 4°C. Then, the elution was then used to transformation the ElectroMAX^{TM} DH10B^{TM} *E. coli* (section

5.1.4). *E. coli* clones were chosen off of selective media plates, grown in an overnight liquid culture and mini-prepped to retrieve the plasmid. The plasmid was restriction enzyme digested and observed on via agarose gel electrophoresis in order to confirm fragment size.

For the pTGR5 plasmid, the transformant confirmation is simpler due to the inserted GFP amplicon within the MCS. Transformants were chosen from the selective media plates and used to inoculate 2 mL of LB Milller/Glc media containing 25 mg L^{-1} Kan in a culture tube. Cells were induced with 1.0 mM IPTG (~2.5 hours after inoculation). Tubes were replaced to the shaking incubator and harvested after 16 h. 1.5 mL of culture was centrifuged at 5000 rpm for 5 min and the supernatant was discarded. Cell pellet was resuspended in 1 mL sterile 0.9% NaCl solution (saline). Two negative control tubes were prepared: one tube contained 1 mL of saline solution, and a separate tube contained a saline resuspension of wild type *C. glutamicum* cells picked off of an LB Miller/agar plate. A positive control tube was prepared in the same way using the *E. coli* transformed pTGR5 cells. Tubes were observed for green fluorescent emission under blue light illumination (λ excitation = 488 nm), due to the expressed and folded *eGFP* within the cells.

4.2. Results

4.2.1. Transformation efficiency of *C. glutamicum* cells using different methods of cell preparation and electrotransformation

The data obtained from all experiments dedicated to the genetic manipulation of *C. glutamicum* is combined in Table 2. All of the methods described had previously published data [95-98] to confirm high efficiency of transformation of *C. glutamicum* cells. Certain methods, mainly the van der Rest et al. method and the Handbook of *Corynebacterium glutamicum* method, were frequently referenced by other groups attempting to transform these bacteria and had confirmed their success [96-105]. However, in this project reproduction of the published results was never achieved. Multiple attempts of both the Gene Pulser method and the Handbook of *Corynebacterium glutamicum* method were completely unsuccessful with the pCGE-04 plasmid and wt *C. glutamicum* cells. It was apparent that with low concentrations of kanamycin in the out-growth plates in addition to low transformation efficiency of the bacteria, the risk of contamination was far higher. This is seen throughout the subsequent methods as well. As a rule, the transformation efficiency of any one trial that had both confirmed transformants AND the contaminating strain on the negative control plates did not factor into the transformation efficiency for that particular method. A 16s RNA sequencing of a reoccuring contaminating bacteria strain prepared by Hirak Saxena (MSc thesis work in progress) revealed it to be of the *Pseudomonas* family, an opportunistic pathogen found in soil. Experiments to discover where this particular contaminant was originating from within the lab were unsuccessful. The contaminating strain did not appear on the out-growth plates containing the pTGR5-transformed species.

The van der Rest et al. method was promising as it yielded some confirmed

transformants. The trials attempted to insert the pCGE-04 plasmid into both of the *C. glutamicum* strains, ATCC 13032 and R163. The trials with the ATCC 13032 strain yielded confirmed transformants, although due to the appearance of the contaminating strain on the negative control plates, the transformation efficiency of these trials was not valid. In an effort to reduce the probability of contamination the van der Rest method was modified, the result was minimally successful. Contamination was still seen on two of the three trials indicating that the changes made to the protocol were not producing the desired affect. One trial produced confirmed colonies but very few were observed, resulting in a transformation efficiency of 56 cfu per µg plasmid DNA (as compared to 10^8 cfu per µg, published data) [97]. In addition, the method was unreliable and therefore of no use within the scope of the project. The trials with the R163 strain returned a better transformation efficiency but a similar probability of contamination demonstrating this method to also be unreliable and efficient. The average (Avg) transformation efficiency was calculated from the two plates containing confirmed transformants, from the one (of four) trials that produced results. The final result was just double the number from the modified van der Rest et al method, at 111 cfu per µg plasmid DNA. Still the numbers are much lower than expected; further experiments were completed to understand this occurrence (refer to section 4.2.2).

Table 2: Accumulation of the data obtained in the attempt to genetically modify *C. glutamicum* cells with xenogenic plasmid DNA. Different methods were tested and are denoted below by section number from the experimental procedures segment of this chapter. Only three of the five methods produced a valid transformation efficiency value, and only one method, section 4.1.5. (Lessard et al. Method) produced reliable results. Any trials where the contaminating strain was observed on the negative control plate were deemed invalid. The pTGR5 has a much higher transformation efficiency then any of the results with the pCGE-04 plasmid inferring that the method itself may have not been the only issue with these trials. The "(Avg)" indicates the averaged transformation efficiency of each media plate containing confirmed transformants. Thus, one trial can produce more than one transformation efficiency and an Avg value in this column. For future experiments involving the

electrotransfromation of *C. glutamicum* cells the pTGR5 plasmids was used exclusively with the section 4.1.5 method from Lessard et al.

Method (by section)	*C. glutamicum* strain	Shuttle Vector	# trials with confirmed transformants (total trials performed)	Transformation efficiency (cfu / µg plasmid DNA)	# trials with contaminating strain on (-) CTRL plate
4.1.2. (Gene Pulser)	ATCC 13032	pCGE-04	0 (2)	0	2
4.1.3. (Handbook)	ATCC 13032	pCGE-04	0 (3)	0	2
4.1.4. (van der Rest et al.)	ATCC 13032	pCGE-04	2 (5)	0	3
	R163	pCGE-04	1 (4)	(Avg) 111	2
4.1.4.1. (modified van der Rest et al.)	ATCC 13032	pCGE-04	1 (3)	56	2
4.1.5. (Lessard et al.)	R163	pCGE-04	1 (2)	2.9×10^2	2
	R163	pTGR5	2 (2)	(Avg) 3.6×10^4	0

The Lessard et al. method of electrotransformation produced the highest transformation efficiency, and with the pTGR5 plasmid produced reliable (and uncontaminated) results. For these trials the *C. glutamicum* R163 strain was used with the two different plasmids, pCGE-04 and pTGR5. Results from the Lessard et al. method using the pTGR5 plasmid included higher transformation efficiency, no observable contaminant, and confirmed transformants for each trial. Due to the fact that all trials from this method only differed by one variable, the plasmid, it was clear that the pCGE-04 plasmid was the affecting the reliability of the other methods, especially the probability of contamination. The highest transformation efficiency achieved during this objective was 3.6×10^4 cfu per µg plasmid DNA, which is an acceptable number for this project. Results from the transformation of *C. glutamicum* R163 using the Lessard et al. method and the pTGR5 vector constructs (Table 6), containing the target proteins, produced similar results (not shown).

4.2.2. The effect of different concentrations of isoniazid (antibiotic) on the growth of *C. glutamicum*

One of the issues apparent during the van der Rest et al. transformation trials was the inconsistency of the *C. glutamicum* cell growth in the presence of isoniazid (principal culture). During this part of the experiment it is vital that the cells are healthy and have enough innate integrity to survive the subsequent electroporation and heat shock. Many trials were aborted due to the inability of the principal culture to achieve an OD_{600} of 1.0 within the time span of 28 h. To understand this irregularity, growth curves of the *C. glutamicum* cells in the presence of isoniazid were plotted. This specific antibiotic acts as an inhibitor of the synthesis of mycolic acid [98], and weakens the cell wall structure of *Mycobacteria* and close relatives (other Actinobacteria). The results from the cultures containing 4 g L^{-1} or less isoniazid in the media are shown in Figure 8. 4 g L^{-1} isoniazid is the suggested published concentration for the van der Rest et al. method. The cells were grown at 30°C instead of 18°C (published) in an effort to study only the effect of the antibiotic on otherwise healthy cells. The result revealed that isoniazid even in low concentrations (1 g L^{-1}) inhibits healthy cell growth and therefore greatly minimizes the probability of success for transformation.
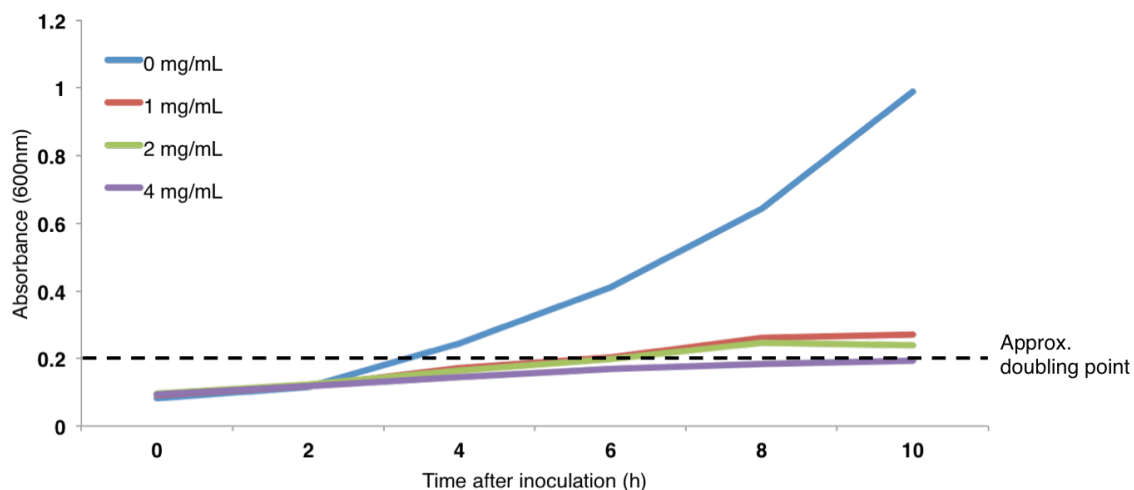
Figure 8: Growth curve of wt *C. glutamicum* cells showing the effect of different concentrations of isoniazid on doubling time of culture. Media was inoculated to an OD_{600} of 0.1 and would reach a doubling point at 0.2 (dotted line). Cells were grown at 30°C, and the OD_{600} of the culture was measured every 2 h for a span of 10 h. Cells grown in 4 g L^{-1} (purple line) isoniazid did not reach a doubling point in the time span. Cells grown in 1 g L^{-1} (red) and 2 g L^{-1} (green) isoniazid were around two times slower to reach a doubling point than the control cells (blue line) containing 0 g L^{-1} isoniazid.

The cells within the van der Rest et al. method suggested growth conditions of 4 g L^{-1} isoniazid did not reach a doubling point in the span of 10 hours (after inoculation). Decreasing the temperature to the suggested 18°C would only further decrease the growth of the cells. It was apparent that the *C. glutamicum* cells used in this project did not react the same way as in the published works, and that isoniazid was the cause of the aforementioned cell growth inconsistencies. Due to this result, the use of this antibiotic in any further trials regarding genetic manipulation of *C. glutamicum* was discontinued. The control culture containing 0 g L^{-1} isoniazid was within the expected range of doubling time (~3-4 h).

4.2.3. Confirmation of *C. glutamicum* molecular clones

4.2.3.1. Confirming *C. glutamicum* clones containing the pCGE-04 plasmid

An adapted method of retrieving plasmid DNA from gram-positive cells was developed and used to confirm *C. glutamicum* clones due to the inconsistency of the results from the basic Qiagen method. During protocol development, agarose gels with the plasmid DNA retrieved directly from the *C. glutamicum* cells were not consistent or clear (not shown). Sanger sequencing of the samples that did show on the agarose gels were not successful. Other groups reported difficulties when proving *C. glutamicum* transformants by observation of plasmid retrieval from the transformed strain. This problem is most likely due to low concentration of recovered plasmid in the elution. Two different methods were attempted to overcome this issue (1) Use PCR, and prepared primers to amplify a certain section of the plasmid and (2) Use the *C. glutamicum* recovered plasmid to transform the ElectroMAX™ DH10B™ *E. coli* cells in order to propagate the plasmid and make for reliably clear mini-preps and DNA agarose gels (Figure 9). The PCR method, once optimized, was effective (not shown). However, *E. coli* transformation method is more efficient and reliable, and the parameters of the experiment do not change with different vector constructs, as they may in the PCR method. The *E. coli* transformation method was used for all confirmations of *C. glutamicum* transformants in this project; Figure 9 shows an example of a DNA agarose gel with the *C. glutamicum* to *E. coli* recovered pCGE-04 plasmid from two different selected colonies (1 and 2).
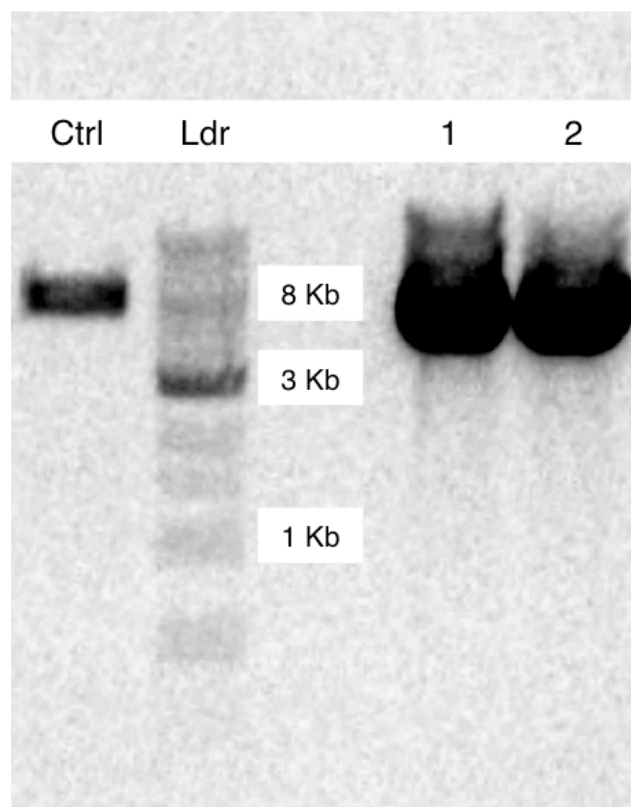
Figure 9: DNA agarose gel confirming the transformation of *C. glutamicum* R163 cells with the pCGE-04 plasmid. The pCGE-04 shuttle vector is 7981 bp in size. Ctrl lane denotes the positive control and is the pCGE-04 stocked plasmid retrieved from *E. coli* cells and digested with NdeI restriction enzyme (section 5.1.5). Lanes 1 and 2 are plasmid preps from separate colonies chosen off of the *C. glutamicum* selective media plates after electroporation. These preps were used separately, to transform ElectroMAX™ DH10B™ *E. coli* cells, therefore there were two trials. One of the transformed *E. coli* from each of the trials was chosen, mini-prepped and run (undigested) on a DNA agarose (0.8%) gel. This result confirms that the two originally chosen *C. glutamicum* cells contained the pCGE-04 plasmid.

4.2.3.2.   Confirming *C. glutamicum* clones containing the pTGR5 plasmid

The pTGR5 plasmid contains the *egfp* gene between restriction enzyme sites NdeI and EcoRI (see Figure 7b), therefore confirmation of *C. glutamicum* clones was far faster and easier than in the pCGE-04 method shown above (section 4.2.3.1.). Figure 10 shows a pTGR5 transformed *C. glutamicum* strain expressing *eGFP*, one tube contains non-induced cells and the other contains IPTG induced

cells. Both tubes show fluorescent cells with only a slight difference (by eye), the IPTG induced cells have observably more protein than non-induced.

(-) CTRL   Transformants
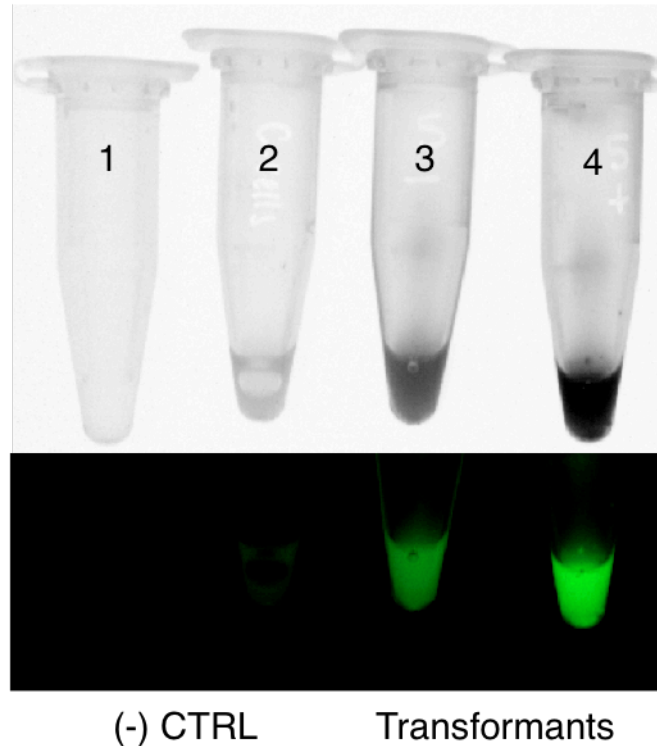
Figure 10: *C. glutamicum* R163 strain with pTGR5 plasmid confirmed by the emission of green fluorescence under blue light illumniation. The two images above are the tubes as they appear under blue light illumination. Tube 1 contains only 0.9% saline solution. Tube 2 contains *C. glutamicum* wt cells. Tube 3 and 4 contain the transformed strain, non-induced plasmid (3) and IPTG induced plasmid (4). Cells were centrifuged from a liquid culture and suspended in 0.9% saline before imaging to diminish the background fluorescence caused by the media.

Plasmid DNA recovered from transformed cells shown in Figure 9 by mini-prepping was used to transform *E. coli* as mentioned in sections 4.1.6 and 4.2.3.1. The plasmid recovered from these *E. coli* cells was used as a secondary measure to confirm the *C. glutamicum* transformants by restriction enzyme digest and agarose gel electrophoresis. The results are shown in Figure 11. The pTGR5 plasmid digested with NdeI and EcoRI releases the *egfp* amplicon, shown below, the undigested plasmid is also shown for comparison.
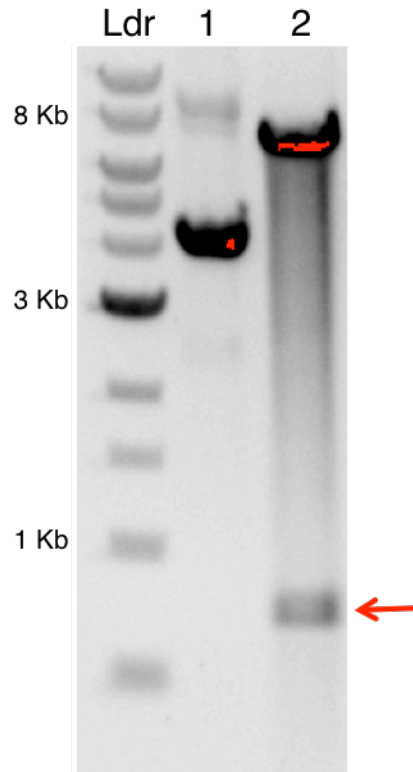
Figure 11: Secondary confirmation of *C. glutamicum* R163 strain with pTGR5 plasmid. pTGR5 plasmid DNA was recovered from *E. coli* cells that had been transformed with the recovered plasmid after *C. glutamicum* transformation using the Lessard et al. method. Lane 1 shows the undigested plasmid and Lane 2 shows the NdeI/EcoRI digested plasmid. The red arrow denotes the fragment released by the double digestion, the *egfp* gene. The pTGR5 plasmid is 7520 bp, and the *egfp* fragment is 720 bp.

4.3. Discussion

4.3.1. Unexpected issues regarding the genetic manipulation of *C. glutamicum*

The genetic manipulation of *C. glutamicum* was a difficult objective to attain. Although this is one of the model gram-positive organisms and many studies have been published which suggest an easily transformed bacteria [96-105], the components and tools involved in this work did not yield the expected outcomes. Careful analysis and comparison of the results revealed that the issue regarding reliable transformation of the *C. glutamicum* electrocompetent cells was probably due to the manipulated plasmid pCGE-04 and not, as was thought, due to the defects of the protocols used for preparing and transforming cells. Except for in the instance of the isoniazid growth curve, which displayed that, inconsistent with literature [97, 98], this antibiotic was preventing the healthy growth of the *C. glutamicum* cells. Therefore, the issues with this specific protocol may have been two-fold.

The pCGE-04 plasmid was altered from its obtained sequence using cycles with NEB restriction enzymes and T4 ligase. Specific changes include the removal of a DNA segment of ~35 bp containing the NdeI restriction site between the aminoglycoside 3-N-acetyltransferase (aph) segment [106] (Kanamycin resistance) and the replicase segment of the plasmid (see Figure 7a) and insertion of a DNA segment containing the RBS site and NdeI site to produce a translational fusion vector near the MCS of the plasmid. It is likely that this manipulation altered the base sequence in such a way that was not detected or expected. After the development of the pCGE-04 vector, primers were designed for the sequencing of the areas that were changed, the results did not reveal any delineation from the expected (not shown) but it is possible that the primers used did not allow the sequence observation of the mutated section. The design of the origin expression vector, pEKEx1, was based on the replication sequences of the

native *Corynebacterial* pBL1, and the *E. coli* ColE1 plasmids [107]. The polymerase from pBL1 necessary for the replication of the plasmid within the transformed cells is close to the DNA segment that was removed from the plasmid in order to make pCGE-04. It is hypothesized that this sequence was altered in a way that prevented most of the activity of the polymerase, but not all. It is unlikely that the *E. coli* ColE1 segment was altered because the pCGE-04 plasmid seemed to replicate as expected in *E. coli*. The full sequencing of the pCGE-04 plasmid would require the design of 9-10 primers spaced evenly around the vector and would probably reveal a mutated segment that was interfering with proper function. This was not completed in this works. However, to further the hypothesis mentioned above, some of the protocols attempted prior to obtaining the pTGR5 shuttle vector, for the preparation and transformation of electrocompetent *C. glutamicum* cells could be repeated with the pTGR5 plasmid. If the pCGE-04 plasmid were the source of the errors, then the transformation results would be dramatically different for pTGR5. In addition, all work completed with regards to the insertion of the pCGE-04 into *C. glutamicum* cells lacked a positive control. A shuttle vector, unaltered, and reliable would have probably revealed the difficulties of the pCGE-04 plasmid sooner than experienced.

4.3.2. The effect of the antibiotic isoniazid on the growth of *C. glutamicum*

It is quite clear that, even in low concentrations, the antibiotic isoniazid inhibits the healthy growth of *C. glutamicum* cells [108]. Isoniazid was originally used as an antibiotic against *M. tuberculosis*, as it inhibits the enoyl ACP reductase, InhA [109, 110]. InhA enzyme is involved in the biosynthesis of mycolic acids, and therefore isoniazid can disrupt the normal composition of the thick gram-positive cell wall membrane [109, 110]]. *C. glutamicum* also contains mycolic acids in the cell wall and it would follow that Isoniazid would have the same effects on the cell

wall as with *M. tuberculosis*. The absence of most mycolic acids could potentially create enough cell wall disruption to allow a large plasmid through to the cytoplasm of the cell, while maintaining the overall fitness of the cell itself – this was the theory. In papers published since the time of the van der Rest et al. electroporation method it has been reported that *C. glutamicum* does not contain the InhA enzyme or other paralogs [111]. It was shown that isoniazid is still a very effective inhibitor of the growth of *C. glutamicum*, however without an endogenous InhA enzyme, the mechanisms of how this works is unknown [111]. The *C. glutamicum* strains that were used for these experiments confirmed the potency of Isoniazid. *C. glutamicum* wt cells were cultivated in optimal conditions, with low concentrations of isoniazid and still the data did not represent what was displayed in the van der Rest et al. paper and others like it. It is likely that a *C. glutamicum* strain with some type of isoniazid resistance, again mechanism unknown, exists and was/is in use in labs doing similar work.

4.3.3. Confirming *C. glutamicum* non-*egfp* clones after transformation

Protocols designed for gram-negative cells, specifically *E. coli*, need to be adjusted to deal with the characteristics of gram-positive cells – mainly the thick, complex defensive barrier that makes up the cell wall. The recovery of the plasmid from *C. glutamicum* cells is notoriously low (Personal communication Dr. Pablo Ravasi). The elutions from the mini-prep procedure will not yield reliable results on an agarose gel, or when subjected to any type of sequencing. This is possibly due to the fact that standard mini-prep columns and reagents are designed for use with *E. coli* cells and do not accommodate the excess cell wall of the *C. glutamicum* cells. The columns often will become clogged with the this debris and cause most of the agarose gels to run as a smear (not shown), while the little DNA that was in the sample remains encumbered in the wells [112]. It was necessary to develop a protocol to reliably recover and observable amount

of plasmid DNA from the *C. glutamicum* transformants using techniques and tools already available in the standard laboratory. With a long lysozyme treatment of the gram-positive transformants in addition to a modified resuspension buffer and multiple preps of the same clone, the mini-prep eluate should contain a small amount of recovered plasmid even when using standard mini-prep columns. Therefore, techniques that work to amplify this recovered plasmid, such as subsequent *E. coli* transformation or PCR amplification of a specific segment can be used to confirm its presence. The minimum amount of DNA required to be observed on a gel is ~20 ng [113], and for Sanger sequencing is ~250 ng (TCAG), Negative results (this work) with both of these techniques using the gram-positive mini-preps indicate that the eluate recovery is often a lot less, probably somewhere between 1 – 10 ng of plasmid DNA. Notably, there were some *C. glutamicum* mini-preps that had the minimum amount of DNA to show on an agarose gel, less than 10% of the mini-preps showed recovered plasmid (results not shown). On the other hand, the minimum amount of DNA required for *E. coli* transformation and PCR amplification is ~1 pg, significantly less than the *C. glutamicum* recovery estimation above. Completing a second transformation with *E. coli*, just to confirm the first transformation, is time consuming but simple and interchangeable without any adjustments between different types of plasmids. *E. coli* transformation is a fundamental technique that is practiced in all molecular biology labs, and the tools are readily available. The adapted method of confirming the plasmid DNA sequence in *C. glutamicum* transformants can be practiced immediately with little money and training time investments, and most importantly extreme reliability.

