Priming and Processing Glycosyltransferases in Bacterial N-linked Glycosylation Pathways

by
Vinita Lukose

B.A. Chemistry
Pomona College, 2008

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

at the

Massachusetts Institute of Technology

September 2015

Signature of Author: Signature redacted

Department of Chemistry
July 6th, 2015

Certified by: Signature redacted

Barbara Imperiali
Class of 1922 Professor of Chemistry and Biology
Thesis Supervisor

Accepted by: Signature redacted

Robert W. Field
Haslam and Dewey Professor of Chemistry
Chairman, Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Signature redacted

Catherine L. Drennan
Professor of Chemistry and Biology
Howard Hughes Medical Institute Investigator and Professor
Thesis Committee Chair

Signature redacted

Barbara Imperiali
Class of 1922 Professor of Chemistry and Biology
Thesis Supervisor

Signature redacted

JoAnne Stubbe
Novartis Professor of Chemistry and Biology

Signature redacted

Matthew Shoulders
Assistant Professor of Chemistry

Priming and Processing Glycosyltransferases in Bacterial N-linked Glycosylation Pathways

by
Vinita Lukose

Submitted to the Department of Chemistry
on July 6th, 2015 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

Abstract

Bacterial cell surfaces prominently feature a variety of complex glycoconjugates. Although these structures are very diverse, the biosynthesis of these glycoconjugates shares common themes. In particular, a priming glycosyltransferase or phosphoglycosyltransferase (PGT) initiates the biosynthesis of glycans by transferring a C1′-phosphosugar onto a polyprenol phosphate substrate. The membrane-bound polyprenol diphosphosugar product is then further elaborated by additional glycosyltransferases, flipped across the inner bacterial membrane, and ultimately used as glycosyl donor substrate in transfer to a final acceptor substrate, which, is a lipid, a protein or another glycan.

This thesis focuses on the priming GTs that catalyze the first step in glycoconjugate biosynthesis. This class of enzymes is structurally diverse and includes several distinct families of enzymes with different structures and membrane topologies. PglC, the smallest and structurally, the simplest PGT found in *Campylobacter jejuni*, represents a tractable model with which to investigate this class of enzymes. First, efforts to structurally characterize PglC are discussed, with a focus on the challenges associated with expressing, purifying and characterizing integral membrane proteins.

The next section of this thesis presents the functional characterization of PglC using a variety of biochemical techniques. In addition, bioinformatics analysis of PglC and related PGT families was employed to provide insight into critical residues required for catalytic activity. To complement this, a predicted structure of PglC was developed, based on the wealth of information contained in the sequences of homologs of PglC. Kinetic analysis of PglC was used to investigate the mechanism of the PGT reaction. Together, these studies aided in the design and evaluation of inhibitors of PglC, inspired by the nucleoside antibiotics tunicamycin and mureidomycin.

In the final chapter, the enzymes in the N-linked protein glycosylation pathway in *Campylobacter jejuni* were evaluated for their tolerance for azide-modified UDP-sugar substrates. *In vitro* experiments were employed to investigate the potential of these enzymes for executing the chemoenzymatic synthesis of the *C. jejuni* glycan with azide-modified sugars at discrete targeted positions. Attempts to metabolically label *C. jejuni* with azide-modified sugars for *in vivo* incorporation into the N-glycan are also presented.

Thesis Supervisor: Barbara Imperiali
Class of 1922 Professor of Chemistry and Biology

Acknowledgements

I would like to start by thanking my advisor, Barbara Imperiali. Her enthusiasm for science is what drew me to her lab, and I am thankful to her for introducing me to the fascinating world of sugars and for helping me appreciate the importance of studying difficult problems. I am extremely grateful for her thorough and thoughtful editing of every chapter of this thesis. Finally, I am thankful to her for bringing together a truly wonderful group of colleagues with whom I have had the pleasure of working for the past six years.

I am also thankful to my committee members for their input and support. Cathy has been a wonderful and encouraging thesis chair, and somehow always knew when I needed a pep talk during our annual meetings. I am also thankful to her for fun times with Shep. My experiences with JoAnne as a teaching assistant for 5.07 greatly shaped my understanding of biochemistry, and I thank her for her influence and guidance since then. Although Matt only joined my thesis committee recently, I am grateful for his input and for his interest in my research.

The Imperiali lab has been a wonderful place to work, and I have many colleagues to acknowledge. I am thankful to James, Mike, Marcie, Daniel, Brenda, and Wendy, for welcoming me into the lab. I am especially thankful to Jay for many helpful conversations about making heptasaccharides. Cliff and Thais were wonderful podmates in the old lab, and I have many fond memories with them. I am thankful for Angelyn's help while we overlapped in graduate school and for all her advice about the future. I greatly appreciated Elke's cheerful attitude and her sense of adventure. Thanks to Philipp for his help in our shared struggles with PglC and isoprene purifications. I am grateful to Andrew for his patience in teaching me how to do my first (and last) chemical sugar synthesis in the lab. Carsten's optimism is inspiring and I thank him for reminding me of the importance of a good attitude. I am grateful to Kasper for his calming presence in the lab, especially during late night experiments. I have enjoyed many thoughtful conversations with Julie, and I am thankful for her friendship and for her knowledge of all things brunch. Many thanks to Joris for always making sure the lab was well-stocked with candy. I am grateful to Cristy for her help with the HPLC, allowing me to get a figure for my thesis at the very last minute. Monika has been a wonderful friend and colleague and I will miss our feminist, environmentalist and political discussions. I am lucky to have overlapped with Michelle for many years in the lab and I am thankful for her company in fixing fraction collectors and for early lunches. I will always remember Austin's kindness in helping me move to a new apartment the day after he did a triathlon. I am happy that Sonya joined the lab, and I am sure that together with the rest of Team PglC, she will make big discoveries about this obstinate little protein. Thanks to the three musketeers, Silvano, Debasis and Jean-Marie for being my gym buddies and for encouraging me to take pizza breaks. I am extremely grateful to Elizabeth Fong for all her help and for everything she does to keep the lab running.

I also have to acknowledge the colleagues I worked with most closely. Meredith (MDogg) was my mentor when I first joined the lab, and I am indebted to her for everything she has taught me. She continued to mentor me long after she left the lab, and I am so touched that she flew across the country to be at my thesis defense. I have also had the opportunity to collaborate with two wonderful post docs in our lab. Garrett is possibly the nicest person I have ever met and I am honored to have had him as my podmate for three years. He is also to be credited for my new-found tolerance of country music. I was also fortunate to have worked closely with Marthe and I am thankful for everything she taught me about science and life. I am so excited for her future as a professor and I wish her, Marnix and Fenna all the best.

I am also thankful to many other people I have interacted with over the course of graduate school. First, I have to thank Jingnan and Tengfei for being wonderful classmates and friends and for sharing all the highs and lows of graduate school with me. Thanks to Marco, Michael, Ken, Yifeng and Hongik for their company all these years, preparing for qualifying exams and for fun dinners at Mulan. Women in Chemistry (WIC) has been an integral part of my MIT experience, and I am thankful to have worked with so many amazing women. Thanks to Christina, Amy, Vicky, Allena, Nootaree, and Leigh for many good memories. It was a pleasure being Co-President with Whitney for two years; she has become a close friend and I will miss her terribly.

I am thankful to my amazing and inspiring professors at Pomona College for sparking my interest in chemistry and encouraging me to pursue graduate school. Special thanks to Professors Matthew Sazinsky, E.J. Crane, Cynthia Selassie, Chuck Taylor, and Frederick Grieman, for their support and encouragement all these years. I am especially grateful to Matt ("Saz") for his help through many graduate school hurdles and for all his advice about life after graduate school. I am also thankful to Katie and Lisa for their company during my fellowship year. Pomona is truly a special place; distance and time have only made me appreciate it more.

I owe a great deal to my "real world" friends for reminding me that there is indeed a world outside of the MIT bubble. Thanks to Bobby and Palak for exploring Boston with me these past six years, and for providing distractions when I needed them. I am grateful to Julia, Shreya and Simrat for always making time for me whenever I needed to chat or vent about graduate school. Vandy has been my biggest cheerleader throughout this process and I am so grateful for her friendship. Sonali, my best Craigslist find ever, has been a wonderful roommate and I thank her for late-night chats over hot chocolate and for reminding me that it's okay to take a day off once in a while.

This thesis is dedicated to my mom, without whose support I would not have made it through graduate school.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 1

List of Abbreviations

Standard 3-letter and 1-letter codes are used for the 20 natural amino acids.
Standard 1-letter codes are used for the 4 common DNA bases

2-AB	2-aminobenzamide
AcCoA	acetyl coenzyme A
ATP	adenosine triphosphate
CEF	cell envelope fraction
CPS	capsular polysaccharide
DDM	n-dodecyl β-D-maltoside
DMSO	dimethylsulfoxide
Dol-P	dolichol phosphate
EDTA	ethlyenediaminetetraacetic acid
ER	endiplasmic reticulum
ESI-MS	electrospray ionization-mass spectrometry
GB1	protein G B1 domain
GPT	GlcNAc-1-P transferase
GT	glycosyltransferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IC_{50}	half maximal inhibitory concentration
IPTG	isopropyl β-D-1-thiogalactopyranoside
K_i	inhibition constant
K_M	Michaelis constant
LB	Luria-Bertani broth
LMNG	lauryl maltose neopentyl glycol
LOS	lipooligosaccharide
MGL	human macrophage galactose-type lectin
NDP	nucleotide disphosphate
Ni-NTA	nickel nitrilotriacetic acid
OGNG	octyl glucose neopentyl glycol
Pgl	Protein glycosylation
PglB(*Ng*)	Bifunctional enzyme in N. gonorrhoeae with a C-terminal PGT domain and an N-terminal acetyltransferase domain
PglC(*Hp*)	PglC from *Helicobacter pullorum*
PGT	phosphoglycosyltransferase
PIRS	polyisoprenol recognition sequence
PSUP	pure solvent upper phase
SUMO	Small Ubiquitin-like Modifier
TMHMM	Tied Hidden Markov Model for transmembrane prediction
UDP	uridine diphosphate
UDP-4Azido-diNAcBac	uridine diphospho- *N*-acetyl,*N'*-azidoacetylbacillosamine
UDP-diNAcBac	uridine diphospho- *N*,*N'*-diacetylbacillosamine

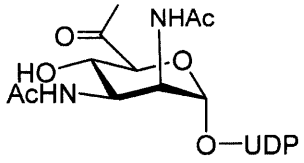
UDP-GalNAc	uridine diphospho-*N*-acetyl-D-galactosamine
UDP-GalNAz	uridine diphospho-*N*-azidoacetyl-D-galactosamine
UDP-GlcNAc	uridine diphospho-*N*-acetyl-D-glucosamine
UDP-Glucose	uridine diphospho-*N*-acetyl-D-glucose
UMP	uridine monophosphate
Und-P	undecaprenol phosphate
Und-PP	undecaprenol diphosphate
UTP	uridine triphosphate

Chapter 1

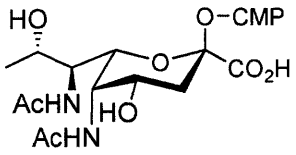
Phosphoglycosyltransferases: Gatekeepers of Glycan Biosynthesis at the Membrane Interface

Introduction

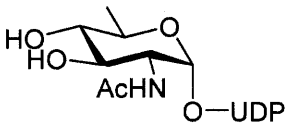
Bacterial cell-surface glycoconjugates play critical roles in establishing interactions with host cells. Key cell-surface glycans include those presented on capsular polysaccharide (CPS), lipopolysaccharide (LPS), lipooligosaccharide, peptidoglycan (in the case of Gram-positive bacteria), and glycoproteins. Investigating and controlling the interactions between these complex molecules and their binding partners requires a detailed understanding of the pathways and processes involved in the biosynthesis of these glycoconjugates. The bacterial cell-surface glycome presents additional challenges for study due to the presence of highly modified "exotic" carbohydrates embedded within diverse glycan architectures. For example, as shown in Figure 1-1, standard sugars such as UDP-GlcNAc can be enzymatically modified by processes such as deoxygenation, amination, and acylation.



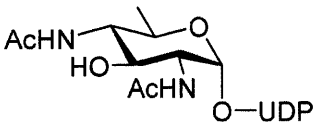
UDP-Man2,3NAcA
Pseudomonas aeruginosa,
Bordatella pertusis



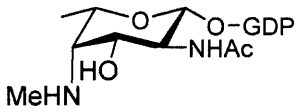
CMP-Pseudaminic acid
Helicobacter pylori,
Campylobacter jejuni,
Pseudomonas aeruginosa



UDP-QuiNAc
Shigella boydi,
Vibrio cholera



UDP-diNAcBac
Campylobacter jejuni,
Neisseria gonorrhoea,
Neisseria meningitides



GDP-Fuc2NAc4NMe
Bordatella pertusis

Figure 1-1 Examples of unique carbohydrates found in prokaryotes.

A shared feature in biosynthesis of these complex glycans is the family of enzymes that catalyzes the synthesis of the first membrane-associated intermediate in glycan assembly, called phosphoglycosyltransferases (PGTs). These integral membrane proteins are also referred to as "priming" or "initiating" glycosyltransferases due to their pivotal roles in commencing the biosynthesis of glycoconjugates. The reaction involves the transfer of a C1′-phosphosugar from a nucleotide diphosphate sugar donor to a polyprenol-phosphate substrate, forming the first membrane-linked intermediate in the synthesis of the full glycan. The products of PGT reactions are then elaborated by glycosyltransferases (GTs), to produce glycans that play key roles in bacterial and eukaryotic physiology. Figure 1-2 illustrates glycoconjugates found in prokaryotes; the lipopolysaccharide (LPS) and lipooligosaccharides (LOS) are characteristic of Gram-negative bacteria, while lipotechoic acid (LTA) and wall techoic acid (WTA) are unique to Gram-positive bacteria.

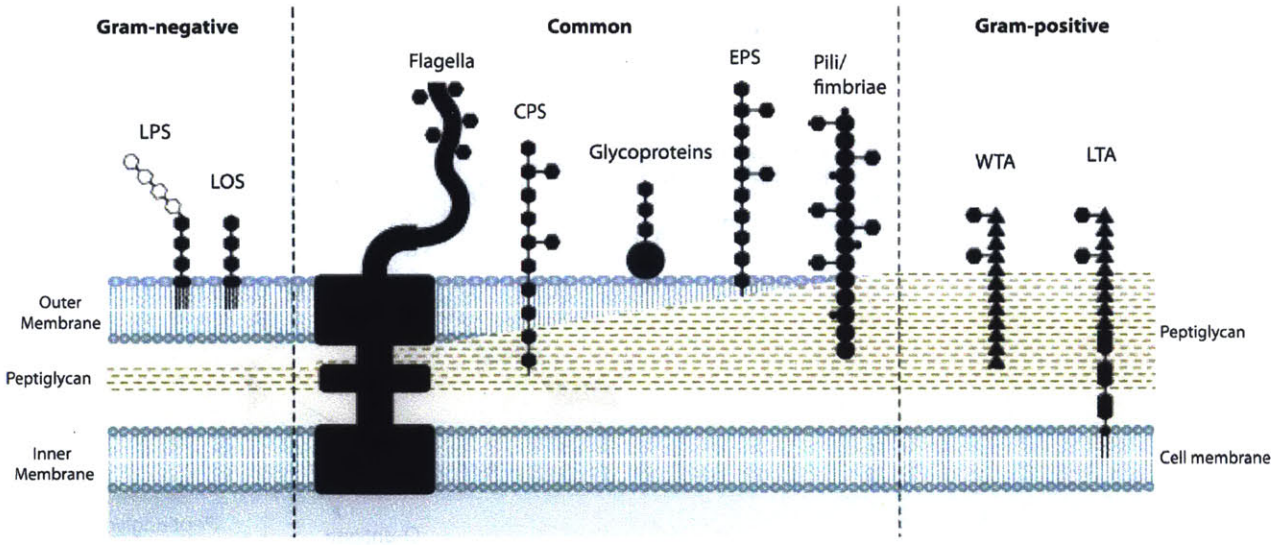


Figure 1-2 Examples of glycoconjugates on the cell surface of bacteria, the biosynthesis of many of which are initiated by PGTs. Structures unique to Gram-negative and Gram-positive bacteria as well as shared structures are shown. Proteins are represented by round dots, hexagons represent carbohydrates, and triangles represent ribitol phosphate or glycerol phosphate moieties. Taken from Tytgat et al[1].

This chapter presents an overview of the different families of PGTs and covers recent developments made in the characterization of this important class of enzymes. The centrality of PGTs in bacterial physiology is discussed as well as the practical challenges associated with studying PGTs.

Common Themes in Glycan Synthesis at the Membrane Interface

The biosynthesis of bacterial cell-surface glycoconjugates such as lipopolysaccharides and glycoproteins share many key steps. First, glycan assembly is initiated by the transfer of a phospho-sugar from a nucleotide diphosphate sugar donor to undecaprenol phosphate (Figure 1-3). Next, the glycan that has been biosynthesized on the undecaprenol diphosphate carrier is translocated across the cell membrane. Finally, the glycan is transferred *en bloc* to protein substrates or lipids, to form the final glycoconjugate products.

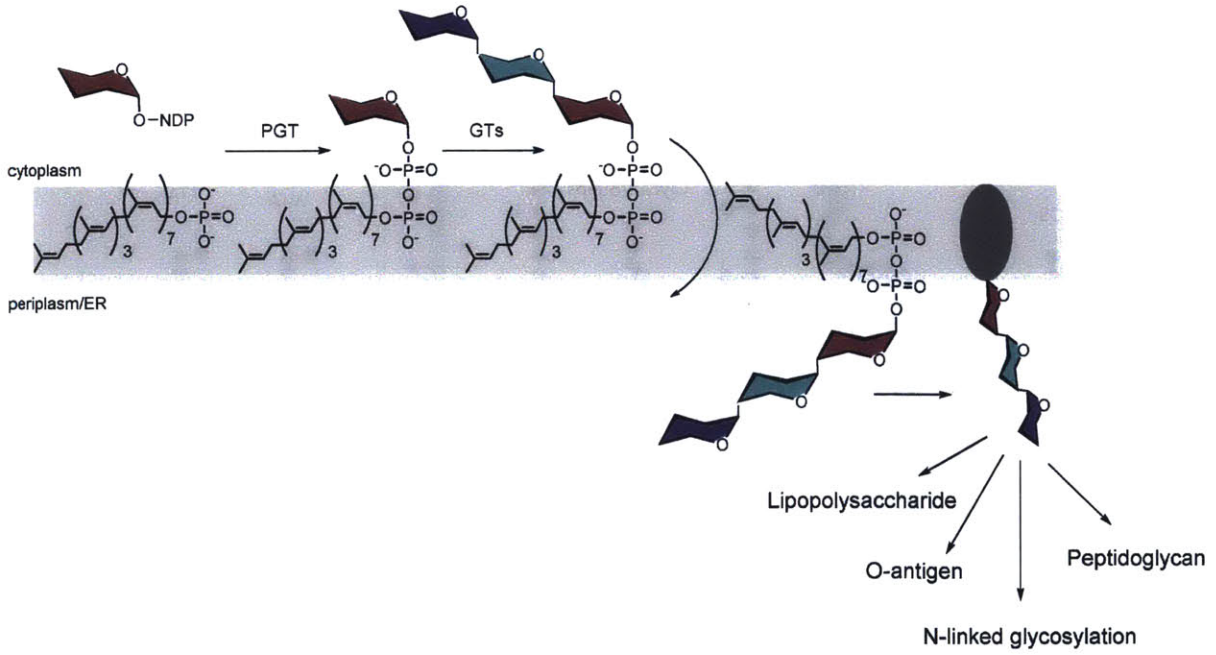


Figure 1-3 Phosphoglycosyltransferases (PGTs) initiate the biosynthesis of glycoconjugates by forming Und-PP-linked sugars that are elaborated by glycosyltransferases (GTs). The glycoconjugate is flipped into the periplasm or endoplasmic reticulum and transferred to an acceptor, which participates in a variety of biological processes.

Glycan intermediates in these processes are assembled on polyprenol phosphates, which serve as membrane anchors. Polyprenol phosphates are essential substrates for important cellular processes listed above, and are classified based on the saturation of the bond at the α-isoprene unit. Undecaprenol and related polyprenols of different lengths contain a linear polymer of unsaturated isoprene units, while dolichols contain a saturated isoprene unit in the α-position of the polymer (Figure 1-4). Bacteria and plants typically contain polyprenols with unsaturated isoprene units, while archaea and eukaryotes use dolichols.

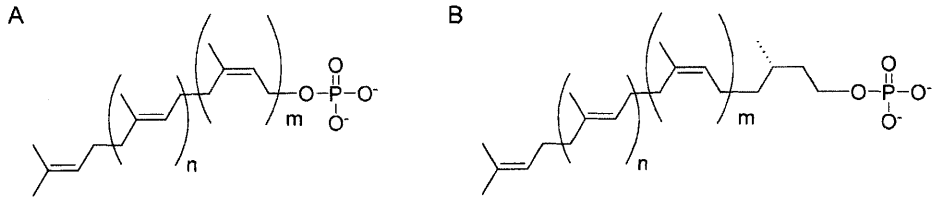


Figure 1-4 Structures of polyprenol phosphate substrates of PGTs. (A) Unsaturated polyprenol phosphate and (B) Dolichol phosphate. The length of the polyprenol and the number of isoprene units varies. 'm' denotes the number of *cis* isoprene units and 'n' is the number of *trans* isoprene units. The value of 'm' range from 12 - 37 in eukaryotes, and the value of 'n' ranges between 2-3 in both prokaryotes and eukaryotes.

In eukaryotic N-linked glycosylation, polyprenols serve an additional role beyond acting as the carrier for glycan assembly. In this case, polyprenol phosphate monosaccharides are synthesized on the cytoplasmic surface of ER membranes using nucleotide diphosphate (NDP) sugar donors.[2] An important difference with these polyprenol-linked sugars is that nucleophilic attack occurs at the anomeric position of the sugar; thus, the carbohydrates are linked to polyprenols by a monophosphate bond and not a diphosphate bond, as in the products of PGT reactions. The polyprenol phosphate-linked sugars are translocated across the endoplasmic reticulum (ER) membrane and serve as glycosyl donors in the ER lumen where nucleotide diphosphate sugars are not available (Figure 1-5).

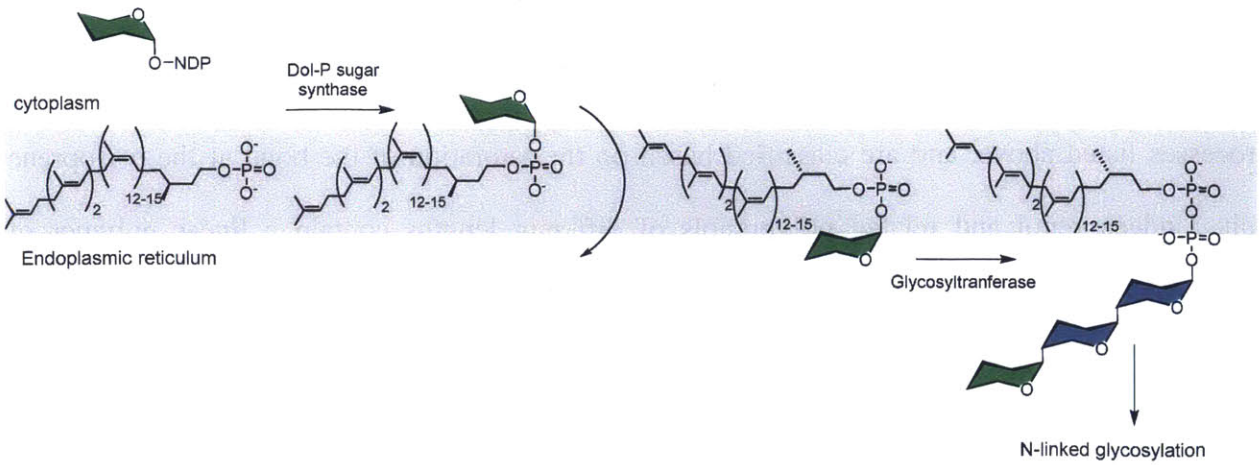


Figure 1-5 Synthesis of dolichol phosphate sugar donors in the endoplasmic reticulum.

Classification of Phosphoglycosyltransferases

Phosphoglycosyltransferases are functionally classified into polyisoprenyl-phosphate hexose-1-phosphate transferases (PHPT) and polyisoprenyl-phosphate *N*-acetylaminosugar-1-phosphate transferases (PNPT) based on the structure of their sugar substrates.[3,4] PHPT enzymes generate Und-PP-hexose products, while PNPT enzymes afford Und-PP-HexNAc products (Figure 1-6). Classification of PGTs into these two families based on sugar substrates was initially helpful in order to predict the function of enzymes. However, as more PGTs with increasingly diverse substrates are discovered, this relationship no longer seems to hold. Thus, it has been suggested that the designations of these families may require a revision.[5]

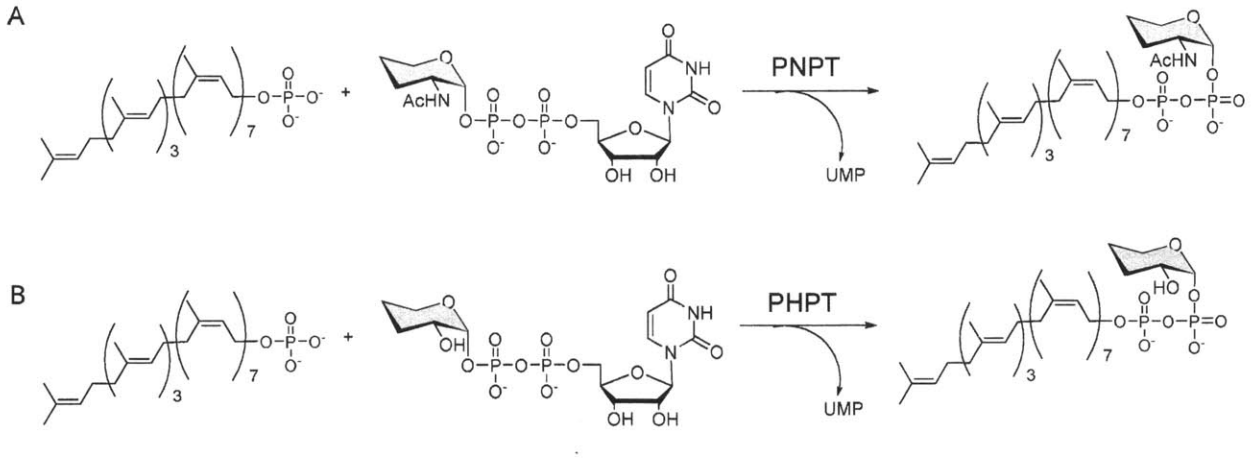


Figure 1-6 Reactions catalyzed by PGTs. (A) The PNPT family (B) The PHPT family.

PGTs can also be classified according to their membrane topologies and number of transmembrane helices, which can vary greatly among families (Figure 1-7). The WecA family contains integral membrane proteins with 11 transmembrane helices. This family of PGTs transfers GlcNAc-1-phosphate to a polyprenol phosphate acceptor. Bacterial members of this family initiate O-antigen biosynthesis, and transfer the phospho-sugar to undecaprenol phosphate,[6] while eukaryotic members such as GlcNAc-1-phosphate transferase (GPT), which initiates N-linked glycosylation in mammals, transfer the phospho-sugar to dolichol phosphate.[7] Next, the MraY family of enzymes comprises membrane proteins with 10 transmembrane helices.[8] Members of this family catalyze the transfer of 1-phospho-MurNAc-pentapeptide (L-Ala-γ-D-Glu-diaminopimelic acid/L-Lys-D-Ala-D-Ala) to undecaprenol phosphate, initiating the biosynthesis of peptidoglycan. In contrast to these multipass membrane proteins, enzymes in the PglC family have a single transmembrane domain and a globular cytoplasmic domain, and initiate processes such as N-linked glycosylation in bacteria.[9] The PglB(*Ng*) family of proteins are bifunctional, with a C-terminal PGT domain and an N-terminal acetyltransferase domain.

Proteins in this family, for example, initiate O-linked glycosylation in bacteria, which occurs through the stepwise assembly of glycan on a polyprenol diphosphate intermediate followed by en bloc transfer. In contrast many other identified O-linked glycosylation systems, particularly in eukaryotes, occurs by transferring single saccharide units from nucleotide or polyprenyl-phosphate activated glycan donors onto protein substrates.[10] The WbaP family contains proteins with five transmembrane domains and initiate O-antigen synthesis in bacteria.[11] The C-terminal domain of WbaP, containing one transmembrane helix and the globular domain, is sufficient for activity. [12]

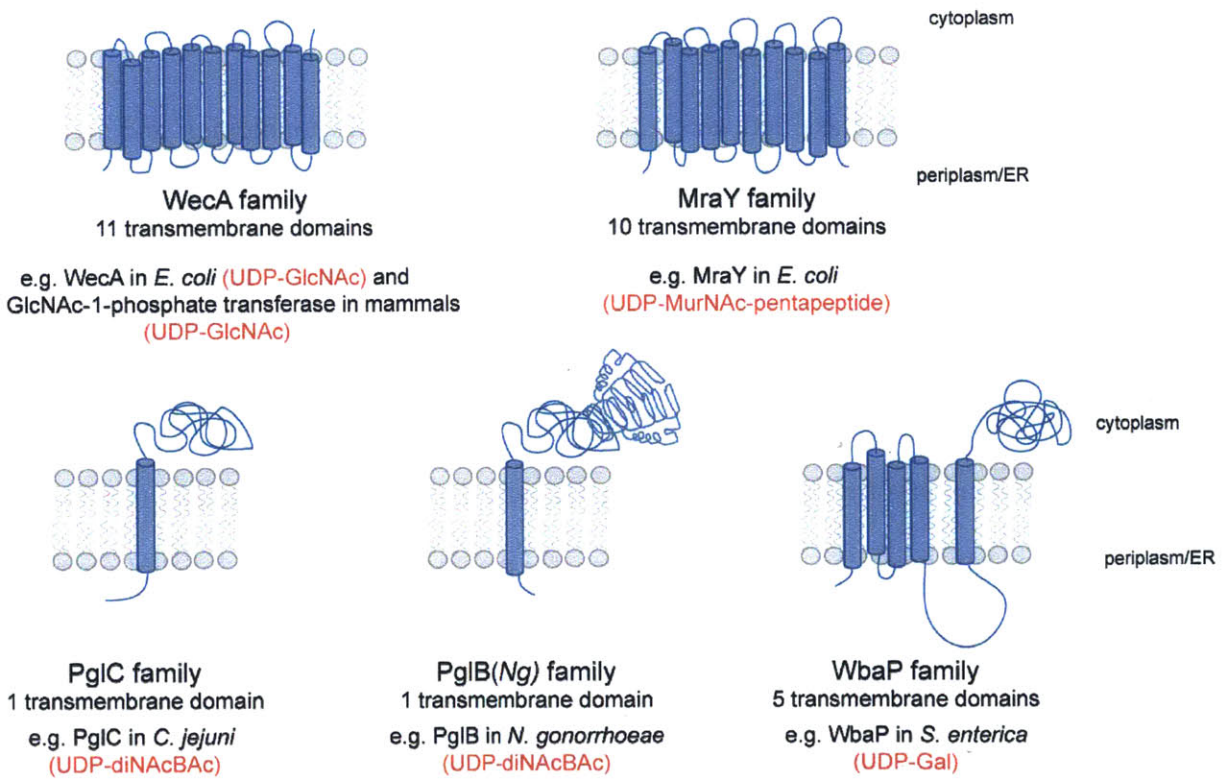


Figure 1-7 Families of PGTs classified by membrane topology. Sugar substrates of examples of each family are shown in red.

Bioinformatics analysis of PGTs has revealed conserved signature sequences in loops between the transmembrane helices in the WecA and MraY families. The significance of these motifs was determined by mutating conserved residues and analyzing the activity of the resulting mutant proteins. For example, both families have a conserved DDXXD motif, in which the adjacent aspartate residues are believed to be important for binding the metal ion cofactor.[6,13] A recent crystal structure of MraY supports this hypothesis.[8] Both families also have conserved motifs with multiple histidine residues (HIHH in the WecA family, and MAPIHHHFEL in the MraY family), which are thought to play a role in recognition of the carbohydrate substrate, although the exact interactions between these motifs and the substrates have not been established.[14,15] Similar studies with the C-terminal domain of proteins in the WbaP family of enzymes have revealed residues important for catalytic activity.[12] Additionally, bioinformatics studies and functional analysis described in Chapter Four have provided insight into residues required for the catalytic activity of PglC.

Overexpression and Purification of Phosphoglycosyltransferases

Phosphoglycosyltransferases have great potential as antibiotic targets due to the significance of their glycoconjugate products, particularly in pathogenic bacteria.[1] However, PGTs are integral membrane proteins, and are particularly intractable to heterologous overexpression. Obtaining pure, stable protein remains a significant challenge in the characterization of these enzymes, as many of the steps require empirical optimization (Figure 1-8). First, overexpression of membrane proteins in host organisms such as *Escherichia coli* results in the accumulation of large amounts of a non-native protein in cell membranes, which can be cytotoxic.[16] Additionally, since integral membrane proteins contain many hydrophobic domains, they are extremely prone to aggregation at many steps during the course of overexpression and

purification. The most destabilizing purification step is often the extraction of proteins from native membranes and subsequent solubilization into detergent micelles. This step is made more complicated by the fact that the best detergent for extraction may not necessarily be the detergent which affords the most active form of the protein. In these cases, a detergent exchange step is required to transfer the protein into micelles of the preferred detergent for activity. This is illustrated in reported strategies to purify WecA, where *n*-dodecyl β-D-maltoside (DDM) was found to be the most efficient detergent for extraction from the membrane fraction, but resulted in the lowest levels of activity when compared to Tween 20, Triton X-100, and *N*-lauryl sarcosine. Triton X-100 yielded the highest activity and was thus used as the detergent of choice in activity assays.[17]

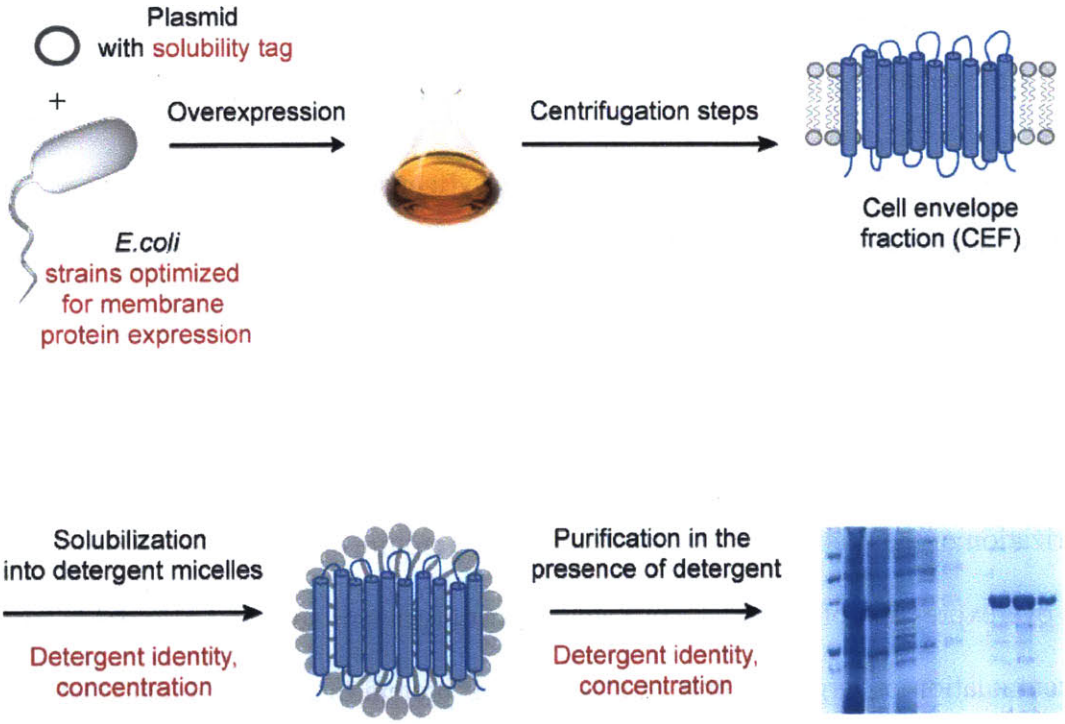


Figure 1-8 Scheme for the heterologous overexpression and purification of membrane proteins. Optimization points in the process are described in red.

Due to the complications encountered in obtaining pure, stable protein, many of the initial studies to characterize PGTs were performed with crude membranes or detergent-solubilized membrane fractions. Although these studies have provided a great deal of insight into this family of enzymes, because the exact quantity and purity of the proteins are unknown, precise quantitative data could not be obtained from these studies. The difficulties associated with obtaining stable protein have impeded crystallography studies with PGTs, and to date MraY is the only PGT with a solved crystal structure.

Phosphoglycosyltransferase Substrate Specificity

The specificity of PGTs for their isoprene substrates has been tested by varying the carbon chain length of the substrates. The WecA family requires a polyprenol phosphate with a minimum of 35 carbons for activity, and shows reduced activity when tested with polyprenols longer than the native C_{55} substrate.[17] Interestingly, WecA appears to achieve 60% of the activity compared to undecaprenol-phosphate when tested with dolichol (C_{55}) phosphate.[17] MraY shows even broader substrate specificity, forming products with dolichol (C_{40}) phosphate, isoprenol (C_{35}) phosphates, and even water-soluble prenyl substrates with as few as 10 carbons; however, the shorter prenol-diphosphate-linked substrates are reported to show much slower turnover.[18] Analysis of the substrate specificity of PglC reveals that the double-bond geometry and α-unsaturation at the alcohol terminus of the polyisoprenols are more important for substrate recognition than the number of isoprene units.[19] Studies with the C-terminal domain of WbaP, which contains the active site and strongly resembles the architecture of PglC, demonstrated that this domain appears to be very selective for undecaprenol phosphate, and is unable to tolerate polyprenol-linked substrates with fewer than 50 carbons (10 isoprene units).[20] Both PglC and the C-terminal domain of WbaP show minimal activity with longer dolichol phosphates.[19,20]

Phosphoglycosyltransferases: Proposed Reaction Mechanisms

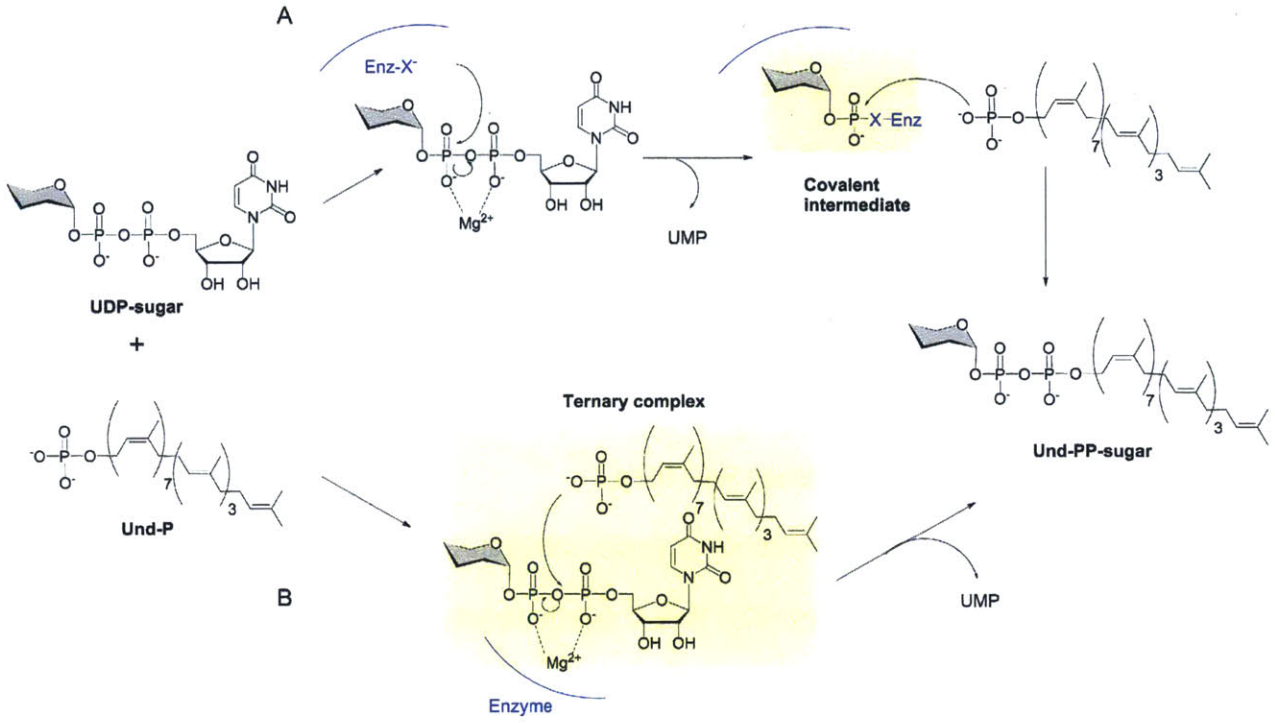


Figure 1-9 Phosphoglycosyltransferase reaction mechanisms. (A) Ping Pong mechanism, with formation of a covalent enzyme intermediate. (B) Bi Bi random mechanism, with formation of a ternary complex.

The phosphoglycosyltransferase reaction is isoenergetic, involving the cleavage of a diphosphate bond in the UDP-sugar substrate, and formation of a diphosphate bond with the undecaprenol phosphate product. The reaction proceeds via nucleophilic attack on the nucleotide-diphosphate-sugar donor and subsequent transfer of a sugar-1-phosphate to a polyprenol phosphate acceptor. Two different mechanisms can be envisioned for this bisubstrate reaction – a Bi Bi Ping-Pong mechanism with the formation of a covalent intermediate or a Bi Bi random or ordered mechanism with intermediate formation of a ternary complex (Figure 1-9). Importantly, both mechanisms result in retention of stereochemistry at the anomeric center of the sugar since the reaction occurs at the adjacent phosphate center. Kinetic analysis of the

eukaryotic GlcNAc-1-P transferase (GPT) suggests a Bi Bi random mechanism for this enzyme, with formation of a ternary complex.[21] In contrast, studies with a crude solubilized membrane fraction of MraY suggest that the transfer MurNAc-pentapeptide occurs via the formation of an acyl phosphate intermediate with an aspartate residue of the enzyme.[13] To date, none of the reported kinetic analysis of PGT reactions to distinguish between these mechanisms has been performed with pure protein.

Inhibition of PGTs by Nucleoside Antibiotics

Nucleoside antibiotics are promising compounds in the quest for new antimicrobial drugs, with the rise of bacterial strains resistant to current antibiotics. Examples of nucleoside antibiotics include tunicamycin, mureidomycin and liposidomycin (Figure 1-10), which are isolated from the fermentation broth of various strains of *Streptomyces.*[22–24]

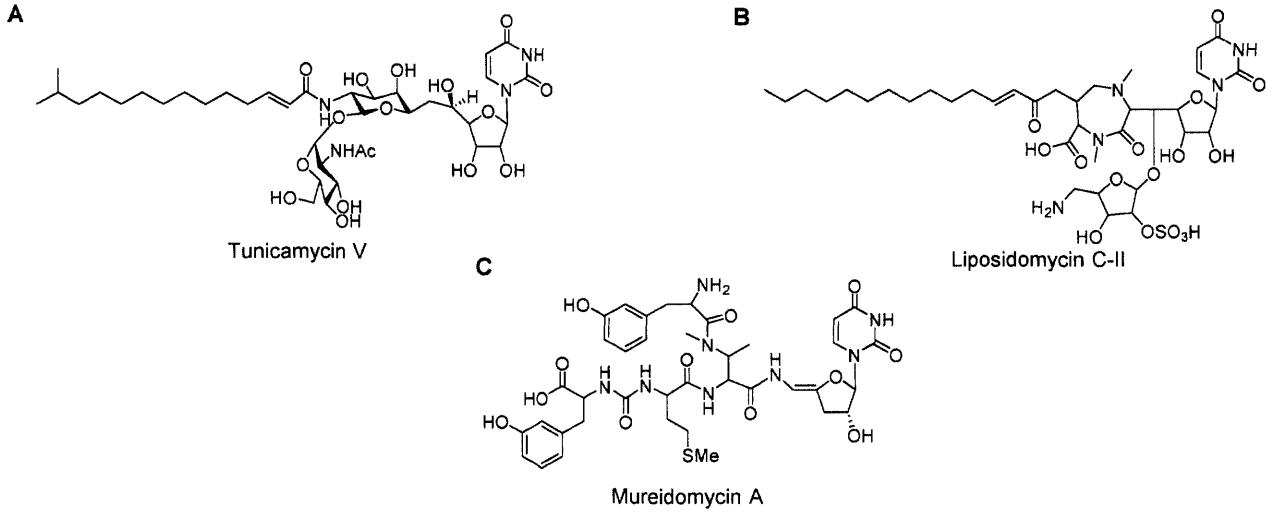


Figure 1-10 Structures of nucleoside antibiotics. (A) Tunicamycin V, (B) Liposidomycin C-II, (C) Mureidomycin A.

Nucleoside antibiotics are potent inhibitors of PGTs, and a great deal of effort has gone into exploring the mechanisms of inhibition by these diverse compounds. In one study, Brandish

et al compared the inhibition of MraY by tunicamycin, mureidomycin and liposidomycin, and found that although these three classes of compounds shared some common structural features, in particular the uridine fragment, the inhibitory properties of these antibiotics varied greatly.[25] For example, tunicamycin was found to be competitive for the UDP-MurNAc pentapeptide substrate, but not the polyprenol phosphate substrate even though it has a fatty acid component that could bind in the active site, while mureidomycin was found to be competitive for both substrates. Additional studies have been performed to determine IC_{50} and K_i values of these compounds for different classes of PGTs (Table 1-1). Many of these studies have provided insight into the modes of inhibition by these natural products, and have served as inspiration in the design of synthetic inhibitors against PGT targets.[26–28] Again, it is important to note that many of these studies were performed with either crude membrane fractions or detergent-solubilized fractions containing unpurified enzyme.