5. Molecular cloning with *Corynebacterium glutamicum* and *Escherichia coli*

5.1. Experimental procedures

The genes that encode the putative mannosylated *C. fimi* proteins, the control lipoprotein *AmyE* from *C. glutamicum*, and the *pmt* homologs from *C. fimi*, *C. flavigena*, and *C. glutamicum*, were amplified by polymerase chain reaction (PCR) and inserted into the MCS of the two *C.glutamicum*/*E. coli* shuttle vectors, pCGE-04 and pTGR5 (Figure 7). The primers used were designed for insertion into the pCGE-04 shuttle vector, subsequently the amplicons were transferred from pCGE-04 to pTGR5 with some exceptions (noted below). Sequences were confirmed by agarose gel electrophoresis after restriction enzyme digest, as well as Sanger sequencing, performed by The Centre for Applied Genomics (The Hospital for Sick Children, Toronto, Canada).

5.1.1. Isolation of genomic DNA from Actinobacteria

This protocol describes the isolation of C. fimi genomic DNA, but has been successful in this project for the isolation of C. flavigena and C. glutamicum genomic DNA. This method is adapted from the DNAzol® Genomic DNA Isolation Reagent (Molecular Research Center Inc., Cincinnati, USA) product information sheet.

C. fimi wild-type (ATCC 484) was streaked onto a plate containing low sodium lennox broth agar media (LSLB/agar) containing (per litre) 10 g trypton, 5 g yeast extract, 0.5 g NaCl, 15 g agar. Plate was incubated at 30°C for 48 hours. One colony from the plate was used to inoculate 10 mL of LSLB with 0.2% xylan and 1 mg L^{-1} Ampicillin (amp) in a 125 mL baffled flask. The flask was placed in the

shaking incubator set to 180 rpm, 30°C for 24-30 hours or until cells have reached terminal OD$_{600}$ (~2.0 for *Cellulomonas*, ~3.0 for *C. glutamicum*). The following steps were done in duplicate. 1.5 mL of cells were transferred to a microcentrifuge tube and centrifuged at 14 000 g for 5 min. The supernatant was discarded and the cells were resuspended in 200 μL TE buffer containing 2 g L^{-1} lysozyme. The cell lysis solution was mixed at 37°C for 2 h. 500 μL DNAzol® was added to the tube and the mixture was incubated at 65°C for 5 min. Due to high viscosity, the mixture was ethanol (EtOH) precipitated to separate genomic DNA from cell lysate. 0.5 mL of 95% EtOH was added to the tube. Tube was inverted to mix, and incubated at room temperature for 3 min. DNA was spooled out of tube and transferred to a new tube using a sterile inoculating loop and spinning gesture. The precipitate was washed twice with 0.8 mL 70% EtOH using the following steps: suspend in EtOH by inversion, incubate at room temperature for 1 min, and centrifuge at 14 000 g for 2 min, repeat. All EtOH was removed and the remaining DNA was suspended in 300μL of 8mM NaOH solution. DNA solution was stored at 4°C. A diagnostic sample was observed via agarose gel electrophoresis.

5.1.2. Bioinformatics and primer design

Bioinformatics on all of the designated genes was completed to determine the restriction enzyme sites best suited for the amplicon insertion into the pCGE-04 plasmid MCS. BioEdit software (Ibis Biosciences, Carlsbard, USA) was used for this task, along with genome data of all three Actinobacteria species from National Center fro Biotechnology Information (NCBI). All amplified genes contain a 5' NdeI restriction enzyme site, CATATG. This sequence contains an inherent translation start site (ATG), which codes for the amino acid methionine (met). The 3' restriction enzyme site varies between genes and is either (BamHI, KpnI, or EcoRI). Primers were designed using the DNA sequence, and then added

restriction sites for cloning on both ends.. The 5' primer contained the NdeI restriction enzyme site, and the 3' primer contained a restriction enzyme site, a 6Histidine repeat ([CATCAC] X3) (6His), and an extra stop codon (after the endogenous one). Primers were prepared by either Bio Basic Canada Inc. (Markham, Ontario, Canada) or Integrated DNA Technologies Inc. (Coralville, Iowa, USA).

The primers to amplify the gene for the control lipoprotein *AmyE* were designed for attachment to a plasmid sequence, pCGL-482, already containing the *amyE* insert. This plasmid was generously donated by Professor Nicolas Bayan (Université de Paris Sud XI) and used in this project for this purpose only.

5.1.3. Polymerase chain reaction amplification of *Cellulomonas* target genes

This protocol was modified from the New England BioLabs® Inc. (NEB) protocol for PCR using Q5® High-Fidelity DNA Polymerase. A 50 µL reaction was set-up: 22.5 µL nuclease free water, 10 µL Q5 Reaction Buffer (5X), 10 µL Q5 High GC Enhancer (5X), 200 µM dNTPs, 0.5 µM each primer (5' and 3'), 3.5 ng genomic DNA (or 200 ng pCGL-482), and 0.5 µL Q5® High-Fidelity DNA polymerase. After components were combined on ice, PCR tubes were transferred to a thermocylcer (T100^{TM} Bio-Rad) prewarmed to 98°C. Table 3 displays the conditions for efficient amplification of the *Cellulomonas* target genes.

Table 3: Thermocycling conditions for PCR amplification of Cellulomonas target genes. Note: *amyE* gene amplification from pCGL-482 is most effective at annealing temperature of 71°C

Time	Temperature (°C)	Step
2 min	98	Initial Denaturation
10 sec	98	Denaturation
30 sec	75	Annealing
30 sec	72	Extension
2 min	72	Final Extension
~	4	Hold

(Denaturation, Annealing, and Extension steps run for 30 CYCLES)

Entire contents of tube were mixed with 10X DNA Loading Dye containing (per 20 mL) 0.025 g Xylene cyanol, 0.025 g Bromophenol Blue, 1.25 mL 0.625% (w/v) SDS, 12.5 mL glycerol, and 6.25 mL nuclease free water and loaded onto a 0.8% agarose gel, containing Gel Green (Biotium, Inc. California, USA) . The gel was subjected to electrophoresis observed under UV light. Gel extraction was completed with the GeneClean Turbo™ Kit (MP Biomedicals LLC, Santa Ana, CA, USA). The eluate was stored in TE Buffer at 4°C. Table 4 below displays the target genes acquired using the above protocol with the exception of two genes, Cg_1014 and Celf_2022, which were synthetically prepared (including the necessary restriction enzyme sites and 3' 6His repeat) by Bio Basic Canada Inc. (Markham, Ontario, Canada).

Table 4: Target genes with the respective origin host strain acquired for this work. Genes denoted hLP – hypothetical lipoprotein or hP – hypothetical protein contain multiple mannosylated sites (refer to Table S1 & S2). Genes highlighted in red have similar sequence identities to known protein families (refer to Table S1 & S2). Genes with (*) were synthesized.

	C. fimi	*C. glutamicum*	*C. flavigena*
Protein *o*-mannosyltransferase (*PMT*), GT39	Celf_3080 (1809 bp)	Cg_1014 (1339 bp)*	Cflav_0843 (1779 bp)
Control Protein (*AmyE*)		Cg_2705 (804 bp)	
(hLP) 1	Celf_1421 (342 bp)		
(hLP) 2	Celf_0777 (792 bp)		
(hLP) 3	**Celf_3336 (1591 bp)**		
hP 1	**Celf_2022 (804 bp)***		

5.1.4. Transformation of *E. coli* and plasmid retrieval using shuttle vector (pCGE-04 and pTGR5)

As stated, ElectroMAX^{TM} DH10B^{TM} *E. coli* were used for the purpose of generating high copy number plasmid stocks of shuttle vectors and new plasmid constructs. The transformation of these cells was completed with the two shuttle vectors (pCGE-04 and pTGR5) as per the transformation protocol in the provided product information sheet. The transformation yielded, on average, 10^8 cfu µg^{-1} plasmid DNA. Transformed colonies were chosen from selective media LB Lennox/agar + 50 µg kanamycin mL^{-1} and mini-prepped to retrieve plasmid. Confirmed transformed colonies were grown in 2 mL LB Lennox + 50 mg L^{-1} Kan and stored in a 1:1 sterile glycerol solution at -80°C.

All plasmid DNA retrieval and stocks were prepared using the GeneAid Plasmid Mini Prep Kit (FroggaBio, Toronto, Canada). The protocol outlined on the provided product information sheet was followed with one exception; centrifugation of the lysed and SDS-precipitated cells took 10 min instead of 3 min (as outlined in the protocol). The time was increased to prevent genomic DNA contamination and cell particle interference in the subsequent steps. The

plasmid DNA concentration in the eluate (TE Buffer) was measured and recorded with spectrophotometer at A_{260} and stored at 4°C.

5.1.5. Plasmid construction using restriction enzymes and T4 DNA ligase

5.1.5.1. Plasmid construction using the pCGE-04 shuttle vector

This protocol describes the construction of one plasmid named CFI-33, which is pCGE-04 containing the gene for the hypothetical glycolipoprotein Celf_3336. The remaining plasmid constructs were produced in the same way, using the restriction enzymes that correspond with the individual gene (see Table 5 and Table 6 for the complete list of plasmid constructs)

The following steps were done separately with the PCR amplified and gel extracted Celf_3336 amplicon and the mini-prep stock of the pCGE-04. The DNA concentration of the two solutions was measured using the Beckman Spectrophotometer at A_{260}. This value in (ng/uL) was used to prepare the digest reaction solution with the adequate amount of NdeI restriction enzyme (NEB). **All restriction enzyme reactions were set-up as per the product information sheet provided with the respective enzyme**. After the digest, enzyme activity was terminated by heat denaturation (if possible), The reaction was purified to remove the small undesirable DNA fragments using the GenElute^{TM} PCR Clean-Up Kit (Sigma-Aldrich, Oakville, ON, Canada). The above protocol was repeated with the PCR clean-up eluate but using the BamHI restriction enzyme (NEB). For the reaction containing the shuttle vector it is necessary to do a gel extraction after the second digest, since the PCR clean-up protocol will not remove fragments greater than 100 bp from the solution. After both digests, a small diagnostic sample was subjected to agarose gel electrophoresis (0.8% agarose) and UV imaging in order to confirm (1) fragment size indicating successful digest

reactions, and (2) approximate molar concentration for the preparation of the ligation reaction.

A ligation reaction was set-up using the two digested DNA fragments and T4 DNA Ligase (NEB) enzyme. The protocol outlined on the provided product information sheet was followed using the option to incubate the ligation reaction overnight (10-12 h) at 16°C. The 3:1 molar ratio of insert to vector was approximated based on the results from the diagnostic agarose gel above. The ligation product was used to transform ElectroMAX^{TM} DH10B^{TM} *E. coli* cells (Section 5.1.4). After transformation, colonies were chosen off of the selective media plates and mini-prepped (Section 5.1.4). As stated, the ligated and mini-prepped plasmid construct was verified in two ways (1) restriction enzyme digest and agarose gel electrophoresis with UV imaging, and (2) Sanger sequencing (TCAG). Table 5 below outlines all pCGE-04 constructs prepared using an arbitrary naming convention as well as the base pair (bp) size of the newly prepared constructs. All strains containing a verified sequence were grown overnight in 2 mL LB Lennox + 50 mg L^{-1} Kan and stored in a 1:1 sterile glycerol solution at -80°C.

Table 5: All target genes inserted into the pCGE-04 plasmid with internal naming convention used throughout the project. The plasmid size refers to the number of base pairs in each new vector construct. The bold text is the example used in the above protocol.

Gene Insert	Naming Convention	Plasmid Size (bp)
Celf_3336	**CFI-33**	**9201**
Celf_0777	CFI-32	8652
Celf_1421	CFI-31	8274
Cg_2705 (AmyE)	CGP-01	9423
Celf_3080	CFT-14	9753
Cflav_0843	CFT-15	9723

5.1.5.2. Plasmid construction using the pTGR5 shuttle vector

The most efficient way of generating the pTGR5 + gene insert vector constructs is to replace the endogenous *egfp* gene with the target genes using the 5' NdeI restriction enzyme site and the 3' EcoRI restriction enzyme site (Figure 12). The genes can be removed from the pCGE-04 plasmid using the above restriction enzymes and ligated into the pTGR5 plasmid with T4 DNA ligase.
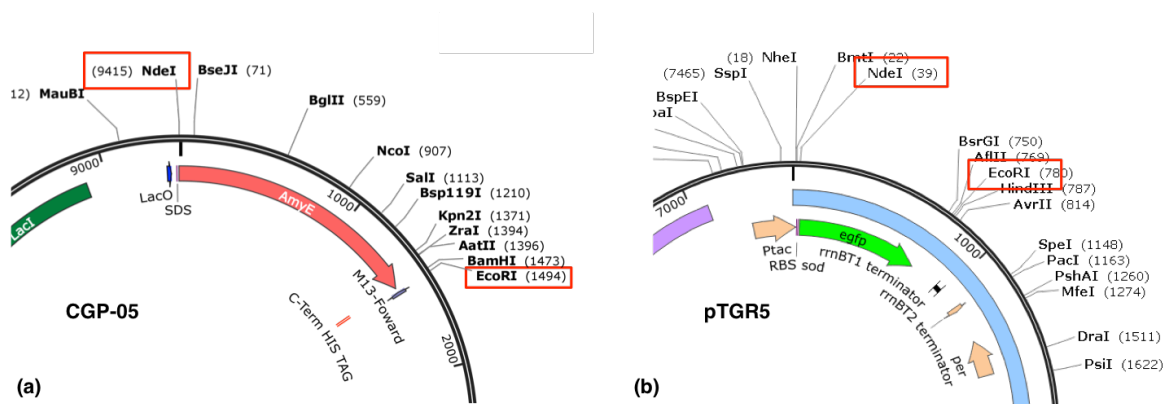
Figure 12: Multiple cloning sites of the plasmids used in this project. The CGP-05 (a) plasmid is the pCGE-04 plasmid containing the *amyE* gene. The EcoRI restriction enzyme site is downstream of all of the restriction enzyme sites that were used for target gene insertion. The genes inserted into the pCGE-04 plasmid can be removed via NdeI and EcoRI restriction enzymes, and ligated into pTGR5 (b) after the *egfp* genes is removed with the same enzymes.

This protocol describes the construction of one plasmid named CFI-36, which is pTGR5 containing the gene for the hypothetical glycolipoprotein Celf_3336. The remaining pTGR 5 plasmid constructs were produced in the same way (see Table 5 and Table 6 for the complete list of plasmid constructs), with the exception of the synthetically prepared genes, celf_2022 and cg_1014 (see Table 4).

Similar steps to those described in 5.1.5.1. were done separately with the plasmid stocked CFI-33 and pTGR5. The restriction enzymes used in this protocol are Thermo Scientific™ FastDigest™ (Thermo Fisher Scientific Inc., Waltham, USA.), which seemed to be more amenable to double digest reactions. **All restriction enzyme reactions were set-up as per the product information sheet provided**. Plasmids retrieved from selected transformed colonies were confirmed only by restriction enzyme digest and agarose gel electrophoresis with UV imaging. Table 6 below outlines all pTGR5 constructs prepared using an arbitrary naming convention as well as the base pair (bp) size of the newly prepared constructs. All strains containing a verified sequence were grown overnight in 2 mL LB Lennox + 50 mg L^{-1} Kan and stored in a 1:1 sterile glycerol solution at -80°C.

Table 6: All target genes inserted into the pTGR5 plasmid with internal naming convention used throughout the project. The plasmid size refers to the number of base pairs in each new vector construct. The bold text is the example used in the above protocol. The (*) denotes synthetically prepared genes.

Gene Insert	Naming Convention	Plasmid Size (bp)
Celf_3336	**CFI-36**	**8059**
Celf_0777	CFI-35	7510
Celf_1421	CFI-34	7123
Cg_2705 (AmyE)	CGP-05	8603
Celf_3080	CFT-16	8932
Cflav_0843	CFT-17	8902
Celf_2022*	CFI-37	7927
Cg1014*	CGP-06	8462

5.1.6. Preparation of *E. coli* and *C.glutamicum* expression strains

It was apparent that protein expression was more successful with the constructs containing the synthetically made shuttle vector (pTGR5), outlined in Table 7. For the remaining experimental procedures in this thesis, bacteria strains containing constructs from only Table 6 will be described. All strains containing a verified sequence were grown overnight in 2 mL LB Miller + 25 mg L^{-1} Kan and stored in a 1:1 sterile glycerol solution (BL21 only 8% glycerol) at -80°C.

5.1.6.1. *C. glutamicum* expression strains (using *C. glutamicum* R163)

The *C. glutamicum* R163 electrocompetent cells were prepared and transformed with the protocol outlined in section 4.1.5. The constructs listed in Table 6 were used to prepare the *C. glutamicum* expression strains. Molecular clones were verified using the protocol outlined in section 4.1.6.

5.1.6.2. *E. coli* expression strains (using *E. coli* BL21)

Commercial BL21 chemically competent *E. coli* cells (NEB) were suitable for transformation and protein expression (this strain does not express the T7 RNA

polymerase) using the vector constructs from Table 7. The protocol for chemical transformation of this strain was described on the provided product information sheet. Molecular clones were verified by plasmid retrieval (section 5.1.4), and subsequent restriction enzyme digest, agarose gel electrophoresis with UV imaging (section 5.1.5.2.).

5.2. Results

5.2.1. Isolation of genomic DNA from Actinobacteria

The DNAzol® reagent was used to isolate the genomic DNA from *C. fimi*, *C. flavigena*, and *C. glutamicum*. Additional steps were added to the suggested protocol to accommodate the tough exterior of these bacteria. First, cells were grown in the presence of sub lethal doses of ampicillin, which will weaken the cell wall by irreversibly binding to transpeptidase and interrupting the final stages of cell wall synthesis [114]. Second, an extended incubation of the cells with lysozyme helps to weaken the cell wall even further by breaking down glycosidic linkages in the peptidoglycan layer [115]. Lastly, to reduce viscosity of the sample for ease of loading into gels, or using in a PCR reaction the samples were sheared through a small gauge needle. Figure 13 (a) and (b) show clear examples of genomic DNA extracted from *C. fimi* and *C. glutamicum,* respectively. Genomic DNA is large and therefore slow to run through the agarose during electrophoresis. In the time span of the electrophoresis, some of the unsheared DNA will not have left the loading wells (Figure 13a). Figure 13b shows some well-sheared samples of genomic DNA. This adapted protocol is a simple and reliable method to extract genomic DNA from thick-walled gram-positive bacteria. This method was also used to extract *C. flavigena* genomic DNA and produced similar results (not shown).
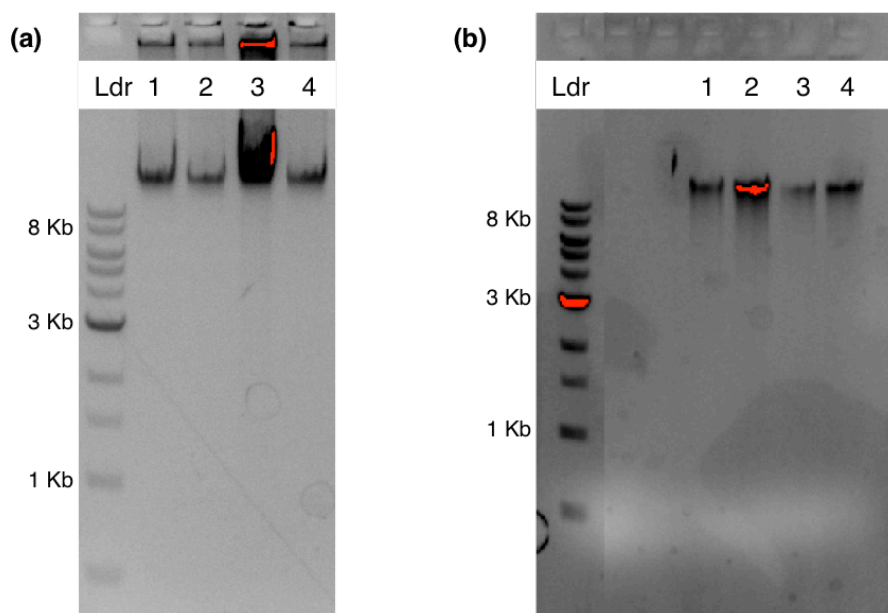
Figure 13: Confirmation of recovered genomic DNA. Both images show genomic DNA on a 0.8% agarose gel after electrophoresis, (a) *C. fimi* and (b) *C. glutamicum*. A protocol was developed to recover genomic DNA from the bacteria using DNAzol®. The protocol was done in duplicate, lanes 1 and 2 in each image are DNA samples from Sample 1 cells after DNAzol® treatment, 9 μL and 4 μL loaded respectively. Lanes 3 and 4 in each image are samples from Sample 2 cells after DNAzol® treatment, 9 μL and 4 μL loaded respectively.

5.2.2. Primer design for PCR amplification of target genes

Figure 14 shows an example of a pair of primers designed for the amplification of a target gene. The example displayed is the PCR blue print for the amplification of the *C. fimi* gene Celf_3336. The primers were designed to bind to the start and finish of the target gene in opposite directions, and promote the action of the DNA polymerase to make a duplicate of the gene. The primers also contain the necessary sequons for the project's objectives: (1) insertion of the amplicon into the plasmid via restriction enzyme sites NdeI and KpnI (or BamHI), and (2) enrichment or observation of recombinant proteins via 6His repeat before the ribosome detected stop codon. Due to the high G and C content of the target

genes, the primers were slightly shorter than what is standard (22 – 25 bp) in order to retain an optimal melting temperature between 69°C and 75°C.
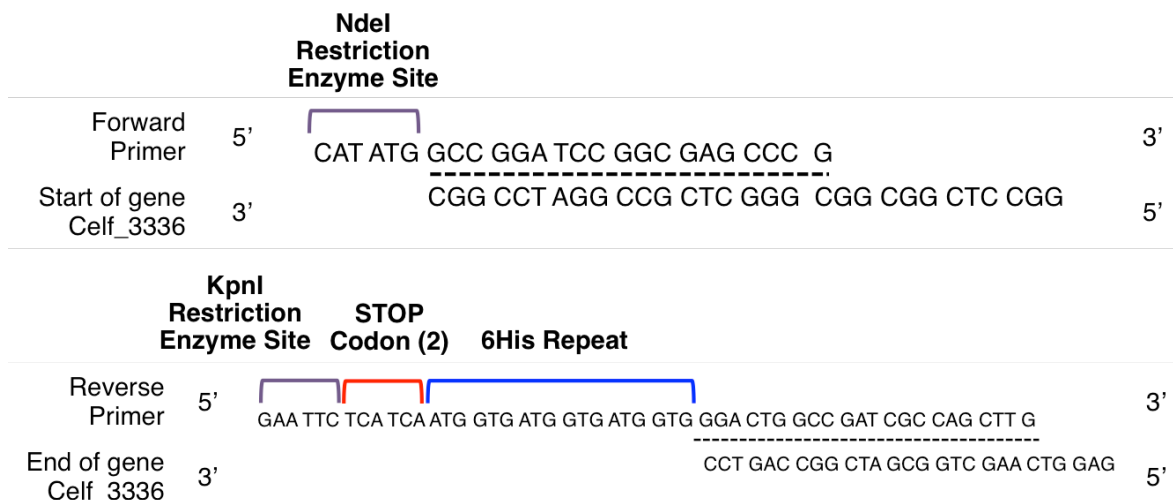
Figure 14: Example of the primer design to amplify a *C. fimi* target gene using PCR. The forward primer contains the NdeI restriction enzyme site upstream of the beginning of the target gene, and the reverse primer contains the KpnI restriction enzyme site downstream of the stop codon of the gene. Primer pairs for other target genes may contain a different reverse primer restriction enzyme. The reverse primer also has a 6His repeat in order to add a protein tag to the final cell expressed recombinant protein. For stronger stop signal, one stop codon was added between the restriction enzyme site and the endogenous stop codon of the gene.

The complete list of acquired primers can be found in the supplementary materials (Table S3). Primers were designed and prepared for the PCR amplification all of the target genes from Table 4 in section 5.1.3, except for gene Celf_2022 which was synthetically prepared.