Table 1-1 Summary of IC_{50} and K_i values determined for different PGTs and nucleoside antibiotics. Values that are not available are shown as n/a. References are shown as superscripts for each reported value.

	Tunicamycin		Liposidomycin		Mureidomycin	
	IC_{50}	K_i	IC_{50}	K_i	IC_{50}	K_i
WecA	11 nM[17]	n/a	n/a	n/a	n/a	n/a
GPT	7 nM[21]	n/a	396 μM[29]	n/a	117 μM[30]	n/a
MraY	12 μM[17]	550 nM[25]	29 pM[29]	n/a	0.04 μM[31]	36 nM[25]

Conclusions

Phosphoglycosyltransferases are an extremely important class of enzymes, initiating key glycan assembly processes in both prokaryotes and eukaryotes. Despite their significance and potential as antibiotic targets, this class of enzymes has been underexplored, largely due to difficulties in expressing and purifying these integral membrane proteins. Nonetheless, the advancement of technologies such as Nanodiscs[32] and lipid superstructures[33] as alternative platforms for stable expression of membrane proteins in a more native-like environment provide promising opportunities for the characterization of PGTs.

The structural diversity observed among bacterial phosphoglycosyltransferases is an example of convergent evolution, where integral membrane proteins with vastly different structures catalyze the same basic reaction. From a biochemical perspective, the structural variation raises important questions about how some PGTs use residues on cytoplasmic loops for the same purposes that other PGTs use residues in cytoplasmic globular domains. The diversity in membrane topologies of this class of enzymes also raises questions about the roles of the transmembrane domains in the mechanisms of these reactions, especially with respect to binding to the membrane-bound polyprenol phosphate substrate. Bioinformatics analysis and deletion studies have provided insight into some of these questions[11,34,35], but there are many questions still to be answered. Fortunately, advances in full-genome sequencing have increased the pool of data available for such types of analysis, and has demonstrated that the PGTs discussed in this chapter are representative of thousands of homologs in other pathways.

Finally, PGTs offer promising opportunities in the field of glycoengineering. Bacterial genes encoding enzymes responsible for the synthesis and processing of polyprenol-diphosphate linked glycans are generally clustered in loci in the bacterial genome. Using bioinformatics

analysis, these gene clusters can be identified[36] and expressed heterologously in more tractable bacterial expression systems for large-scale biosynthesis of complex glycoconjugates. Thus, a thorough characterization of PGTs and related glycosyltransferases of a given pathway from pathogenic bacteria could provide access to a diverse set of glycoconjugates with a myriad of therapeutic applications.

References

(1) Tytgat, H. L. P.; Lebeer, S. The Sweet Tooth of Bacteria: Common Themes in Bacterial Glycoconjugates. *Microbiol. Mol. Biol. Rev.* **2014**, *78* (3), 372–417.

(2) Lommel, M.; Strahl, S. Protein O-Mannosylation: Conserved from Bacteria to Humans*. *Glycobiology* **2009**, *19* (8), 816–828.

(3) Valvano, M. A. Export of O-Specific Lipopolysaccharide. *Front. Biosci.* **2003**, *8*, s452–s471.

(4) Price, N. P.; Momany, F. A. Modeling Bacterial UDP-HexNAc: Polyprenol-P HexNAc-1-P Transferases. *Glycobiology* **2005**, *15* (9), 29R – 42R.

(5) Hug, I.; Feldman, M. F. Analogies and Homologies in Lipopolysaccharide and Glycoprotein Biosynthesis in Bacteria. *Glycobiology* **2011**, *21* (2), 138–151.

(6) Lehrer, J.; Vigeant, K. A.; Tatar, L. D.; Valvano, M. A. Functional Characterization and Membrane Topology of Escherichia Coli WecA, a Sugar-Phosphate Transferase Initiating the Biosynthesis of Enterobacterial Common Antigen and O-Antigen Lipopolysaccharide. *J. Bacteriol.* **2007**, *189* (7), 2618–2628.

(7) Lehrman, M. A. Biosynthesis of N-Acetylglucosamine-P-P-Dolichol, the Committed Step of Asparagine-Linked Oligosaccharide Assembly. *Glycobiology* **1991**, *1* (6), 553–562.

(8) Chung, B. C.; Zhao, J.; Gillespie, R. A.; Kwon, D.-Y.; Guan, Z.; Hong, J.; Zhou, P.; Lee, S.-Y. Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis. *Science* **2013**, *341* (6149), 1012–1016.

(9) Glover, K. J.; Weerapana, E.; Chen, M. M.; Imperiali, B. Direct Biochemical Evidence for the Utilization of UDP-Bacillosamine by PglC, an Essential Glycosyl-1-Phosphate Transferase in the Campylobacter Jejuni N-Linked Glycosylation Pathway. *Biochemistry* **2006**, *45* (16), 5343–5350.

(10) Hartley, M. D.; Morrison, M. J.; Aas, F. E.; Børud, B.; Koomey, M.; Imperiali, B. Biochemical Characterization of the O-Linked Glycosylation Pathway in Neisseria Gonorrhoeae Responsible for Biosynthesis of Protein Glycans Containing N,N'-Diacetylbacillosamine. *Biochemistry* **2011**, *50* (22), 4936–4948.

(11) Saldías, M. S.; Patel, K.; Marolda, C. L.; Bittner, M.; Contreras, I.; Valvano, M. A. Distinct Functional Domains of the Salmonella Enterica WbaP Transferase That Is Involved in the Initiation Reaction for Synthesis of the O Antigen Subunit. *Microbiology* **2008**, *154* (2), 440–453.

(12) Patel, K. B.; Furlong, S. E.; Valvano, M. A. Functional Analysis of the C-Terminal Domain of the WbaP Protein That Mediates Initiation of O Antigen Synthesis in Salmonella Enterica. *Glycobiology* **2010**, *20* (11), 1389–1401.

(13) Lloyd, A. J.; Brandish, P. E.; Gilbey, A. M.; Bugg, T. D. H. Phospho-N-Acetyl-Muramyl-Pentapeptide Translocase from Escherichia Coli: Catalytic Role of Conserved Aspartic Acid Residues. *J. Bacteriol.* **2004**, *186* (6), 1747–1757.

(14) Anderson, M. S.; Eveland, S. S.; Price, N. P. Conserved Cytoplasmic Motifs That Distinguish Sub-Groups of the Polyprenol phosphate:N-Acetylhexosamine-1-Phosphate Transferase Family. *FEMS Microbiol. Lett.* **2000**, *191* (2), 169–175.

(15) Amer, A. O.; Valvano, M. A. Conserved Amino Acid Residues Found in a Predicted Cytosolic Domain of the Lipopolysaccharide Biosynthetic Protein WecA Are Implicated in the Recognition of UDP-N-Acetylglucosamine. *Microbiology* **2001**, *147* (11), 3015–3025.

(16) Wagner, S.; Baars, L.; Ytterberg, A. J.; Klussmeier, A.; Wagner, C. S.; Nord, O.; Nygren, P.-Å.; Wijk, K. J. van; Gier, J.-W. de. Consequences of Membrane Protein Overexpression in Escherichia Coli. *Mol Cell Proteomics* **2007**, *6* (9), 1527–1550.

(17) Al-Dabbagh, B.; Mengin-Lecreulx, D.; Bouhss, A. Purification and Characterization of the Bacterial UDP-GlcNAc:Undecaprenyl-Phosphate GlcNAc-1-Phosphate Transferase WecA. *J Bacteriol* **2008**, *190* (21), 7141–7146.

(18) Breukink, E.; van Heusden, H. E.; Vollmerhaus, P. J.; Swiezewska, E.; Brunner, L.; Walker, S.; Heck, A. J. R.; de Kruijff, B. Lipid II Is an Intrinsic Component of the Pore Induced by Nisin in Bacterial Membranes. *J. Biol. Chem.* **2003**, *278* (22), 19898–19903.

(19) Chen, M. M.; Weerapana, E.; Ciepichal, E.; Stupak, J.; Reid, C. W.; Swiezewska, E.; Imperiali, B. Polyisoprenol Specificity in the Campylobacter Jejuni N-Linked Glycosylation Pathway. *Biochemistry* **2007**, *46* (50), 14342–14348.

(20) Patel, K. B.; Ciepichal, E.; Swiezewska, E.; Valvano, M. A. The C-Terminal Domain of the Salmonella Enterica WbaP (UDP-galactose:Und-P Galactose-1-Phosphate Transferase) Is Sufficient for Catalytic Activity and Specificity for Undecaprenyl Monophosphate. *Glycobiology* **2012**, *22* (1), 116–122.

(21) Keller, R. K.; Boon, D. Y.; Crum, F. C. N-Acetylglucosamine-1-Phosphate Transferase from Hen Oviduct: Solubilization, Characterization, and Inhibition by Tunicamycin. *Biochemistry* **1979**, *18* (18), 3946–3952.

(22) Takatsuki, A.; Arima, K.; Tamura, G. Tunicamycin, a New Antibiotic. I. Isolation and Characterization of Tunicamycin. *J. Antibiot.* **1971**, *24* (4), 215–223.

(23) Isono, F.; Katayama, T.; Inukai, M.; Haneishi, T. Mureidomycins A-D, Novel Peptidylnucleoside Antibiotics with Spheroplast Forming Activity. III. Biological Properties. *J. Antibiot.* **1989**, *42* (5), 674–679.

(24) Isono, K.; Uramoto, M.; Kusakabe, H.; Kimura, K.; Isaki, K.; Nelson, C. C.; McCloskey, J. A. Liposidomycins: Novel Nucleoside Antibiotics Which Inhibit Bacterial Peptidoglycan Synthesis. *J. Antibiot.* **1985**, *38* (11), 1617–1621.

(25) Brandish, P. E.; Kimura, K. I.; Inukai, M.; Southgate, R.; Lonsdale, J. T.; Bugg, T. D. Modes of Action of Tunicamycin, Liposidomycin B, and Mureidomycin A: Inhibition of Phospho-N-Acetylmuramyl-Pentapeptide Translocase from Escherichia Coli. *Antimicrob. Agents Chemother.* **1996**, *40* (7), 1640–1644.

(26) Gentle, C. A.; Bugg, T. D. H. Role of the Enamide Linkage of Nucleoside Antibiotic Mureidomycin A: Synthesis and Reactivity of Enamide-Containing Analogues. *J. Chem. Soc., Perkin Trans. 1* **1999**, No. 10, 1279–1286.

(27) Nigel I Howard, T. D. H. B. Synthesis and Activity of 5′-Uridinyl Dipeptide Analogues Mimicking the Amino Terminal Peptide Chain of Nucleoside Antibiotic Mureidomycin A. *Bioorganic & medicinal chemistry* **2003**, *11* (14), 3083–3099.

(28) Auberger, N.; Frlan, R.; Al-Dabbagh, B.; Bouhss, A.; Crouvoisier, M.; Gravier-Pelletier, C.; Le Merrer, Y. Synthesis and Biological Evaluation of Potential New Inhibitors of the Bacterial Transferase MraY with a B-Ketophosphonate Structure. *Org. Biomol. Chem.* **2011**, *9* (24), 8301–8312.

(29) Kimura, K.; Ikeda, Y.; Kagami, S.; Yoshihama, M.; Suzuki, K.; Osada, H.; Isono, K. Selective Inhibition of the Bacterial Peptidoglycan Biosynthesis by the New Types of Liposidomycins. *J. Antibiot.* **1998**, *51* (12), 1099–1104.

(30) Inukai, M.; Isono, F.; Takatsuki, A. Selective Inhibition of the Bacterial Translocase Reaction in Peptidoglycan Synthesis by Mureidomycins. *Antimicrob. Agents Chemother.* **1993**, *37* (5), 980–983.

(31) Dini, C. MraY Inhibitors as Novel Antibacterial Agents. *Current Topics in Medicinal Chemistry* **2005**, *5* (13), 1221–1236.

(32) Bayburt, T. H.; Sligar, S. G. Membrane Protein Assembly into Nanodiscs. *FEBS Lett.* **2010**, *584* (9), 1721–1727.

(33) Barauskas, J.; Johnsson, M.; Tiberg, F. Self-Assembled Lipid Superstructures: Beyond Vesicles and Liposomes. *Nano Lett.* **2005**, *5* (8), 1615–1619.

(34) Furlong, S. E.; Valvano, M. A. Characterization of the Highly Conserved VFMGD Motif in a Bacterial Polyisoprenyl-Phosphate N-Acetylaminosugar-1-Phosphate Transferase. *Protein Science* **2012**, *21* (9), 1366–1375.

(35) Amer, A. O.; Valvano, M. A. The N-Terminal Region of the Escherichia Coli WecA (Rfe) Protein, Containing Three Predicted Transmembrane Helices, Is Required for Function but Not for Membrane Insertion. *J. Bacteriol.* **2000**, *182* (2), 498–503.

(36) Kumar, M.; Balaji, P. V. Comparative Genomics Analysis of Completely Sequenced Microbial Genomes Reveals the Ubiquity of N-Linked Glycosylation in Prokaryotes. *Mol. BioSyst.* **2011**, *7* (5), 1629–1645.

Chapter 2

Towards the Structural Characterization of PglC, a Bacterial Phosphoglycosyltransferase

Introduction

Phosphoglycosyltransferases (PGTs) are a class of enzymes that transfer C1′-α-linked phosphosugars from their nucleotide-diphosphate-activated forms to a membrane-associated polyprenol phosphate substrate. Bacterial phosophoglycosyltransferases are involved in peptidoglycan and cell wall biosynthesis and generally catalyze the biosynthesis of polyprenol diphosphate-linked glycans that are involved in glycolipid and glycoprotein biosynthesis. Bacterial PGTs are classified into four major classes based on their membrane topologies and overall architecture.[1] The WecA subfamily (Figure 2-1A) has 11 predicted transmembrane helices and no soluble domain. Enzymes in this subfamily are homologous to the eukaryotic Alg7 (25% similarity), which forms the first membrane-bound intermediate in the eukaroyotic N-linked protein glycosylation pathway.[2] Enzymes in the WecA subfamily are involved in lipopolysaccharide O-antigen biosynthesis and catalyze the transfer of UDP-N-acetyl-glucosamine to undecapreol phosphate. Second, the MraY subfamily of enzymes (Figure 2-1B) has 10 transmembrane helices and no soluble domain. These enzymes transfer a phosphorylated disaccharide-pentapeptide conjugate to form the first membrane-bound intermediate in peptidoglycan biosynthesis. The third subfamily of PGTs, is represented by WbaP, (Figure 2-1C), which is involved in the initiation of O-antigen biosynthesis. These enzymes are predicted to have five transmembrane helices, a large periplasmic loop, and a soluble C-terminal domain. Finally, the PglC subfamily (Figure 2-1D) is topologically the most simple and has only one N-terminal transmembrane helix and a C-terminal soluble domain. This subfamily has thus far been the least characterized structurally and biochemically.

As the topological representations in Figure 2-1 show, all phosphoglycosyltransferase families are integral membrane proteins. These enzymes catalyze the reaction between a soluble

substrate and a membrane-anchored substrate to form a membrane-anchored product; thus, membrane association is to be expected. However, the variability in the number of transmembrane domains for enzymes that catalyze the same type of reaction raises questions about the importance of these domains in binding substrate and stabilizing protein structure. Extensive biochemical and topological studies with the first three classes of PGTs have provided insight into conserved carbohydrate recognition and metal-binding domains in the soluble regions and loops of these enzymes.[1] It is important to note that very few of these conserved motifs are found in PglC-type PGTs.

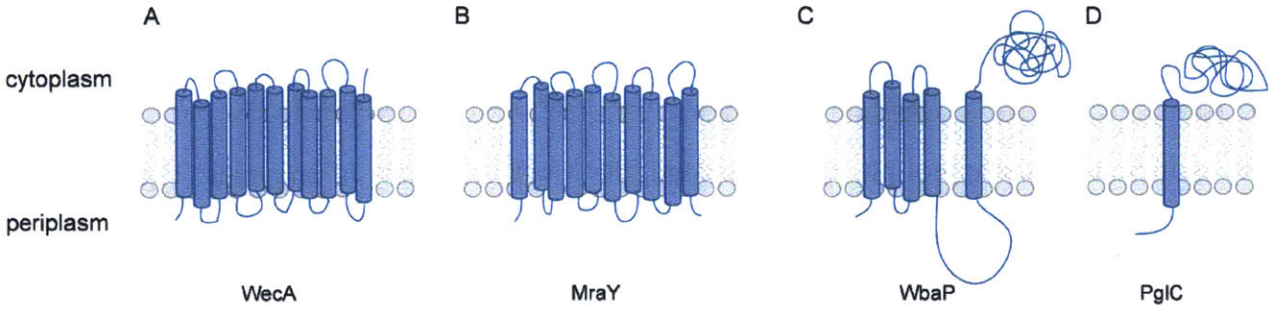


Figure 2-1 Membrane topologies of the different families of phosphoglycosyltransferases. (A) WecA subfamily (B) MraY subfamily (C) WbaP subfamily (D) PglC subfamily.

PglC from *Campylobacter jejuni* has been biochemically characterized as a phosphoglycosyltransferase that transfers *N,N'*-diacetylbacillosamine-1-phosphate from UDP-*N,N'*-diacetylbacillosamine (UDP-diNAcBac) to undecaprenol phosphate.[3] PglC is a 23 kDa protein that is predicted to have a single transmembrane helix at the N-terminus and a soluble cytoplasmic domain at the C-terminus (Figure 2-2). The small size and comparatively simple topology of PglC make it an intriguing candidate protein for biochemical characterization of phosphoglycosyltransferases, and, potentially, a tractable candidate for structural studies. PglC

contains the minimal architecture required to catalyze the phosphoglycosyltransferase reaction, and a crystal structure of this enzyme would provide important insight into this class of proteins. PglC is essential for the biosynthesis of the *C. jejuni* N-glycan, a known virulence factor.[4] PglC is representative of many other homologs found in pathogenic bacteria. Thus, a structural understanding of PglC could guide the synthesis of inhibitors for this class of enzyme, and understand if inhibitors of PglC can be used to attenuate the pathogenicity of *C. jejuni*. There are currently no structures in the Protein Data Bank with adequate sequence homology to PglC to use as templates for protein structure prediction.

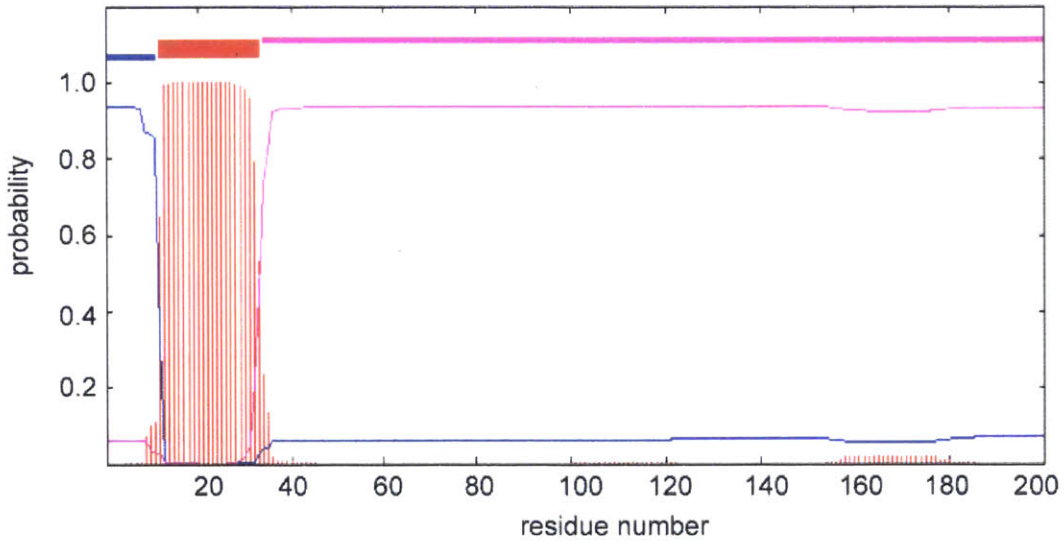


Figure 2-2 Hydropathy plot of PglC generated by the TMHMM server.[5] Residue numbers are represented on the x-axis, and the probability that each residue is in a transmembrane helix is plotted on the y-axis.

Membrane protein X-ray crystallography is challenging for a variety of reasons. Membrane proteins are highly hydrophobic in nature, which makes them poorly soluble. Membrane proteins that are expressed heterologously in *E. coli* can either insert into the cytoplasmic membrane or form inclusion bodies during overexpression. Proteins that are

expressed into inclusion bodies can be isolated and refolded, although the correct refolding of a membrane protein can be challenging, low yielding and may not accurately reflect the native folded structure. Although insertion into the cytoplasmic membrane is preferred and eliminates the need for refolding, the accumulation of large amounts of a non-native protein in membranes can be cytotoxic, which can reduce protein yields greatly.[6] The purification of membrane proteins from the cytoplasmic membrane involves a low-speed centrifugation step to remove the cell debris, followed by a high-speed centrifugation step to isolate the membrane fraction, also known as the cell envelope fraction (CEF). Once the membrane fraction is isolated, the protein of interest is extracted from native lipids by solubilizing the protein in detergent. A final centrifugation step removes unsolubilized protein, and the supernatant from this step is carried forward for further purification. Detergents are used to mimic the membrane environment but they often have complex properties in solution, and need to be carefully selected for crystallography purposes. The choice of detergent can have a significant effect on final protein behavior – not only for the initial extraction from the CEF, but also for stability of the protein, and for catalytic activity. For example, proteins may be need to be solubilized in one detergent and then exchanged to a different detergent during purification for optimal protein stability, catalytic activity and crystallization properties.[7,8] Obtaining stable protein at the high concentrations required for crystallography is also a challenge, as these proteins are prone to aggregation in non-native environments. The method of protein concentration also needs to be carefully selected, as most of these methods also increase the detergent concentration, which can have negative consequences such as causing phase separation.[9] Membrane proteins offer additional difficulties to the crystallography process – the presence of detergent micelles results in crystals with high solvent content, making the crystals fragile, and susceptible to radiation

damage.[10] To date, MraY is the only PGT whose crystal structure has been determined. It is important to note that 19 different homologs had to be screened before a structure was obtained.[11] The relatively simple structure of PglC presents a different type of difficulty – although it would be expected to overexpress more easily because it has a single transmembrane domain, it has been extremely challenging to purify and crystallize. Membrane proteins with a single transmembrane domain are notably underrepresented in the Protein Data Bank, possibly due to difficulties with ordered crystallization for structures with this topology. [12]

This chapter summarizes our attempts and ultimate success with obtaining pure, monodispersed PglC at the high protein concentrations required for X-ray crystallography. The expression and purification of PglC is discussed, detailing the optimization of components ranging from the method of overexpression to choice of detergent as well as solubility and coexpression tags. This chapter first addresses our efforts to express PglC with different coexpression tags and to optimize the detergents used at different stages of purification. Next, the importance of the transmembrane domain was explored by working with the soluble domain of PglC for structural and functional studies. Lastly, PglC homologs from other *Campylobacter* species were investigated as additional targets for structural studies. This chapter highlights major developments, and a more detailed description of incremental progress can be found in the Appendix.

Results and Discussion

PglC was initially expressed and characterized as a T7-PglC-His_6 construct, solubilized in Triton X-100 detergent.[3] While catalytically active, this construct yielded low amounts of aggregated protein. Additionally, Triton X-100 is not an ideal detergent for X-ray crystallography as it is heterogeneous. A His_6-TEV-GB1-PglC construct was engineered by

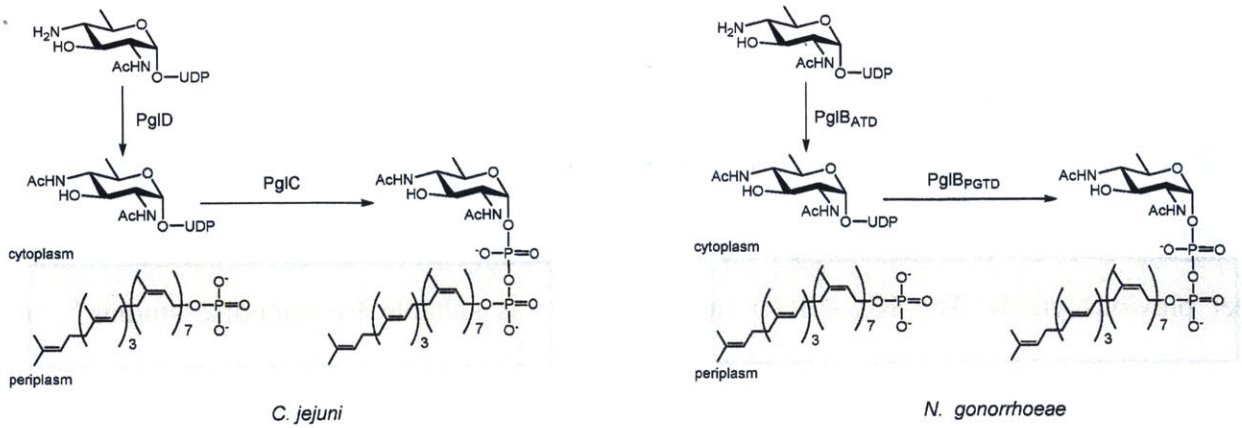
Meredith Hartley, as the GB1 (B1 immunoglobulin) fusion has been successfully used as a coexpression tag in the expression of small proteins.[13] The presence of the GB1 domain appeared to make the His_6 tag more accessible and exhibited improved binding to Ni-NTA resin. The GB1-PglC fusion was purified with a variety of detergents to assess their abilities to effectively solubilize protein. Ultimately, dodecyl maltoside (DDM) was used as the detergent of choice, as it is compatible with protein crystallography. While these developments improved the overall yields and stability of PglC greatly, the protein still aggregated at concentrations between 2-5 mg/mL. Further optimization was required to maintain it in a soluble state at these concentrations.

Optimization of Coexpression Tags and Detergents

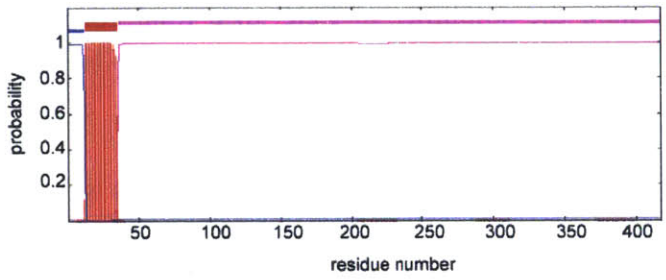
Homology to PglB from Neisseria gonorrhoeae

In *C. jejuni*, PglD and PglC catalyze consecutive reactions i.e. the acetyltransferase reaction in the synthesis of UDP-diNAcBac and the transfer of phospho-diNAcBac to undecaprenol phosphate (Figure 2-3A). These reactions are the initiating steps in the synthesis of the ultimate glycan donor (GalNAc-α1,4-GalNAc-α1,4-(Glcβ1,3)-GalNAc-α1,4-GalNAc-α1,4-GalNAc- α1,3-diNAcBac) for N-linked protein glycosylation in *C. jejuni*. PglB(*Ng*) is a bifunctional protein from *Neisseria gonorrhoeae* which catalyzes the same reactions as PglC and PglD.[14] PglB(*Ng*) has an N-terminal phosphoglycosyltransferase domain and a C-terminal acetyltransferase domain, which exhibit high homology to *C. jejuni* PglC and PglD, respectively (Figure 2-3B, C). The acetyltransferase domain has been isolated and characterized biochemically and structurally.[15]

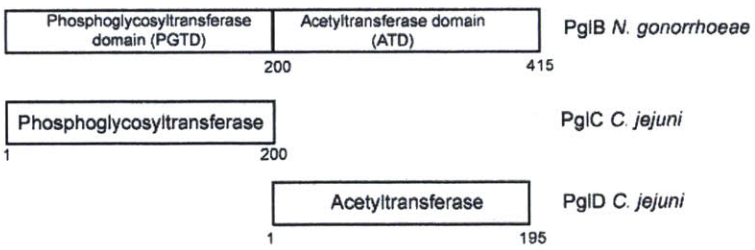
A



B



C



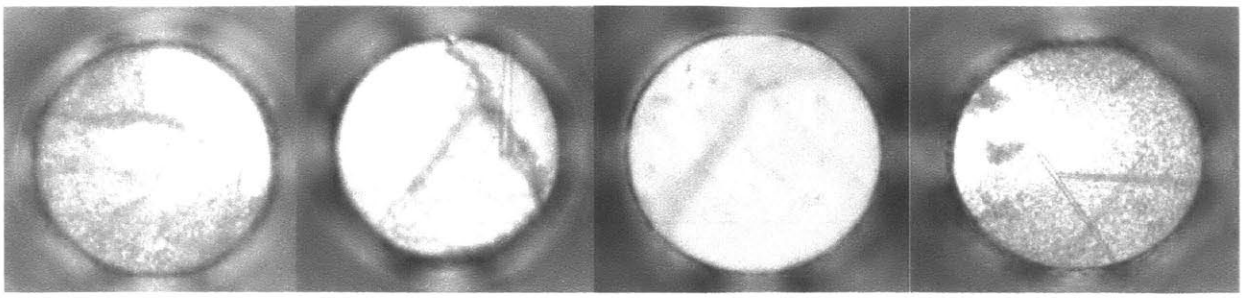
D

	% Similarity	% Identity
PglB-PGTD and PglC	52%	27%
PglB-ATD and PglD	34%	30%

Figure 2-3 Homology comparison of *C. jejuni* and *N. gonorrhoeae* PGTs. (A) Reactions catalyzed by PglC and PglC in *C. jejuni* and the corresponding reaction catalyzed by PglB in *N. gonorrhoeae.* (B) Hydropathy plot of PglB(*Ng*) predicted by the TMHMM server. Residue numbers are represented on the x-axis, and the probability that each residue is in a transmembrane helix is plotted on the y-axis (C) Schematic representation of the domains of PglB depicting regions homologous to PglC and PglD. (D) Table comparing the similarity and identity between these domains and PglC and PglD.

Due to these similarities, it was hypothesized that the *C. jejuni* PglC and PglD might interact with each other in vivo to form a complex that resembles PglB(*Ng*), and that this interaction might help stabilize PglC *in vitro*. The expression of PglC was optimized to afford a construct which contained an N-terminal His_6 and GB1 (B1 immunoglobulin) tag yielding soluble, monodisperse PglC with yields of approximately 1-2 mg/L of culture, a 10-fold increase over previous yields. The His_6-GB1-PglC construct was suitable for nanodisc studies[16], but aggregates at the high concentrations required for crystallography (> 3mg/mL). PglD can be readily expressed and purified in high quantities[17], and its addition to PglC at equimolar concentrations appeared to help maintain the solubility of PglC to a final concentration of 3 mg/mL of each protein (6 mg/mL total protein).

A sample of an equimolar amount of His_6-GB1-PglC and PglD was submitted to the Hauptman-Woodward Institute to be used with their membrane protein-specific crystallography screens [18]. A few crystals were obtained in these initial screens (Figure 2-4) but these results could not be reproduced to ensure that the crystals were of the complex of PglC and PglD or just PglC. As a control to ascertain whether the stabilizing effects of PglD were specific, BSA was added to PglC in equimolar amounts and appeared to have a similar stabilizing effect. Based on these results, it appeared that PglC was not being stabilized by PglD by the formation of a complex that resembles PglB.



0.1 M NaH_2PO_4 0.1 M HEPES pH 7.0 32% PEG 4000	0.05 M NaCl 0 1 M HEPES pH 7.0 20% PEG 6000	0.1 M Bis Tris pH 6.5 45% PEG 400	0.1 M KCl 0.1 M HEPES pH 7.0 24% PEG 4000

Figure 2-4 Crystals obtained from the membrane protein screen at the Hauptman-Woodward Institute

PglB was also expressed as a GB1 fusion to assess its stability and solubility in this form. This construct contained an N-terminal GB1 solubility tag in addition to a C-terminal acetyltransferase domain that is known to be extremely soluble. The presence of both of these domains was hypothesized to improve the yield of the overall protein, as the intractable PglC-like domain would be sandwiched between two well-expressing and stable solubilizing domains. However, the improved expression of this construct was not observed (Figure 2-5); the total yield of His_6-GB1-TEV-PglB was <0.1 mg/L of culture.

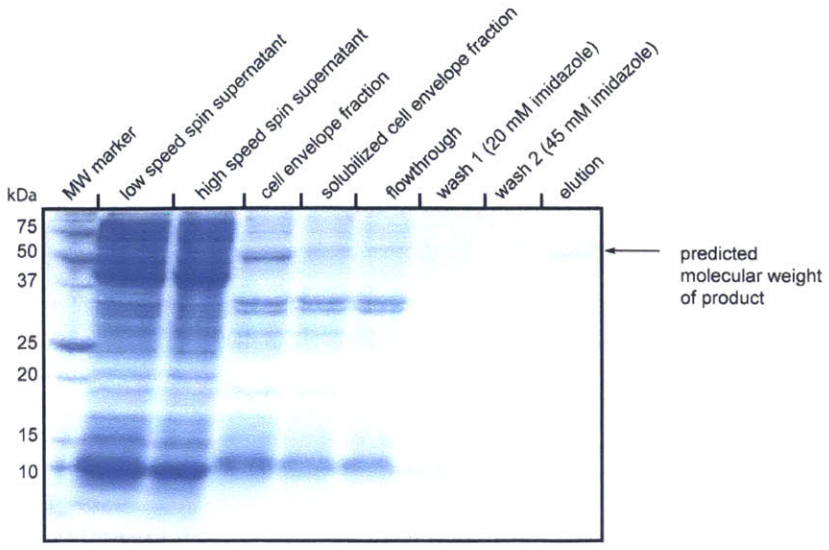


Figure 2-5 SDS-PAGE analysis of the overexpression and purification of His6-GB1-TEV-PglB. A faint band at ~50 kDa is observed in the elution lane, corresponding to the expected molecular weight of the product.

PglB was also cloned into the pET-24 a(+) vector, which provides an N-terminal T7 tag and a C-terminal His_6 tag. The overexpression of this construct as well as the GB1-fusion were tested with the TaKaRa chaperone plasmid set which contains different combinations of DnaK, DnaJ, GrpE, GroES, GroEL, DnaK and DnaK chaperones.[19] A small-scale screen was done with all five chaperone lines (Figure 2-6). The presence of chaperones did not appear to improve the expression of the GB1-fusion. However, some of the chaperone-expressing cell lines seemed to increase the total expression of the T7-PglB-His_6 construct. This construct was grown in BL21-T2, BL21-T3, BL21-T5 and BL21-RIL on a larger scale to purify protein from the CEFs and to compare the amount of detergent-solubilized protein that could be isolated from each sample (Figure 2-7). In all cases, a large quantity of protein was still present in the cell debris pellet, suggesting that the protein misfolded during overexpression and was most likely in inclusion bodies. The cell line that expressed supplemented chaperones were unable to increase the final

amount of soluble protein, as indicated by the low amount of protein present in the lane corresponding to the solubilized CEF.

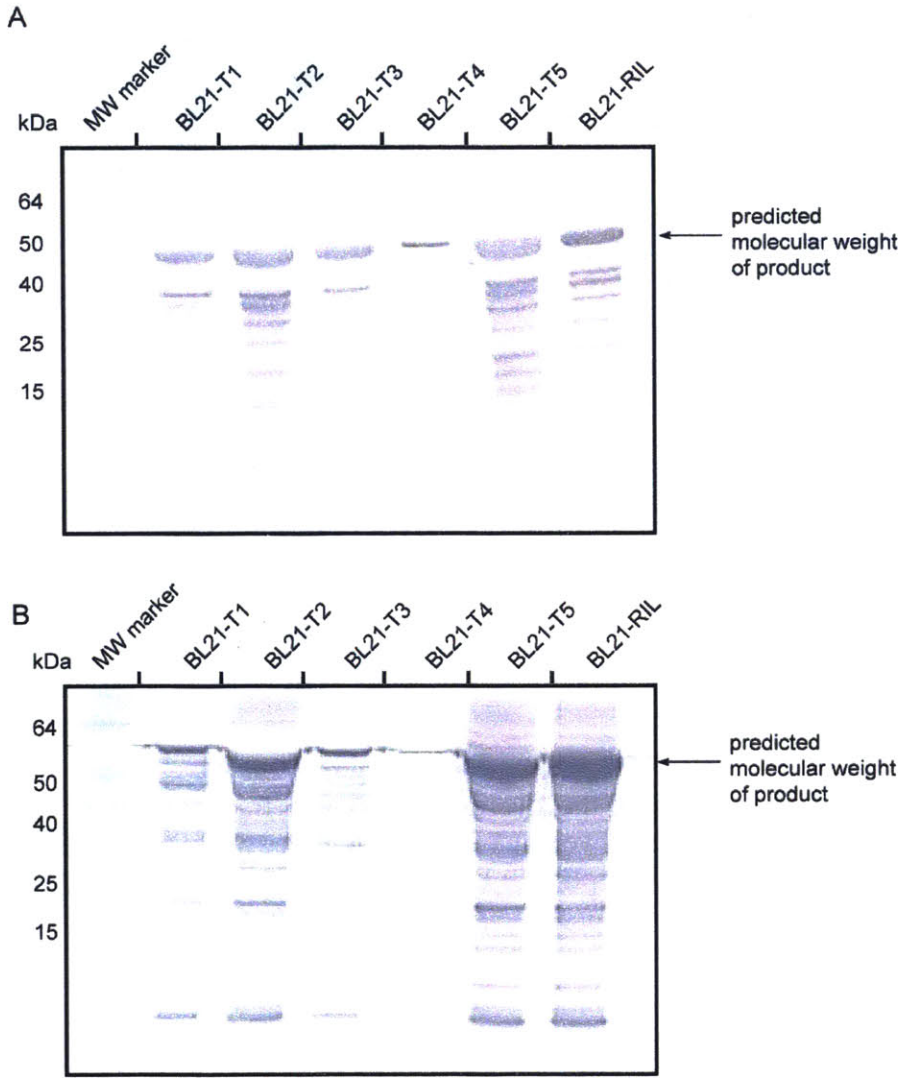


Figure 2-6 Western blot analysis of a small-scale screening of the overexpression of (A) T7-PglB-His_6 and (B) His_6-GB1-TEV-PglB with five different chaperone cell lines. The analysis was performed with cell lysates. BL21-T1: DnaK, DnaJ, GrpE, GroES, GroEL; BL21-T2:

GroES, GroEL; BL21-T3: DnaK, DnaJ, GrpE; BL21-T4: GroES, GroEL, Trigger factor; BL21-T5: Trigger factor. BL21-RIL cells were included as a control.

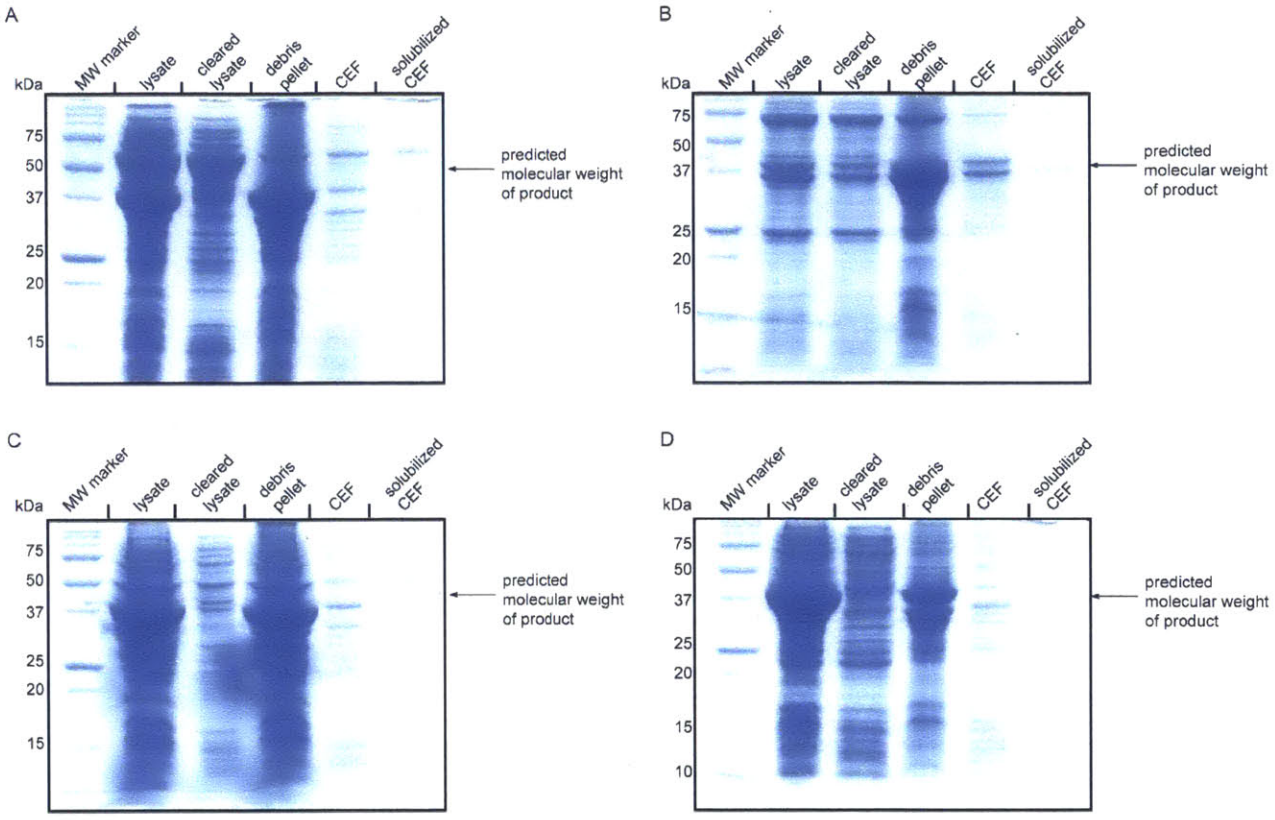


Figure 2-7 SDS-PAGE analysis of the overexpression and fractionation of T7-PglB-His_6 in chaperone cell lines. The molecular weight of the protein is ~50 kDa. (A) BL21-T2: GroES, GroEL (B) BL21-T3: DnaK, DnaJ, GrpE; (C) BL21-T5: Trigger factor (D) BL21-RIL.

SUMO-fusion of PglC

The small ubiquitin-related modifier (SUMO) tag was chosen as it has been successful in solubilizing other intractable proteins.[20] An additional advantage is that the SUMO protease can then be used to cleave the SUMO tag in a highly specific fashion to obtain native protein. A SUMO-fusion of PglC was constructed and this fusion was tested in cell lines designed to

improve protein solubility (Figure 2-8). When overexpressed in BL21-RIL cells and purified with 0.03% DDM, this construct was stable at concentrations less than 2 mg/mL. However these protein concentrations are not high enough for crystallographic studies and attempts to increase the concentration of PglC in DDM resulted in protein aggregation.

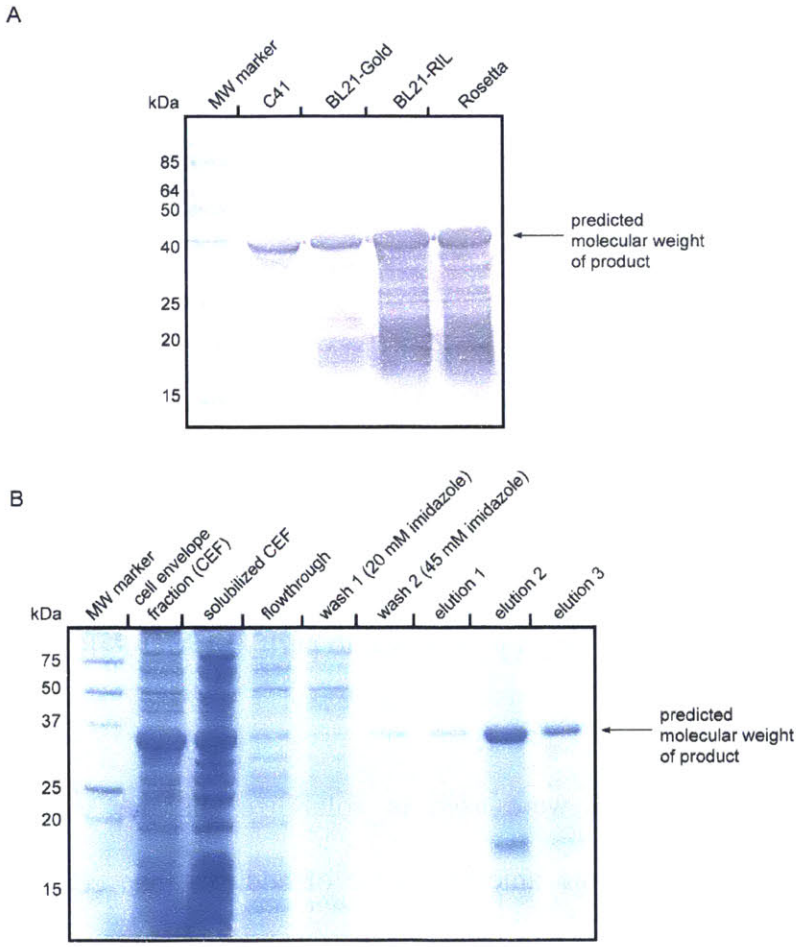
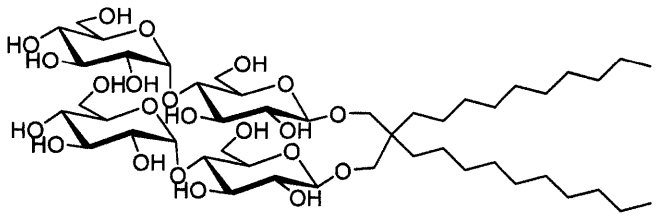


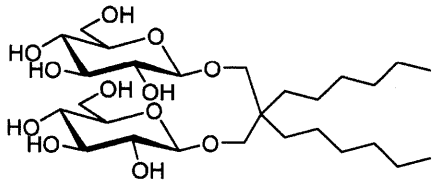
Figure 2-8 Optimization of the expression of SUMO-PglC. (A) Western blot analysis of SUMO-PglC expressed in different cell lines, analyzed with an α-His antibody. (B) SDS-PAGE analysis of the purification of SUMO-PglC in BL21-RIL cells.