5.2.3. Polymerase chain reaction (PCR) amplification of *Cellulomonas* and *C. glutamicum* target genes

PCR can be a useful tool for the isolation of genes through the massive amplification of any one segment of DNA. Due to the fact that DNA is double

stranded the number of copied genes increases exponentially as the process proceeds. Figure 15 shows two examples of PCR amplified genes using two different backbones and three different primer pairs. The *C. fimi* genomic DNA was used in Figure 15a to amplify two of the target genes, Celf_3336 is 1331 bp (Lane 1) and Celf_0777 is 803 bp. The smears in each lane are commonly seen when using a genomic DNA backbone. Because it is a very expansive stretch of DNA, the primers can sometimes bind non-specifically to a sequence that is close-to the target sequence. This creates a smaller amount of amplified segments at many different sizes. Different parameters of the PCR process can be adjusted to minimize this effect, such as extending the primers to make them more specific to the target gene, or changing the primer annealing time or temperature [116].
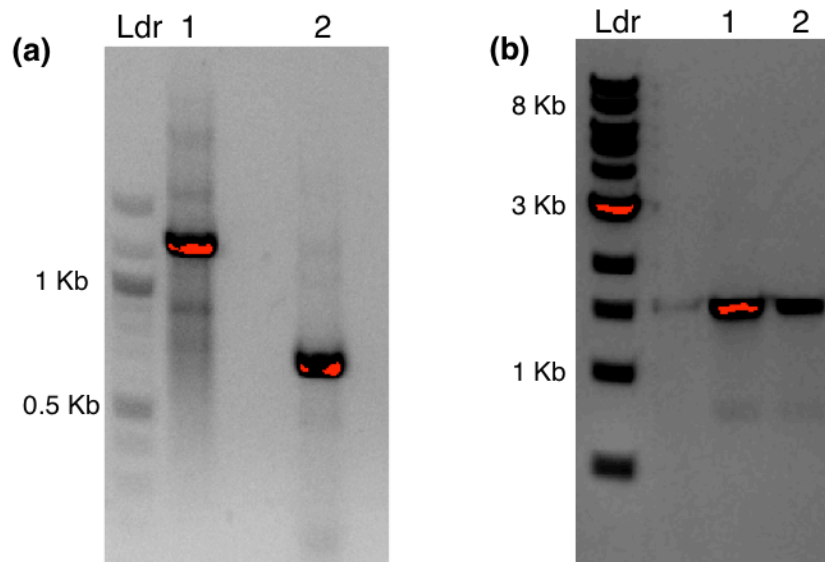
Figure 15: *C. fimi* and *C. glutamicum* PCR amplified target genes. *C. fimi* genomic DNA was used as a backbone (a) to amplify two target genes Celf_3336 is 1331 bp (lane 1) and Celf_0777 is 803 bp (lane 2). The plasmid pCGL-482 was used as the backbone (b) to amplify the *C. glutamicum* native *amyE* gene, which is 1480 bp. Figure (b) lanes 1 and 2 show two different annealing temperatures used during the PCR process, 71°C and 75°C respectively. The genomic DNA was extracted by DNAzol® and primers were designed to contain necessary restriction enzyme sites for insertion into the pCGE-04 plasmid, and a 6His repeat for enhanced enrichment of recombinant proteins.

The pCGL-482 plasmid contained the native *C. glutamicum* gene for the protein, *AmyE*. Primers were designed to amplify the *amyE* gene using this plasmid as the backbone, shown in Figure 15b. The *amyE* gene is 1480 bp. When a plasmid is used as a backbone as opposed to the entire genome of a bacterium, the non-specific binding is minimal, and therefore the gel lanes are a lot clearer with minimal smearing. The only variable between the two samples in Figure b is the annealing temperature of the primer to the unraveled plasmid DNA. Due to the overexposure of the band in lane 1 indicating a higher concentration of *amyE* in that sample, the 71°C annealing temperature is slightly more prolific than the 75°C sample (lane 2). All PCR amplified target genes (Table 4) produced results similar to Figure 15 (not shown).

5.2.4. Transformation of *E. coli* and plasmid retrieval of shuttle vector (pCGE-04 and pTGR5)

ElectroMAX™ DH10B™ *E. coli* was an optimal strain for stocking plasmids due to a mutation in the genome of the *endA1* gene resulting in an increased plasmid yield and copy number per cell. Both shuttle vectors used in this work were cloned into this strain of *E. coli* and recovered by mini-prep for future use in all of the new vector constructs. In addition, these cells were commercially prepared and guaranteed to result in a higher than average transformation efficiency for *E. coli*. These cells also produced the best results for the propagation of new vector constructs within the cells, but worse results when attempting to express proteins as compared to the two other strains of *E. coli* used in this project, BL21 and JM109 (not shown).

The mini-prepping protocol was slightly modified (original from FroggaBio gene aid kit) to produce a higher concentration of DNA in the eluate. Mainly, more plasmid DNA was eluted if (1) the high-speed centrifugation of cell material was

extended to allow a higher degree of separation between plasmid DNA and all other material, and (2) EDTA was added to the elution buffer. EDTA helps to sequester the metal ions from the solution to inhibit any DNA degrading enzymes. Figure 16 shows two examples of DNA mini-preps. Figure 16a is the pCGE-04 plasmid recovered from the DH10B *E. coli* cells, showing the undigested plasmid (lane 1), and the NdeI digested plasmid (lane 2). The pCGE-04 plasmid is 7981 bp (undigested plasmids often show on a gel in a smeary or unexpected pattern due to the alternate forms of circular DNA). Figure 16b shows the pTGR5 plasmid from the same cells, undigested plasmid only for confirmation of prior transformation. The pTGR5 plasmid is 7520 bp. Two colonies (i and ii) containing this plasmid were chosen from the selective media plate after transformation, mini-prepped and subjected to agarose gel electrophoresis. A separate plasmid called pTGR6 which is the same as plasmid pTGR5 but contains a different RBS binding site (pTGR6 - *lacZ*, pTGR5 - *sod*) was also cloned into *E. coli* cells, and subsequently mini-prepped out and stocked. Results from the pTGR6 trial are shown in Figure 16b, lanes 3 and 4. The pTGR6 plasmid achieves better expression in *E. coli* cells than pTGR5 (not shown) but was not used in this project because it has weaker expression in *C. glutamicum* cells.
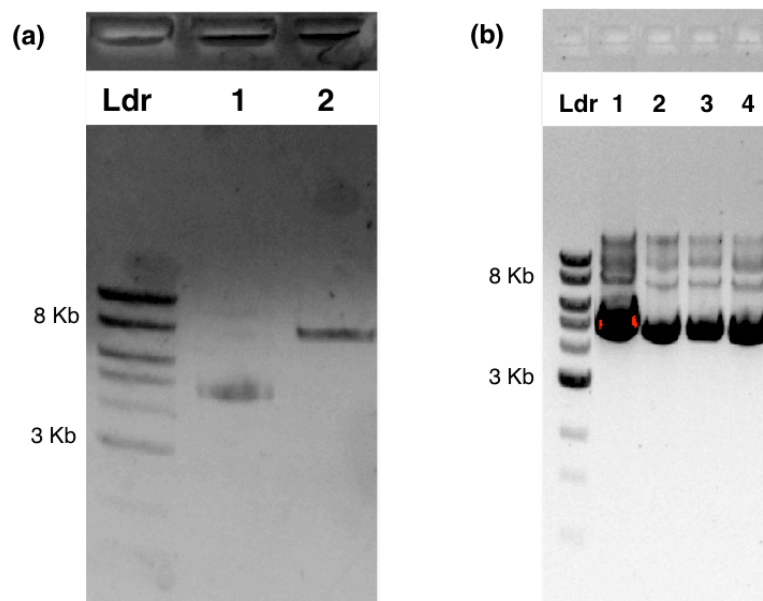
Figure 16: DNA agarose gels confirming transformation of *E. coli* cells and modified mini-prep protocol with two different shuttle vectors, pCGE-04 and pTGR5. The pCGE-04 plasmid was recovered from *E. coli* cells (a), Lane 1 – undigested pCGE-04 plasmid and Lane 2 – NdeI digested pCGE-04 plasmid. The pCGE-04 plasmid has 7981 bp. The pTGR5 and pTGR6 plasmid was recovered from *E. coli* cells (b). Both plasmids have 7520 bp. Two transformed colonies were chosen from the selective media plates from each strain. Figure 16b - Lanes 1 and 2 show the pTGR5 plasmid recovered from the chosen colonies (i and ii respectively). Lanes 3 and 4 show the pTGR6 plasmid recovered from the chosen colonies (i and ii respectively). The samples in (b) are all undigested plasmid and therefore have a smeary appearance on the gel due to alternate forms of the circular DNA.

By observation of Figure 16 the pTGR5 and pTGR6 plasmid have a high copy number (as compared to the pCGE-04 plasmid). This pattern was noticed throughout the project. When recovering the pCGE-04 plasmid itself and the pCGE-04 new constructs the DNA concentration as measured by spectrophotometry was significantly lower, often a 2-fold difference, than the pTGR plasmid and pTGR5 new constructs.

5.2.5. Plasmid construction

5.2.5.1. Plasmid construction using the pCGE-04 shuttle vector

The PCR amplified genes (section 5.2.3.) were extracted from the gel (Figure 15). The entire eluate of amplified gene and a sample of the pCGE-04 stocked plasmid (5.2.4.) were double digested in sequential reactions with a solution clean-up in between to remove the small segments of DNA that are released by the digestion. This created complimentary base pair overhangs that would accommodate the insertion of the gene into the pCGE-04 plasmid using the ligating enzyme T4 DNA ligase (NEB). The NdeI restriction enzyme site was always on the 5' end of the gene where as the 3' end either had BamHI or KpnI restriction enzymes sites. Figure 17 shows the diagnostic gel that was prepared and run (the loaded DNA was electrophoresed). The figure contains NdeI/KpnI

digested pCGE-04 plasmid (lane 1), the NdeI/KpnI digested Celf_3336 (lane 2), and the NdeI/BamHI digested Celf_0777 (lane 3). Each band has clean edges indicating a complete digestion of the DNA, and with minimal smearing an indication that the enzymes were cutting DNA without star activity.
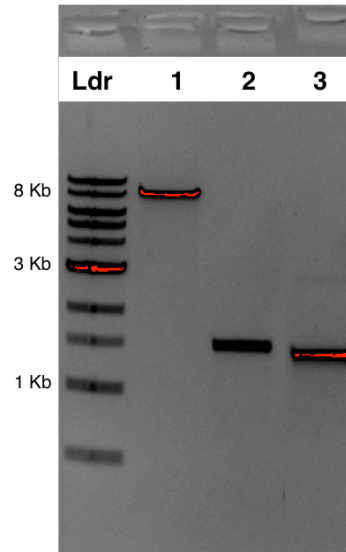
Figure 17: Diagnostic gel showing restriction enzyme digested plasmid and target genes before ligation. DNA agarose gel (0.8%) with digested plasmid and target genes indicates that the samples contain the necessary building blocks for a precise ligation and creation of the expression plasmid construct. The pCGE-04 plasmid (lane 1) and Celf_3336 (lane 2) gene were digested with NdeI and KpnI. These two samples were combined in the subsequent T4 ligase reaction to create the vector CFI-33. The Celf_0777 (lane 3) was digested with NdeI/BamHI and combined in a T4 ligase reaction with the pCGE-04 plasmid digested with the same enzymes (not shown) to create the plasmid CFI-32.

All of the ligase reactions were used to transform the ElectroMAX™ DH10B™ *E. coli* cells. Colonies chosen from the selective media were propagated in liquid media and mini-prepped to retrieve the plasmid DNA. A sample of the plasmid was double digested to release the ligated insert and observed on a DNA agarose gel. The example shown in Figure 18 is the vector construct CFI-33, which contains the Celf_3336 gene in the pCGE-04 plasmid. The image displays the undigested (#a) and digested (#b) plasmid retrieved from 6 separate colonies chosen off of the selective media plate. Also loaded onto the gel were the double

digested DNA segments before ligation (Figure 17) to be used as positive
controls for the digested vector constructs.
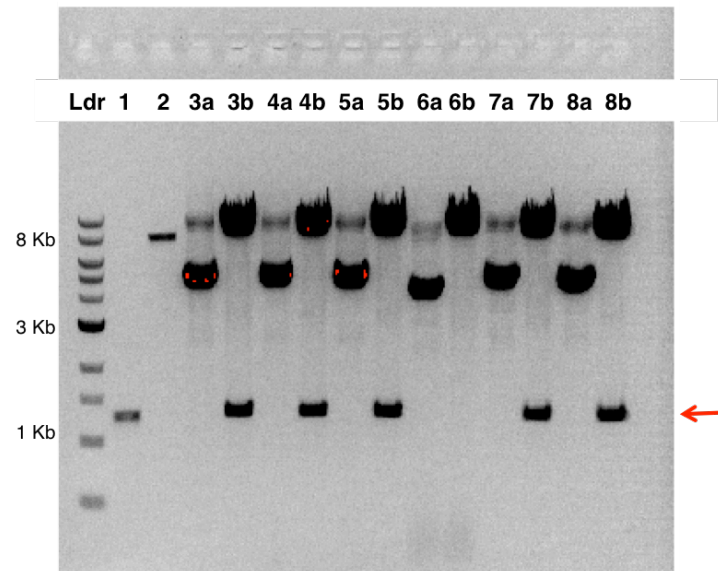
Figure 18: New plasmid construct, CFI-33, retrieved and digested from transformed
E. coli cells. All lanes denoted #a contain undigested plasmid, all lanes denoted #b
contain NdeI/KpnI digested plasmid. 6 colonies were chosen from the selective
media, each colony is represented sequentially in lanes 3 – 8 on a DNA agarose gel
(0.8%). Lanes 1 and 2 contain the positive controls for the digested insert Celf_3336
(lane 1) and the digested plasmid pCGE-04 (lane 2). The red arrow denotes the
Celf_3336 gene at 1336 bp. Due to supercoiling of the plasmid DNA most of the
undigested plasmid shows on the gel at a bp size smaller than the expected 7981
bp. Lane 6b clearly shows that this specific colony did not contain the correct
plasmid. All other chosen colonies appear to contain the expected plasmid.

It is clear from Figure 18 above that 5 of the 6 selected colonies contained the

Celf_3336 insert, confirmed when compared to the base pair ladder (Ldr) and the

positive control samples loaded, NdeI/KpnI digested pCGE-04 (lane1), and

NdeI/Kpni digested Celf_3336 (lane 2). Lane 6b, shows the digested pCGE-04

plasmid did not release the insert, therefore it was somehow ligated to itself via

incompatible ends. Over the course of this work all of strains containing the

correct plasmid (lane 3, 4, 5, 7, 8) were Sanger sequenced to confirm the

sequence is correct and unmutated (not shown). Due to the observably large

amount of DNA loaded onto the gel, the plasmid was not completely digested by

the enzymes in each reaction. For this type of diagnostic gel an incomplete digestion is acceptable. Similar results were seen with all of the pCGE-04 based constructs (Table 4) (not shown).

5.2.5.2.   Plasmid construction using the pTGR5 shuttle vector

The construction of the pCGE-04 MCS was advantageous for easy insertion of the target genes into the pTGR5 plasmid. The Actinobacterial genes replaced the *egfp* gene in the pTGR5 plasmid between the NdeI/EcoRI sites to generate the new vector constructs (Table 12). Using all of the molecular cloning tools described in this chapter, all of the target gene inserts could be removed from pCGE-04, gel extracted and purified, and ligated into a digested and purified pTGR5 vector. Figure 19 displays clear results from each of those steps: the pTGR5 plasmid NdeI/EcoRI digested (Figure 19a), the CFI-33 plasmid NdeI/EcoRI digested (Figure 19b), and the diagnostic gel of the digested pTGR5 and Celf_3336 after gel extraction, before T4 DNA ligation (Figure 19c). The red arrows on the figure indicate the DNA segment that was gel extracted and purified from that particular agarose gel.
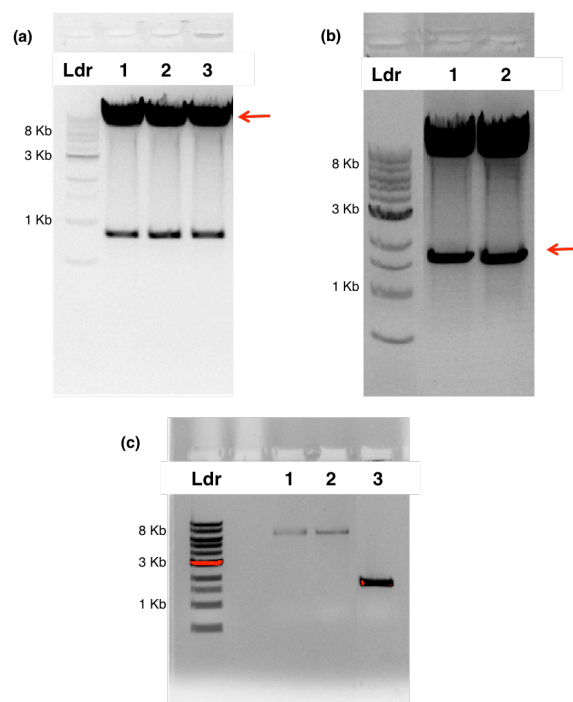
Figure 19: The DNA agarose gel collection displaying results from the transfer of the target genes from the pCGE-04 plasmid into the pTGR5 plasmid. Figure (a) shows the NdeI/EcoRI digested pTGR5 plasmid and Figure (b) shows the NdeI/EcoRI digested CFI-33 plasmid, the red arrow denotes the segments of DNA that were gel extracted and purified. Each lane in these two figures is the same digested sample aliquoted into separate lanes due to high volume. Figure (c) shows the diagnostic gel after the DNA purification and before the ligation reaction. Lanes 1 and 2 show two separate samples of purified pTGR5 plasmid from (a), and lane 3 shows the purified Celf_3336 plasmid from (b).

In similar fashion to the results described in section 5.2.5.1. the ligated plasmid was transformed into the ElectroMAX^TM DH10B^TM *E. coli* cells and retrieved via mini-prepping. Figure 20 shows the results of the NdeI/EcoRI mini-prepped plasmid from 3 chosen colonies. The results clearly indicated that all of the colonies contain the expected bp size of DNA segments and therefore the new vector construct, CFI-36.
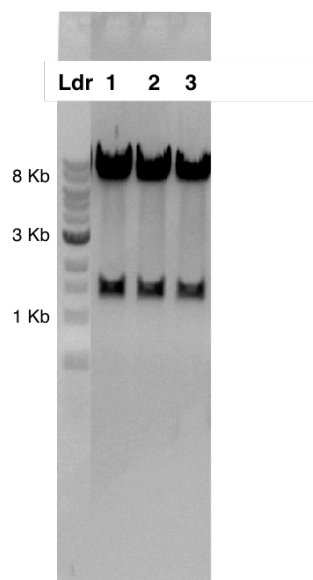
Figure 20: New vector construct, CFI-36, retrieved and digested from transformed *E. coli* cells. 3 colonies were chosen from the selective media plate after transformation, each colony is represented sequentially in lanes 1 – 3 on a DNA agarose gel (0.8%). The digested CFI-36 plasmid releases the Celf_3336 insert (1336 bp). All colonies chosen have the expected vector.

Similar results were seen with all of the pTGR5 based constructs (Table 6). The CFI-37 and CGP-06 plasmids, the pTGR5 vector containing the Celf_2022 and Cg_1014 (*C. glutamicum* GT39) respectively contained the two synthetically prepared genes. Although the primers were available for the PCR amplification of the Cg_1014 gene, it was found to contain an internal EcoRI site and therefore needed to be synthetically prepared with an adjusted sequence.

5.3. Discussion

5.3.1. Amplification of target genes from *C. fimi*, *C. flavigena*, and *C. glutamicum*

This chapter outlines a clear-cut collection of procedures to produce an expression plasmid containing a target gene amplified from an isolated bacterial genome. The protocols developed above are contributory for this work through to the final realization for the project; that is for the POM mechanisms and putative *PMT* substrates from any related Actinobacteria to be reconstructed in the host strain *C. glutamicum* and successively confirmed and studied.

The amplification of the target genes from the three Actinobacteria requires first, the isolation of the genomic DNA and designing and obtaining the appropriate primers, and second, PCR amplification of the target gene while adding appropriate restriction enzyme sites and 6His enrichment tag. Genomic DNA isolation is less finicky than plasmid DNA retrieval and therefore can be relatively easy to adjust for different types of bacteria. A good cell lysis reagent is required [118], similar to the proprietary ingredients in DNAzol®, and for *C. glutamicum* and other tough-walled cells it must contain lysozyme. Then, purification of the DNA may be done by any usual means such as ethanol precipitation or phenol-chloroform [112, 118]. In *C. glutamicum* the removal of some of the unwanted pieces of membrane before the purification yields better results. The DNA can then be sheared via syringe or restriction enzyme digest, which will make the DNA run more smoothly on an agarose gel in addition to making the PCR reaction more efficient – smaller segments of DNA makes for easier melting and separation of strands [119].

The primers can be designed for any target gene using any required sequence for restriction enzyme digest, or protein tag enrichment. Most of the primer design is stated in section 5.2.2. One additional note is that the primers were designed to amplify the gene starting at the initiation codon downstream of the leader

sequence. In this type of work, the leader sequence or 5' untranslated region (5' UTR), is not necessary due to the inherent and optimized RBS site in the plasmid. In prokaryotes, the 5' UTR after transcription can form unwanted secondary structures that may interfere with the translation of the protein [119], in addition they have a high GC content which make the PCR reaction more difficult to manage [120]. For the synthetic proteins Celf_2022, and Cg_1014 the leader sequence was maintained because PCR was not the technique required to obtain these genes. In the event that the leader sequence is interfering with the overexpression of the target protein in the cell, it would be straightforward to create a new vector with the protein minus the leader sequence using the skills highlighted in this chapter.

Lastly, for the PCR step, the Q5® DNA polymerase allowed for easy amplification of the Actinobacteria genes that characteristically have a high GC content. GC rich DNA is harder to melt and separate due to the higher frequency of hydrogen bonds between the strands [120]. Recent studies show that the optimal annealing times and temperatures have much more narrow ranges in GC-rich templates (~75%) vs. normal GC templates (~50%) [121]. The paper denotes a theoretical approach explicit for GC-rich primers and templates to decipher the optimal annealing time and temperature for a specific instance [121]. Fortunately, a standard reaction make-up and a simple annealing temperature gradient during the thermocycling step managed to expose the optimal annealing temperature for both *Cellulomonas* and *C. glutamicum* primer/template pairs, and the results were sufficient for the subsequent work. Although, the pattern in the above work appears to be that the annealing temperature is distinctive to the sequence, it can change based on sequence and primer combination and therefore a gradient temperature PCR experiment is recommended for any first reaction with new primers.

5.3.2. Construction and sequence confirmation of new shuttle vectors using
pCGE-04 and pTGR5

As mentioned above, these experiments were generally clear-cut and simple since the tools and protocols used were well established. Arguably, the two most important factors for achieving a correct and plentiful ligation are the 3:1 ratio of insert to vector (best estimation) in the ligation reaction, and the health and competency of the *E. coli* cells. In the beginning stages when the techniques were not so practiced, a reaction with a 1:1 or 2:1 insert to vector ratio would yield little to no ligated plasmid (not shown). In addition, internally prepared chemically competent *E. coli* cells would often results in only 3-5 colonies on the outgrowth plate, which greatly decreases the probability that one of those colonies is the correctly ligated plasmid. As skills developed it became clear that well-made (commercial or internal) electrocompetent DH10B *E. coli* cells, and an accurate estimation of a 3:1 insert to vector ratio based on an DNA agarose gel result is the most reliable way to obtain the correct ligation product propagated in an abundant amount of transformed *E. coli* cells.

For the confirmation of the new vector constructs its was advantageous to complete a restriction enzyme digest of the mini-prepped vector to remove the insert and view by DNA agarose gel electrophoresis, in addition to sending a small sample of the mini-prep to be Sanger sequenced. In the instances where the plasmid was not a ligation product with a PCR amplicon the mini-prepped DNA can be confirmed solely by DNA agarose gel electrophoresis without doing a restriction enzyme digest. There were cases early on in the project where the Sanger sequencing results were analyzed poorly and caused a false positive. Expression tests therefore were negative (not shown). Upon revalidation of the sequencing a base pair deletion was found just upstream of the 6His-tag causing a frame shift mutation that exchanged the His amino acids for others, and abolished the stop codon. Originally only two of the five *E. coli* mini-preps (Figure

72

18) were sequenced. After the frame shift mutation was discovered, the remaining three mini-preps were sent for sequencing and the analysis found that just one of these contained an unmutated sequence. The protein expression from this strain yielded a strong band on the western blot with anti-6His-Peroxidase. Thus, from the Figure 18, which showed five positive results for the ligation of Celf_3336 into pCGE-04, only one gene was the correct sequence. All other analyzed sequencing was re-checked for similar mistakes. In the case of CFI-31 a different frame shift mutation was discovered, and sequencing of the remaining mini-preps revealed the same mutation in all of them. Therefore, the PCR amplification of the Celf_1421 gene was repeated, and the vector CFI-34 was prepared, sequenced, and confirmed to be mutation-free.

Modern DNA technology allows sequencing to be inexpensive and accessible to most laboratories, without this direct confirmation of DNA sequence base-by-base new ligated products may not be valid. Therefore, the cause of negative expression tests in un-sequenced expression plasmids is not necessarily due to protease degradation of the foreign protein as is the common theory.

6. Recombinant protein expression and glycosylation in engineered *E. coli* and *C. glutamicum* strains

6.1. Experimental procedures

Troubleshooting experiments for the expression of recombinant proteins in both *E. coli* and *C. glutamicum* sought to uncover an effective and reliable way to purify the target proteins. In this process, the efficiency of the pTGR5 plasmid was tested, along with the Lac operon for IPTG induction. The purification of the recombinant protein with immobilized metal affinity chromatography (IMAC) Ni-charged resin columns is dependent on an accessible C-terminal 6His tag. The target proteins were detected and observed by Coomassie Brilliant Blue staining, immunoblotting, and lectin blotting. Recombinant proteins produced by *E. coli* are not glycosylated (*O-mannosylated*), where as *C. glutamicum* recombinant proteins could be glycosylated and be detected by the ConA lectin.

6.1.1. Bacterial cell growth conditions, harvest, and cell fractionation

6.1.1.1. *Escherichia coli*

Engineered BL21 *E. coli* were grown on LB Lennox/agar + 50 mg L^{-1} Kan plates and incubated at 30°C overnight. Single colonies were chosen from plates and used to inoculate a small volume (2 mL – 5 mL) LB Lennox + 50 mg L^{-1} Kan. The cells were grown at 30°C with shaking at 180 rpm overnight. For protein expression larger cultures were prepared in volumes of 50 mL, 100 mL or 200 mL in baffled flasks. Larger cultures contained LB Lennox + 50 mg L^{-1} Kan and were inoculated using the liquid starter culture to a final OD$_{600}$ of 0.1. Cells were grown at 30°C, 180 rpm and induced with 1.0 mM IPTG once the OD$_{600}$ of the culture was 0.25 (~1.5 h after inoculation). Cells were allowed to continue growth

at 25°Cfor 4-6 hours or until an OD_{600} of ~3.0 to 4.0 (stationary phase) was reached. Cells were harvested by centrifugation at 4000 g for 15 min and supernatant was discarded. The cell pellets were washed 2X with 20 mL phosphate buffered saline (PBS, containing per liter: 8 g NaCl, 0.2 g KCl, 1.42 g Na_2HPO_4 and 0.24 g KH_2PO_4, pH 7.2). Washed pellets were divided into 0.5 g aliquots and frozen at -20°C until needed.