Purification of SUMO-PglC Using Amphiphiles

Reports of membrane protein crystal structures using amphiphiles suggest that these detergent mimics are more effective at stabilizing membrane proteins than conventional detergents.[21] In particular, a new class of amphiphiles, derived from neopentyl glycol, with hydrophilic groups derived from maltose show favorable behavior relative to conventional detergents. Lauryl maltose neopentyl glycol (LMNG) and octyl glucose neopentyl glycol (OGNG) are members of this maltose–neopentyl glycol (MNG) amphiphile family (Figure 2-9).



Lauryl Maltose Neopentyl Glycol (LMNG)



Octyl Glucose Neopentyl Glycol (OGNG)

Figure 2-9 Chemical structures of amphiphiles used to purify SUMO-PglC.

Both LMNG and OGNG were used to solubilize and purify and SUMO-PglC to determine whether this was a more amenable type of additive than conventional detergents. Purification with LMNG resulted in adequate solubilization of PglC from the CEF (Figure 2-10A), however, the protein eluted in the void volume of the column during size exclusion chromatography, indicating that it formed an aggregate (Figure 2-10C). OGNG was unable to efficiently solubilize SUMO-PglC from the CEF (Figure 2-10B). Attempts to improve yields by solubilizing the protein out of the CEF with Triton X-100 and switching to the amphiphiles afterwards were also unsuccessful (data not shown).

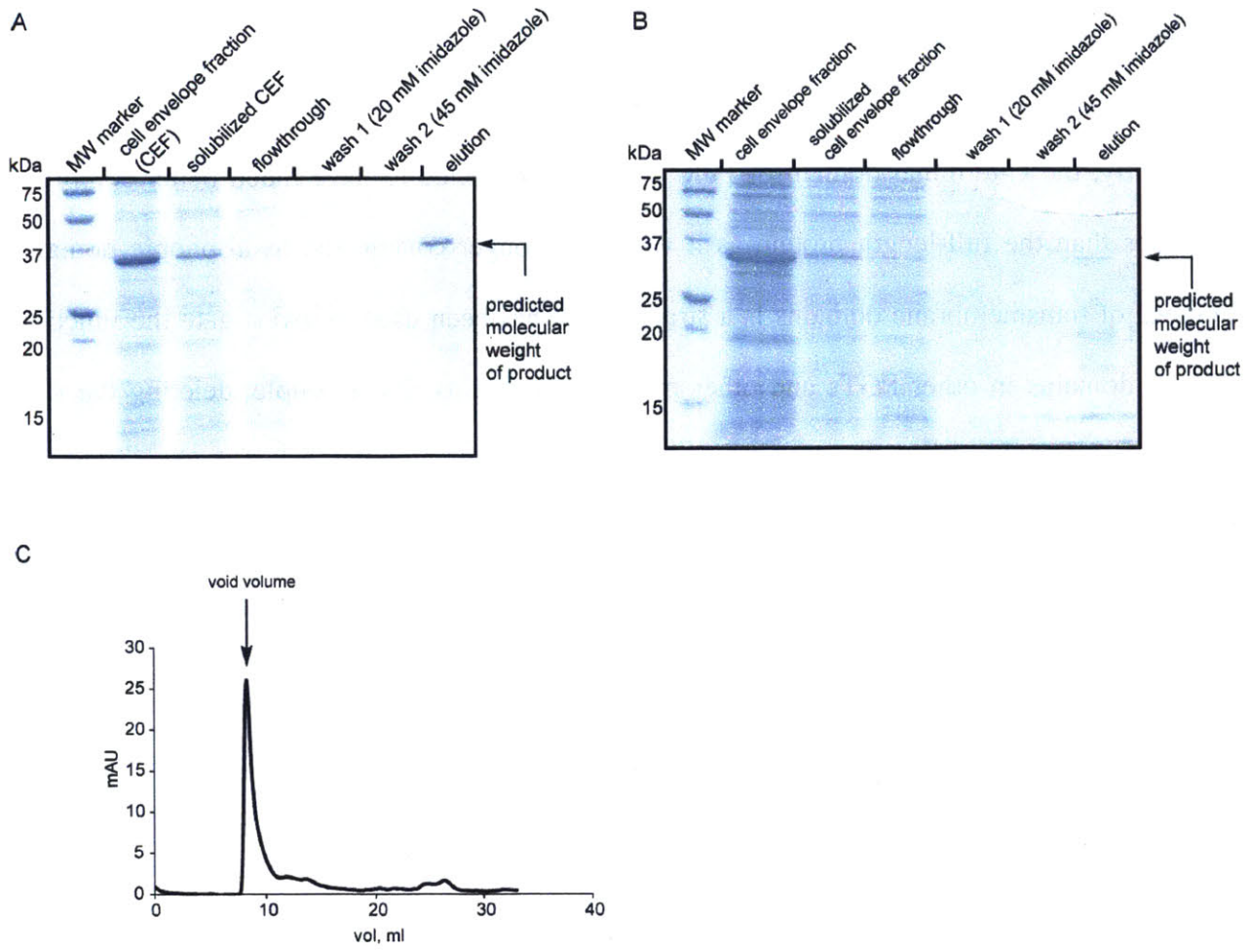


Figure 2-10 Purification of SUMO-PglC with amphiphiles. (A) SDS-PAGE analysis of the purification of SUMO-PglC with LMNG. (B) SDS-PAGE analysis of the purification of SUMO-PglC with OGNG. (C) Gel filtration analysis of protein purified using LMNG.

Expression of the Soluble Domain of PglC and PglB(*Ng*)

It was hypothesized that the aggregation of both PglC and PglB(*Ng*) observed under these variety of conditions was largely due to the N-terminal transmembrane helix. The role of the transmembrane domain in the catalytic activity of PglC is not clear. One hypothesis is that the transmembrane domain simply anchors PglC into the membrane. If this is the primary purpose of the transmembrane domain, constructs of the C-terminal soluble domain of PglC - lacking the

transmembrane helix - could potentially be used to characterize the activity of PglC further, and structural studies of this domain would still provide insight into its catalytic mechanism. Most importantly, the C-terminal domain of PglC would be expected to have much better solubility properties than the full-length protein as it would no longer contain the hydrophobic domain. Removal of transmembrane domains is a strategy that has been used to investigate the function of these domains in other PGTs and other membrane proteins. For example, deleting the first three transmembrane domains of WecA was found to affect function but does not affect insertion of the protein into the bacterial membrane.[22] In the case of WbaP, the first four transmembrane domains are not required for catalytic activity. [23] Cytochrome P450s, which also contain a single transmembrane helix, have been successfully crystallized with truncated forms of the transmembrane domain.[24]

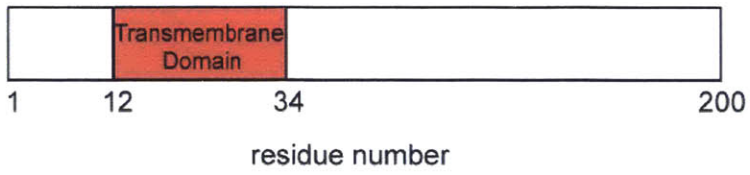


Figure 2-11 Schematic representation of PglC. The transmembrane domain, shown in red, is predicted to span residues 12 to 34.

The transmembrane helix of PglC is predicted to be between residues 12 and 34 (Figure 2-11). Three constructs were prepared with truncations at different residues after the transmembrane helix, as subtle differences in the distance from the helix could potentially influence protein solubility. The three constructs of soluble PglC ($sPglC_{36}$, $sPglC_{39}$ and $sPglC_{42}$) were designed in the pET-NO vector, which provides a TEV-cleavable C-terminal His_8 tag. A solubility tag was believed to be unnecessary at the time, as this globular domain was predicted

to be quite soluble. However, the protein overexpressed in a misfolded form - in inclusion bodies. Protein could be extracted out of inclusion bodies using 6 M urea (Figure 2-12), and was refolded by gradually decreasing the amount of urea by dialysis. When analyzed by size exclusion chromatography (Figure 2-13), the protein obtained by this method appeared to be aggregated and eluted in the void volume (8 mL) of the size exclusion column. This was observed for both the $sPglC_{36}$ and $sPglC_{39}$ constructs.

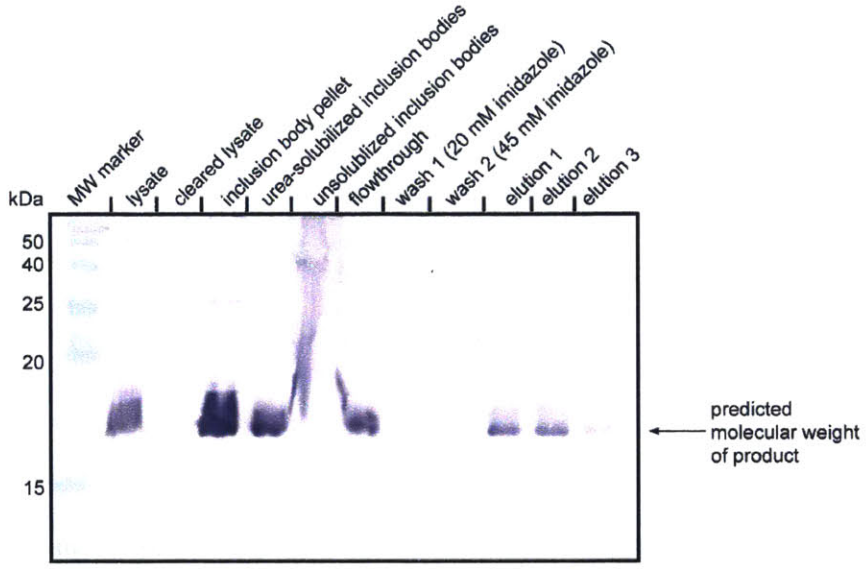


Figure 2-12 Western blot analysis of the overexpression and purification of $sPglC_{36}$, analyzed with an α-His antibody.

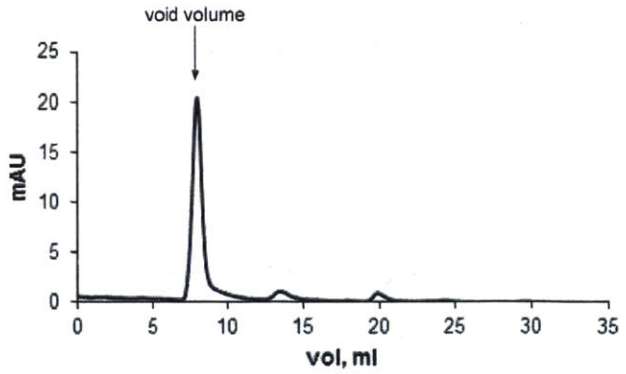


Figure 2-13 Gel filtration analysis of the soluble domain of PglC ($sPglC_{36}$) isolated from inclusion bodies.

Attempts to purify the soluble domain of PglB were also unsuccessful. Three constructs were designed, each starting at different distances from the end of the predicted transmembrane helix ($sPglB_{33}$, $sPglB_{38}$ and$sPglB_{41}$; the transmembrane domain is predicted to end at residue number 32). The absence of the hydrophobic transmembrane domain and the presence of a very soluble C-terminal acetyltransferase domain were expected to help maintain the solubility of the protein. However, when overexpressed, the majority of the protein was present in the cell debris pellet, suggesting that it had misfolded into inclusion bodies (Figure 2-14).

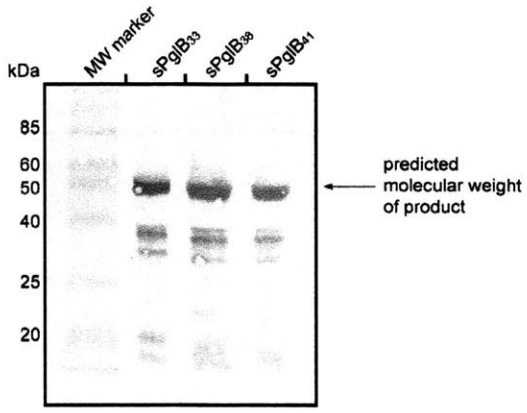


Figure 2-14 Western blot analysis of the cell debris fraction of different constructs of soluble PglB. The major band corresponds to the expected molecular weight of the product (~50 kDa).. The western blot was probed with an α-His antibody.

A solubilizing fusion tag was implemented to increase the solubility of these constructs. The small ubiquitin-related modifier (SUMO) tag was chosen as it has been successful in solubilizing the full-length PglC. Three constructs of the soluble domain of PglC (SUMO-$sPglC_{36}$, SUMO-$sPglC_{39}$ and SUMO-$sPglC_{42}$) were designed and expressed, as it was once again hypothesized that small differences in the length of the construct might affect the stability of the protein and its ability to crystallize. The SUMO tag enhanced the solubility of these constructs, resulting in the presence of protein in the cleared lysate fraction (Figure 2-15A). However, additional wash steps were insufficient to completely remove impurities, and, importantly, the

SUMO-sPglC$_{36}$ fusion appeared to be catalytically inactive (Figure 2-15B). This result suggests that the transmembrane domain must have a more important role than simply localizing protein to the membrane and may have important implications in the mechanism of PglC.

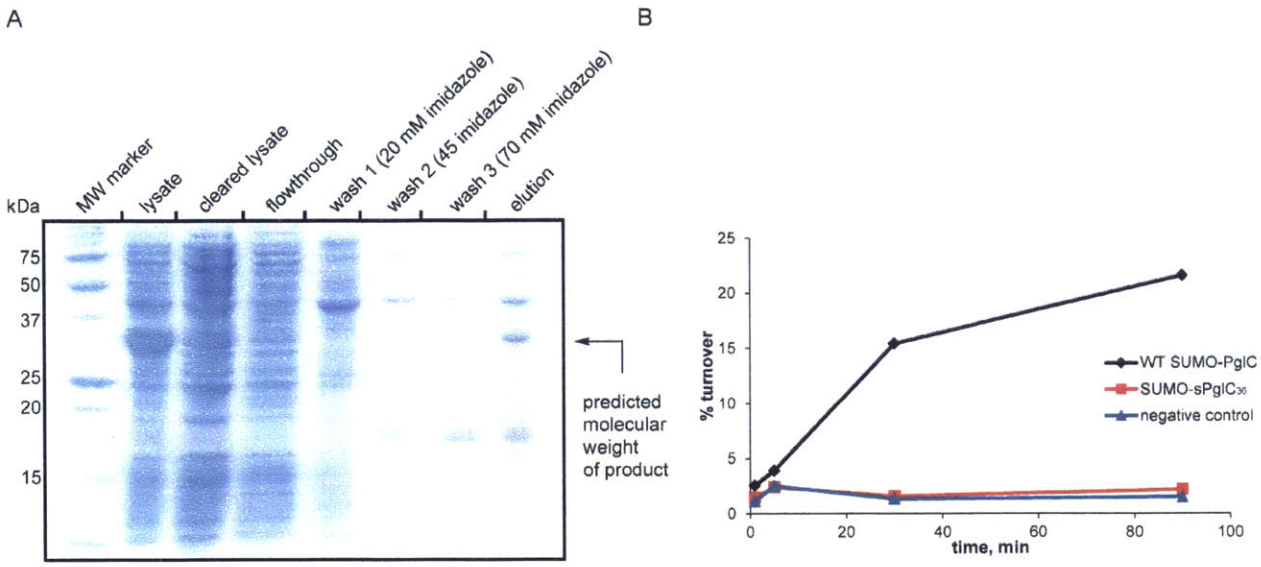


Figure 2-15 Analysis of SUMO-sPglC$_{36}$. (A) SDS-PAGE analysis of the overexpression and purification of SUMO-sPglC$_{36}$. The molecular weight of the product is approximately 35 kDa. (B) Activity assay with purified SUMO-sPglC$_{36}$. The activity was compared to a SUMO-fusion of the full-length PglC (described in the next section). A negative control with no enzyme was included.

PglC Homologs

The SUMO-fusion of PglC was found to still aggregate at high concentrations even with the SUMO solubility tag, making it a challenging candidate for structural analysis by X-ray crystallography. In light of this observation, the hydropathy plot of the *C. jejuni* PglC was examined in closer detail, and a hydrophobic "patch" was observed between residues 160-180 (Figure 2-16).

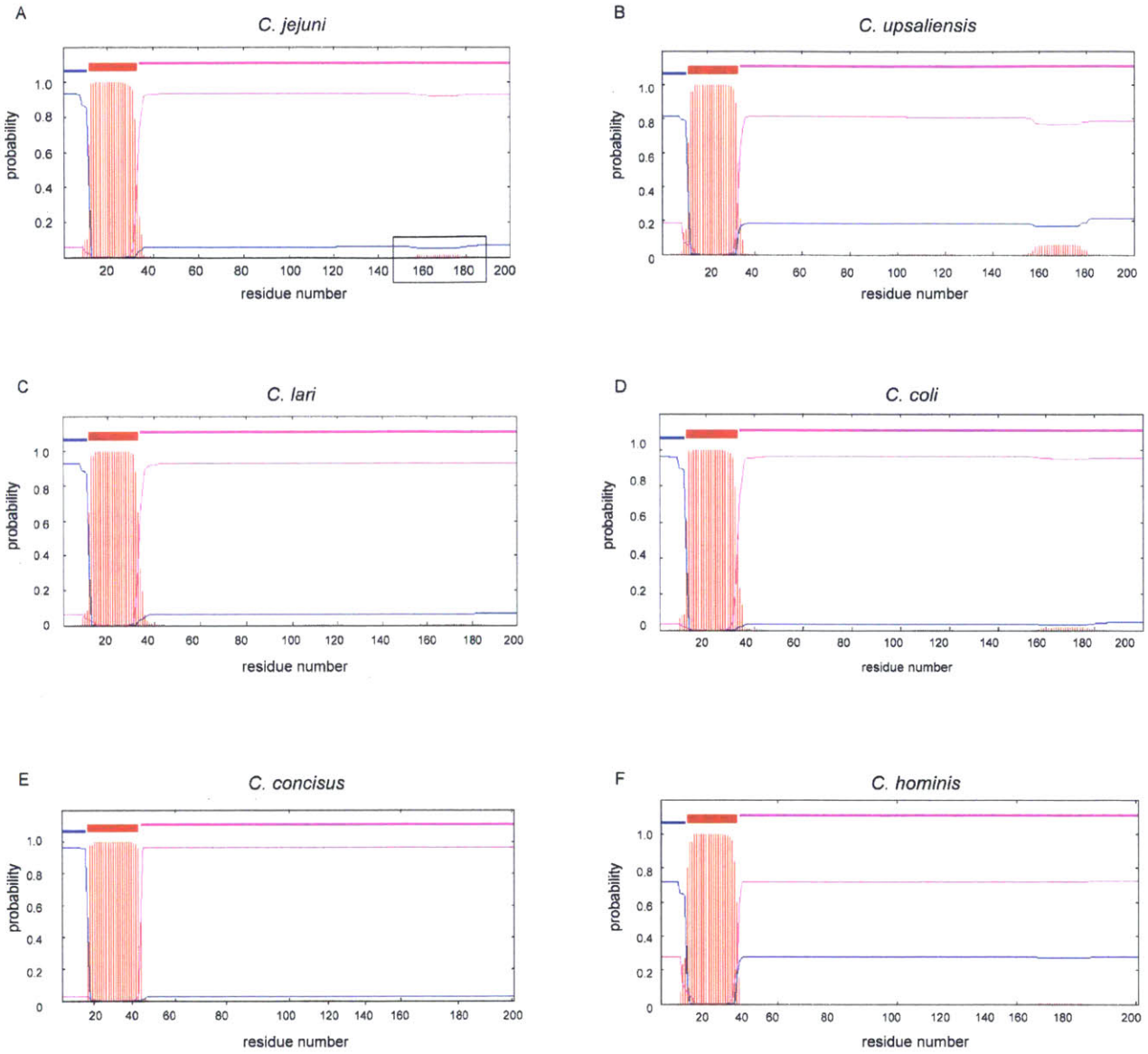


Figure 2-16 Hydropathy plots of PglC homologs from different *Campylobacter* species generated using the TMHMM server.[5] Residue numbers are plotted on the x-axis and the probability that each residue is present in a transmembrane helix is plotted on the y-axis. (A) *C. jejuni* (B) *C. upsalienses* (C) *C. lari* (D) *C. coli* (E) *C. concisus* (F) *C. hominis.* The hydrophobic "patch" for PglC from *C. jejuni* is marked with a box.

Based on the hydropathy plot, the region between residues 160 and 180 in PglC from *C. jejuni* does not appear to be hydrophobic enough to be a transmembrane helix. It is possible that these residues form an amphipathic helix that would be associated with the membrane. A helical

wheel was constructed to assess the distribution of hydrophobic residues in this region, and the predicted helix does appear to have the hydrophobic residues on one face of the helix (Figure 2-17).

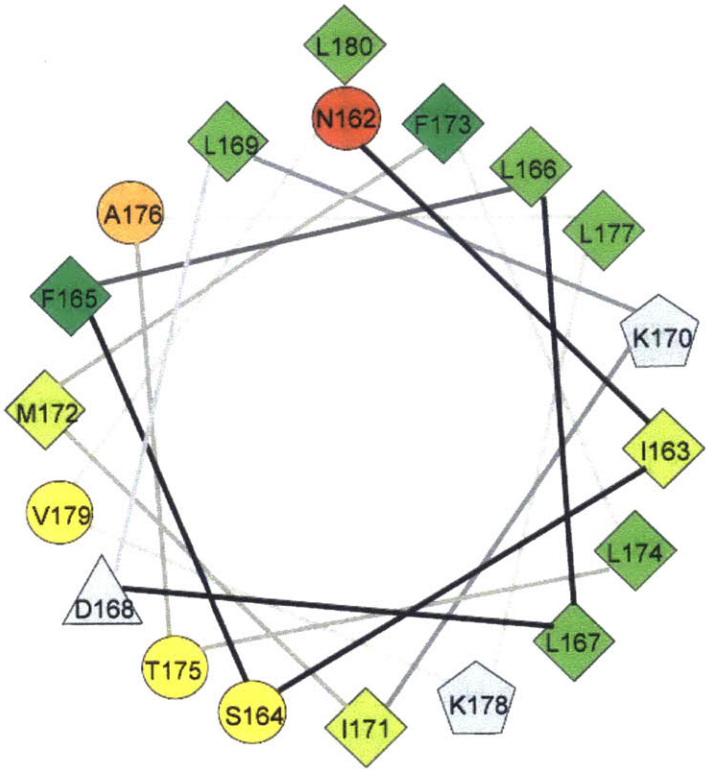


Figure 2-17 Helical wheel prediction constructed using prediction software created by Don Armstrong and Raphael Zidovetzki: Version Id: wheel.pl,v 1.4 2009-10-20 21:23:36 don Exp. Hydrophilic residues are represented as circles, hydrophobic residues as diamonds, potentially negatively charged as triangles, and potentially positively charged as pentagons. The most hydrophobic residue is shown in green, and the amount of green decreases proportionally to the hydrophobicity, with zero hydrophobicity coded as yellow. Hydrophilic residues are coded red with pure red being the most hydrophilic (uncharged) residue, and the amount of red decreasing proportionally to the hydrophilicity. The potentially charged residues are light grey.

It was hypothesized that the presence of this hydrophobic patch could be a cause of the aggregation seen in PglC from *C. jejuni*, and may explain why even the cytoplasmic domain was prone to aggregation. PglC homologs from different *Campylobacter* species were analyzed to

assess their hydropathy plots and it was observed that the hydrophobicity of the region between residues 160 – 180 varied considerably among homologs (Figure 2-16). The *Campylobacter* species selected for this analysis were all known to use UDP-diNAcBac as a substrate.[25] Based on this comparison, the PglC homologs from *Campylobacter lari* and *Campylobacter concisus* were selected as the most suitable candidates to pursue, as their hydropathy plots showed minimal probability of hydrophobicity in the region between residues 160-180 as well as the rest of the cytoplasmic domain of the protein. This approach of using a homolog from a different *Campylobacter* species was successful in the case of the oligosaccharyl transferase, PglB, where the protein from *C. jejuni* is also prone to aggregation at high concentrations, and the homolog from *C. lari* was eventually used to obtain the first crystal structure of this protein.[26]

PglC from *C. lari* and *C. concisus* were expected to be more soluble and better behaved than the *C. jejuni* PglC due to the absence of the hydrophobic patch. Therefore, they were initially expressed without any solubility fusion tag. They were expressed as His_6-TEV-PglC constructs, with a TEV-cleavable His_6-tag at the N-terminus (Figure 2-18A,B). Unfortunately, the proteins did not express at high yields (<1 mg/L of culture) and were aggregated when analyzed by size exclusion chromatography (Figure 2-18C, D).

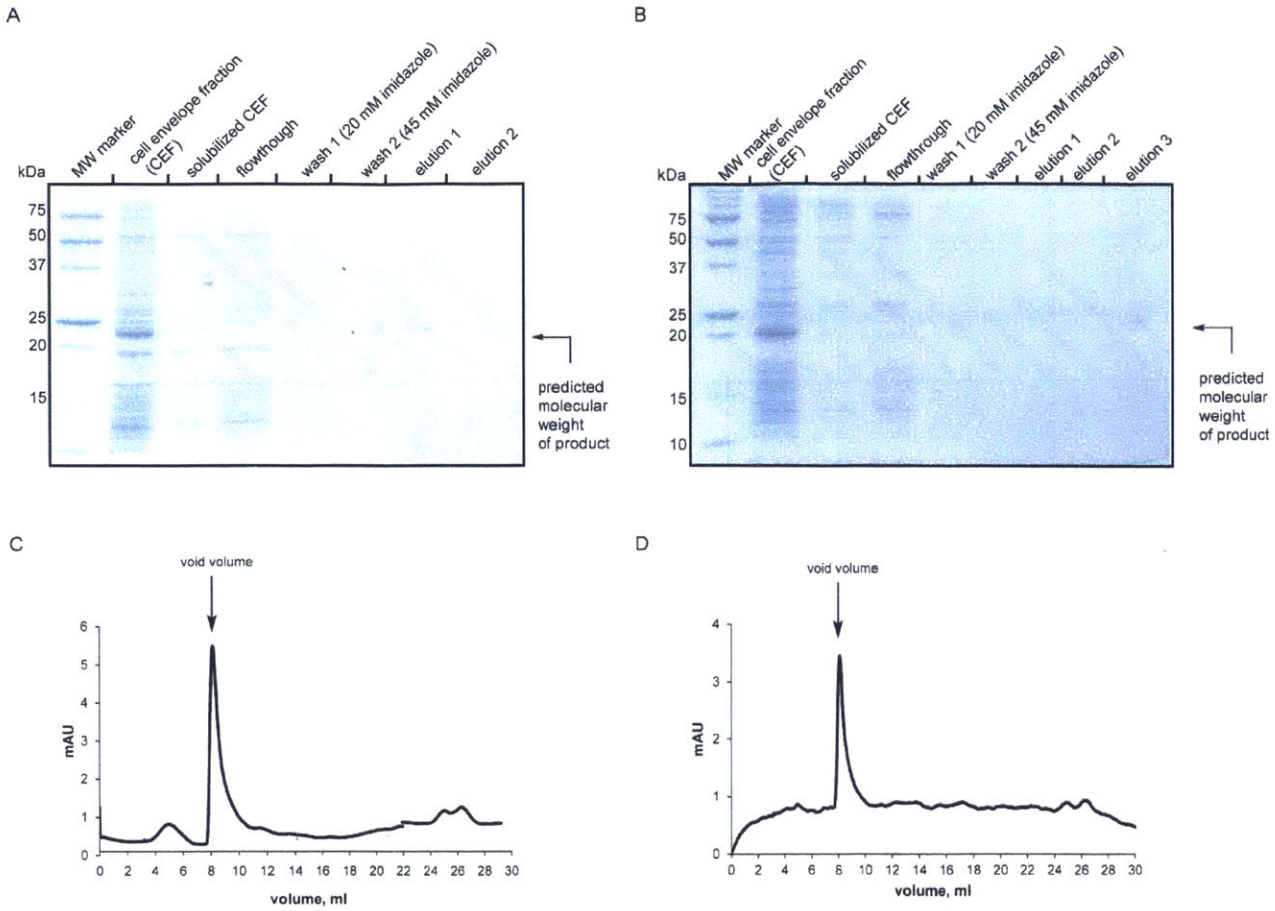


Figure 2-18 Purification of PglC homologs. (A) SDS-PAGE analysis of PglC from *C. concisus* (B) SDS-PAGE analysis of PglC from *C. lari.* Faint bands can be observed at the expected molecular weight of the product (~20 kDa). (C) Gel filtration analysis of PglC from *C. concisus* (D) Gel filtration analysis of PglC from *C. lari*. The void volume of the column is 8 mL.

SUMO-fusions of the PglCs from *Campylobacter lari* and *Campylobacter concisus* were then constructed to determine whether the presence of a SUMO tag would increase yields as it had with PglC from *C. jejuni*. The presence of the SUMO tag greatly increased levels of expression (Figure 2-19A,B). These SUMO fusions were purified and analyzed by gel filtration (Figure 2-19C). Importantly, it also reduced the aggregation of PglC at high concentrations (~3 mg/mL) when analyzed by gel filtration, and compared to a similar concentration of a SUMO fusion of PglC from *C. jejuni* (Figure 2-19C). These results support the hypothesis that the

presence of hydrophobic patches in the cytoplasmic domain of PglC were contributing to its aggregation at high protein concentrations.

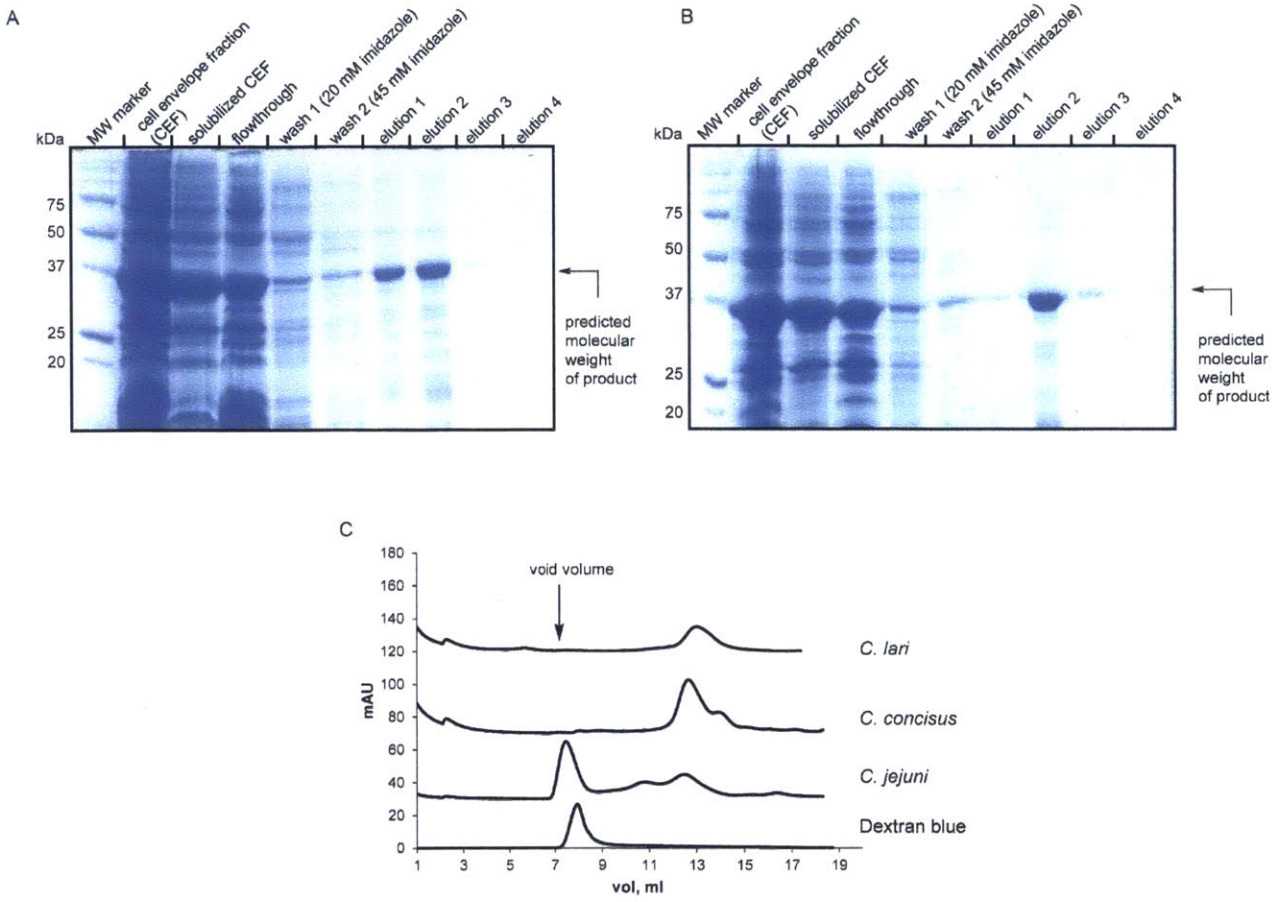


Figure 2-19 Purification of PglC homolog proteins. SDS-PAGE analysis of (A) SUMO-PglC from *C. concisus* and (B) SUMO-PglC from *C. lari*. (C) Gel filtration analysis of PglC from *C. jejuni, C. concisus and C. lari.* Dextran blue is used to mark the void volume of the column.

Crystallography with SUMO-PglC from *C. concisus* and *C. lari*

Based on these results, SUMO-PglC from both *C. concisus and C. lari* showed great promise for protein crystallography. SUMO-PglC from *C. concisus* and *C. lari* were overexpressed and purified on a large scale for X-ray crystallography purposes. Attempts to

remove the SUMO domain using the SUMO protease were unsuccessful (Figure 2-20). The SUMO protease recognizes the tertiary structure of the SUMO domain instead of a specific amino acid sequence. The inability of the protease to remove the SUMO domain from the SUMO-PglC fusion suggested that the fusion was structured in a way that prevented the protease from recognizing the SUMO domain and the adjacent cleavage site. At this stage, it was decided to proceed with crystallography, as the SUMO fusions were much less aggregated than PglC without the SUMO fusion tag. Crystal trays were set up using the MembFac, Index, MemGold and MemGold2 Screens at protein concentrations of 6.4 mg/mL (SUMO-PglC *C. lari*) and 7.4 mg/mL (SUMO-PglC *C. concisus*). However no crystals were observed in any of these screens, even after a year.

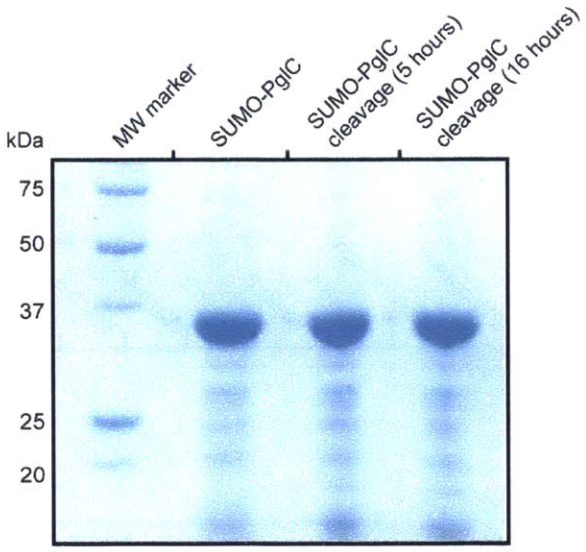


Figure 2-20 SDS-PAGE analysis of SUMO protease cleavage reactions with SUMO-PglC. The fusion protein migrates at ~35 kDa, and the cleaved PglC migrates at ~20 kDa.

Optimization of SUMO protease cleavage of SUMO-PglC

The failure of SUMO-PglC to crystallize led us to believe that the presence of the SUMO domain might be inhibiting crystal formation. The SUMO-PglC constructs were modified to

include short amino acid linkers between the SUMO domain and PglC, in order to provide more flexibility to make this region more accessible to the SUMO protease. Attempts to insert Ala_3 and Ala_5 linkers by Quickchange methods were unsuccessful; this was attributed to the high GC content of these primers that may have had high levels of secondary structure. A second strategy to insert Ser-Gly-Ser-Gly linkers between the SUMO domain and PglC by amplification and ligation into the pET SUMO vector was successful (Figure 2-21A). This construct was cleaved by the SUMO protease, affording PglC without a solubility tag. The cleavage reaction was optimized by Dr. Debasis Das and Sonya Entova, and cleaved PglC was readily purified from the protease reaction mixture (Figure 2-21B).

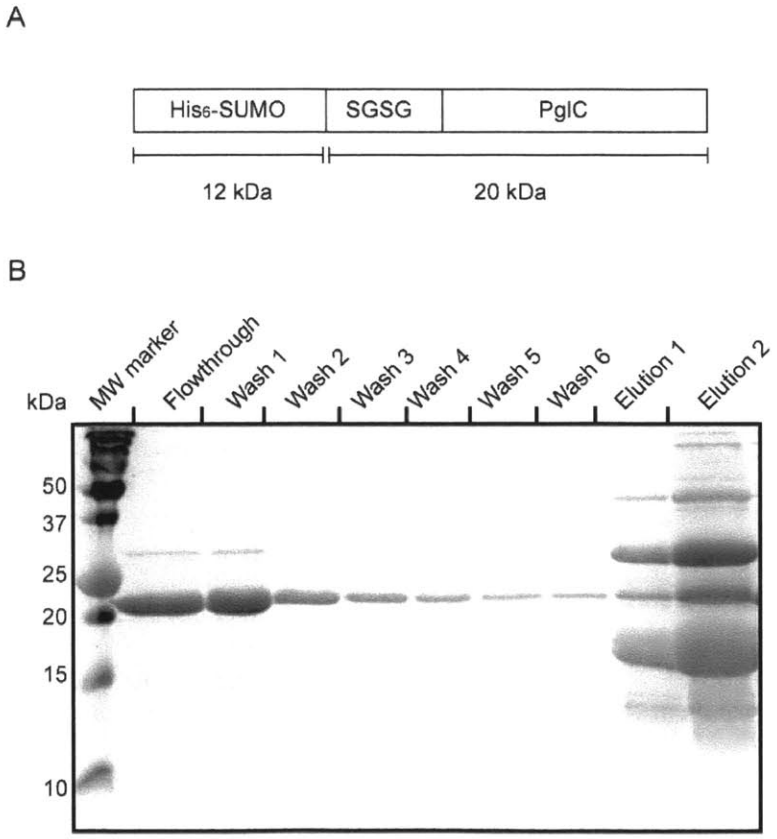
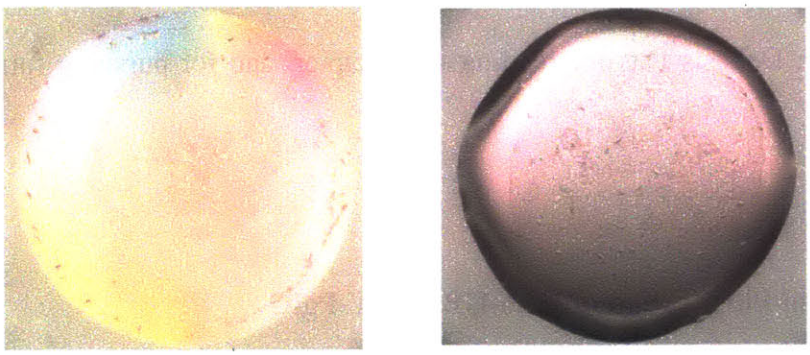


Figure 2-21 Addition of a linker between SUMO domain and PglC. (A) Schematic representation of the modified SUMO-PglC fusion protein with the Ser-Gly-Ser-Gly linker. (B) SDS-PAGE analysis of Ni-NTA purification after the SUMO protease reaction. The flowthrough and wash fractions contain cleaved PglC, and the elution fractions contain uncleaved SUMO-PglC and the cleaved SUMO domain.

Preliminary Crystallography Results

The procedure for overexpression of these SUMO-PglC fusions has been optimized using autoinduction, which increased yields greatly to 6-8 mg/L of culture. After proteolytic cleavage, PglC can be concentrated to concentrations as high as 6 mg/mL without aggregating. Samples of PglC were submitted to the Hautpman Woodward Institute for screening with their crystallography conditions specific for membrane proteins. In collaboration with Andrew Lynch, reproducible crystals have been obtained by optimizing these conditions, in the presence of magnesium and uridine diphosphate. Europium chloride and Iodo-uridine diphosphate are also being supplemented to use as heavy atoms for phasing (Figure 2-22). These crystals are currently being analyzed and further optimized to eventually obtain the crystal structure of PglC.



100 mM BisTris 6, 300 mM $MgCl_2$, 35% PEG3350, 2 mM Iodo-UDP, 1 hour soak

100mM BisTris 6, 200 mM $MgCl_2$, 1mM $EuCl_3$

Figure 2-22 Preliminary crystals of PglC obtained by Andrew Lynch.

Conclusions

Despite significant developments in the field, membrane protein crystallography remains a challenging endeavor. The first major challenge in membrane protein crystallography is the production of large amounts of stable, pure protein. This chapter describes efforts to overcome

the difficulties associated with membrane protein expression and purification, and demonstrates that reported successful strategies for some membrane proteins may not be applicable to others. Structural information about phosphoglycosyltransferases would greatly improve our understanding of the mechanisms of this diverse and important family of enzymes. Future work involves optimizing crystallography conditions in order to obtain stable crystals, which is currently underway in collaboration with the group of Professor Karen Allen at Boston University.

Acknowledgements

We are grateful to Professor Karen Allen and Andrew Lynch for their crystallography expertise. I am greatly indebted to Dr. Meredith Hartley for teaching me how to express and purify PglC when I began this project, and for helpful discussions ever since. Finally, I am thankful to Dr. Debasis Das for reading this chapter, and for pursuing future structural studies with PglC.

Experimental Methods

Expression and Purification of His6-GB1-TEV-PglC

PglC was cloned into the pGBH vector and overexpressed as described previously.[16] PglC cell pellets were thawed in 5% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, supplemented with 80 mg lysozyme and 40 μL protease inhibitor cocktail (Calbiochem). The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with 1 second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65 minutes to pellet the membranes, or cell envelope fraction

(CEF). The CEF was isolated and homogenized into 1% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, 1% Triton X-100, using a glass homogenizer. This sample was incubated with gentle rocking at 4 °C for 2-3 hours, after which it was centrifuged again (145,000 x g) to remove insoluble material. The resulting supernatant was incubated with 1 mL Ni-NTA resin for 1 hour. The resin was washed with 30 mL Wash 1 buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.05% Triton X-100, 20 mM imidazole), followed by a wash with 30 mL Wash 2 buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.05% Triton X-100, 45 mM imidazole). PglC was eluted in 4 x 3 mL fractions of elution buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.05% Triton X-100, 300 mM imidazole). Fractions containing PglC were dialyzed into 4 liters of 50 mM Tris pH 8.0, 150 mM NaCl, using 10 kDa molecular weight cut off Slide-A-lyzer dialysis cassettes (Thermo Scientific), and concentrated using 3 kDa Amicon Ultra centrifugal filters (Millipore).

Addition of PglD for X-ray crystallography

PglD was purified as previously described.[17] The PglC purification protocol was modified so that the CEF was homogenized into 50 mM Tris pH 8.0, 150 mM NaCl, 1% n-dodecyl beta-D-maltoside (DDM), and the wash and elution buffers contained 0.03% DDM instead of 0.05% Triton X-100. An equimolar amount of PglD was added to PglC immediately after elution, and the protein mixture was dialyzed using 10 kDa molecular weight cut off Slyde-A-lyzer dialysis cassettes (Thermo Scientific).

Expression and Purification of His_6-GB1-TEV-PglB

PglB was cloned into pGBH using the P1 and P2 primers (Table 2-1). The PglB-pGBH vector was transformed into BL21-Gold cells (Agilent) for overexpression, using carbenicillin for selection. Starter cultures of 5 mL were used to inoculate 1 L cultures of LB, which were incubated at 37 °C with shaking until they reached an O.D. of 0.6-0.8. Cultures were induced

with 0.5 mM IPTG, and incubated at 16 °C for 16 hours. Cells were harvested by centrifuging at 9000 x g, and cells were stored at -80 °C. Cell pellets were thawed in 5% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, supplemented with 80 mg lysozyme and 40 μL protease inhibitor cocktail (Calbiochem). The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65 minutes to pellet the CEF. The CEF was isolated and homogenized into 1% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, 1% n-dodecyl beta-D-maltoside (DDM), using a glass homogenizer. This sample was incubated in a beaker with gentle stirring at 4 °C for 12 hours , after which it was centrifuged again (145,000 x g) to remove insoluble material. The resulting supernatant was incubated with 1 mL Ni-NTA resin for 1 hour. The resin was washed with 30 mL Wash 1 buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.03% DDM, 20 mM imidazole), followed by a wash with 30 mL Wash 2 buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.03% DDM, 45 mM imidazole). PglC was eluted in 4 x 3 mL fractions of elution buffer (50 mM Tris pH 8.0, 150 mM NaCl, 0.03% DDM, 300 mM imidazole).

Expression of PglB in TaKaRa Chaperone Cell Lines

The T7-PglB-His_6 plasmid was cloned using primers P1 and P2 (Table 2-1), and transformed into cells containing the TaKaRa chaperone set (TaKaRa Bio)[19], using the recommended antibiotics for selection. The small scale test expressions were performed in 5 mL LB cultures, which were inoculated with 25 μL of starter cultures. Arabinose or tetracyclin were added to induce the expression of the chaperones. The cultures were induced with 1 mM IPTG at

an optical density between 0.6-0.8. Cultures were incubated at 37 °C for 3 hours, and centrifuged to harvest the cells. The cells were boiled in a set volume of SDS loading buffer, and analyzed by SDS-PAGE.

The large scale LB cultures (500 mL) were inoculated with 2 mL of the starter cultures. Arabinose was added to induce the expression of the chaperones, and the plasmids were selected with kanamycin and chloramphenicol. The cultures were induced with 1 mM IPTG at O.D. = 0.6-0.8, incubated at 37 °C for 2 hours, and then at 16 °C for 16 hours. Cells were harvested by centrifugation at 9000 x g and cells were stored at -80 °C. Cell pellets were thawed and resuspended in 10% of the original culture volume of 50 mM Tris pH 8.0, 100 mM NaCl. The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65 minutes to pellet the CEF. The CEF was isolated and homogenized into 1% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, 1% Triton X-100, and incubated at 4°C for 1 hour with gentle rocking. The samples were centrifuged again (145,000 x g) to remove insoluble material. Samples from each fraction were analyzed by SDS-PAGE.