All volumes below are used per 0.5 g frozen cell pellet but can be scaled up accordingly. Frozen pellets were thawed and lysed using the CelLytic^{TM} cell lysis reagent (CLb) (Sigma-Aldrich, Oakville, ON, Canada). CLb is a proprietary blend of zwitterionic detergents. The protocol outlined on the provided product information sheet was followed to release solubilized proteins into the lysate. The membrane fraction (insoluble material) was solubilized in 1.0 mL of 1X CLb containing 4 M urea. The lysate and insoluble fractions were stored at 4°C. Protein quantification was completed using the Pierce® BCA Protein Assay Kit (Thermo Fisher Scientific Inc., Waltham, MA, USA).

6.1.1.2.   *Corynebacterium glutamicum*

Engineered R163 *C. glutamicum* cells were grown on LB Miller/agar + 25 mg L^{-1} Kan plates and incubated at 30°C overnight. Single colonies were chosen and used to inoculate 10 mL LB Miller + 25 mg L^{-1} Kan. Cells were grown at 30°C with shaking at 180 rpm overnight. For protein expression larger cultures were prepared in volumes of 50 mL, 100 mL or 200 mL in baffled flasks. Larger cultures contained LB Miller + 25 mg L^{-1} Kan and were inoculated using the liquid starter culture to a final OD_{600} of 0.1. Cells were grown at 30°C, 180 rpm and induced with 1.0 mM IPTG (section 7.1.2.2.) once the OD_{600} of the culture was 0.30 (~3.5 h after inoculation). Cells were allowed to continue growth at 30°C, 180 rpm for 12–16 hours or until an OD_{600} of ~3.0 to 4.0 (stationary phase) was

reached. Cells were harvested by centrifugation at 4000 g for 15 m and supernatant was discarded. Cells were washed 2X with PBS. Washed pellets were divided into 0.5 g aliquots and frozen at -20°C until needed.

All volumes below are used per 0.5 g frozen cell pellet but can be scaled up accordingly. Frozen pellets were thawed and mechanically lysed using a cooled mortar and pestle; a 1:1 ratio of g celite to g cell pellet was added to the mortar before lysing. The cell/celite mixture was suspended 3 mL of CHAPS detergent solution (CHAPS) containing 100 mM NaCl, 50 mM Tris HCl, 25 mM glucose, 15% (v/v) sterile glycerol, 600 units (U) DNase I (Roche, Mannheim, Germany), 20 mM $MgCl_2$, 1 mM $MnCl_2$, 2% CHAPS detergent and 1X COMPLETE protease inhibitor tablet (Roche), pH 7.2. Suspension was transferred to a falcon tube and mixed on an end-to-end rotator at 4°C for 4 – 8 h. Mixed suspension was low-speed centrifuged at 500 g for 2 min, to remove celite from samples. Lysate was removed from the falcon tube and transferred to 1.5 mL microcentrifuge tubes. Lysate was centrifuged at 16 000 g for 20 min to separate soluble fraction from membrane fraction (insoluble material). Soluble fraction was removed and transferred to a new microcentrifuge tube. The insoluble material was solubilized in 1.0 mL of CHAPS containing 4 M urea. The lysate and insoluble fractions were stored at 4°C. Protein quantification was completed using the Pierce® BCA Protein Assay Kit.

6.1.2. Determination of optimal IPTG concentration for recombinant protein expression

6.1.2.1. IPTG induced protein expression in *E. coli*

A 10 mL liquid starter culture of (CGP-05) *E. coli* was prepared and used to inoculate 5 x 50 mL media in 0.5 L baffled flasks. Different concentrations of

IPTG were used in each flask: 0 mM, 0.5 mM, 1 mM, 2 mM, and 4 mM to induce the cells at the OD_{600} stated in section 6.1.1.1. Cells were grown, harvested, and fractionated as per section 6.1.1.1. Protein concentration was determined by BCA as stated in section 6.1.1.1. Proteins to be analyzed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) were standardized to 25 μg per well. Protein from all cell lysates (soluble and insoluble) was mixed with 4X SDS Loading Dye (supplementary methods) and heated at 75°C for 5 min. Samples were analyzed on a 12% polyacrylamide gel run at 190V for 50 min with Coomassie Brilliant Blue staining. A second SDS-PAGE gel was performed and the proteins were transferred to a PVDF membrane by electrophoresis at 100V for 2 h. The 6His-tagged proteins were observed on PVDF after incubation with the Anti-His6-Peroxidase (anti-His6-P) (Sigma-Aldrich, Oakville, ON, Canada) using western blotting technique (supplementary methods).

6.1.2.2.   IPTG induced protein expression in *C. glutamicum*

A 50 mL liquid starter culture of (CGP-05) *C. glutamicum* was prepared and used to inoculate 5 x 50 mL media in 0.5 L baffled flasks. Different concentrations of IPTG were used in each flask: 0 mM, 0.5 mM, 1 mM, 2 mM, and 4 mM to induce the cells at the OD_{600} stated in section 6.1.1.2. Cells were grown, harvested, and fractionated as per section 6.1.1.2. Protein concentration was determined by BCA as stated in section 6.1.1.2. Protein from all cell lysates (soluble and insoluble) was observed via SDS-PAGE and immunoblotting with anti-His6-P as described above.

6.1.3. Purification of recombinant 6His-tagged proteins in engineered *E. coli* and *C. glutamicum*

0.5 mL of the soluble and insoluble fractions from the *E. coli* and *C. glutamicum* cell fractionations were 10X diluted with denaturing wash buffer 1 and filtered through 0.45 µM syringe filter. Filtrates were loaded separately onto an IMAC Ni-charged column (Bio-Scale™ Mini Nuvia™ 5 mL cartridge) previously equilibrated with denaturing wash buffer 1. All buffer formulations used in IMAC purification can be found in Table S4 and are optimized based on the suggested compositions in the provided product information sheet. The column was washed with 6 column volumes (CV) of denaturing wash buffer 1 containing 5 mM of imidazole. A second wash was completed with 6 CV of denaturing wash buffer 2 containing 25 mM of imidazole. 6His-tagged proteins were eluted with 5 CVs of denaturing elution buffer containing 500 mM of imidazole.

Fractions collected: flow-through (FT), Wash 1 (W1), Wash 2 (W2), and Eluate (E) were concentrated from ~25 – 30 mL to 5 mL by membrane ultrafiltration (Vivaspin 20 MWCO 10 000) (GE Healthcare Life Science, Mississauga, ON, Canada). Ultrafiltration (1.25 mL/min) was completed by centrifugation with a fixed angle rotor at 5000 g for 19 min at 4°C. The resulting protein concentrate fractions were analyzed by SDS-PAGE with Coomassie Brilliant Blue staining, immunoblotting with anti-His6-P, and lectin blotting with ConA.

6.1.4. Observing glycosylated proteins with ConA lectin blotting

ConA (Sigma-Aldrich, Oakville, ON, Canada) is pre-conjugated to horse radish peroxidase (ConA-P). Once the purified proteins were confirmed with western blotting, another SDS-PAGE was performed with the same samples. The proteins were transferred to a PVDF membrane by electrophoresis. The following steps were completed: membrane wash in PBS for 15 min, protein blocking by 2% (v/v) TWEEN-20 (Sigma-Aldrich) in PBS (PBST) for 1 h, membrane wash 2X in PBS for 15 min, membrane incubation with ConA-P solution containing (per 20

mL) 0.05 mg L^{-1}, 0.05% (v/v) TWEEN-20, 1 mM of each CaCl$_2$, MnCl$_2$, and MgCl$_2$, in PBS for 16 h or longer, membrane wash 4X in PBST for 10 min. The lectin bound proteins were visualized by chemiluminescence using enhanced chemiluminescent (ECL) substrate incubation of the membrane for 5 min. Images were captured using BioRad Imager. *Note: certain protein lysates from C. glutamicum were first enriched with ConA resin batch binding (supplementary methods) before ConA-P detection.*

6.2. Results

6.2.1. Protein Quantification with BCA Assay

For the Bicinchoninic Acid Assay (BCA) bovine serum albumin (BSA) protein standards were prepared immediately prior to the protein assay experiment for increased accuracy of protein quantification in the cell lysates. Figure 21 shows an example of a BCA standard graph prepared after the assay was complete. The equation of the line was used to calculate the protein concentration of the cell lysates collected just prior using the absorbance of the lysate solution at 540nm. Cell lysates collected using the methods developed in this work produce an average protein concentration of 10.5 mg/mL.

y = 0.001x + 0.0146
R^2 = 0.99932
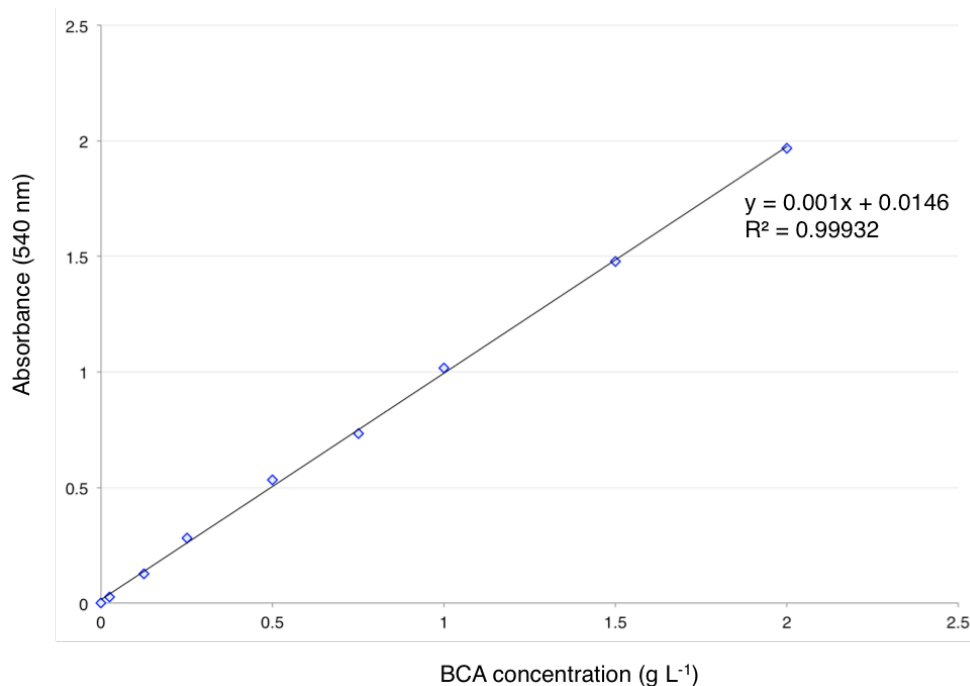
Figure 21: Absorbance at 540nm of standard solutions with known BSA concentration (g L^{-1}). The slope-intercept equation of the above line is y=0.001x + 0.0146. The absorbance (540 nm) of the prepared cell lysates was measured and substituted in the equation as the y value to find the protein concentration of the

solution. The R^2 value is close to 1 indicating that the prediction of the protein concentration of a solution via its absorbance (540 nm) is linear.

6.2.2. Optimal IPTG concentration for protein expression in bacteria liquid culture

In order to verify the functionality of the pTGR5 lac operon to promote the expression of the inserted proteins in the presence of IPTG, cells were grown in different concentrations of IPTG, lysed, and subjected to SDS-PAGE with Coomassie Brilliant Blue staining and immunoblotting with anti-His6-P. This experiment was completed with the same vector construct, CGP-05, in *E. coli* and *C. glutamicum*.

6.2.2.1. IPTG induction in *E. coli*

BL21 *E. coli* cells with CGP-05 plasmid expressed recombinant *AmyE* in the presence of IPTG. The suggested concentration of IPTG is 0.5 mM (BL21 provided product information sheet). The effect of this concentration of IPTG on the production of *AmyE* was observed, other concentrations 0, 1, 2, and 4 mM IPTG were observed as well (Figure 22). The results clearly show that in the absence of IPTG, the lac repressor (bound to the lac operon) is effective at preventing expression of *AmyE* in the cells. Little difference can be seen in the cells with the different concentrations of IPTG added. For future experiments in which recombinant protein expression and purification was the objective, 1 mM IPTG was used in the cell culture
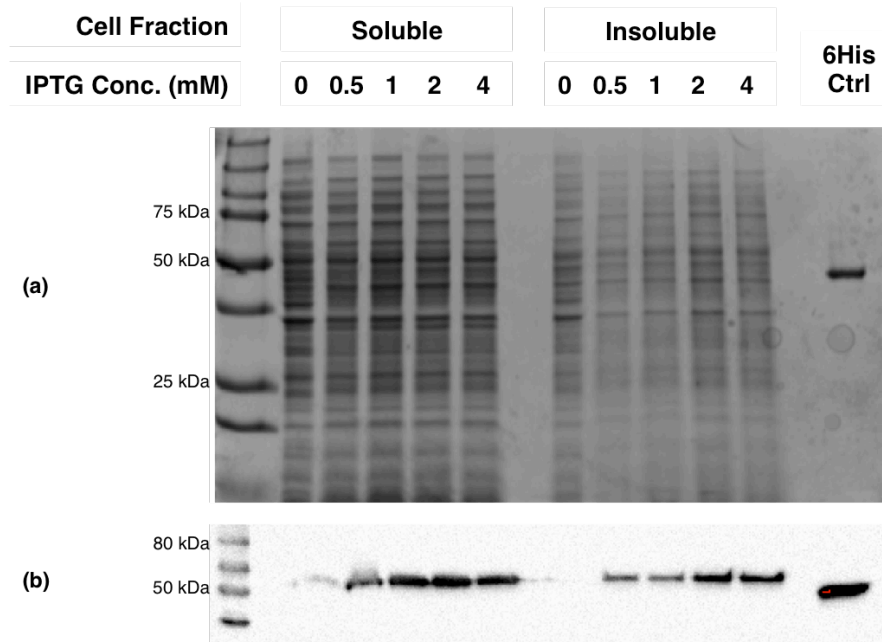
Figure 22: A Coomassie Blue stained gel (a) and an anti-6His-P treated protein blot (b) of the same samples IPTG treated *E. coli* cells expressing *AmyE* from the CGP-05 plasmid. The CLb soluble samples contain about 2X the total protein that are in the insoluble sample, and also observably contain a more diverse profile of proteins. The western blot shows that the *AmyE* protein is expressed in *E. coli* only in IPTG treated cells indicating that the plasmid is functional within the cell.

6.2.2.2. IPTG induction in *C. glutamicum*

Results similar to those shown in the prior section are seen here. The lac operon in the *C. glutamicum* cells appears to be slightly more leaky than the *E. coli* cells, as the band representing the non-induced recombinant *AmyE* on the immunoblot is more substantial. The suggested concentration of IPTG to be used was the same as the *E. coli* cells at 0.5mM. The *AmyE* expression at this concentration is suitable for further experiments, and is not significantly different from the expression at the higher concentrations of IPTG. In this case, the developed lipoprotein extraction protocol (this work) was not used and so the majority of the *AmyE* recombinant protein is in the insoluble fraction. For subsequent experiments in which recombinant protein expression and purification was the objective, 0.5 mM IPTG was used in the cell culture.
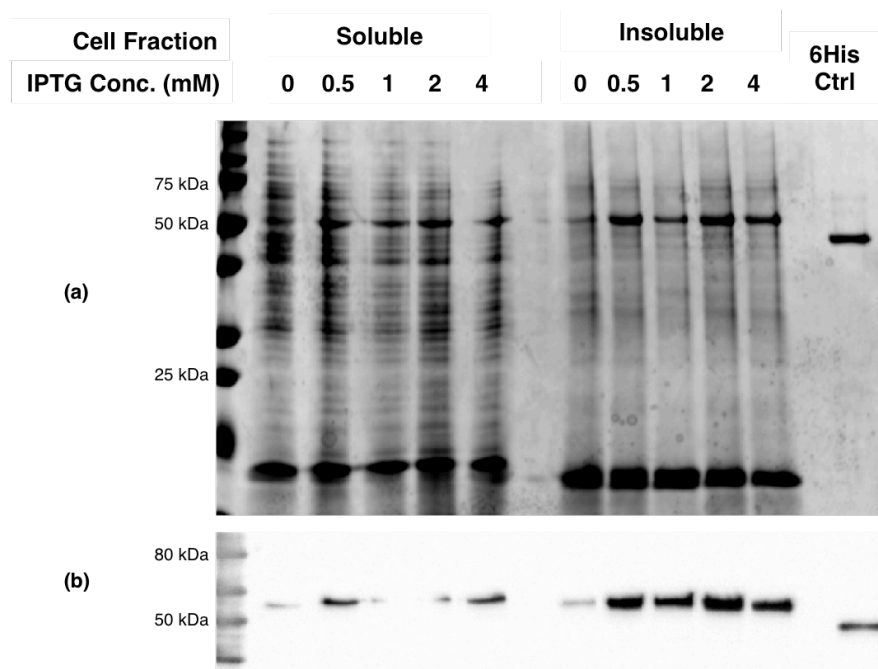
Figure 23: A Coomassie Blue stained gel (a), and an anti-6His-P treated protein blot (b) of the same samples IPTG treated *C. glutamicum* cells expressing *AmyE* from the CGP-05 plasmid. SDS-PAGE. The lysate buffer was modified to accommodate this issue in later trials. The western blot shows that the *AmyE* protein is expressed well in *C. glutamicum* in IPTG treated cells as opposed to non-IPTG treated cells indicating that the plasmid is functional within the cell. *Note: the cells were lysed using an early recipe CHAPS detergent buffer, which seemed to be interfering with the resolution of the SDS-PAGE. The lysate buffer was modified to accommodate this issue in later trials.*

6.2.3. Using IMAC with Ni-charged pre-packed columns to enrich 6His-tagged recombinant proteins from bacteria cell lysate

IMAC columns were used to enrich the recombinant proteins from the cell lysate. The columns that we had purchased had the advantage of being compatible with low concentrations of CHAPS, EDTA, and DTT. This is useful for the overexpression of foreign proteins as it helps to decrease the activity of the proteases, DNAse and RNAse in the lysate.

6.2.3.1. Recombinant protein enrichment in *E. coli* and *C. glutamicum*

Preliminary results for the enrichment of a recombinant *C. fimi* Celf_2022 isomerase is shown below in Figure 24. Figure 24a is the Coomassie Blue stained gel of the collected fractions from the 6His-tagged enrichment of Celf_2022 in both organisms and Figure 24b is the western blot with the anti-6His-P antibody of the same samples. The samples loaded are the denatured insoluble sample (Inslb), the IMAC column flow through (FT), wash (W), and elution (E). The Celf_2022 isomerase has a high sequence similarity to a natively expressed *C. glutamicum* isomerase (45%) and thus more likely to fold properly within the cell, mimicking (a) endogenous protein(s). In addition, this protein lacks a lipidation signal sequence and has the advantage of increased solubility over other putative glyco-lipoproteins. The red arrows in the figure denote the recombinant Celf_2022 protein in the elution fraction at ~27 kDa. The band appears on both the Coomassie Blue stained gel and the western blot confirming twice that the protein contains the detectable 6 His tag. This figure validates two crucial points: 1) both model organisms *E. coli* and *C. glutamicum* can express putative *C. fimi* mannosylated proteins, which can be detected and enriched via the 6His-tag, and 2) POM may be more easily characterized with soluble *C. fimi* proteins that maintain endogenous paralogs within *C. glutamicum*. The soluble enriched fraction of both organisms shows the Celf_2022 at the expected mass value, on the western blot where as the insoluble fraction shows a band at a higher mass value (~32 kDa) and in the *E. coli* insoluble fraction two bands appear at ~27 kDa and ~32 kDA. The higher molecular weight bands in the insoluble fractions may be a result of isomer formation possibly due to further glycosylation. This result can be later confirmed by mass spectrometry.
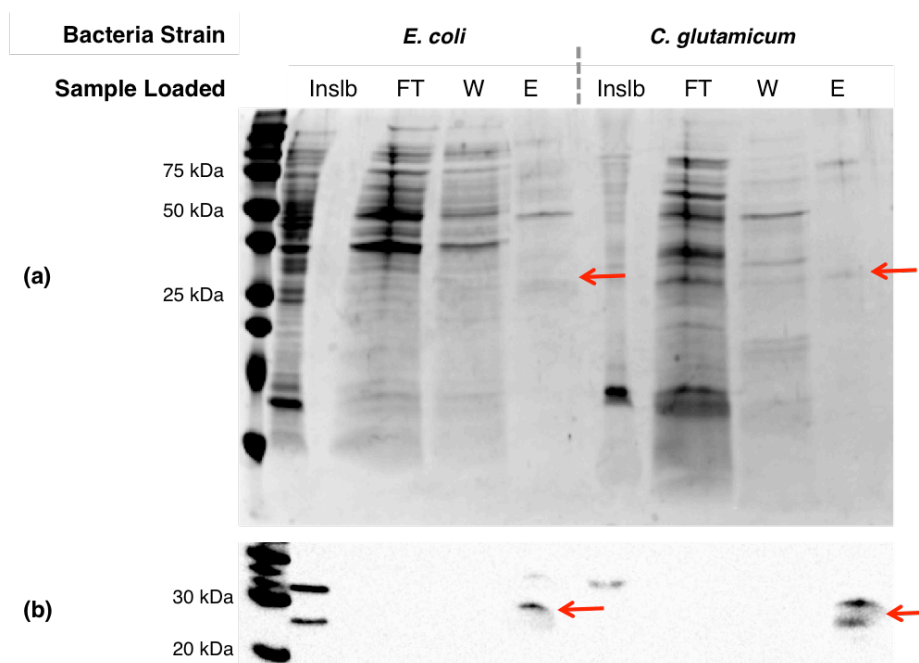
Figure 24: *E. coli* and *C. glutamicum* strains expressing recombinant Celf_2022 protein, a putative mannosylated protein from *C. fimi*, from the CFI-37 plasmid. Cells were lysed using mechanical and chemical means. The insoluble fraction (Inslb) was denatured by a detergent/urea mix. The soluble fraction was enriched for 6His-tagged proteins via IMAC columns. Fractions observed were the flow through (FT), wash (W), and elution (E). The samples were observed by SDS-PAGE (a) and western blot analysis with anti-6His-peroxidase (b). The red arrows denote the recombinant Celf_2022 protein enriched in the soluble fraction from both organisms and detected by the antibody on the western blot in the insoluble and soluble fractions.

6.2.4. Observing glycosylated proteins with ConA lectin

The ConA lectin is a useful tool for observing proteins containing terminal mannose sugar chains. ConA can be used via enrichment by batch binding and with lectin blotting as well as in combination, examples of these are shown below.

6.2.4.1. *C. glutamicum* endogenous mannosylated proteins

It is beneficial to use ConA affinity chromatography in conjunction with ConA lectin-blotting to observe the profile of mannosylated proteins in a sample. Figure 25 below shows an example. The figure displays *C. glutamicum* lysate samples from pCGE-04 (1) and CGP-05 plasmid (2) containing cells, neither of these plasmids expresses putative mannosylated *C. fimi* proteins and thus the results shown below are profiles of the endogenously expressed mannosylated proteins. The chromatography flow through and elution were observed by SDS-PAGE (Figure 25a) and lectin blot with ConA (Figure 25b). Currently, in development for this project is a protocol to remove the mannolipids from the lysate samples, as these would interfere with the *O*-mannose protein to ConA binding. The ConA lectin blotting membrane below shows in the lower molecular weight area, large and smeary bands that are believed to be mannolipid molecules of which are naturally abundant in Actinobacteria including *C. glutamicum* [122]. The elution material after the affinity chromatography is showing a few more bands than expected in both the Coomassie Blue stained gel and the ConA lectin blot indicating that the method needs to be optimized, as there is some non-specific binding. The expected number of bands is four to six as reported by Mahne et al. (2006), at various migration points: Rpf2 (cell surface associated protein) with three isoforms at 41/46/49 kDa (ia, ib, ic), LppS (secreted lipoprotein) at 49 kDa (ii), Cg_1859 (unknown function) at 22 kDa (iii), and lastly an unreported mannosylated protein at 40 kDa (iv) [52]. Due to the interference of the mannolipids with the ConA lectin the arrows indicating the suggested proteins are only estimations at this point. The image of the flow through material on the lectin blot suggests that some of the mannosylated proteins are not efficiently binding to the beads during chromatography, again possibly due to the interference with the mannolipids.
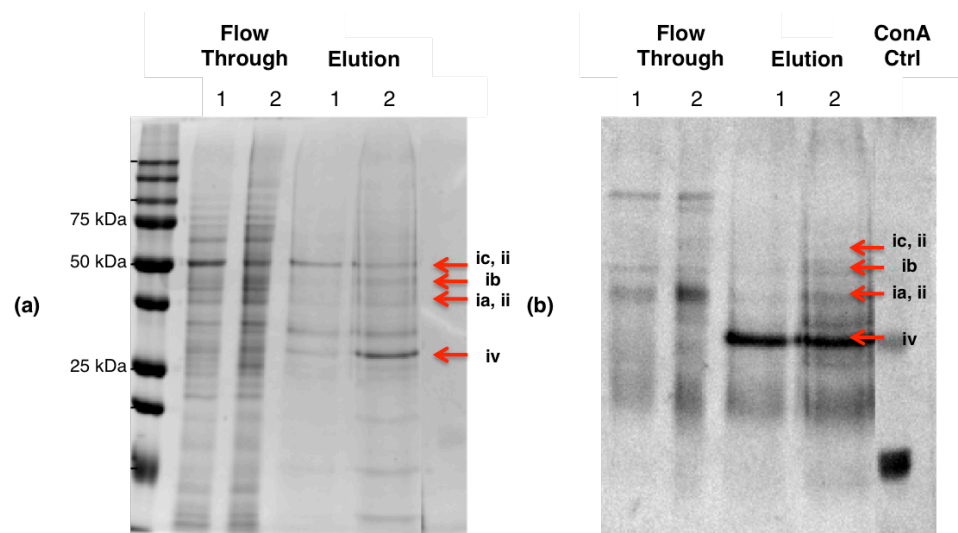
Figure 25: *C. glutamicum* endogenous mannosylated proteins. *C. glutamicum* cells were lysed using mechanical and chemical means. The soluble fraction enriched for mannosylated proteins using ConA affinity chromatography. The flow through and elution fractions were collected and observed by SDS-PAGE (a) and lectin blot with ConA. (b). The ConA ctrl used was RNase B protein. The red arrows in the elution column indicate estimated locations of the *C. glutamicum* native mannosylated proteins as reported previously [52]. Legend: (ia)/(ib)/(ic) isoforms of Rpf2 at 41/46/49 kDa, (ii) LppS at 49 kDa, (iii) Cg_1859 at 22 kDa, and (iv) unreported.