Expression, Purification and Refolding of soluble PglC ($sPglC_{36}$, $sPglC_{39}$, $sPglC_{42}$)

The soluble domain of PglC was cloned into the pET-NO vector, which provides a His_8-TEV-sPglC construct, using primers P3, P4, P5 and P6 (Table 2-1). The plasmids were transformed into BL21-Gold cells (Agilent) for overexpression, using carbenicillin for selection. 5 mL starter cultures were used to inoculate 1 L cultures of LB, which were incubated at 37 °C with shaking until they reached an O.D. of 0.6-0.8. Cultures were induced with 0.5 mM IPTG,

and incubated at 16 °C for 16 hours. Cells were harvested by centrifuging at 9000 x g, and cells were stored at -80 °C. Cell pellets were thawed in 5% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl. The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. The cell lysate was centrifuged at 142,000 x g for 65 minutes. The resulting pellet was homogenized into 50 mM Tris pH 8.0, 150 mM NaCl, 6M urea, and incubated overnight at 4 °C with gentle rocking, to isolate protein from inclusion bodies. The sample was centrifuged at 37,000 x g to pellet unsolubilized material. The resulting supernatant was incubated with 2 mL Ni-NTA resin for 1 hour at 4 °C. The resin was washed with 30 mL 50 mM Tris pH 8.0, 150 mM NaCl, 6 M urea, 20 mM imidazole and then with 30 mL 50 mM Tris pH 8.0, 150 mM NaCl, 6 M urea, 45 mM imidazole. The protein was eluted with 4 x 3 mL fractions of 50 mM Tris pH 8.0, 150 mM NaCl, 6 M urea, 300 mM imidazole.

Refolding by Dialysis

Fractions containing sPglC were dialyzed using a 3 kDa molecular weight cut off Slide-A-lyzer dialysis cassette (Thermo Scientific) into 4 L 50 mM Tris pH 8.0, 150 mM NaCl, 6 M urea. The buffer was exchanged to buffer with reducing amounts of urea (6M → 3M → 1 M → 0 M) every 2 hours, to promote the refolding of the protein. Gel filtrations analysis was performed using a Superdex S200 10/300 column (GE Healthcare), equilibrated in 50 mM Tris pH 8.0, 150 mM NaCl.

Expression and Purification of soluble PglC SUMO Constructs (SUMO-sPglC$_{36}$, SUMO sPglC$_{39}$, SUMO- sPglC$_{42}$)

The three constructs of the soluble domain of PglC were cloned into the pET SUMO vector (Life Technologies) using primers P10, P11, P12 and P13 (Table 2-1). The plasmids were transformed into BL21-RIL cells (Agilent) for overexpression, using carbenicillin and chloramphenicol for selection. 5 mL starter cultures were used to inoculate 1 L cultures of LB, which were incubated at 37 °C with shaking until they reached an O.D. of 0.6-0.8. Cultures were induced with 0.5 mM IPTG, and incubated at 16 °C for 16 hours. Cells were harvested by centrifuging at 9000 x g and cells were stored at -80 °C. Cell pellets were thawed in 5% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl. The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. The cell lysate was centrifuged at 142,000 x g for 65 minutes. The resulting supernatant was incubated with 2 mL Ni-NTA resin for 1 hour. The resin was washed with 30 mL 50 mM Tris pH 8.0, 150 mM NaCl, 20 mM imidazole and then with 30 mL 50 mM Tris pH 8.0, 150 mM NaCl, 45 mM imidazole. The protein was eluted with 4 x 3 mL fractions of 50 mM Tris pH 8.0, 150 mM NaCl, 300 mM imidazole.

Activity Assay with SUMO-sPglC$_{36}$

SUMO-sPglC$_{36}$ was assayed using a coupled assay with PglA, the next enzyme in the pathway. Assays contained 20 μM Und-P, 3% DMSO, 1% Triton X-100, 30 mM Tris pH 8, 20 mM $MgCl_2$, 40 μM UDP-Bac, 18 nM PglC, 2 μM PglA, 20 μM [^{3}H]-UDP-GalNAc. 10 μL aliquots were taken at timepoints and quenched in 1 mL $CHCl_3$:MeOH. The organic layer was washed three times with 400 μL PSUP (Pure Solvent Upper Phase, composed of 15 mL $CHCl_3$,

240 mL MeOH, 1.83 g KCl in 235 mL H_2O). The resulting aqueous layers were combined with 5 mL EcoLite (MP Biomedicals) liquid scintillation cocktail. Organic layers were combined with 5 mL OptiFluor (PerkinElmer). Both layers were analyzed on a Beckman Coulter LS6500 scintillation counting system.

Expression and Purification of SUMO-PglC with DDM at pH 7.5

PglC was cloned into the pET SUMO vector using the P13 and P14 primers (Table 2-1). The plasmid was transformed into BL21-RIL cells for overexpression, and was selected using kanamycin and chloramphenicol. Starter cultures of 5 mL were used to inoculate 1 L cultures of LB, which were incubated at 37 °C with shaking until they reached an O.D. of 0.6-0.8. Cultures were induced with 1 mM IPTG, and incubated at 16 °C for 16 hours. Cells were harvested by centrifuging at 9000 x g, and cells were stored at -80 °C. Cell pellets were thawed in 5% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl, supplemented with 80 mg lysozyme and 40 μL protease inhibitor cocktail (Calbiochem). The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65 minutes to pellet CEF. The CEF was isolated and homogenized into 1% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl, 1% n-dodecyl beta-D-maltoside (DDM), using a glass homogenizer. The homogenized solution was incubated at 4 °C for 12 hours with gentle rocking, after which it was centrifuged again (145,000 x g) to remove insoluble material. The resulting supernatant was incubated with 2 mL Ni-NTA resin for 1 hour. The resin was washed with 30 mL Wash 1 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 20 mM

imidazole), followed by a wash with 30 mL Wash 2 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 45 mM imidazole). PglC was eluted in 4 x 3 mL fractions of elution buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 300 mM imidazole).

Expression and Purification of SUMO-PglC with LMNG and OGNG Amphiphiles

The pET SUMO vector containing PglC was transformed into BL21-RIL cells. 5 mL starter cultures were used to inoculate 1 L cultures of LB, which were incubated at 37 °C with shaking until they reached an O.D. of 0.6-0.8. Cultures were induced with 1 mM IPTG, and incubated at 16 °C for 16 hours. Cells were harvested by centrifuging at 9000 x g, and cells were stored at -80 °C. Cell pellets were thawed in 5% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl, supplemented with 80 mg lysozyme and 40 μL protease inhibitor cocktail (Calbiochem). The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65 minutes to pellet CEF. The CEF was isolated and homogenized into 1% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl, 1% Triton X-100 or 1% LMNG or 60 mM OGNG, using a glass homogenizer. This sample was incubated with gentle rocking at 4 °C for 12 hours, after which it was centrifuged (145,000 x g) to remove insoluble material. The resulting supernatant was incubated with 2 mL Ni-NTA resin for 1 hour. The resin was washed with 30 mL Wash 1 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.003% LMNG or 2.55 mM OGNG, 20 mM imidazole), followed by a wash with 30 mL Wash 2 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.003% LMNG or 2.55 mM OGNG, 45 mM imidazole). PglC was eluted in 4 x 3 mL fractions of elution buffer (50 mM HEPES pH 7.5,

100 mM NaCl, 0.003% LMNG or 2.55 mM OGNG, 300 mM imidazole). The purification was analyzed by SDS-PAGE. Gel filtrations analysis was performed using a Superdex S200 10/300 column (GE Healthcare), equilibrated in 50 mM HEPES pH 7.5, 100 mM NaCl, 0.003% LMNG.

Expression and Purification of T7-PglC-His_6 from *C. lari* and *C. consisus*

PglC from *C. lari* was cloned into the pET-NO vector using primers P15 and P16 (Table 2-1). PglC from *C. concisus* was cloned into the pET-NO vector using primers P18 and P19 (Table 2-1). The plasmid was transformed into BL21-RIL cells (Agilent) for overexpression, using kanamycin and chloramphenicol for selection. Starter cultures of 5 mL were used to inoculate 1 L cultures of LB, which were incubated at 37 °C with shaking until they reached an O.D. of 0.6-0.8. Cultures were induced with 0.5 mM IPTG, and incubated at 16 °C for 16 hours. Cells were harvested by centrifuging at 9000 x g, and cells were stored at -80 °C. Cell pellets were thawed in 5% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, supplemented with 80 mg lysozyme and 40 µL protease inhibitor cocktail (Calbiochem). The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65 minutes to pellet the CEF. The CEF was isolated and homogenized into 1% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl, 1% n-dodecyl beta-D-maltoside (DDM), using a glass homogenizer. This sample was incubated at 4 °C with gentle rocking for 16 hours, after which it was centrifuged again (145,000 x g) to remove insoluble material. The resulting supernatant was incubated with 1 mL Ni-NTA resin for 1 hour. The resin was washed with 30 mL Wash 1 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 20

mM imidazole), followed by a wash with 30 mL Wash 2 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 45 mM imidazole). PglC was eluted in 4 x 3 mL fractions of elution buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 300 mM imidazole).

Expression and Purification of SUMO-PglC from *C. lari* and *C. consisus*

PglC from *C. lari* was cloned into the pET-SUMO vector using primers P16 and P17 (Table 2-1). PglC from *C. concisus* was cloned into the pET-SUMO vector using primers P19 and P20 (Table 2-1). The plasmid was transformed into BL21-RIL cells (Agilent) for overexpression, using kanamycin and chloramphenicol for selection. Starter cultures of 5 mL were used to inoculate 1 L cultures of LB, which were incubated at 37 °C with shaking until they reached an O.D. of 0.6-0.8. Cultures were induced with 0.5 mM IPTG, and incubated at 16 °C for 16 hours. Cells were harvested by centrifuging at 9000 x g, and cells were stored at -80 °C. Cell pellets were thawed in 5% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, supplemented 40 µL protease inhibitor cocktail (Calbiochem). The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65 minutes to pellet the CEF. The CEF was isolated and homogenized into 1% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, 1% n-dodecyl beta-D-maltoside (DDM), using a glass homogenizer. This sample was incubated at 4 °C with gentle rocking for 16 hours, after which it was centrifuged again (145,000 x g) to remove insoluble material. The resulting supernatant was incubated with 3 mL Ni-NTA resin for 1 hour. The resin was washed with 30 mL Wash 1 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 20 mM imidazole),

followed by a wash with 30 mL Wash 2 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 45 mM imidazole). PglC was eluted in 4 x 3 mL fractions of elution buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 300 mM imidazole). Gel filtrations analysis was performed using a Superdex S200 10/300 column (GE Healthcare), equilibrated in 50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM.

Cloning of linker regions into SUMO-PglC (*C. lari* and *C. consisus*) constructs

Quikchange primers (P21, P22, P23, P24) were designed to insert Ala_5 and Ala_3 linkers between the SUMO domain and PglC. Primers P25 and P26 contained Ser-Gly-Ser-Gly linkers in the sequence of the primers, and were used to amplify the PglC genes. These PCR products were ligated into the pET SUMO vector to provide PglC from *C. concisus* with a Ser-Gly-Ser-Gly linker between the SUMO domain and PglC.

Overexpression of SUMO-SGSG-PglC (*C. consisus*) by autoinduction

The pET SUMO-SGSG-PglC plasmid was transformed into BL21-RIL cells (Agilent) for overexpression, using kanamycin and chloramphenicol for selection. Overexpression was performed using the Studier method.[27] In this method, 1 mL of an overnight cell culture was added to expression media containing 30 μg/mL kanamycin and 30 μg/mL chloramphenicol in 1 L of autoinduction media (0.1% (w/v) tryptone, 0.05% (w/v) yeast extract, 2 mM MgSO4, 0.05% (v/v) glycerol, 0.005% (w/v) glucose, 0.02% (w/v) α-lactose, 2.5 mM Na_2HPO_4, 2.5 mM KH_2PO_4, 5 mM NH_4Cl, 0.5 mM Na_2SO_4). Cells were allowed to grow with shaking for 3 h at 37 °C. After 3 h, the temperature was decreased to 16 °C, and the cells were incubated for 16 hours. Cells were harvested by centrifuging at 9000 x g, and cells were stored at -80 °C. Cell pellets were thawed in 10% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, 40 μL protease inhibitor cocktail (Calbiochem). The cells were lysed by two

rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65 minutes to pellet the CEF. The CEF was isolated and homogenized into 10% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl, 1% n-dodecyl beta-D-maltoside (DDM), using a glass homogenizer. This sample was incubated at 4 °C with gentle rocking for 16 hours, after which it was centrifuged (145,000 x g) to remove insoluble material. The resulting supernatant was incubated with 1 mL Ni-NTA resin for 16 hours. 20 µL protease inhibitor cocktail was added to prevent proteolysis The resin was washed with 30 mL Wash 1 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 20 mM imidazole), followed by a wash with 30 mL Wash 2 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 45 mM imidazole). PglC was eluted in 4 x 1 mL fractions of elution buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 300 mM imidazole).

SUMO protease cleavage of SUMO-PglC fusion proteins

Eluted protein was desalted using pre-packed HiTrap desalting columns (GE Healthcare) to remove imidazole. Desalted fractions were incubated with SUMO protease at 16 °C with gentle shaking at 80 rpm, for 6 hours. The reaction mixture was incubated with 250 µl resin for 45 minutes. The flow through and 6 x 200 µL wash fractions were collected. The Ni-NTA resin was washed with 500 mM imidazole to elute uncleaved protein and the His-tagged SUMO domain. All fractions were analyzed by SDS-PAGE to determine which fractions contained cleaved PglC.

Table 2-1 List of primers used in this study. Restriction sites are depicted in bold font.

Primer #	Primer Name **(Restriction Site)**	
P1	PglB F **(BamHI)**	5′-CGC**GGATCC**ATGAGTAAAGCCGTCAAACG-3′
P2	PglB R **(XhoI)**	5′-ATCG**CTCGAG**TTATGCCGTCCCGG TCTTGGGG-3′
P3	sPglC$_{36}$ F **(BamHI)**	5′-CGC**GGATCC**ACTCAAGGAAGTGTG-3′
P4	sPglC$_{39}$ F **(BamHI)**	5′-CGC**GGATCC**AGTGTGATTTTTA-3′
P5	sPglC$_{42}$ F **(BamHI)**	5′-CGC**GGATCC**TTTACCCAAAA-3′
P6	sPglC R **(XhoI)**	5′-ATCG**CTCGAG**TCAGTTCTTGCCA-3′
P7	sPglB$_{33}$ F **(BamHI)**	5′-CGC**GGATCC**CGCAAGAATCTG-3′
P8	sPglB$_{38}$ F **(BamHI)**	5′-CGC**GGATCC**CTCGCCCCGT-3′
P9	sPglB$_{41}$ F **(BamHI)**	5′-CGC**GGATCC**TTCTTCATTCGGGA-3′
P10	SUMO-sPglC$_{36}$ F **(BsaI)**	5′-CGCC**GGTCTCCAGGT**ACTCAAGGAAGTGTG-3′
P11	SUMO-sPglC$_{39}$ F **(BsaI)**	5′-CGCC**GGTCTCCAGGT**AGTGTGATTTTTA-3′
P12	SUMO-sPglC$_{42}$ F **(BsaI)**	5′-CGCC**GGTCTCCAGGT**TTTACCCAAAA-3′
P13	SUMO-PglC R **(XhoI)**	5′-ATCG**CTCGAG**TTATGCCGTCCCGGTCTT-3′
P14	SUMO-PglC F **(BsaI)**	5′-CGCC**GGTCTCCAGGT**ATGTATGAAAAA-3′
P15	PglC (*C. lari*) F **(BamHI)**	5′-CGC**GGATCC**ATGTATAAAAACGGTTTAAAG-3′
P16	PglC (*C. lari*) R **(XhoI)**	5′-ATCG**CTCGAG**TTAGTTGTGTCCATTGAATTT-3′
P17	SUMO-PglC (*C. lari*) F **(BsaI)**	5′-CGCC**GGTCTCCAGGT**ATGTATAAAAACGGTT TAA-3′
P18	PglC (*C. concisus*) F **(BamHI)**	5′-CGC**GGATCC**ATGTATAGAAATTTTTTAA-3′ AGAGAGT
P19	SUMO-PglC (*C. concisus*) R **(XhoI)**	5′-ATCG**CTCGAG**TTAGTTTTTGCCATTAAATTTC
P20	SUMO-PglC (*C. concisus*) F **(BsaI)**	5′-CGCC**GGTCTCCAGGT**ATGTATAGAAATTTTTTA AAGAGA-3′
P21	SUMO PglC (*C. lari*) **3 Ala** F	5′-ACCGCGAACAGATTGGAGGT**GCAGCAGCA** ATGTATAAAAACGGTTTAAA-3′

Primer #	Primer Name **(Restriction Sites)**	
P22	SUMO PglC (*C. lari*) **5 Ala** F	5′-ACCGCGAACAGATTGGAGGT**GCAGCAGCAG CGGCG**ATGTATAAAAACGGTTTAAA-3′
P23	SUMO PglC (*C. concisus*) **3 Ala** F	5′ACCGCGAACAGATTGGAGGT**GCAGCAGCA**AT GTATAGAAATTTTTTAAAGA-3′
P24	SUMO PglC (*C. concisus*) **5 Ala** F	5′-ACCGCGAACAGATTGGAGGT**GCAGCAGCAG CGGCG**ATGTATAGAAATTTTTTAAAGA-3′
P25	SUMO PglC (*C. lari*) **SGSG** F **(BsaI)**	5′CGCC**GGTCTCCAGGT**TCTGGCTCTGGGATGT ATAAAAACGGTTTA-3′
P26	SUMO PglC (*C. concisus*) **SGSG** F **(BsaI)**	5′CGCC**GGTCTCCAGGT**TCTGGCTCTGGGATGT ATAGAAATTTTTTAAAGA-3′

References

(1) Price, N. P.; Momany, F. A. Modeling Bacterial UDP-HexNAc: Polyprenol-P HexNAc-1-P Transferases. *Glycobiology* **2005**, *15* (9), 29R – 42R.

(2) Burda, P.; Aebi, M. The Dolichol Pathway of N-Linked Glycosylation. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1999**, *1426* (2), 239–257.

(3) Glover, K. J.; Weerapana, E.; Chen, M. M.; Imperiali, B. Direct Biochemical Evidence for the Utilization of UDP-Bacillosamine by PglC, an Essential Glycosyl-1-Phosphate Transferase in the Campylobacter Jejuni N-Linked Glycosylation Pathway. *Biochemistry* **2006**, *45* (16), 5343–5350.

(4) Szymanski, C. M.; Burr, D. H.; Guerry, P. Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infect Immun* **2002**, *70* (4), 2242–2244.

(5) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete genomes1. *Journal of Molecular Biology* **2001**, *305* (3), 567–580.

(6) Wagner, S.; Baars, L.; Ytterberg, A. J.; Klussmeier, A.; Wagner, C. S.; Nord, O.; Nygren, P.-Å.; Wijk, K. J. van; Gier, J.-W. de. Consequences of Membrane Protein Overexpression in Escherichia Coli. *Mol Cell Proteomics* **2007**, *6* (9), 1527–1550.

(7) Dong, M.; Baggetto, L. G.; Falson, P.; LeMaire, M.; Penin, F. Complete Removal and Exchange of Sodium Dodecyl Sulfate Bound to Soluble and Membrane Proteins and Restoration of Their Activities, Using Ceramic Hydroxyapatite Chromatography. *Analytical Biochemistry* **1997**, *247* (2), 333–341.

(8) McGregor, C.-L.; Chen, L.; Pomroy, N. C.; Hwang, P.; Go, S.; Chakrabartty, A.; Privé, G. G. Lipopeptide Detergents Designed for the Structural Study of Membrane Proteins. *Nat Biotech* **2003**, *21* (2), 171–176.

(9) Wiener, M. C. A Pedestrian Guide to Membrane Protein Crystallization. *Methods* **2004**, *34* (3), 364–372.

(10) Carpenter, E. P.; Beis, K.; Cameron, A. D.; Iwata, S. Overcoming the Challenges of Membrane Protein Crystallography. *Current Opinion in Structural Biology* **2008**, *18* (5), 581–586.

(11) Chung, B. C.; Zhao, J.; Gillespie, R. A.; Kwon, D.-Y.; Guan, Z.; Hong, J.; Zhou, P.; Lee, S.-Y. Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis. *Science* **2013**, *341* (6149), 1012–1016.

(12) Monk, B. C.; Tomasiak, T. M.; Keniya, M. V.; Huschmann, F. U.; Tyndall, J. D. A.; O'Connell, J. D.; Cannon, R. D.; McDonald, J. G.; Rodriguez, A.; Finer-Moore, J. S.; et al. Architecture of a Single Membrane Spanning Cytochrome P450 Suggests Constraints That Orient the Catalytic Domain Relative to a Bilayer. *PNAS* **2014**, *111* (10), 3865–3870.

(13) Cheng, Y.; Patel, D. J. An Efficient System for Small Protein Expression and Refolding. *Biochemical and Biophysical Research Communications* **2004**, *317* (2), 401–405.

(14) Hartley, M. D.; Morrison, M. J.; Aas, F. E.; Børud, B.; Koomey, M.; Imperiali, B. Biochemical Characterization of the O-Linked Glycosylation Pathway in Neisseria Gonorrhoeae Responsible for Biosynthesis of Protein Glycans Containing N,N'-Diacetylbacillosamine. *Biochemistry* **2011**, *50* (22), 4936–4948.

(15) Morrison, M. J.; Imperiali, B. Biochemical Analysis and Structure Determination of Bacterial Acetyltransferases Responsible for the Biosynthesis of UDP-N,N'-Diacetylbacillosamine. *J Biol Chem* **2013**, *288* (45), 32248–32260.

(16) Hartley, M. D.; Schneggenburger, P. E.; Imperiali, B. Lipid Bilayer Nanodisc Platform for Investigating Polyprenol-Dependent Enzyme Interactions and Activities. *Proc Natl Acad Sci U S A* **2013**, *110* (52), 20863–20870.

(17) Olivier, N. B.; Imperiali, B. Crystal Structure and Catalytic Mechanism of PglD from Campylobacter Jejuni. *J Biol Chem* **2008**, *283* (41), 27937–27946.

(18) Koszelak-Rosenblum, M.; Krol, A.; Mozumdar, N.; Wunsch, K.; Ferin, A.; Cook, E.; Veatch, C. K.; Nagel, R.; Luft, J. R.; DeTitta, G. T.; et al. Determination and Application of Empirically Derived Detergent Phase Boundaries to Effectively Crystallize Membrane Proteins. *Protein Sci.* **2009**, *18* (9), 1828–1839.

(19) Nishihara, K.; Kanemori, M.; Kitagawa, M.; Yanagi, H.; Yura, T. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in Escherichia Coli. *Appl Environ Microbiol* **1998**, *64* (5), 1694–1699.

(20) Marblestone, J. G.; Edavettal, S. C.; Lim, Y.; Lim, P.; Zuo, X.; Butt, T. R. Comparison of SUMO Fusion Technology with Traditional Gene Fusion Systems: Enhanced Expression and Solubility with SUMO. *Protein Sci* **2006**, *15* (1), 182–189.