6.2.4.2. Recombinant proteins expressed in *C. glutamicum* are *O*-mannosylated

The results from section 6.2.4.1. are confirmation that the mannosylated proteins in *C. glutamicum* are recovered by experiments affinity the ConA lectin. Thus, experiments to discover the mannosylation of recombinant expressed proteins could proceed. Figure 26 below is the counter part to Figure 24 in section 6.2.3.1. The same samples were observed by lectin blot using ConA lectin. Note that the affinity chromatography fractions shown below are from IMAC columns and not ConA resin. The figure confirms that the *C. glutamicum* elution fraction contains a mannosylated protein in the same location as denoted on the western blot (Figure 24). A ConA blot representing the negative control is provided in Figure S4. This confirms that this protein contains the synthetic 6His-tag as well as an

unknown number of attached terminal mannose sugars therefore this protein is most likely the recombinantly expressed *C. fimi* Celf_2022 isomerase and likely mannosylated by the *C. glutamicum* native *PMT* enzyme. To further corroborate this fact, the *E. coli* expressed version of the protein is not displayed in the elution column on the blot indicating that the same protein as displayed on the western blot is not mannosylated. The FT fraction does not show a band at the expected mass of the Celf_2022 recombinant protein indicating like previous results that the Ni-charged affinity chromatography is effective at binding the 6His-tagged protein, and subsequently eluting. In addition, the proteins below in the *C. glutamicum* FT column are somewhat of a duplicate of the results shown in Figure 25 from 6.2.4.1. but showing a much clearer protein profile. At this point an acceptable glycoprotein ladder is not available and therefore the first lane in the figure below was loaded with a standard prestained protein ladder, which needed to be imaged separately and placed together afterward.
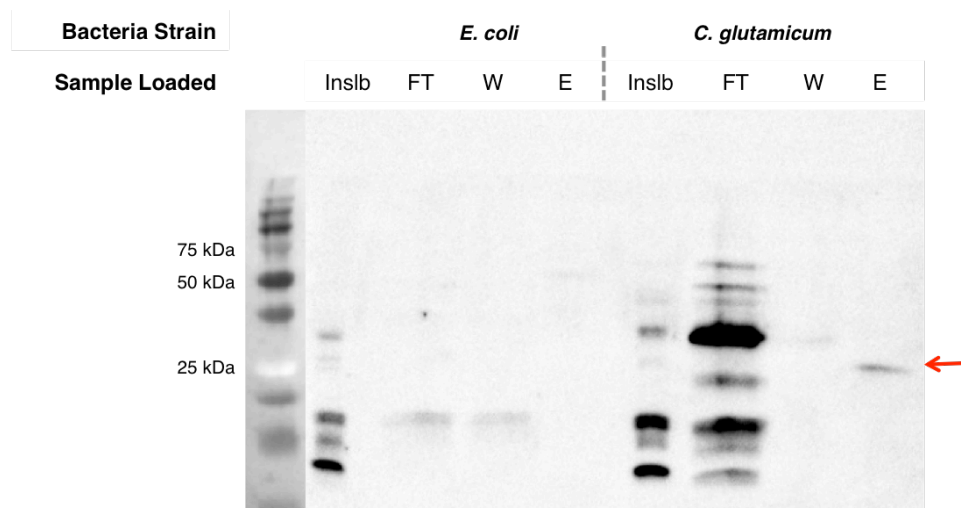
Figure 26: ConA blot of the *E. coli* and *C. glutamicum* strains expressing recombinant Celf_2022 protein, a *C. fimi* putative mannosylated isomerase from the CFI-37 plasmid. Cells were lysed using mechanical and chemical means. The insoluble fraction (inslb) was denatured by a detergent/urea mix. The soluble fraction was enriched for 6His-tagged proteins via IMAC columns. Fractions observed were the flow through (FT), wash (W), and elution (E). The samples were observed by lectin blotting with the ConA blot. The red arrow denotes the recombinant Celf_2022

protein enriched in the soluble fraction and mannosylated in *C. glutamicum* by the native *PMT*. The *E. coli* E fraction does not contain the mannosylated version of Celf_2022. The concentration of protein loaded in each lane as measured by BCA assay is 24 µg.

6.3. Discussion

6.3.1. IPTG induction of the pTGR vector constructs to induce expression of the recombinant proteins in *E. coli* and *C. glutamicum*

The use of IPTG in conjunction with a lac operon containing plasmid to induce recombinant protein expression has been an available technology for over three decades [123], it is a reliable and tested function used throughout all cell and molecular biology works. In previous work completed with the under-performing and potentially mutated pCGE-04 vector constructs the IPTG induction was leaky and produced uncontrollable recombinant protein expression. This was indicated by the fact that more of the recombinant protein was found in the non-induced culture as opposed to the induced culture (not shown). Due to these results, when the pTGR-5 plasmid was obtained, the fundamental IPTG experiments were duplicated to confirm the results from the published paper by Ravasi et al. [94]. The IPTG induced expression of the recombinant control protein *AmyE* from the pTGR derived plasmid CGP-05 was efficient and reliable as expected. This is most prominently demonstrated by the lack of recombinant protein from the cells that were not induced by IPTG, indicating that the potential for unintended protein expression with this plasmid is low. This result matched the work generated in the original paper. This particular objective could be elaborated to include a loading control and therefore a more valid form of standardizing the concentration of loaded protein, however the focus of this particular experiment was only to confirm that the IPTG induced recombinant protein expression performed as expected, and an estimation of the appropriate IPTG concentration per strain. The *C. glutamicum* cells with the CGP-05 plasmid do display a small amount of expression in absence of the IPTG, but there is significant concentration difference between the induced and non-induced recombinant protein. The *C. glutamicum* lysates were smeary and displayed bands with low resolution

probably due to salt ions from the buffer interfering with the SDS-PAGE analysis. The lysing protocol was further optimized to attain the recombinant protein in the soluble solution as well as higher resolution on the SDS-PAGE using a combination of NaBr chaotrophic agent and CHAPs detergent as well as dialysis to remove the interfering salts from the buffer.

6.3.2. The enrichment of 6His-tagged recombinant proteins from bacteria cell lysate

The IMAC columns were used for the enrichment of the recombinantly expressed *C. fimi* Celf_2022 putative mannosylated isomerase from *E. coli* and *C. glutamicum* bacteria. The eluate from both strains displays the Celf_2022 protein at the expected size (27 kDa) on both the Coomassie Blue stained gel as well as the anti-6His-P treated western blot. The absence of a similar sized band on the western blot in the flow through column from both strains confirms that the Ni-charged column enrichment is effective at binding to the recombinant protein 6His-tag. This is the optimized protocol, as prior results found that the 6His-tagged proteins were not so efficiently bound to the resin (not shown). The concentration of the recombinant protein in the eluate is low as the band in both images is faint and the exposure time for the western blot image was lengthened by a factor a four from previous results (not shown). This can be improved by a longer centrifugation cycle during the protein concentration step. The protein concentrators are theoretically effective for concentration of solutions to up to 30X, during this trial a concentration of only 5-8X was achieved for the elution fraction, taking the protein concentration of the total sample from 5 μg mL^{-1} to 25 – 40 μg mL^{-1} . It might also be improved by repetition of the experiment on a larger scale. Observed on the western blot the insoluble fractions from both strains contain some of the Celf_2022 protein at a higher molecular weight on the western blot. This may be an indication of a Celf_2022 isoform, a result of

extended glycosylation. The extension of the glyan chains after the initial mannose is attached to the protein is a poorly understood function in bacteria. Other GT enzymes, besides *PMT* in *C. glutamicum*, may be responsible for this further modification contributing to the mass increase of the protein. *E. coli* endogenous GT enzymes may produce the same result although no significant sequence similarity of Celf_2022 to any native *E. coli* protein was discovered by BLAST® query. Further elucidation of this result is possible by homogenous purification of both proteins that appear on the western blot, peptide digest and subsequent mass spectrometry analysis. The expression of Celf_2022 in *C. glutamicum* had more tangible success than the other putative hLPs expressed during this work (not shown), this may be due to its low content of non-polar AAs, thereby increasing its functional solubility in addition to *C. glutamicum* endogenous paralogs increasing the potential for proper folding of the recombinant protein Celf_2022. It is possible that the most important factor when attempting to use this collection of protocols is to first choose a protein that has a similar sequence to a native *C. glutamicum* protein. It has been observed during this work that an improved system for the expression of foreign proteins in *C. glutamicum* is required to express completely heterogeneous proteins, but before this can proceed, the clarification of the exact cause of the hLP protein expression difficulty is necessary.

6.3.3. ConA lectin blotting of *C. glutamicum* and *E. coli* cell lysates

6.3.3.1. Endogenous *C. glutamicum* proteins

To test the efficacy of the ConA protocols two experiments involving this lectin were combined to produce a profile of the endogenous *C. glutamicum* mannosylated proteins, affinity chromatography using ConA bound resin, and lectin blotting with ConA lectin bound peroxidase. The ConA lectin is a primary

source for discovery of putative mannosylated proteins in an organism [75, 84]. Reports by Mahne et al. (2006) show 4 mannosylated proteins in *C. glutamicum* 2 are confirmed by other means, and 2 are putative [56]. Although there are only four proteins so far, previous reports have observed multiple bands on *C. glutamicum* ConA lectin specific blots indicating the appearance of multiple isoforms of a specific protein Rpf2. Multiple isoforms seem to be a function of the extent of the glycosylation of the protein. More than four bands were observed in the elution fraction from the ConA affinity chromatography on the Coomassie Blue stained gel and the lectin blot. This results is caused by one or a combination of all of the following: mannosylated lipid molecules that are present within the cell wall of *C. glutamicum* [16] are also binding to the ConA resin and being eluted with the mannosylated proteins, isoforms of other glycoproteins besides Rpf2 which have not been reported are expressed, some of the bands contain proteins which have a terminal glucose and not a mannose, and/or the ConA chromatography protocol is not optimized and maintains some weak binding of other glycosylated proteins to the resin, which are then eluted by the α-methylglucoside. The location of the known mannosylated proteins and their isoforms were estimated on Figure 25, however only more advanced molecular techniques can validate this suggestion. For the purposes of this objective, which was to observe the protein profile of the endogenous *C. glutamicum* mannosylated proteins and test the efficacy of both the ConA resin and ConA-peroxidase conjugate this result is acceptable. Further optimization of the protocol involved achieving a higher resolution of the ConA blots, which was completed by buffer exchange. A protocol to remove the mannosylated lipids from the cell lysates is being developed.

6.3.3.2. ConA blotting to observe mannosylated Celf_2022 in *E. coli* vs. *C. glutamicum*

The *C. fimi* Celf_2022 isomerase expressed recombinantly in *C. glutamicum* shows evidence of mannosylation by the native *C. glutamicum* PMT enzyme, as shown in the eluate fraction after Ni-charged column chromatography. The Celf_2022 isomerase is a glycoprotein discovered in the preliminary findings it was shown to be more robustly expressed on cellulose containing media over xylan containing media and is shown to have seven associated mannose molecules (Section 3.2). It has homology to a cyclophilin type peptidyl-prolyl cis-trans isomerase in other bacteria and eukarya strains and shows 45% sequence similarity to a specific *C. glutamicum* isomerase. It has yet to be fully structurally and functionally characterized with respect to the placement of the mannose moieties and the functional role of the sugar chains. In many species, cycolophilin type isomerases are often indicated in virulence [124], a reoccurring theme in the study of protein mannosylation. It is likely that the successful expression and mannosylation of the foreign Celf_2022 is due to the high sequence similarity of the protein to a native *C. glutamicum* protein and thus the proper folding of the foreign protein and subsequent *PMT* interaction.

This is the first acquired evidence that the *C. glutamicum* PMT is able to mannosylate an expressed foreign protein; this is indication as hypothesized that the POM systems in Actinobacteria *C. glutamicum* and *Cellulomonas* are related and that the *PMT* enzymes may be interchangeable. However, this experiment is preliminary and must be repeated to confirm the results. In addition the yield of the elution would be necessary data for future experiments. The sequence similarity between the PMT enzymes from *C. glutamicum* and *Cellulomonas* are quite low (< 20%), but still are functionally similar and able to accept similar substrates for mannosylation. The unsolved difficulty may lie in the ability of the substrates to be expressed and folded properly in heterogeneous strains that would allow PMT interaction as well as 6His-tag detection. The results obtained in Figure 26 from the *E. coli* insoluble fraction may be due to glucosylated proteins/lipids that can also bind to the ConA lectin, *E. coli* cells do not have the

94

required *PPM* to mannosylate proteins although there are confirmed glycosylated proteins have been previously reported in *E. coli* cells [17]. The *C. glutamicum* FT fraction shows a clearer profile of the endogenous *C. glutamicum* proteins, although the locations of the specific proteins would be only estimations at this point. The mannosylated lipids are also observed in the < 15kDa area on the blot as expected. Further confirmation of all of the above theories can be done by trypsin digest and mass spectrometry of any specific protein that can be purified via a stronger form of chromatography such as high performance liquid chromatography. A full purification of the required protein from *C. glutamicum* and *E. coli* is also needed to complete the cell-free assay of PMT activity. Needless to say, while optimization of certain protocols would have been ideal, this result was a necessary accomplishment to the full scope of the project being realized and was a milestone achievement in this project.

7. Optimized protocol for *C. glutamicum* recombinant lipoprotein enrichment using control protein *AmyE*

7.1. Experimental Procedures

The target hypothetical lipo-glycoproteins, Celf_1421, Celf_0777, and Celf_3337, as well as the three GT39 enzymes are all membrane bound proteins. Lipoproteins are not easily solubilized due the hydrophobic portions of the protein that allow it to be inserted into the membrane. The protocol development below sought to find an efficient and reliable method to solubilize and observe 6His-tagged lipoproteins from gram-positive bacteria cells. The method incorporates the combined use of a chaotrophic agent and zwitterionic detergent to solubilize the membrane proteins.

7.1.1. Bacterial cell growth conditions and harvest

Engineered R163 *C. glutamicum* cells containing the CGP-05 plasmid were grown and harvested as stated in section 6.1.1.2.

7.1.2. Cell fractionation and lipoprotein enrichment from Actinobacteria cells

All samples obtained in this section were observed by SDS-PAGE with Coomassie Brilliant Blue staining and immunoblotting with anti-His6-P.

7.1.2.1. Optimized protocol using NaBr and CHAPS detergent

Cells were mechanically lysed as described previously in section 6.1.1.2. Cell/celite mixture was suspended in 10 mL of a membrane solubilizing solution containing the chaotrophic agent NaBr (This is made up of 2.5 M NaBr, 20 mM MgCl$_2$, 1 mM MnCl$_2$, 2000 U DNase I (Roche), and 1X COMPLETE protease inhibitor tablet (Roche), dissolved in 1 X PBS, pH 7.2. The suspension was transferred to a falcon tube and mixed on an end-to-end rotator at 4°C for 24 h. Mixed suspension was clarified first with centrifugation at 500 g for 2 min, to remove celite from samples. The lysate was transferred to 1.5 mL microcentrifuge tubes, was and further fractionated by centrifugation at 16 000 g for 20 min to separate soluble fraction from membrane fraction (insoluble material). The soluble fraction was removed, transferred to a new falcon tube, and stored at 4°C.

The insoluble fractions were resuspended in 10 mL CHAPS detergent solution (CHAPS 2) containing 0.2 g L^{-1} (fresh) lysozyme (Bio Shop, Burlington, Canada), 0.01 g L^{-1} RNase A, 1000 U DNase I (Roche), and 1X COMPLETE protease inhibitor tablet (Roche), dissolved in 1X PBS, pH 7.2. Suspension was transferred to a falcon tube and mixed on an end-to-end rotator at 4°C for 24 h. Mixed suspension was transferred to 1.5 mL microcentrifuge tubes and centrifuged at 16 000 g for 20 min. The solubilized membrane fraction (detergent fraction) was removed from tubes and stored at 4°C. Insoluble material was solubilized in 1 mL of CHAPS 2 containing 4 M urea. The insoluble fractions were stored at 4°C.

7.1.2.2.   Comparing the efficacy of zwitterionic detergents on lipoprotein extraction

Multiple zwitterionic detergents were tested for their ability to extract lipoproteins from the membrane fraction after the NaBr treatment. The above protocol

(6.1.2.1) was used to test the following detergents (a 2% (w/v) concentration of all detergents was used): CHAPS detergent, CHAPS with ASB 14, ASB 16, and ASB-C80. All detergents were purchased from Sigma Aldrich, (ASB – Amidosulfobetaine).

7.1.3. Desalting and buffer exchange

The soluble and detergent fractions were dialyzed with denaturing wash buffer 1 (refer to section 7.1.3) using Float-A-Lyzer® G2 dialysis device (Spectrum Labs Inc. California, USA) as per the product information sheet provided. Buffer was exchanged 3X in 24 h. Dialyzed samples were stored at 4°C.

7.1.4. Purification of recombinant 6His-tagged *AmyE* protein from dialyzed soluble and detergent fractions

6His-tagged recombinant *AmyE* protein expressed in engineered (CGP-05) *C. glutamicum* cells were purified from the soluble and detergent fractions obtained in sections 7.1.2.1 and 7.1.3. by IMAC. The protocol from section 6.2.3. was followed with two exceptions; the initial sample dilution and final observation by lectin blotting were both unnecessary steps in this experiment. ConA lectin blotting would be required in the event of the purification of a glycolipoprotein (refer to section 6.1.4). *Note: the insoluble fraction from section 7.1.2.1 was not purified due to its high viscosity.*

7.1.5. Expression and purification of recombinant AmyE in engineered *E. coli* cells

AmyE protein was expressed and purified successfully from BL21 *E. coli* cells via the methods outlined in Chapter 6 (refer to sections 6.1.1.1. and 6.1.3.).

7.2. Results

7.2.1. Comparing the efficacy of zwitterionic detergents on *C. glutamicum* lipoprotein extraction

Zwitterionic detergents are the preferred kind of detergents for this work. In comparison to other types of detergents the characteristics of zwitterionic detergents make them better at solubilizing and breaking down protein-protein interactions without disrupting the native fold of the protein, and easier to remove from a lysate solution after solubilization [125]. In this experiment, the effect of the chaotrophic agent NaBr at removing peripheral membrane proteins prior to membrane solubilization as well as the efficacy of four different types of zwitterionic detergents were tested. The synthesized zwitterionic detergents are called sulfobentaines and are differentiated by the number of carbon atoms in the non-ionic tail. The sulfobentaines used in this experiment, ASB-14, ASB-16, and ASB-C8Ø, also have an amido group to represent the positive ion in the ionic segment of the molecule [122]. CHAPS detergent is the fourth type of detergent used which has all of the properties of the ASB detergents but instead of long carbon chains it contains a fused ring system that is sterol based in the hydrophobic segment (Figure S2). The results in Figure 27 show that the combined use of the NaBr and a zwitterionic detergent, either CHAPS (lane 2) or a CHAPS/ASB-14 mixed solution (lane 3) are extremely efficient at solubilizing the *AmyE* lipoprotein. The chaotrophic agent used, NaBr, was efficient at removing some of the peripheral membrane proteins and solubilizing a sample of the recombinant *AmyE* protein (lane 1) from the mechanically lysed fraction. Further treatment of the insoluble fraction after NaBr incubation using CHAPS detergent allows most of the remaining *AmyE* to be solubilized. Interestingly, the CHAPS/ASB-14 detergent mixture also solubilizes what looks like an equal amount of *AmyE* to the CHAPS treatment, but the ASB-14 detergent alone solubilizes none. Thus, in the mixture solution the recombinant

AmyE shown on the western blot (Figure 27b) is likely due to the CHAPS detergent only.
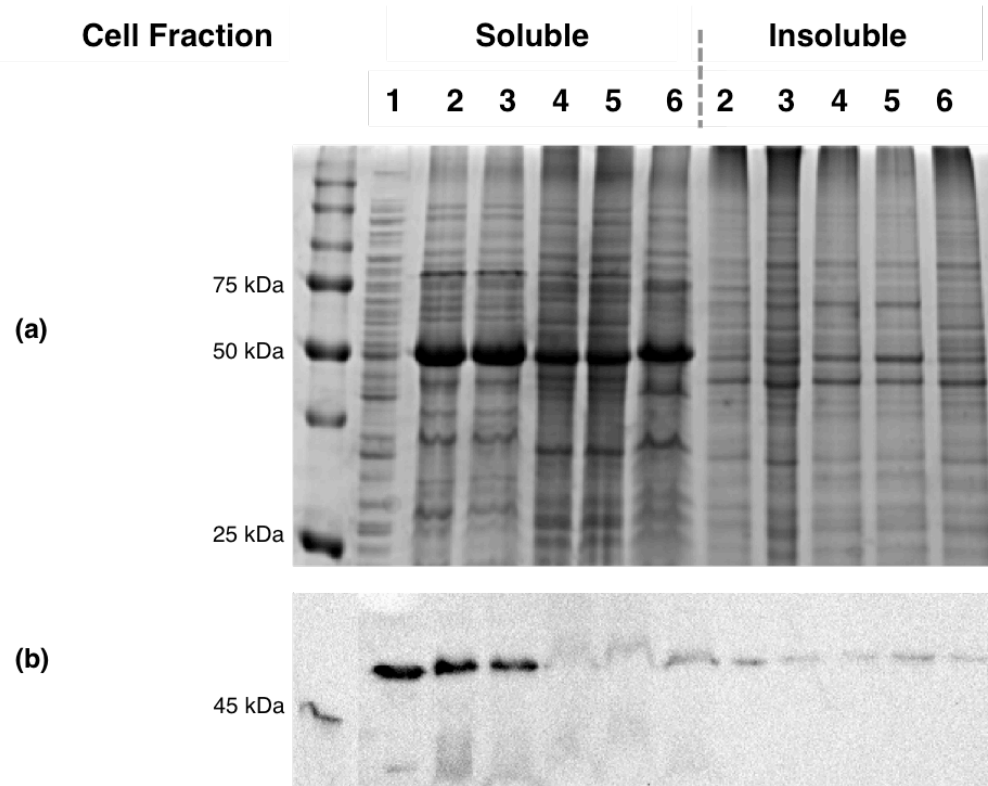
Figure 27: Mechanically lysed and chemically treated *C. glutamicum* cells expressing AmyE from the CGP-05 plasmid. Different chemical treatments were prepared and used on the cells to determine the efficacy to solubilize the overexpressed lipoprotein. All cells had a prefractionation treatment with NaBr, and afterward the chosen detergent. Chemical Legend: (1) NaBr only, (2) + CHAPS, (3) + CHAPS/ASB-14, (4) + ASB-14, (5) + ASB-16, (6) +ACB**.** Samples were subjected to SDS-PAGE and Coomassie Blue stained, as well as western blot with Anti-6His-P. The results show that the option to best solubilize the recombinant lipoprotein is (2) NaBR + CHAPS detergent.

The questions remains, why the recombinant AmyE after treatment with either of the three amidosulfobentaines alone is not visible on the western blot? If the detergents were not adequate to solubilize the lipoproteins, then it would be expected that they would appear in the insoluble fraction (lanes 4,5, and 6). In addition, a dark/intense band appears on the Coomassie Blue stained gel (Figure 27a) in lanes 4,5, and 6 under the soluble fractions at the same molecular weight

size that *AmyE* is expected (55 kDa), and seen in the first three lanes. It is possible that the ASB detergents can somehow interfere with the exposure of the 6His-tag to the anti-His-P antibody during the western blotting, and therefore the recombinant protein is not being detected. Specifically, the ASB detergent could be interfering with the SDS unfolding of the protein causing it to smear and be incompatible to protein blot transfer and subsequent antibody detection. Although this is a curious result, the information gathered from this experiment is adequate to use in the subsequent work. A dual treatment of NaBr and CHAPS detergent, after mechanical lysis, sufficiently solubilizes lipoproteins from the complex cell wall of gram-positive bacteria.

7.2.2. Enrichment of solubilized recombinant lipoprotein from *C. glutamicum* cells using IMAC with Ni-charged resin

The success of the lipoprotein solubilization work allowed the protocol development for enrichment of the recombinant lipoprotein to move forward. Efficient solubilization allows the lysate fraction to be first dialyzed to remove the NaBr and CHAPS and second subjected to affinity chromatography with Ni-charged resin that binds to the 6His-tag. Figure 28 below shows the results after both of these methods. Before dialysis (bDlys) shows the solubilized samples after the listed treatment, NaBr (soluble fraction), or NaBr and CHAPS (detergent fraction). The soluble fraction is effectively all of the soluble proteins from the cell in addition to the proteins that were removed from the membrane via the chaotrophic agent. The protein profile difference between the bDlys sample and the flow through (FT) sample on the Coomassie Blue stained gel (Figure 28a) should be minimal since the only treatment in between is that a buffer is exchange. Although the recombinant AmyE is maintained in the fraction during the dialysis step as shown on the western blot image (Figure 28b), it seems that there is protein lost in the discarded buffer. Because the soluble fraction is fully

suspended in the chaotrophic agent, it is likely that this fraction has more NaBr ions than the detergent fraction. The affinity chromatography is greatly affected by a high salt concentration and due to the western blot showing that the recombinant protein from the soluble fraction is in the FT, Wash (W), and Elution (E) this shows there was binding interference between Ni^{2+} in the IMAC column and N-terminal poly-His tag on the recombinant protein. Allowing the recombinant protein to flow through the column without binding.

On the other side of the gel the detergent fraction shows that the affinity chromatography was effective. CHAPS can also interfere with Ni-charged resin at high concentrations. However, the concentration of protein in the bDlys sample and E sample on both the Coomassie Blue stained gel and the western blot are similar indicating that the dialysis was efficient at removing the detergent from the solution. Still, most of the recombinant protein was recovered but in comparing the bDlys and the FT samples of the detergent fraction on the Coomassie Blue stained gel there is noticeable protein loss.
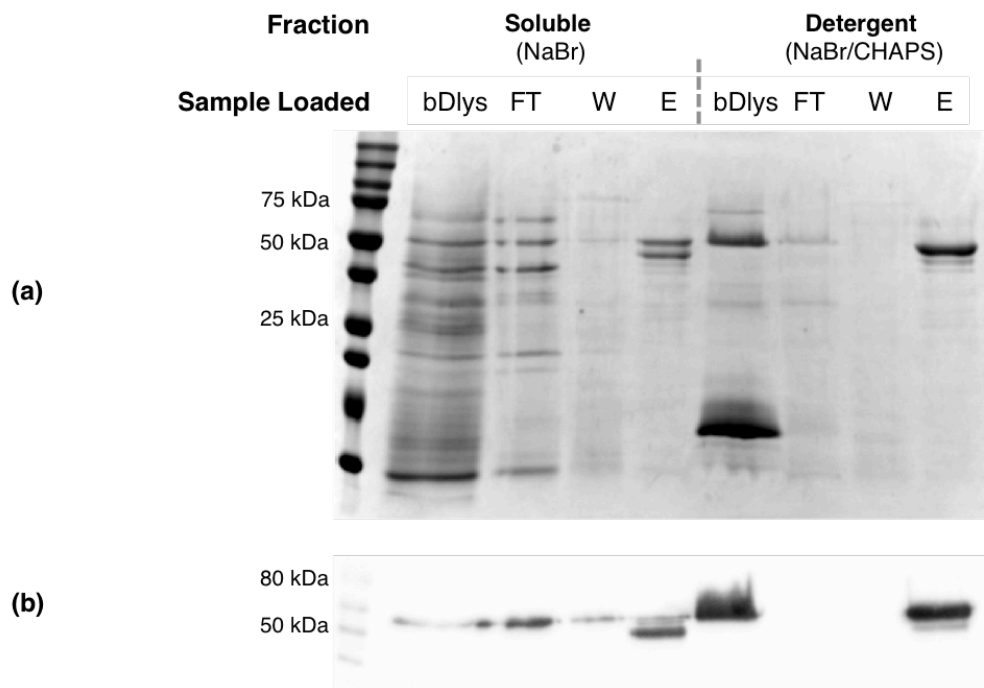
Figure 28: *C. glutamicum* cells expressing recombinant lipoprotein *AmyE* were mechanically lysed. Lysates were NaBr/CHAPs treated, dialyzed to remove salts and detergents, and enriched for the recombinant 6His-tagged lipoprotein by affinity chromatography. Legend: lysate sample after specified treatment (bDlys), IMAC flow through (FT), wash (W), and elution (E). The detergent fraction contains most of the acylated recombinant *AmyE* protein (55 kDa) in the E fraction and the soluble fraction contains an equal amount of lipidated and non-lipidated recombinant protein (47 kDa). The Coomassie Blue stained gel (a) shows the ability of the dialysis to retain the population of proteins in the lysate, some protein loss is observed. The western blot (b) shows the binding of the 6His-tag in the soluble fraction is ineffective as the recombinant protein is maintained in all three eluates FT, W, and E. Like the detergent fraction where the binding was not affected by the contents of the buffer, the recombinant protein should only be seen in the bDlys and E at approximately equal concentrations.

The *AmyE* lipoprotein can be expressed at two different sizes, 47 kDa and 55 kDa. The high molecular weight isoform is owing to the triacylation of the terminal cysteine after cleavage of the lipobox, producing the lipoprotein product, and increasing the size of the molecule. Possibly due to overexpression of the *AmyE*, there is an observable population of non-acylated *AmyE* protein found in the soluble sample just under the expected recombinant *AmyE* band. Without the lipidation, the protein is soluble. The E sample of the soluble fraction shows that most of the recombinant *AmyE* recovered is non-acylated, where as most of the *AmyE* recovered in the E step of the detergent fraction is the full lipoprotein expressing at the expected molecular weight.

7.2.3. Expression and enrichment of native *C. glutamicum* lipoprotein in BL21 *E. coli* cells

E. coli cells are well solubilized with the CLb detergent leaving a very little amount of the insoluble material. Based on the results in Figure 29, this material holds a lot of the overexpressed *AmyE* lipoprotein. The early expression tests of recombinant *AmyE* protein were completed in the *E. coli* transformation strain DH10B, and when lysed yielded soluble recombinant *AmyE* protein (not shown),

likely due to a weakened cell membrane of the DH10B strain. The expression strain BL21 has a healthy membrane, which makes the solubilization of the lipoproteins more challenging, as it was in the *C. glutamicum* cells. An added prefractionation process with NaBr may be necessary in cases where the maximum amount of recombinant protein is the objective. For the purpose of this experiment to determine that recombinant *AmyE* protein from *C. glutamicum* can be expressed and enriched in *E. coli* with a detectable 6His-tag, the result below is acceptable.

The soluble fraction after the detergent treatment contains enough recombinant *AmyE* to prove the success of the 6His-tag with the IMAC. The Coomassie Blue stained gel (Figure 29a) shows multiple proteins in the enrichment, although only one is picked up on the western blot by the anti-His6-P (Figure 29b). The multiple proteins in the E column could be due to insufficient number of wash steps before the elution, or low concentration of the imidazole in the wash buffer, which does not remove the weakly bound proteins from the IMAC. A wash/elution gradient involving different concentrations imidazole could be completed to find the concentration optimal for protein enrichment. The results below confirm that an observable population of the recombinant lipoprotein *AmyE* can be recovered with the CLb (an easy tool for *E. coli* lysis and protein recovery). The *eGFP* columns represent a negative control for the protein expression and 6His-tag monitoring by western blot.
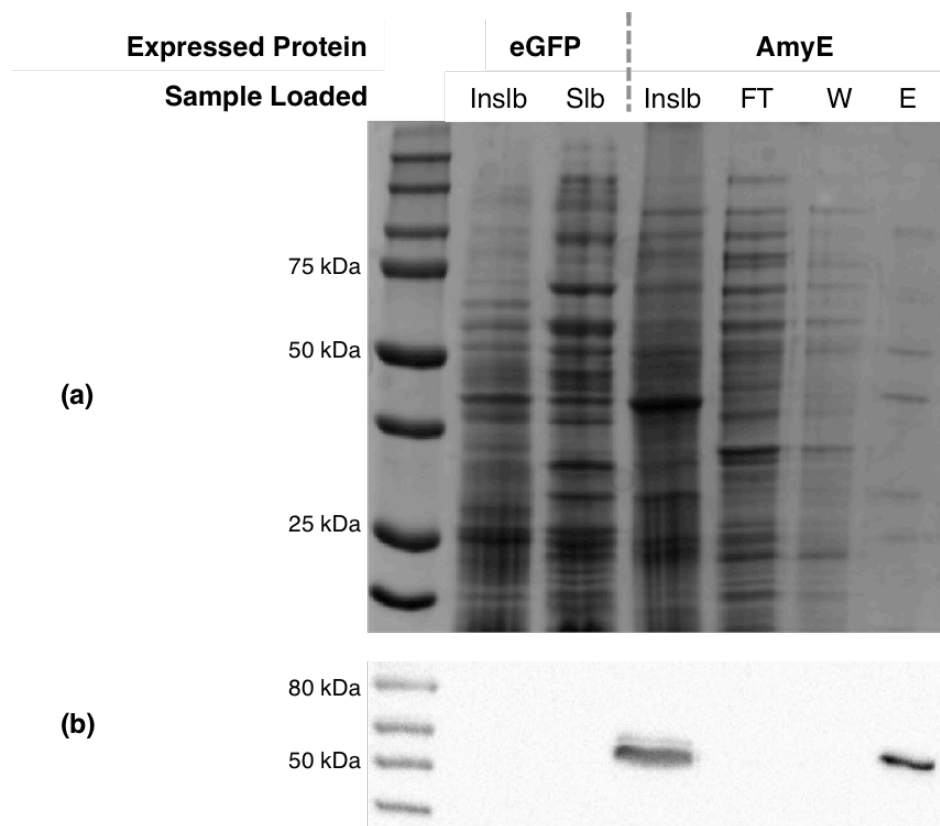
Figure 29: *E. coli* cells expressing recombinant protein *AmyE* were lysed with CLb and enriched for the protein containing the 6His-tagged by affinity chromatography. Legend: Insoluble particulate was solubilized by urea and 1X CLb detergent (Inslb), Soluble fraction was enriched through IMAC flow through (FT), wash (W), and elution (E). The *eGFP* columns display represents a negative control for the recombinant protein; the Inslb and Slb fractions show diverse protein expression in the Coomassie Blue stained gel (a) without any 6His-tagged peptides in the western blot (b). The recombinant clones express *AmyE*, as shown by the western blot (b). The insoluble fraction contains some recombinant protein indicating the CLb detergent treatment isn't as effective as expected at membrane solubilization but for the purpose of this experiment these results are acceptable. The recombinant *AmyE* band appears to be ~55kDa on the western blot (b) indicating that protein is triacylated and that the lipidation signal sequence is conserved between the two species.

7.3. Discussion

7.3.1. Comparing the efficacy of zwitterionic detergents on *C. glutamicum* lipoprotein extraction

Lipoproteins in gram-positive bacteria are often difficult to analysis via SDS-PAGE because they are in low abundance in the cell and poorly soluble in buffer due to their hydrophobic nature and/or lipid modifications [122, 126]. However, these proteins comprise 22% of the total identified proteins in the *C. glutamicum* genome, consequently there is value in deciphering the best methods to breakdown the lipid-rich exterior and solubilize the proteins within it, which is necessary for the subsequent characterization [122]. Prior to the development of this protocol, results using common lysing methods and buffers were not able to remove any of the recombinant *AmyE* lipoprotein from the insoluble fraction (not shown) and therefore continuing with the objective to enrich the protein would not be possible. The results displayed above most prominently show that an overexpressed lipoprotein in *C. glutamicum* can be solubilized using these methods and can be useful for future experiments of a similar nature. Nevertheless, it is still possible that the most effective detergent to choose is on a protein-by-protein basis as reported elsewhere [122]. This theory is confirmed by the fact that the ASB detergents, although used successfully and often elsewhere for the same types of experiments [122], did not produce the anticipated results in this case. Before the detergent treatment all samples contained roughly the same amount of *AmyE* protein, as judged by the Coomassie Blue stained gel, however the 6His-tag and anti-His6-P binding is affected by the ASB detergent and the recombinant *AmyE* does not appear on the western blot. This may be due to the SDS interfering with the unfolding of the protein causing a spear as opposed to a tight bind. For these purposes these ASB detergents are not useful to solubilize and maintain the native form of the lipoprotein. Arguably, the most

valuable take away is that this detergent test can be carried out rather quickly and reliably with any number of detergents that the user is interested in in order to decipher the best one for a specific case.

For the *C. glutamicum* cells, and due to the cell wall composition containing many complex molecules, a pre-fractionation treatment using a chaotrophic agent like NaBr is recommended before the detergent is added. NaBr is relatively mild as compared to urea or guanidine thiocyanate, which is beneficial to the preservation of the proteins in the sample. As a chaotrophic agent the NaBr weakens the hydrophobic effect of macromolecules in the membrane, which in turn compromises the membrane integrity, and tertiary folding of the proteins [122]. This wash releases approximately 40% of the integral membrane proteins from the membrane fraction as compared to 70% - 90% using the harsher agents [122]. However, more qualitative analyses of the cell lysates after treatment with the harsher agents revealed some soluble and integral membranes were missing, as reported by Shcluesener et al. 2005 [122].

7.3.2. Enrichment of solubilized recombinant lipoprotein from *C. glutamicum* and *E. coli* cells using IMAC.

A lipoprotein with an available 6His-tag can be efficiently enriched using the IMAC, as validated by the results shown above. 6His-tags can be an extremely useful tool, and seem to be reliable in strains, like *E. coli* BL21, that have been modified for foreign protein expression. The putative *C. fimi* mannosylated lipoproteins that were chosen for expression using the pTGR5 vector (Celf_0777 in CFI-35 and Celf_3336 in CFI-36) do not have any sequence similarity to proteins natively found in *E. coli* or *C. glutamicum*. The expression tests using these vector constructs worked fairly well in *E. coli*, the 3336 and 0777 proteins were observed on the western blot using the anti-6His-P at the appropriate sizes

(Figure S3). The one caveat is that the Celf_0777 contains a potential internal start site, which initiates the translation of the protein downstream of the expected start site producing two different 6His-tagged bands on the western blot. Both proteins were found mostly in the insoluble fraction in the BL21 *E. coli* expression strain, and therefore will require a prefractionation membrane treatment for solubilization before subsequent enrichment.

On the other hand the same experiments using CFI-35 and CFI-36 expressed in *C. glutamicum* lacked any 6His-tagged proteins on the western blots (Figure S3). Reported in this work and by other groups [127, 128] foreign proteins with no sequence similarity show difficulty expressing synthesized 6His-tags, which can be caused by cytotoxicity, inefficient translation, insolubility or misfolding (buried 6His-tag and PTM sequon) or a combination of any of the above. Since the C. fimi Celf_3336 hLPs shows homology to an extracellular substrate binding protein, it may be unstable in the *C. glutamicum* cell without a specific type of ATP-binding cassette (ABC) transporter. It was postulated earlier that the mannosylated proteins in *C. fimi* are differentially expressed based on metabolic carbon available. This indicates that some or all of the hLPs discovered may be associated with other ABC transporters that do not have counter parts in *C. glutamicum* and are unstable upon expression in the cell.

Naturally, this is not an issue for over expressed proteins native to *C. glutamicum* or foreign proteins that show a higher sequence similarity to a specific native protein from the organisms or that are not associated with specific *C. fimi* proteins, as is the case with Celf_2022 (see Chapter 6). Therefore, *AmyE* was efficiently enriched in the results above. A failed enrichment experiment, which returned an eluate without any enriched protein was completed on the *C. glutamicum* fractions that did not undergo a dialysis step, due to the interference of the chaotrophic agent and/or the detergent with the IMAC resin. Buffer exchange by dialysis is a necessary step when using the NaBr and CHAPS and

subsequent IMAC enrichment. Lastly, the protocol in *E. coli* could be optimized to address the two issues: the eluate contains some Ni-charged weakly binding proteins to the IMAC resin, and the BL21 expression strain needs a higher concentration of detergent and/or a prefractionation wash with a chaotrophic agent to release some of the lipoprotein from the insoluble fraction.

The protocols in this chapter addressed and solved an important problem that microbiologists face when attempting to study hydrophobic and lipoproteins and/or working with difficult gram-positive bacteria; the solubilization and enrichment of an insoluble protein that would normally be trapped in the cell membrane. Especially when doing foundational work, it is useful to have protocols that are reproducible and technically easy – which contributes to wide applicability. Moreover, these types of proteins are often involved in the fitness of the cell, whether it is cell defense, taking up/signaling for metabolites, or invading host cells. If the methods for studying these molecules can become as reliable and easy as the existing methods for studying soluble proteins in *E. coli,* it is inevitable that the some of the fundamental biology of the microorganism, from the evolution of the cell to antimicrobial resistance, be solved.

8. Concluding Remarks

Actinobacteria are industrially and medically significant bacteria that use protein glycosylation as a means of virulence, proliferation, metabolite competition and general cell homeostasis to name a few. Glycosylation of specific proteins can affect their physiological functions and roles within the bacterial cell. POM is a specific glycosylation process conserved from archaea to man. The evolution of this process is remarkable, as it has developed from a somewhat nonessential process in bacteria into a required cell process in humans. Studying the POM mechanism in simpler organisms such as bacteria can shed light on this poorly understood process and offer insight to some unanswered questions regarding the evolution of protein glycosylation. This thesis contributes a molecular toolbox, a collection of specific and reliable protocols that can be used to study POM in an accessible and simple system – specifically the gram-positive and gram-negative model organisms, *C. glutamicum* and *E. coli* respectively. A similar type of project for N-linked glycosylation was completed by Wacker et al. (2002) via the functional transfer of the *C. jejuni* N-linked glycosylation pathway into *E. coli* cells [31]. This work was a breakthrough point for the elucidation of N-linked glycosylation in bacteria and allowed for the exploration of this process in a simpler organism. It is the overall goal of this project to offer the same type of progress for the second most common type of glycosylation and the implementation of the protocols in this toolbox offer a reputable starting point to achieve this. With this toolbox all aspects of species related POM systems can be observed including *PMT*s from other organisms, the *PMT* substrates, and the mannose-lipid donors.

For this big picture goal to be realized additional work is required. Future projects aim to (1) optimize certain protocols mentioned in this thesis, (2) work on the characterization of the native *PMT* from *C. glutamicum* and related foreign *PMT*s from *Cellulomonas* bacteria, are they interchangeable? (3) Develop an *in vitro* cell-

free assay to probe the structure function relationship of the PMT enzyme. The developed *O*-mannosylation assay of purified **putative** glycoproteins can then be assessed to test ideas about the function of the GT39 *PMT* on various protein types, and (4) complete mass spectrometry on purified *Cellulomonas* glycoproteins expressed in *C. glutamicum* in order to characterize frequency and placement of the mannose moieties. All of these objectives would individually provide great value to the Glycobiology community, as do the protocols developed in this thesis. More over the completion of this project would impact the work of other more specialized groups effective in globally significant industries including host-pathogen interactions, metabolic fitness of a bacterial cell, human CDGs, antimicrobial resistance, and large-scaled production of useful enzymes, antigens, or other proteins.

Appendices

Supplementary Methods

SM 1. Optimized *C. glutamicum* optimized mini-prep protocol

1. Make Make a LB-Miller 10mL overnight culture of *C. glutamicum* cells from a single colony containing 1% Glucose.
   The culture should reach an A_{600} of 5-8 after 12 hours.

2. Harvest 3 x 500 uL of cells in microcentrifuge tubes at 6000 x g for 5 min. Remove the supernatant completely.

3. Resuspend the cells in 1 mL Buffer B.

   - Buffer B – 50mM Glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0
     - 50mL – 0.45g Glucose, 0.186g EDTA, 1.25mL Tris-HCl (1M, pH 8.0)
     - 250mL – 1.8g Glucose, 0.745g EDTA, 5mL Tris-HCl (1M, pH 8.0)

4. Pellet cells via centrifugation at 6000 x g for 2 min. Remove supernatant completely.

5. Prepare a 1 mL stock of 100mg/mL Lysozyme.

6. Resuspend the pellet in 0.3mL Buffer PD1 containing 20mg/mL FRESH Lysozyme. All mini-prep supplies are taken from the GeneAid Mini-Prep Kit.

   - 60 µL of 100 mg/mL Lysozyme

7. Incubate tubes at room temperature for 2 – 4 hours. Transfer tubes to 4°C (preferably rotating) for 12 – 48 hours.

8. Remove tubes from fridge. *Put PD3 buffer in -20°C freezer*. Add 0.3mL Buffer PD2 to each tube. Mix gently by inverting 4-6 times, and incubate at room temperature for 5 min.

9. Add 0.3mL of chilled Buffer PD3 to each tube, mix immediately and gently by inverting 4-6 times, and incubate on ice or in freezer for 15 min.

10. Centrifuge tubes at maximum speed in a microcentrifuge for 10 min. Remove all three supernatant into ONE PD column + 2 mL Collection Tube. **Be careful not**

to transfer any white pelleted matter into column.

11. Follow instructions **ONLY step 5 to step 7** for High-Speed Plasmid Mini Kit Protocol from GeneAid.

12. Elute DNA into 50 µL total volume. Measure concentration of final elution at A_{260} with spectrophotometer.

SM 2. Western Blotting Protocol for Immunodetection His-Tagged Proteins

1. Run the samples on an acrylamide gel as per usual. If time permits, make two gels and load both with the same samples. Then, Coomassie stain one gel and use the other gel for your blot.
    a. Use Precision Plus All Blue (prestained) Standard – 5 µL

2. Cut 1 piece of PVDF membrane the size of your gel (7.5cm x 10cm) and obtain two pieces of 3 mm filter paper (sometimes may have to cut yourself) for each gel being analyzed. Always wear gloves when handling PVDF membrane (can keep it within blue sheets until the gel is ready).

3. Make up 1.0-L of CAPS electroblotting buffer in 10% MeOH.

    a. **10X CAPS Buffer:** In 1.0-L bottle add 22.1g CAPS and 900 mL of dH$_2$O. Adjust pH to 11 with NaOH. Fill with dH$_2$O to 1.0-L.
    b. **1X CAPS Buffer with 10% MeOH:** In 1.0-L bottle add 100 mL of 10X CAPs Buffer pH 11 and 100 mL of MeOH. Fill with dH$_2$O to 1.0-L.

4. Add approx. 5 mL of MeOH to a 50mL falcon tube and submerge the PVDF membrane (with blue protecting sheets). Roll falcon tube to completely cover PVDF membrane, this activates the membrane.

5. Transfer the PVDF (with blue protecting sheets) to a clean Ziploc container, pour approx. 20mL 1X CAPS/10% MeOH Buffer. Add filter paper on top of PVDF. Add both sponges on top of filter paper. Place container in fridge. Put 1.0-L bottle of 1X CAPS/10% MeOH Buffer in freezer.

6. Once the acrylamide gel has finished running. Prepare your sandwich.

    a. Place the Gel Holder Cassette with hinges down and black side down on a solid and level surface.

b. Remove the PVDF membrane and filter paper from the buffer but make sure they do not dry out. One at a time try to remove air bubbles from the fiber pads while they are still submerged in the liquid.

c. Place one fiber pad on black part of cassette

d. Place one filter paper on fiber pad

e. Place PVDF on the filter paper

f. Place gel on top of PVDF

g. Place filter paper on top of gel

h. Place fiber pad on the filter paper

i. Close cassette carefully as not to jar the contents

7. Place the cassette in the Buffer Tank (Black side facing towards the back and red/black sides matched accordingly). Put Buffer Tank on top of a stirrer.

8. Fill Buffer tank with cold 1X CAPS/10% MeOH Buffer up to the Blot Line. Put in ice packet and small stirring rod. Put lid on. Turn on stirrer.

9. Run blot at 50V, 100mA, 100W for 1 hour.

10. Once completed, open sandwich in the same manner that it was prepared to preserve orientation of the membrane. Cut the corner of the upper left-hand side of the membrane to establish orientation. In blue-ink record the date in the bottom right-hand corner of the membrane.

11. Remove membrane and place in container with approx. 80mL of blocking solution (TBST with 1% BSA (albumin)). Leave for 2hrs or overnight with gentle shaking.

a. **1X TBS Buffer:** Dissolve 6.05g Tris (50mM) and 8.76g NaCl (150mM) in 800 mL dH$_2$O. Adjust pH to 7.5 with HCl. Fill with dH$_2$O to 1.0-L.

b. **1X TBST:** Dissolve 1 mL of Tween 20 in 1.0-L of 1X TBS

c. **Blocking Solution:** Add 1 g of Albumin to 100 mL of 1X TBST (use only 80mL for blocking and save 20 mL for primary antibody detection)

12. After Blocking, wash the membrane a few times with fresh 1X TBST. Add approx. 20mL of Anti-His$_6$-Peroxidase Solution to the blot. Incubate for 1 hr with shaking.

a. **Anti-His$_6$-Peroxidase Solution:** Dissolve 0.3 μL (or volume containing requested amount of units of enzyme) in 20 mL TBST + 1% BSA.

13. Remove antibody solution and rinse with 1X TBST Buffer for 1 hr while changing the buffer every 10m.

14. Remove rinsing solution. Combine the ECL Solutions (1) Luminol and (2) H_2O_2 in a 1:1 ratio (5 mL of each) and pour onto blot. **Leave on blot for only 1 minute!** Pour off ECL solution and cover blot with saran wrap.

   a. **Luminol Solution:** Dissolve 112.5mg luminol in 2.5 mL DMSO. Dissolve 18.5mg p-Coumaric Acid in 1.25 mL DMSO. Add each to 0.1M Tris pH 8.8 in a final volume of 250 mL. Store at 4°C.
   b. **H_2O_2 Solution:** Add 153 μL Hydrogen Peroxide Sigma H – 1009 to 250mL 0.1M Tris pH 8.8. Store at 4°C.

SM 3. ConA Affinity Batch Binding

1. Prepare Buffers:
   a. ConA Buffer A, pH 7.4: 20 mM Tris, 0.5 M NaCl, 1.0 mM $CaCl_2$, 1.0 mM $MnCl_2$
   b. ConA Buffer B, pH 7.4: 20 mM Tris, 0.5 M NaCl, 0.3 M α-methylglucoside

2. Obtain ~ 100 µL ConA resin per 1 mL cell lysate suspended in ConA Buffer A.

3. Rinse ConA in ConA Buffer A 3X by low speed centrifuge

4. Incubate cell lysate with rinsed ConA Buffer A at RT for 2 hr. Low speed centrifuge and remove supernatant as flow through fraction (FT).

5. Add 1 mL of ConA Buffer A to beads and incubate at RT for 10 m. Low speed centrifuge and remove supernatant as Wash 1 (W1). Repeat step and remove fraction as Wash 2 (W2).

6. Add 500 µL ConA Buffer B to beads and incubate at RT for 30 m. .Low speed centrifuge and remove supernatant as Elution 1 (E1). Repeat step and incubate for 4 – 6 hr and label supernatant as Elution 2 (E2).

7. Analysis fractions on SDS-PAGE gels.

SM 4. Preparing 4X Loading Dye Recipe

1. Combine 1.5 mL of 1 M Tris-HCl pH 6.8, 3 mL of 1 M DTT (dithiothreitol), 0.6 g of SDS (sodium dodecyl sulfate), 0.03 g of bromophenol blue, 2.4 mL of glycerol

2. Bring final volume to 7.5 mL
   • If solution is orange/yellow in color, add 1 drop of 5 M NaOH to adjust pH

Table S1. Glycoproteins identified from *C. fimi* cell extracts

Gene	Annotation	Carbon Source	Mass	# Hexose Residues
Celf_2022	Peptidyl-prolyl isomerase (cyclophilin)	CMC & Xylan	26 700	7 hex
Celf_0777	Hypothetical protein	CMC & Xylan	28 730	7 hex
Celf_2865	Hypothetical protein	CMC & Xylan	37 840	at least 2hex
Celf_1830	Extracellular ligand binding receptor	CMC & Xylan	43 070	3 hex
Celf_3336	Extracellular solute binding family 1	CMC & Xylan	47 730	7 hex
Celf_0189	Penicillin-binding protein transpeptidase	CMC Tx100 & Xylan?	68 580	8 hex
Celf_1421	Hypothetical protein	Xylan	13 290	at least 8 hex (up to 11)
Celf_2833	Hypothetical protein	Xylan	17 100	2 hex
Celf_3669	Periplasmic binding protein	Xylan	34 060	4 hex
Celf_3229	Periplasmic binding protein	Xylan	35 820	6-7 hex
Celf_0922	Hypothetical protein	Xylan	38 040	at least 4 hex
Celf_1347	Extracellular solute binding family 1	Xylan	46 700	3 hex
Celf_3272	Extracellular solute-binding family 1	Xylan	60 700	2 hex
Celf_2028	Preprotein translocase YajC subunit	Xylan & CMC	15 350	2-3hex
Celf_1573	Extracellular solute-binding family 3	Xylan & CMC?	29 760	4 hex
Celf_1453	Periplasmic binding protein	Xylan & CMC ?	34 920	6-7 hex
Celf_2803	ATP synthase F0 subcomplex subunit B	Xylan & CMC	21 050	at least 1 hex
Celf_1045	Hypothetical protein	Xylan & CMC	21 411	2 hex
Celf_3689	Peptidylprolyl isomerase FKBP-type	Xylan & CMC	31 970	6 hex
Celf_2314	Extracellular solute binding family 1	Xylan & CMC	43 020	4 hex
Celf_0553	NLP/P6 protein	Xylan & CMC	50 990	2 hex
Celf_0591	Periplasmic binding protein	Xylan &CMC?	37 220	3 hex

Table S2. Glyco-lipoproteins identified from sequence containing putative glycosylation and lipobox

Gene	Annotation	Signal Peptidase II cleavage site : LX[A/G]C and Glycopeptide
Celf_3229	Periplasmic binding protein	MRHLRRTAVAASVAVLALALAAACSGSTDDDATAEPGSTASDGTTFPVTIESALGTAVVEEKPER
Celf_0591	Periplasmic binding protein	MKLSLHRRAGAVALTGALALALAAACGSDDPTGSGDTTPGSDESVSTLSGELNGAGASSQEK
Celf_3669	Periplasmic binding protein	MRNLRALPAAALAGTAALLLAACGTTEEAGAAEPSTSADVAGGPVTITDDR
Celf_1347	Extracellular solute binding family 1	MLRFAAVGVAAALTLTACSSGDDTGSDDATDAGTPAPATIK
Celf_3336	Extracellular solute binding family 1	MRHRPALRGRTAVLRSAALLAVGALALTACAGSGEPAAEATSSGPPEPVEIR
Celf_2314	Extracellular solute binding family 1	MRRSIPVVAATLGLALTLAACSGSSGETSDDASASAPAESAGTLTVWVDDTR
Celf_1573	Extracellular solute-binding family 3	MRARRLVALAAAATLGLAACSSGGGDDDTDAGAGGDATETSGGAEGTIR
Celf_1830	Extracellular ligand binding receptor	MIRSTHAVRAAALAGAAALILAACSGGGEGSDDATEGGGGDTAAPLK
Celf_0922	Hypothetical protein	MVRVTGTTRATAGRHGRARAATRWAVPLLGAVLALAACGDGSGQPSPTSVTQR
Celf_0777	Hypothetical protein	MSARRRGVHPGDFSVLRSRLFALPAAVLLTATLAACGGGSDSPSAEKTSASSEKSTPTPTPEVAELTTEDFVAR
Celf_1045	Hypothetical protein	MAARTRHGRPPDRADGSPRPRAAAALVVAVVGTLAGCAGAAPEMT//TDYPVERVEPTAPSGR*
Celf_2865	Hypothetical protein	MKKIIRVAALGGAVALTLAACGSAPEETEETAGTGDAAASDFK
Celf_1421	Hypothetical protein	MKLPATRTRLAVGAATAALAATAFLAGCSDSGAEDEPTAVETTTESSTEDEAASGQVEVIDQTHLLLATAQDTLEAR
Celf_3689	Peptidylprolyl isomerase FKBP-type	MRRTTTARAIAAATTALVLSLSLAACSGDGGGSDDSPSADASASVSAAQAAD

**Table S3** Primers designed for the amplification of target genes by PCR. (F) Forward primer (R) Reverse primer.

Gene		Primer (5'-3')
Celf_3080	F	GAA GTA TTC CAT ATG CCT CCG ACA CGT GAT GAT GAT G
	R	GAA GAA GGT ACC TTA TTA ATG GTG ATG GTG ATG GTG AAT CCA GCT GGT CAG CCA CAT G
Cg_1014	F	GAA GTA TTC CAT ATG AGC CAA GCC CTA CCT GTT CGT G
	R	GAA GAA GGT ACC TTA TTA ATG GTG ATG GTG ATG GTG GCG CCA GCT TGG GAA CCA C
Cflav_0843	F	GAA GTA TTC CAT ATG CCG ACC GAC GGA GAC GAC ACC G
	R	GAA GAA GGA TCC TCA TCA ATG GTG ATG GTG ATG GTG GAT CCA GGT CGG CAG CCA C
Cg_2705	F	GGG GTA TTC CAT ATG ATG AAG TAC TTG GTG GCG C
	R	GGG GAA GGA TCC TCA TCA CCG GTA GTA GTA GTA GTA GTA GTA ATT
Celf_1421	F	GGG GTA TTC CAT ATG TCC GAC TCG GGC GCC GAG G
	R	GGG GAA GGA TCC TCA TCA ATG GTG ATG GTG ATG GTG GGA CGC GGG GTC GTC CGT G
Celf_0777	F	GGG GTA TTC CAT ATG TCC ACG CCG ACG CCG ACC C
	R	GGG GAA GGA TCC TCA TCA ATG GTG ATG GTG ATG GTG GAA CGG CAT CTC CGT GGT GAT CTG
Celf_3336	F	GGG GTA TCC CAT ATG GCC GGA TCC GGC GAG CCC G
	R	GGG GAA GGT ACC TCA TCA ATG GTG ATG GTG ATG GTG GGA CTG GCC GAT CGC CAG CTT G

Table S4. IMAC Ni-charged column buffer recipes

	NaCl, mM	Na Phosphate, mM	Imidazole, mM
Lysis/Wash Buffer 1	300	50	5
Wash Buffer 2	300	50	25
Elution Buffer	300	50	500

Supplementary Figures

Dolichol phosphate

(a)
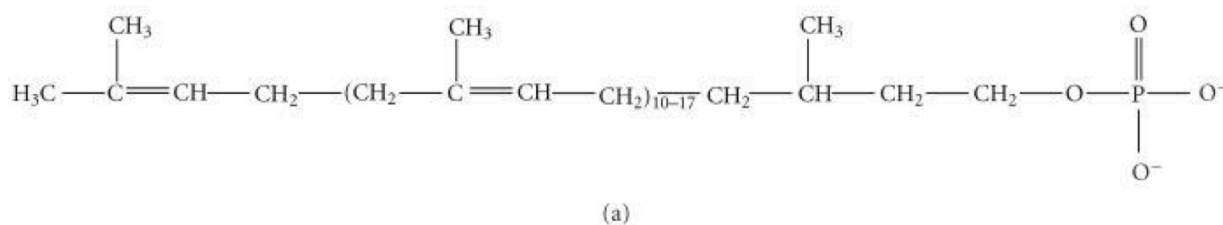
Undecaprenol phosphate
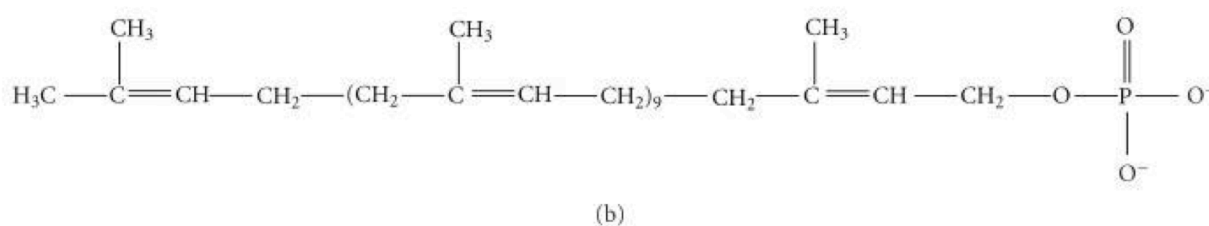
(b)

Figure S1. Structure of dolichol phosphate donor for eukaryotic Dol-P-Man synthase vs polyprenol phosphate donor for prokaryotic polyprenol monophosphomanno-synthase*.*

CHAPS
MW 615

https://www.thermofisher.com/order/catalog/product/28300
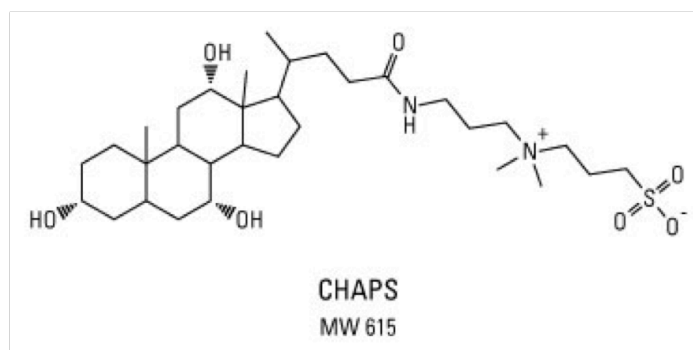
Figure S2. Molecular structure of CHAPS detergent.

(a) CFI-33 (*E. coli*)

(b) CFI-33 (*C. glutamicum*)

(c) CFI-36 (*C. glutamicum*)
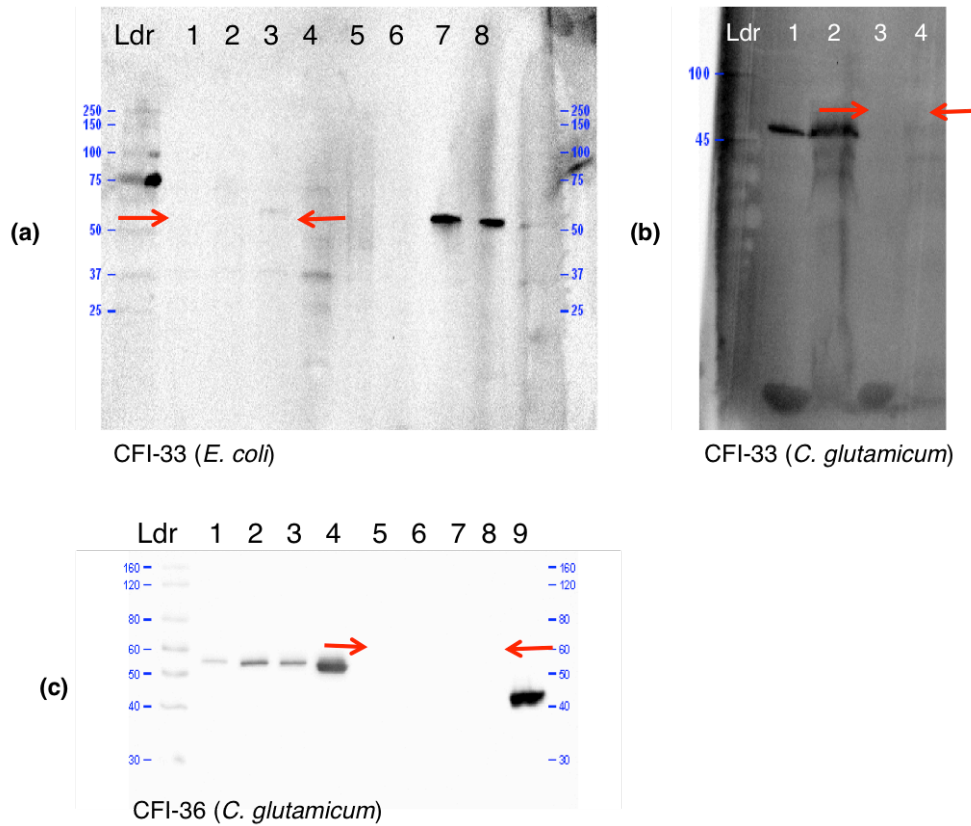
Figure S3. Western blot images showing no detected expression of 6His tag from *C. glutamicum* strains expressing vector with *C. fimi* lipoprotein gene. Between the red arrows denotes where the expected band indicating the tagged expression of the Celf_3336 in *C. glutamicum* (b) (c). *E. coli* BL21 expression Celf_3336 is contained in solely in the insoluble fraction (a). Images (b) and (c) both contain strong bands under lanes that were loaded with the *AmyE* expressing cells (lanes 1 & 2 (b); lanes 1-4 (c)). As expected the 6His tag is intact and *AmyE* acts as a positive control for lipoprotein extraction.

C. glutamicum

Sample Loaded   bDlys   FT   W   E

75 kDa

50 kDa

25 kDa
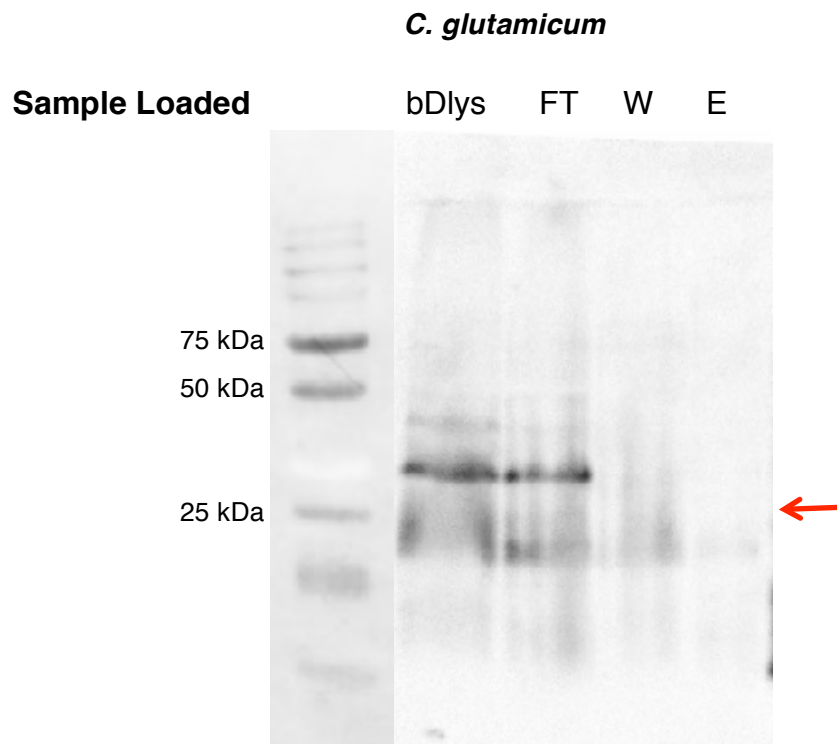
Figure S4. ConA blot representing negative control for the expression and mannosylation of the recombinant Celf_2022 in *C. glutamicum*. *C. glutamicum* strain with CGP-05 expressing AmyE protein (non-mannosylated) soluble fraction is enriched for 6His-tagged proteins via IMAC columns in the same manner as Celf_2022. Cells were lysed using mechanical and chemical means. The soluble fraction was enriched for 6His-tagged proteins via IMAC columns. Fractions observed were the before dialysis (bDlys), flow through (FT), wash (W), and elution (E). The samples were observed by lectin blotting with the ConA blot. The red arrow denotes the absence of the mannosylated protein in the elution fraction. The concentration of protein loaded in each lane as measured by BCA assay is 24 µg.

References

1. Dell, A., Galadari, A., Sastre, F., & Hitchen, P. (2010). Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. *International Journal of Microbiology*, *2010*(7), 1–14. http://doi.org/10.1155/2010/148178

2. Iwashkiw, J. A., Vozza, N. F., Kinsella, R. L., & Feldman, M. F. (2013). Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Molecular Microbiology*, *89*(1), 14–28. http://doi.org/10.1111/mmi.12265

3. Ohtsubo, K., & Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*, *126*(5), 855–867. http://doi.org/10.1016/j.cell.2006.08.019

4. Tytgat, H. L. P., & Lebeer, S. (2014). The Sweet Tooth of Bacteria: Common Themes in Bacterial Glycoconjugates. *Microbiology and Molecular Biology Reviews*, *78*(3), 372–417. http://doi.org/10.1128/MMBR.00007-14

5. Schaffer, C., Graninger, M., & Messner, P. (2001). Prokaryotic glycosylation. *Proteomics*, *1*(2), 248–261. http://doi.org/10.1002/1615-9861(200102)1:2<248::AID-PROT248>3.0.CO;2-K

6. Dube, D. H., Champasa, K., & Wang, B. (2010). Chemical tools to discover and target bacterial glycoproteins. *Chemical Communications*, *47*(1), 87. http://doi.org/10.1039/c0cc01557a

7. Pandhal, J., & Wright, P. C. (2010). N-Linked glycoengineering for human therapeutic proteins in bacteria. *Biotechnology Letters*, *32*(9), 1189–1198. http://doi.org/10.1007/s10529-010-0289-6

8. Eichler, J. (2013). Extreme sweetness: protein glycosylation in archaea. *Nature Reviews Microbiology*, *11*(3), 151–156. http://doi.org/10.1038/nrmicro2957

9. Zafar, S., Nasir, A., & Bokhari, H. (2011). Computational analysis reveals abundance of potential glycoproteins in Archaea, Bacteria and Eukarya. *Bioinformation*, *6*(9), 352–355.

10. Argyros, R., Nelson, S., Kull, A., Chen, M.-T., Stadheim, T. A., & Jiang, B. (2013). A Phenylalanine to Serine Substitution within an O-Protein Mannosyltransferase Led to Strong Resistance to PMT-Inhibitors in Pichia pastoris. *PLoS ONE*, *8*(5), e62229–10. http://doi.org/10.1371/journal.pone.0062229

11. Rosnoblet, C., Peanne, R., Legrand, D., & Foulquier, F. (2012). Glycosylation disorders of membrane trafficking. *Glycoconjugate Journal*, *30*(1), 23–31. http://doi.org/10.1007/s10719-012-9389-y

12. Lommel, M., Schott, A., Jank, T., Hofmann, V., & Strahl, S. (2011). A Conserved Acidic Motif Is Crucial for Enzymatic Activity of Protein O-Mannosyltransferases. *Journal of Biological Chemistry*, *286*(46), 39768–39775. http://doi.org/10.1074/jbc.M111.281196

13. Loibl, M., & Strahl, S. (2013). Protein O-mannosylation: What we have learned from baker's yeast. *BBA - Molecular Cell Research*, *1833*(11), 2438–2446. http://doi.org/10.1016/j.bbamcr.2013.02.008

14. Upreti, R. K., Kumar, M., & Shankar, V. (2003). Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics*, *3*(4), 363–379. http://doi.org/10.1002/pmic.200390052

15. Liu, C. F., Tonini, L., Malaga, W., Beau, M., Stella, A., Bouyssie, D., et al. (2013). Bacterial protein-O-mannosylating enzyme is crucial for virulence of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *110*(16), 6560–6565. http://doi.org/10.1073/pnas.1219704110

16. Espitia, C., Servín-González, L., & Mancilla, R. (2010). New insights into protein O-mannosylation in actinomycetes. *Molecular BioSystems*, *6*(5), 775–781. http://doi.org/10.1039/b916394h

17. Lindenthal, C., & Elsinghorst, E. A. (2001). Enterotoxigenic Escherichia coli TibA Glycoprotein Adheres to Human Intestine Epithelial Cells. *Infection and Immunity*, *69*(1), 52–57. http://doi.org/10.1128/IAI.69.1.52-57.2001

18. Karlyshev, A. V. (2004). The Campylobacter jejuni general glycosylation system is important for attachment to human epithelial cells and in the colonization of chicks. *Microbiology*, *150*(6), 1957–1964. http://doi.org/10.1099/mic.0.26721-0

19. Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., & Logan, S. M. (2003). Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Molecular Microbiology*, *48*(6), 1579–1592. http://doi.org/10.1046/j.1365-2958.2003.03527.x

20. Horzempa, J., Comer, J. E., Davis, S. A., & Castric, P. (2006). Glycosylation Substrate Specificity of Pseudomonas aeruginosa 1244 Pilin. *Journal of Biological Chemistry*, *281*(2), 1128–1136. http://doi.org/10.1074/jbc.M510975200

21. Lengeler, K. B., Tielker, D., & Ernst, J. F. (2007). Protein-O-mannosyltransferases in virulence and development. *Cellular and Molecular Life Sciences*, *65*(4), 528–544. http://doi.org/10.1007/s00018-007-7409-z

22. Varki, A. (2006). Nothing in Glycobiology Makes Sense, except in the Light of Evolution. *Cell*, *126*(5), 841–845. http://doi.org/10.1016/j.cell.2006.08.022

23. Vigerust, D. (2011). Protein glycosylation in infectious disease pathobiology and treatment. *Open Life Sciences*. http://doi.org/10.2478/s11535-011-0050-8

24. Tan, F. Y. Y., Tang, C. M., & Exley, R. M. (2015). Sugar coating: bacterial protein glycosylation and host–microbe interactions. *Trends in Biochemical Sciences*, *40*(7), 342–350. http://doi.org/10.1016/j.tibs.2015.03.016

25. Patterson, J. H., Waller, R. F., Jeevarajah, D., Billman-Jacobe, H., & McConville, M. J. (2003). Mannose metabolism is required for mycobacterial growth. *Biochem. J.*, *372*(1), 77–10. http://doi.org/10.1042/BJ20021700

26. Wehmeier, S., Varghese, A. S., Gurcha, S. S., Tissot, B., Panico, M., Hitchen, P., et al. (2009). Glycosylation of the phosphate binding protein, PstS, in Streptomyces coelicolorby a pathway that resembles protein O-mannosylation in eukaryotes. *Molecular Microbiology*, *71*(2), 421–433. http://doi.org/10.1111/j.1365-2958.2008.06536.x

27. Hug, I., & Feldman, M. F. (2011). Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria. *Glycobiology*, *21*(2), 138–151. http://doi.org/10.1093/glycob/cwq148

28. Lehle, L., & Tanner, W. (1978). Glycosyl transfer from dolichyl phosphate sugars to endogenous and exogenous glycoprotein acceptors in yeast. *Eur. J. Biochem.*, *83*(2), 563–570.

29. Langsford, M. L., Gilkes, N. R., Singh, B., Moser, B., Miller, R. C., Jr., Warren, R. A., & Kilburn, D. G. (1987). Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional

domains. *FEBS Letters*, *225*(1-2), 163–167.

30. Vasudevan, D., & Haltiwanger, R. S. (2014). Novel roles for O-linked glycans in protein folding. *Glycoconjugate Journal*, *31*(6-7), 417–426. http://doi.org/10.1007/s10719-014-9556-4

31. Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., et al. (2002). N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into *E. coli*. *Science*, *298*(5599), 1790–1793. http://doi.org/10.1126/science.298.5599.1790

32. Messner, P., Schffer, C., & Kosma, P. (2013). Bacterial cell-envelope glycoconjugates. *Advances in Carbohydrate Chemistry and Biochemistry* (1st ed., Vol. 69, pp. 209–272). Elsevier Inc. http://doi.org/10.1016/B978-0-12-408093-5.00006-X

33. Lommel, M., & Strahl, S. (2009). Protein O-mannosylation: Conserved from bacteria to humans. *Glycobiology*, *19*(8), 816–828. http://doi.org/10.1093/glycob/cwp066

34. Lehle, L., Strahl, S., & Tanner, W. (2006). Protein Glycosylation, Conserved from Yeast to Man: A Model Organism Helps Elucidate Congenital Human Diseases. *Angewandte Chemie International Edition*, *45*(41), 6802–6818. http://doi.org/10.1002/anie.200601645

35. Jaeken, J. (2010). Congenital disorders of glycosylation. *Annals of the New York Academy of Sciences*, *1214*(1), 190–198. http://doi.org/10.1111/j.1749-6632.2010.05840.x

36. Comstock, L. E., & Kasper, D. L. (2006). Bacterial Glycans: Key Mediators of Diverse Host Immune Responses. *Cell*, *126*(5), 847–850. http://doi.org/10.1016/j.cell.2006.08.021

37. Girrbach, V., & Strahl, S. (2003). Members of the Evolutionarily Conserved PMT Family of ProteinO-Mannosyltransferases Form Distinct Protein Complexes among Themselves. *Journal of Biological Chemistry*, *278*(14), 12554–12562. http://doi.org/10.1074/jbc.M212582200

38. Beltrán-Valero de Bernabé, D., Currier, S., Steinbrecher, A., Celli, J., van Beusekom, E., van der Zwaag, B., et al. (2002). Mutations in the O-mannosyltransferase gene POMT1 give rise to the severe neuronal migration disorder Walker-Warburg syndrome. *American Journal of Human Genetics*, *71*(5), 1033–1043. http://doi.org/10.1086/342975

39. Tabish, S., Raza, A., Nasir, A., Zafar, S., & Bokhari, H. (2011). Analysis of glycosylation motifs and glycosyltransferases in Bacteria and Archaea. *Bioinformation*, *6*(5), 191–195.

40. Wang, Z.-X., Deng, R.-P., Jiang, H.-W., Guo, S.-J., Le, H.-Y., Zhao, X.-D., et al. (2012). Global Identification of Prokaryotic Glycoproteins Based on an Escherichia coli Proteome Microarray. *PLoS ONE*, *7*(11), e49080–11. http://doi.org/10.1371/journal.pone.0049080

41. Tanner, W. (1969). A lipid intermediate in mannan biosynthesis in yeast. *Biochemical and Biophysical Research Communications*, *35*(1), 144–150.

42. Orlean, P., Albright, C., & Robbins, P. W. (1988). Cloning and Sequencing of the Yeast Gene for Dolichol Phosphate Mannose Synthase, an Essential Protein. *Journal of Biological Chemistry*, *263*(33), 17499–17507.

43. Maeda, Y., & Kinoshita, T. (2008). Dolichol-phosphate mannose synthase: Structure, function and regulation. *Biochimica Et Biophysica Acta (BBA) - General Subjects*, *1780*(6), 861–868. http://doi.org/10.1016/j.bbagen.2008.03.005

44. Orlean, P. (1990). Dolichol Phosphate Mannose Synthaste Is Required In Vivo for Glycosyl Phosphatidylinositol Membrane Anchoring, O Mannosylation, and N Glycosylation of Protein in *Saccharomyces cerevisiae. Molecular and Cellular Biology*, *10*(11), 5796–5805.

45. Girrbach, V. (2000). Structure-Function Analysis of the Dolichyl Phosphate-Mannose: Protein O-Mannosyltransferase ScPmt1p. *Journal of Biological Chemistry*, *275*(25), 19288–19296. http://doi.org/10.1074/jbc.M001771200

46. Willer, T. (2003). O-mannosyl glycans: from yeast to novel associations with human disease. *Current Opinion in Structural Biology*, *13*(5), 621–630. http://doi.org/10.1016/j.sbi.2003.09.003

47. Strahl-Bolsinger, S., Gentzsch, M., & Tanner, W. (1999). Protein O-mannosylation. *Biochimica Et Biophysica Acta*, *1426*(2), 297–307.

48. Gurcha, S. S., Baulard, A. R., Kremer, L., Locht, C., Moody, B. D., Muhlecker, W., et al. (2002). Ppm1, a novel polyprenol monophosphomannose synthase from *Mycobacterium tuberculosis. Biochem. J.*, *365*, 441–450.

49. Haselbeck, A. (1989). Purification of GDP mannose: dolichyl-phosphate O-Beta-D-mannosyltransferase from *Saccharomyces cerevisiae. Eur. J. Biochem.*, *181*, 663–668.

50. Spiro, R. G. (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, *12*(4), 43R–56R. http://doi.org/10.1093/glycob/12.4.43R

51. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, *37*(Database), D233–D238. http://doi.org/10.1093/nar/gkn663

52. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., & Henrissat, B. (2013). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, *42*(D1), D490–D495. http://doi.org/10.1093/nar/gkt1178

53. VanderVen, B. C., Harder, J. D., Crick, D. C., & Belisle, J. T. (2005). Export-Mediated Assembly of Mycobacterial Glycproteins Parallels Eukaryotic Pathways. *Science*, *309*(5736), 941–943.

54. Mahne, M., Tauch, A., Palhler, A., & Kalinowski, J. R. (2006). The Corynebacterium glutamicum gene pmt encoding a glycosyltransferase related to eukaryotic protein- O-mannosyltransferases is essential for glycosylation of the resuscitation promoting factor (Rpf2) and other secreted proteins. *FEMS Microbiology Letters*, *259*(2), 226–233. http://doi.org/10.1111/j.1574-6968.2006.00269.x

55. Strahl-Bolsinger, S., & Scheinost, A. (1999). Transmembrane topology of pmt1p, a member of an evolutionarily conserved family of protein O-mannosyltransferases. *Journal of Biological Chemistry*, *274*(13), 9068–9075.

56. Zarschler, K., Janesch, B., Pabst, M., Altmann, F., Messner, P., & Schaffer, C. (2010). Protein tyrosine O-glycosylation--A rather unexplored prokaryotic glycosylation system. *Glycobiology*, *20*(6), 787–798. http://doi.org/10.1093/glycob/cwq035

57. Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyltransferases: Structures, Functions, and Mechanisms. *Annual Review of Biochemistry*, *77*(1), 521–555. http://doi.org/10.1146/annurev.biochem.76.061005.092322

58. Otzen, D. (2012). N for AsN - O for StrOcture? A strand-loop-strand motif for prokaryotic O-glycosylation. *Molecular Microbiology*, *83*(5), 879–883. http://doi.org/10.1111/j.1365-2958.2012.07972.x

59. Posey, J. E., Shinnick, T. M., & Quinn, F. D. (2006). Characterization of the twin-arginine translocase secretion system of Mycobacterium smegmatis. *Journal of Bacteriology*, *188*(4), 1332–1340. http://doi.org/10.1128/JB.188.4.1332-1340.2006

60. Xu, C., Wang, S., Thibault, G., & Ng, D. T. W. (2013). Futile protein folding cycles in the ER are terminated by the unfolded protein O-mannosylation pathway. *Science*, *340*(6135), 978–981. http://doi.org/10.1126/science.1234055

61. Schultz, J. C., & Takayama, K. (1975). The role of mannosylphosphorylpolyisoprenol in glycoprotein biosynthesis in Mycobacterium smegmatis. *Biochimica Et Biophysica Acta*, *381*(1), 175–184.

62. Cooper, H. N., Gurcha, S. S., Nigou, J., Brennan, P. J., Belisle, J. T., Besra, G. S., & Young, D. (2002). Characterization of mycobacterial protein glycosyltransferase activity using synthetic peptide acceptors in a cell-free assay. *Glycobiology*, *12*(7), 427–434. http://doi.org/10.1093/glycob/cwf051

63. Rainczuk, A. K., Yamaryo-Botte, Y., Brammananth, R., Stinear, T. P., Seemann, T., Coppel, R. L., et al. (2012). The lipoprotein LpqW is essential for the mannosylation of periplasmic glycolipids in Corynebacteria. *Journal of Biological Chemistry*, *287*(51), 42726–42738. http://doi.org/10.1074/jbc.M112.373415

64. Korduláková, J., Gilleron, M., Mikusova, K., Puzo, G., Brennan, P. J., Gicquel, B., & Jackson, M. (2002). Definition of the first mannosylation step in phosphatidylinositol mannoside synthesis. PimA is essential for growth of mycobacteria. *Journal of Biological Chemistry*, *277*(35), 31335–31344. http://doi.org/10.1074/jbc.M204060200

65. Haselbeck, A., & Tanner, W. (1983). O-Glycosylation in *Saccharomyces cerevisiae* is initiated in the endoplasmic reticulum. *FEBS Letters*, *158*(2), 335–338.

66. Szymanski, C. M., & Wren, B. W. (2005). Protein glycosylation in bacterial mucosal pathogens. *Nature Reviews Microbiology*, *3*(3), 225–237. http://doi.org/10.1038/nrmicro1100

67. Mohiman, N., Argentini, M., Batt, S. M., Cornu, D., Masi, M., Eggeling, L., et al. (2012). The ppm Operon Is Essential for Acylation and Glycosylation of Lipoproteins in Corynebacterium glutamicum. *PLoS ONE*, *7*(9), e46225. http://doi.org/10.1371/journal.pone.0046225

68. Kämpfer, P. (2010). Actinobacteria. *Handbook of Hydrocarbon and Lipid Microbiology*. http://doi.org/10.1007/978-3-540-77587-4_133

69. Wilson, D. B. (2008). Three Microbial Strategies for Plant Cell Wall Degradation. *Annals of the New York Academy of Sciences*, *1125*(1), 289–297. http://doi.org/10.1196/annals.1419.026

70. Prasad, S., Singh, A., & Joshi, H. C. (2007). Ethanol as an alternative fuel from agricultural, industrial and urban residues. *Resources, Conservation and Recycling*, *50*(1), 1–39. http://doi.org/10.1016/j.resconrec.2006.05.007

71. Kirchner, O., & Tauch, A. (2003). Tools for genetic engineering in the amino acid-producing bacterium Corynebacterium glutamicum. *Journal of Biotechnology*, *104*(1-3), 287–299. http://doi.org/10.1016/S0168-1656(03)00148-2

72. Hermann, T. (2003). Industrial production of amino acids by coryneform bacteria. *Journal of Biotechnology*, *104*(1-3), 155–172. http://doi.org/10.1016/S0168-1656(03)00149-4

73. Gilkes, N. R., Langsford, M. L., Kilburn, D. G., R C Miller, J., & Warren, R. A. (1984). Mode of action and substrate specificities of cellulases from cloned bacterial genes. *Journal of Biological Chemistry*, *259*(16), 10455–10459.

74. Dobos, K. M., Swiderek, K., Khoo, K. H., Brennan, P. J., & Belisle, J. T. (1995). Evidence for glycosylation sites on the 45-kilodalton glycoprotein of Mycobacterium tuberculosis. *Infection and Immunity*, *63*(8), 2846–2853.

75. Espitia, C., & Mancilla, R. (1989). Identification, isolation and partial characterization of Mycobacterium tuberculosis glycoprotein antigens. *Clinical and Experimental Immunology*, *77*(3), 378–383.

76. Garbe, T., Harris, D., Vordermeier, M., Lathigra, R., Ivanyi, J., & Young, D. (1993). Expression of the Mycobacterium tuberculosis 19-kilodalton antigen in Mycobacterium smegmatis: immunological analysis and evidence of glycosylation. *Infection and Immunity*, *61*(1), 260–267.

77. González-Zamorano, M., Mendoza-Hernández, G., Xolalpa, W., Parada, C., Vallecillo, A. J., Bigi, F., & Espitia, C. (2009). Mycobacterium tuberculosisGlycoproteomics Based on ConA-Lectin Affinity Capture of Mannosylated Proteins. *Journal of Proteome Research*, *8*(2), 721–733. http://doi.org/10.1021/pr800756a

78. Mazola, Y., Chinea, G., & Musacchio, A. (2011). Integrating Bioinformatics Tools to Handle Glycosylation. *PLoS Computational Biology*, *7*(12), e1002285–7. http://doi.org/10.1371/journal.pcbi.1002285

79. Michell, S. L., Whelan, A. O., Wheeler, P. R., Panico, M., Easton, R. L., Etienne, A. T., et al. (2003). The MPB83 antigen from Mycobacterium bovis contains O-linked mannose and (1-->3)-mannobiose moieties. *Journal of Biological Chemistry*, *278*(18), 16423–16432. http://doi.org/10.1074/jbc.M207959200

80. Herrmann, J. L., Delahay, R., Gallagher, A., & Robertson, B. (2000). Analysis of post-translational modification of mycobacterial proteins using a cassette expression system. *FEBS Letters*, *473*(3), 358–362. http://doi.org/10.1016/S0014-5793(00)01553-2

81. Herrmann, J. L., O'Gaora, P., Gallagher, A., Thole, J. E., & Young, D. B. (1996). Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from Mycobacterium tuberculosis. *The EMBO Journal*, *15*(14), 3547–3554.

82. Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L., & Brunak, S. (1998). NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate Journal*, *15*(2), 115–130.

83. Córdova-Dávalos, L. E., Espitia, C., González-Cerón, G., Arreguín-Espinosa, R., Soberón-Chávez, G., & Servín-González, L. (2013). Lipoprotein N-acyl transferase (Lnt1) is dispensable for protein O-mannosylation by Streptomyces coelicolor. *FEMS Microbiology Letters*, *350*(1), 72–82. http://doi.org/10.1111/1574-6968.12298

84. Sartain, M. J., & Belisle, J. T. (2009). N-Terminal clustering of the O-glycosylation sites in the Mycobacterium tuberculosis lipoprotein SodC. *Glycobiology*, *19*(1), 38–51.

http://doi.org/10.1093/glycob/cwn102

85. Lara, M., Servín-González, L., Singh, M., Moreno, C., Cohen, I., Nimtz, M., & Espitia, C. (2004). Expression, secretion, and glycosylation of the 45- and 47-kDa glycoprotein of Mycobacterium tuberculosis in Streptomyces lividans. *Applied and Environmental Microbiology*, *70*(2), 679–685. http://doi.org/10.1128/AEM.70.2.679-685.2004

86. Sieling, P. A., Hill, P. J., & Dobos, K. M. (2008). Conserved mycobacterial lipoglycoproteins activate TLR2 but also require glycosylation for MHC class II-restricted T cell activation. *The Journal of Immunology, 180,* 5833-5842*.* http://doi.org/10.4049/jimmunol.180.9.5833

87. Torres, A., Juárez, M. D., Cervantes, R., & Espitia, C. (2001). Molecular analysis of Mycobacterium tuberculosis phosphate specific transport system in Mycobacterium smegmatis. Characterization of recombinant 38 kDa (PstS-1). *Microbial Pathogenesis*, *30*(5), 289–297. http://doi.org/10.1006/mpat.2001.0434

88. Abu-Qarn, M., Eichler, J., & Sharon, N. (2008). Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Current Opinion in Structural Biology*, *18*(5), 544–550. http://doi.org/10.1016/j.sbi.2008.06.010

89. Cain, J. A., Solis, N., & Cordwell, S. J. (2014). Beyond gene expression: The impact of protein post-translational modifications in bacteria. *Journal of Proteomics*, *97*, 265–286. http://doi.org/10.1016/j.jprot.2013.08.012

90. Szymanski, C. M. (2002). Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infection and Immunity*, *70*(4), 2242–2244. http://doi.org/10.1128/IAI.70.4.2242-2244.2002

91. Pitarque, S., Herrmann, J. L., Duteyrat, J.-L., Jackson, M., Stewart, G. R., Lecointe, F., et al. (2005). Deciphering the molecular bases of Mycobacterium tuberculosis binding to the lectin DC-SIGN reveals an underestimated complexity. *Biochem. J.*, *392*(Pt 3), 615–624. http://doi.org/10.1042/BJ20050709

92. Cambi, A., & Figdor, C. G. (2003). Dual function of C-type lectin-like receptors in the immune system. *Current Opinion in Cell Biology*, *15*(5), 539–546. http://doi.org/10.1016/j.ceb.2003.08.004

93. Horn, C., Namane, A., Pescher, P., Riviere, M., Romaine, F., Puzo, G., et al. (1999). Decreased Capacity of Recombinant 45/47-kDa Molecules (Apa) of *Mycobacterium tuberculosis* to Stimulate T Lymphocyte Responses Related to Changes in Their Mannosylation Pattern. *Journal of Biological Chemistry*, *274*(45), 32023–32030. http://doi.org/10.1074/jbc.274.45.32023

94. Ravasi, P., Peiru, S., Gramajo, H., & Menzella, H. G. (2012). Design and testing of a synthetic biology framework for genetic engineering of Corynebacterium glutamicum. *Microbial Cell Factories*, *11*(1), 1–1. http://doi.org/10.1186/1475-2859-11-147

95. Bacterial, gram positive electrotransformation (nd). In *Gene Pulser® Electroprotocols online*. Retrieved from http://biorad-ads.com/transfection_protocols/

96. Eggeling, L., Bott, M. (2005). Experiments. In L. Eggeling & O. Reyes (Eds.), *Handbook of Corynebacterium glutamicum* (535 – 566). Florida: CRC Press.

97. van der Rest, M. E., Lange, C., & Molenaar, D. (1999). A heat shock following electroporation induces highly efficient transformation of Corynebacterium glutamicum with xenogeneic plasmid DNA. *Applied*

Microbiology and Biotechnology, *52*(4), 541–545. http://doi.org/10.1007/s002530051557

98. Jang, K.-H., & Britz, M. L. (2000). Improved electrotransformation frequencies of Corynebacterium glutamicum using cell-surface mutants. *Biotechnology Letters*, *22*(7), 539–545. http://doi.org/10.1023/A:1005629224109

99. Follettie, M. T., Peoples, O. P., Agoropoulou, C., & Sinskey, A. J. (1993). Gene structure and expression of the Corynebacterium flavum N13 ask-asd operon. *Journal of Bacteriology*, *175*(13), 4096–4103.

100. Tauch, A., Kirchner, O., Löffler, B., Götker, S., Puhler, A., & Kalinowski, J. (2002). Efficient electrotransformation of corynebacterium diphtheriae with a mini-replicon derived from the Corynebacterium glutamicum plasmid pGA1. *Current Microbiology*, *45*(5), 362–367. http://doi.org/10.1007/s00284-002-3728-3

101. Liebl, W., Bayerl, A., Schein, B., Stillner, U., & Schleifer, K. H. (1989). High efficiency electroporation of intact Corynebacterium glutamicum cells. *FEMS Microbiology Letters*, *53*(3), 299–303.

102. Li, M., Niu, T., Zhang, J.-H., Su, Z., & Kong, D.-J. (2010). High Electroporation Efficiency of Corynebacterium glutamicum with Xenogeneic Plasmid DNA. *2010 4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE)*, 1–4. http://doi.org/10.1109/ICBBE.2010.5517786

103. Vertès, A. A., Inui, M., & Yukawa, H. (2005). Manipulating corynebacteria, from individual genes to chromosomes. *Applied and Environmental Microbiology*, *71*(12), 7633–7642. http://doi.org/10.1128/AEM.71.12.7633-7642.2005

104. Nešvera, J., & Pátek, M. (2011). Tools for genetic manipulations in Corynebacterium glutamicum and their applications. *Applied Microbiology and Biotechnology*, *90*(5), 1641–1654. http://doi.org/10.1007/s00253-011-3272-9

105. Kirchner, O., & Tauch, A. (2003). Tools for genetic engineering in the amino acid-producing bacterium Corynebacterium glutamicum. *Journal of Biotechnology*, *104*(1-3), 287–299. http://doi.org/10.1016/S0168-1656(03)00148-2

106. Hotta, K., Ishikawa, J., Ichihara, M., Naganawa, H., & Mizuno, S. (1988). Mechanism of increased kanamycin-resistance generated by protoplast regeneration of Streptomyces griseus. I. Cloning of a gene segment directing a high level of an aminoglycoside 3-N-acetyltransferase activity. *The Journal of Antibiotics*, *41*(1), 94–103.

107. Eikmanns, B. J., Kleinertz, E., Liebl, W., & Sahm, H. (1991). A family of Corynebacterium glutamicum/Escherichia coli shuttle vectors for cloning, controlled gene expression, and promoter probing. *Gene*, *102*(1), 93–98.

108. Jang, K. H., Pierotti, D., Kemp, G. W., & Best, G. R. (1997). Mycotic acid composition of Corynebacterium glutamicum and its cell surface mutants: effects of growth with glycine and isonicotinic acid hydrazide. *Microbiology.*

109. Pasqualoto, K. F. M., Ferreira, M. M. C., Santos-Filho, O. A., & Hopfinger, A. J. (2006). Molecular dynamics simulations of a set of isoniazid derivatives bound to InhA, the enoyl-acp reductase fromM. tuberculosis. *International Journal of Quantum Chemistry*, *106*(13), 2689–2699. http://doi.org/10.1002/qua.21055

110. Molle, V., Gulten, G., Vilchèze, C., Veyron-Churlet, R., Zanella-Cléon, I., Sacchettini, J. C., et al. (2010). Phosphorylation of InhA inhibits mycolic acid biosynthesis and growth of Mycobacterium tuberculosis. *Molecular Microbiology*, *78*(6), 1591–1605. http://doi.org/10.1111/j.1365-2958.2010.07446.x

111. Delaine, T., Bernardes-Génisson, V., Quémard, A., Constant, P., Cosledan, F., Meunier, B., & Bernadou, J. (2012). Preliminary Investigations of the Effect of Lipophilic Analogues of the Active Metabolite of Isoniazid Toward Bacterial and Plasmodial Strains. *Chemical Biology & Drug Design*, *79*(6), 1001–1006. http://doi.org/10.1111/j.1747-0285.2012.01374.x

112. Pitcher, D. G., Saunders, N. A., & Owen, R. J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. *Letters in Applied Microbiology*, *8*(4), 151–156. http://doi.org/10.1111/j.1472-765X.1989.tb00262.x

113. Lee, P. Y., Costumbrado, J., Hsu, C. Y., Kim, Y. H. Agarose Gel Electrophoresis for the Separation of DNA Fragments. *J. Vis. Exp.* (62), e3923, doi:10.3791/3923 (2012).

114. Moore, B. A., Jevons, S., & Brammer, K. W. (1979). Inhibition of transpeptidase activity in Escherichia coli by thienamycin. *Antimicrobial Agents and Chemotherapy*, *15*(6), 831–833."

115. Bera, A., Herbert, S., Jakob, A., Vollmer, W., & Götz, F. (2004). Why are pathogenic staphylococci so lysozyme resistant? The peptidoglycan O-acetyltransferase OatA is the major determinant for lysozyme resistance of Staphylococcus aureus. *Molecular Microbiology*, *55*(3), 778–787. http://doi.org/10.1111/j.1365-2958.2004.04446.x

116. Lorenz, T. C. Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. *J. Vis. Exp.* (63), e3998, doi:10.3791/3998 (2012).

117. Chomczynski, P., Mackey, K., Drews, R., & Wilfinger, W. (1997). DNAzol: a reagent for the rapid isolation of genomic DNA. *BioTechniques*, *22*(3), 550–553.

118. Brewster, J. D., & Paoli, G. C. (2013). DNA extraction protocol for rapid PCR detection of pathogenic bacteria. *Analytical Biochemistry*, *442*(1), 107–109. http://doi.org/10.1016/j.ab.2013.07.013

119. van, P. E., van, B. A., & Hays, J. P. (2008). Principles and Technical Aspects of PCR Amplification. Dordrecht, NLD: Springer. Retrieved from http://www.ebrary.com

120. Obradovic, J., Jurisic, V., Tosic, N., Mrdjanovic, J., Perin, B., Pavlovic, S., & Djordjevic, N. (2013). Optimization of PCR Conditions for Amplification of GC-Rich EGFRPromoter Sequence. *Journal of Clinical Laboratory Analysis*, *27*(6), 487–493. http://doi.org/10.1002/jcla.21632

121. Mamedov, T. G., Pienaar, E., Whitney, S. E., TerMaat, J. R., Carvill, G., Goliath, R., et al. (2008). A fundamental study of the PCR amplification of GC-rich DNA templates. *Computational Biology and Chemistry*, *32*(6), 452–457. http://doi.org/10.1016/j.compbiolchem.2008.07.021

122. Schluesener, D., Fischer, F., Kruip, J., Rögner, M., & Poetsch, A. (2005). Mapping the membrane proteome of Corynebacterium glutamicum. *Proteomics*, *5*(5), 1317–1330. http://doi.org/10.1002/pmic.200400993

123. Oehler, S., Eismann, E. R., Krämer, H., & Müller-Hill, B. (1990). The three operators of the lac operon cooperate in repression. *The EMBO Journal*, *9*(4), 973–979.

124. Bell, A., Monaghan, P., & Page, A. P. (2006). Peptidyl-prolyl cis–trans isomerases (immunophilins) and their roles in parasite biochemistry, host–parasite interaction and antiparasitic drug action. *International Journal for Parasitology*, *36*(3), 261–276. http://doi.org/10.1016/j.ijpara.2005.11.003

125. Tastet, C., Charmont, S., Chevallet, M., Luche, S., & Rabilloud, T. (2003). Structure-efficiency relationships of zwitterionic detergents as protein solubilizers in two-dimensional electrophoresis. *Proteomics*, *3*(2), 111–121. http://doi.org/10.1002/pmic.200390019

126. Henningsen, R., Gale, B. L., Straub, K. M., & DeNagel, D. C. (2002). Application of zwitterionic detergents to the solubilization of integral membrane proteins for two-dimensional gel electrophoresis and mass spectrometry. *Proteomics*, *2*(11), 1479–1488. http://doi.org/10.1002/1615-9861(200211)2:11<1479::AID-PROT1479>3.0.CO;2-A

127. Nakashima, N., Mitani, Y., & Tamura, T. (2005). Microbial Cell Factories. *Microbial Cell Factories*, *4*(1), 7–5. http://doi.org/10.1186/1475-2859-4-7

128. Lambertz, C., Garvey, M., Klinger, J., Heesel, D., Klose, H., Fischer, R., & Commandeur, U. (2014). Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. *Biotechnology for Biofuels*, *7*(1), 135. http://doi.org/10.1186/s13068-014-0135-5