(21) Chae, P. S.; Rasmussen, S. G. F.; Rana, R. R.; Gotfryd, K.; Chandra, R.; Goren, M. A.; Kruse, A. C.; Nurva, S.; Loland, C. J.; Pierre, Y.; et al. Maltose-Neopentyl Glycol (MNG) Amphiphiles for Solubilization, Stabilization and Crystallization of Membrane Proteins. *Nat Meth* **2010**, *7* (12), 1003–1008.

(22) Amer, A. O.; Valvano, M. A. The N-Terminal Region of the Escherichia Coli WecA (Rfe) Protein, Containing Three Predicted Transmembrane Helices, Is Required for Function but Not for Membrane Insertion. *J. Bacteriol.* **2000**, *182* (2), 498–503.

(23) Patel, K. B.; Ciepichal, E.; Swiezewska, E.; Valvano, M. A. The C-Terminal Domain of the Salmonella Enterica WbaP (UDP-galactose:Und-P Galactose-1-Phosphate Transferase) Is Sufficient for Catalytic Activity and Specificity for Undecaprenyl Monophosphate. *Glycobiology* **2012**, *22* (1), 116–122.

(24) Williams, P. A.; Cosme, J.; Vinković, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone. *Science* **2004**, *305* (5684), 683–686.

(25) Nothaft, H.; Scott, N. E.; Vinogradov, E.; Liu, X.; Hu, R.; Beadle, B.; Fodor, C.; Miller, W. G.; Li, J.; Cordwell, S. J.; et al. Diversity in the Protein N-Glycosylation Pathways Within the Campylobacter Genus. *Mol Cell Proteomics* **2012**, *11* (11), 1203–1219.

(26) Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P. X-Ray Structure of a Bacterial Oligosaccharyltransferase. *Nature* **2011**, *474* (7351), 350–355.

(27) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr. Purif.* **2005**, *41* (1), 207–234.

Chapter 3

Functional Analysis of a Monotopic Phosphoglycosyltransferase Using Bioinformatics, Mutagenesis, and Structure Modeling

The work described in this chapter was performed in collaboration with Lingqi Luo, a graduate student in the lab of Professor Karen Allen at Boston University. Lingqi performed the sequence alignments and generated the EVfold predicted structural model of PglC.

Introduction

Phosphoglycosyltransferases (PGTs) are a family of enzymes that catalyze the transfer of a sugar-1-phosphate from a nucleotide-activated donor to a polyprenol-phosphate acceptor substrate. This family encompasses the previously described polyisoprenyl-phosphate *hexose*-1-phosphate transferase (PHPT) and polyisoprenyl-phosphate *N*-acetylaminosugar-1-phosphate transferase (PNPT) family of enzymes.[1,2] The products of PGT reactions are elaborated into complex polyprenol-diphospho-glycans, which serve as glycan donors in pathways such as the biosynthesis of glycoproteins and glycolipids. Although most identified PGTs are bacterial, there are important eukaryotic examples such as Alg7, which by catalyzing the biosynthesis of dolichol-PP-GlcNAc, initiates the N-linked protein glycosylation pathway in all eukaryotes from yeast to man.[3]

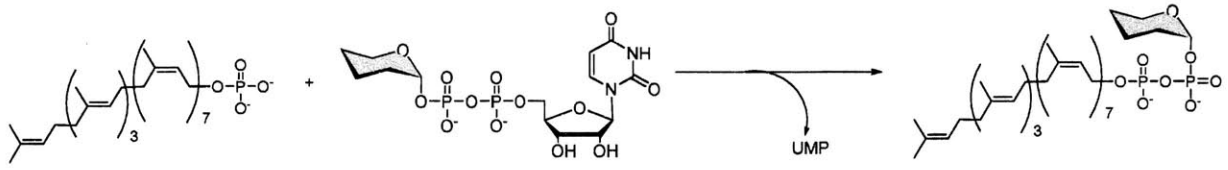


Figure 3-1 Phosphoglycosyltransferase reaction shown with a generic UDP-carbohydrate substrate and undecaprenol phosphate.

PGTs catalyze reactions at the membrane interface, between a soluble sugar substrate and a membrane-bound polyprenol phosphate to form a membrane-bound product (Figure 3-1). PGTs are integral membrane proteins and can be organized into subfamilies based on their membrane topologies. The best known PGTs include members of the WecA subfamily and MraY subfamily. These are polytopic membrane proteins, with multiple transmembrane helices. To date, MraY from *Aquifex aeolicus* is the only PGT with a solved crystal structure.[4]

Conservation mapping onto the crystal structure, revealed three conserved aspartate residues and one conserved histidine residue in the predicted active site. Mutagenesis analysis of the aspartate residues implicates these residues in the mechanism of action of MraY.[5]

In conjunction with studies on prokaryotic protein glycosylation we have identified a distinct structural family of prokaryotic PGTs, typified by the *Campylobacter jejuni* PGT known as PglC. PglC catalyzes the reaction between UDP-diNAcBac and undecaprenol-phosphate, forming undecaprenol-P-P-diNAcBac with the release of UMP as a by-product.[6] PglC-like proteins are Type I membrane proteins, classified by the location of the C-terminal domain in the cytosol. Homologs of PglC are typified by monotopic structures with a short N-terminal periplasmic domain, a single transmembrane helix, and a globular C-terminal domain. In contrast with polytopic PGTs such as MraY and WecA, the PglC subfamily of PGTs are small (approximately 200 amino acids in length). Importantly, this subfamily does not contain many of the conserved carbohydrate recognition domains or metal binding motifs found in the larger PGT families.

Functional and bioinformatics analysis reveals two other related families, which include the primary sequence of the small PGT embedded into a more complex framework (Figure 3-2). One of these families is typified by the bifunctional PglB(*Ng*) from *N. gonorrhoeae*, which features an N-terminal PGT domain, highly homologous to the *C. jejuni* PglC, and a C-terminal domain acetyltransferase domain that catalyzes acetylation of the UDP-4-amino sugar precursor to UDP-diNAcBac.[7] The other family is exemplified by WbaP from *Salmonella enterica*, which catalyzes the transfer of galactose-1-P to undecaprenol-phosphate.[8] Enzymes in this family are polytopic, with a total of five transmembrane helices; however, the first four transmembrane helices are not essential for catalytic activity.[9,10] The functional C-terminal domain contains a

single transmembrane helix and a cytosolic globular domain, strongly resembling the predicted architecture of PglC. A recent study with WcaJ, an *E. coli* enzyme that is a member of the WbaP family, used PhoA/LacZ fusions and cysteine residue labeling to investigate the cellular localization of loops and the globular domain.[10] The results of this study suggest that the fifth transmembrane helix forms a hairpin bend in the membrane. As similar studies have not been performed with PglC, it is unclear whether this is also the case for the PglC family of PGTs.

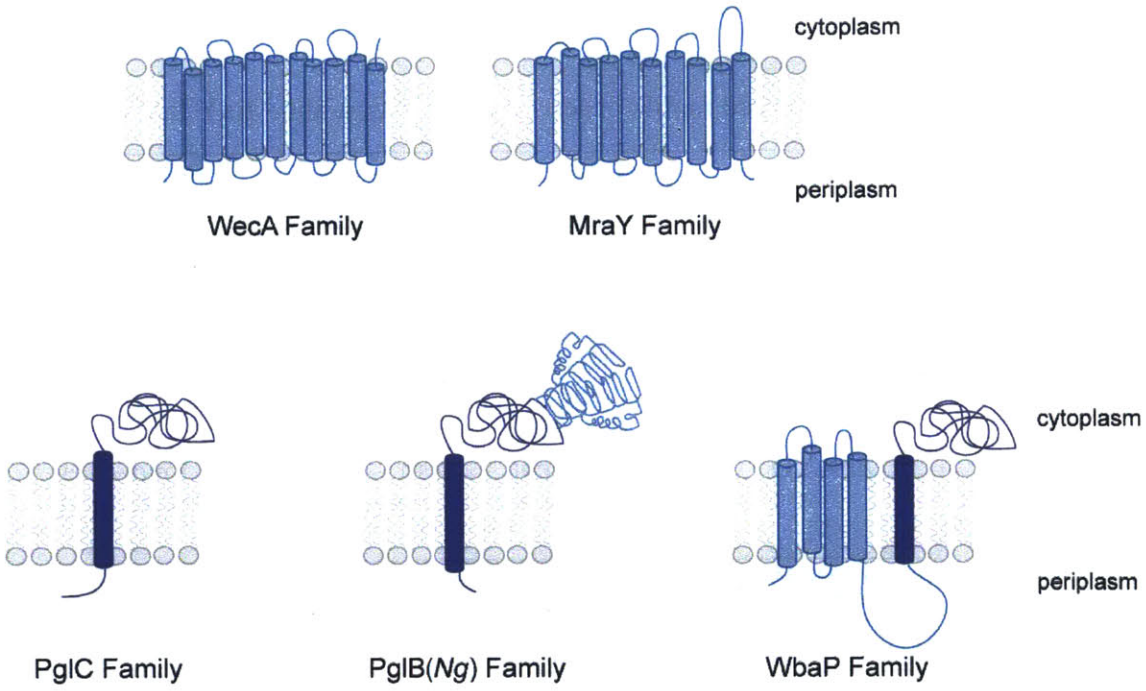


Figure 3-2 Predicted topologies of different families of phosphoglycosyltransferases. The PglC-like domain is shown in dark blue.

This chapter summarizes our bioinformatics investigations of PglC and the additional information gained by expanding our analysis to related PGT families that contain monotopic PglC-like domains. This analysis sets the stage for mutagenesis studies, allowing us to identify residues critical for catalytic activity.

Results and Discussion

Bioinformatics analysis of the WecA and MraY families of PGTs have revealed conserved signature sequences in loops between the transmembrane helices, which are implicated in both substrate binding and catalysis (Table 3-1). The significance of these motifs was confirmed by mutating key residues in these conserved sequences and analyzing the activity of the resulting mutant proteins. It is important to note that none of these motifs are observed in PglC.

Table 3-1 Summary of conserved motifs observed in the WecA and MraY familes of PGTs

PGT Family	Motif	Proposed Function	Reference
WecA	DDXXD	Binding to metal cofactor	Lehrer et al[11]
	HIHH	Carbohydrate recognition	Anderson et al[12], Amer et al[13]
	VFMGD	Binding and/or recognition of the nucleotide moiety of the nucleoside phosphate precursor	Furlong et al[14]
MraY	DDXXD	Adjacent Asp residues proposed to bind metal cofactor	Lloyd et al[5], Chung et al[4]
	MAPIHHHFEL	Carbohydrate recognition	Price et al[2]
	VFMGD	Asp proposed to be a catalytic nucleophile	Lloyd et al[5]

An alignment of sequences from the PglC family reveals that roughly 25% of the sequence is conserved (data not shown). However, sequence alignments of the PglC domain from the PglC, PglB(*Ng*) and WbaP families reveals highly conserved residues which would not have been apparent from an alignments of any of the families individually (Figure 3-3). Based on these results, we selected conserved residues for analysis by mutagenesis (Table 3-2). Our selection includes negatively charged aspartate and glutamate residues (D92, E93, D156, D168) as there is evidence that these residues may coordinate metal ion cofactors or be important

nucleophiles in the WecA and MraY families of PGTs.[5,15] We also selected conserved positively charged arginine residues (R87, R111), which may be involved in interactions with the phosphate groups of substrates and products. In addition, we were intrigued by the presence of a highly conserved proline residue in the transmembrane helix (P24), as prolines are traditionally believed to be helix breaking residues. A conserved methionine residue (M62) was also investigated, as it may serve an important structural role. Finally, two non-conserved glutamate residues (E65, E116) were mutated to glutamine residues, as controls for mutagenesis analysis.

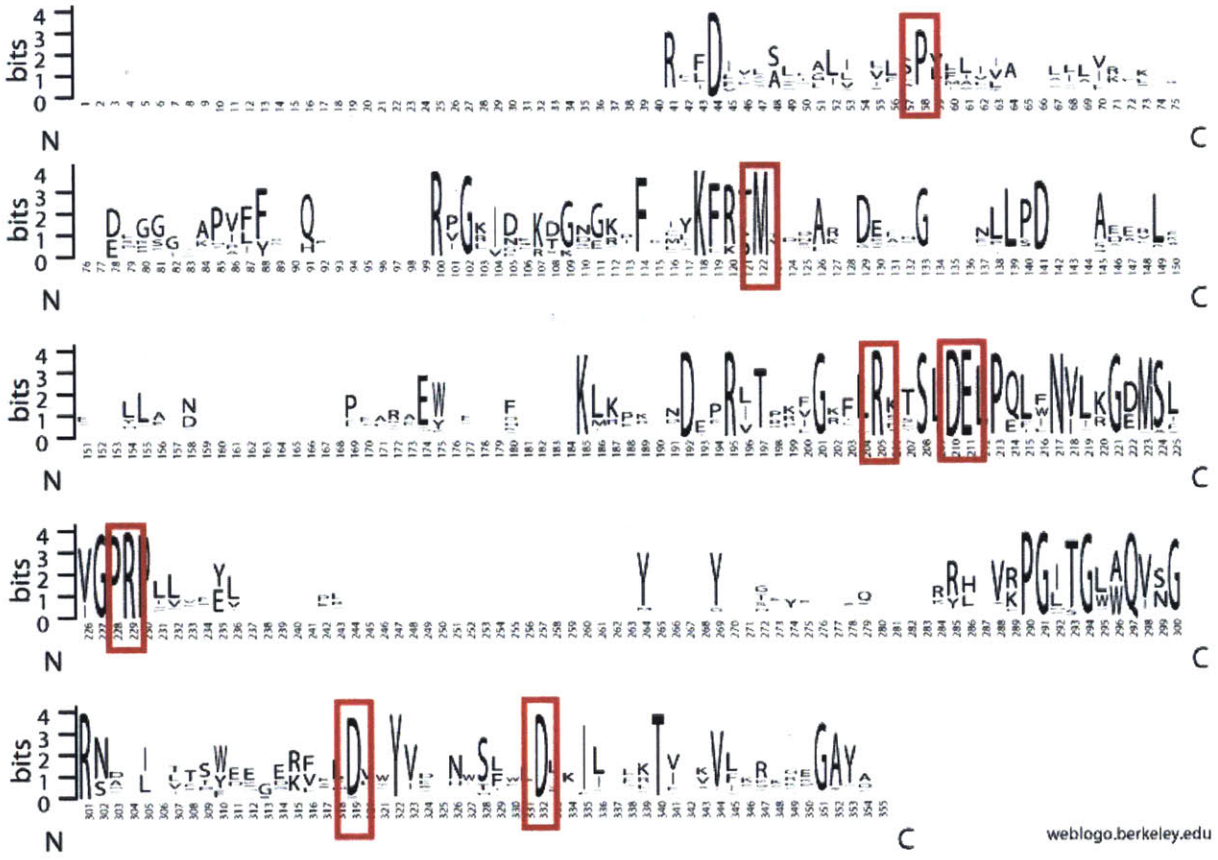


Figure 3-3 Sequence alignment of the PglC-like domain from the PglC, PglB(*Ng*) and WbaP families of PGTs. The alignment was created using WebLogo.[16] Residues marked with a red box were pursued for mutagenesis studies.

Table 3-2 Summary of residues selected for mutagenesis analysis

Mutation	Predicted Role of Mutated Residue
P24A, P24W	Interactions with Und-P substrate
M62Q, M62I	Structural
E65Q	Control
R87Q	Interactions with phosphate groups
D92A	Coordination of metal cofactor
E93Q	Coordination of metal cofactor
R111Q	Interactions with phosphate groups
D156A	Coordination of metal cofactor/Catalytic residue
D168A	Coordination of metal cofactor/Catalytic residue
E116Q	Control

Wild-type and mutants of PglC were expressed as a SUMO fusion to help maintain protein solubility during overexpression and purification. The SUMO tag greatly improved the solubility of wild-type PglC and the SUMO-PglC fusion exhibits good catalytic activity. All proteins were purified to homogeneity and analyzed by SDS-PAGE for purity, and were analyzed by gel chromatography to assess their levels of aggregation (Figure 3-4). Enzyme activity was quantified by an extraction based assay using undecaprenol phosphate and [^{3}H]-UDP-diNAcBac. Purified PglC mutants were tested at a maximum of hundred times the concentration of the wild-type enzyme to determine whether activity could be restored at these higher concentrations.

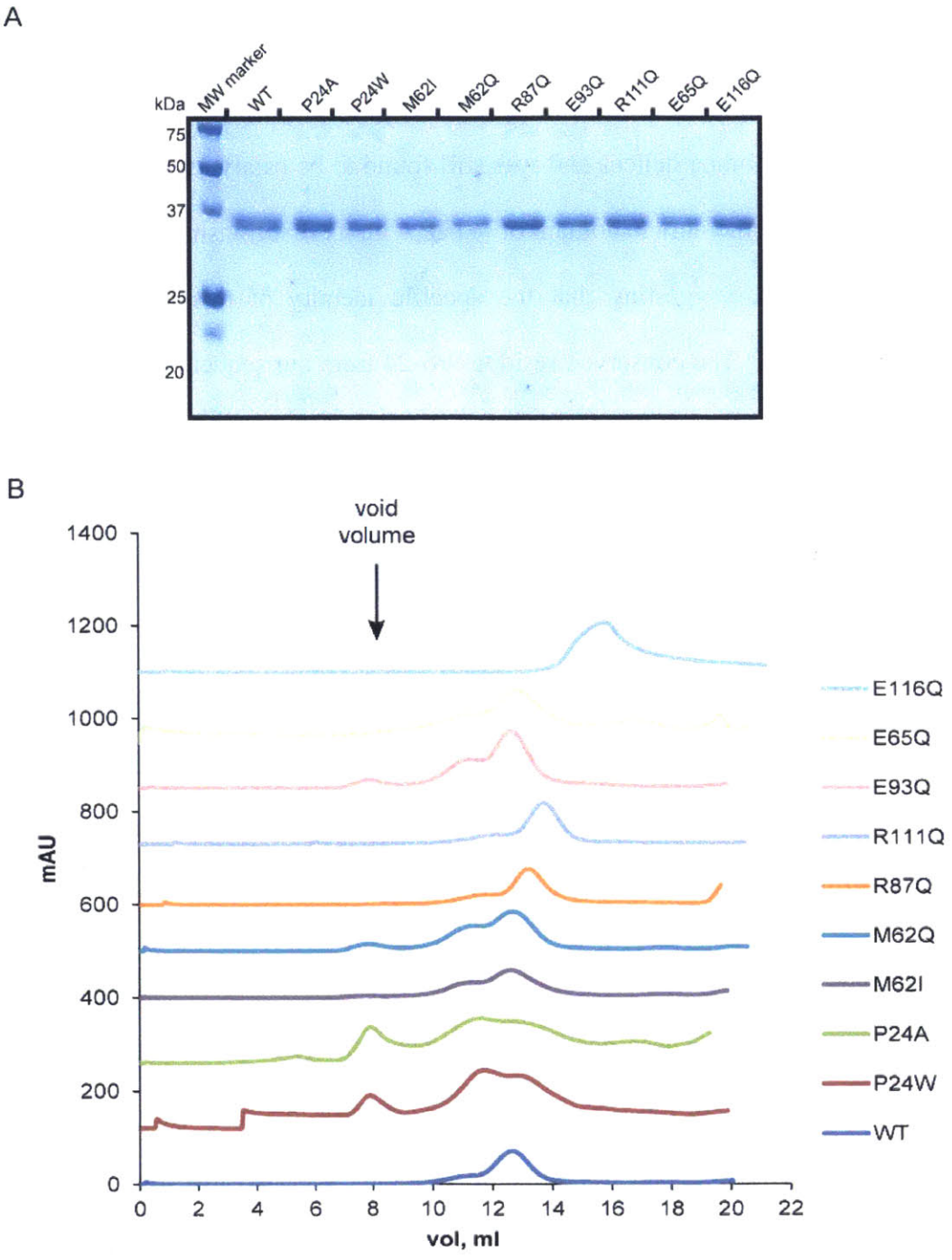


Figure 3-4 Analysis of PglC mutants. (A) SDS-PAGE analysis to assess the purity of each mutant. (B) Gel filtration analysis of the mutants.

Deletion studies have previously been used to provide insight into the importance of the transmembrane domains in many families of PGTs. For example, WbaP was expressed without the first four transmembrane helices and was still found to be catalytically active.[17] Importantly, replacing the fifth transmembrane domain of WbaP with its first transmembrane helix resulted in non-functional protein, suggesting that the specific identity of this transmembrane helix is important for catalysis.[8] The conserved residue Pro-24 from our sequence alignments is part of a sequence that strongly resembles the polyisoprene recognition sequence (PIRS), LL(F/I)IXFXXIPFXFY, which is known to be important for interactions with polyprenol phosphates.[4,18] Interestingly, our analysis also revealed that a tryptophan residue replaces this proline residue in approximately 5% of the sequences. We therefore decided to mutate Pro-24 to both alanine and tryptophan to observe the effects of these mutations on enzyme activity. Our biochemical assay results suggest that this proline residue is important, as the P24A mutant showed no activity. However, substituting the proline with tryptophan results in retention of activity at slightly lower levels than the wild-type enzyme (Table 3-3).

Mutagenesis analysis of the conserved arginine residues, Arg-87 and Arg-111 resulted in catalytically inactive PglC. These positively charged residues may interact with the negatively charged phosphate groups of the substrates and products of this reaction, or may participate in salt bridges required to maintain the structure of PglC. We also investigated the role of the conserved methionine residue (M62), hypothesizing that it may play a role in maintaining protein structure. When mutated to glutamine, a polar residue, activity decreased significantly; however, mutating Met-62 to isoleucine retained activity at almost the same level as the wild-type protein (Table 3-3).

Table 3-3 Summary of activity assays performed with PglC mutants. Activity is defined as follows: Wild-type levels of activity (+++), active at 30 nM enzyme (++), active at 300 nM enzyme (+), inactive at 300 nM enzyme (-). The aspartate residues were used as cell envelope fractions (CEFs) in activity assays.

Mutation	Predicted Role of Mutated Residue	Activity
WT		+++
P24A	Interactions with Und-P	-
P24W	Interactions with Und-P	++
M62Q	Structural	+
M62I	Structural	++
E93Q	Coordination of metal cofactor	-
R87Q	Interactions with phosphate groups	-
R111Q	Interactions with phosphate groups	-
E116Q	Control	+
E65Q	Control	++

Aspartic acid residues are implicated in metal ion cofactor coordination in the WecA and MraY families of proteins.[15] Sequences in both families contain adjacent aspartic acid residues (D90/91 in *E. coli* WecA and D115/D116 in *E. coli* MraY). The adjacent aspartate residues resemble the conserved DDXXD motif found in other enzymes with diphosphate substrates, such as prenyl transferases, where the DD pair coordinates the Mg^{2+} cofactor.[19] The WbaP family does not contain this DD motif, but does have sequential aspartate and glutamate residues (D382/E383), which when mutated, abolishes enzyme activity. Our sequence alignment highlighted the corresponding residues in PglC (D92/E93), and our biochemical analysis also shows a complete loss of activity when either of these residues are mutated.

Aspartate residues may also have nucleophilic roles in PGTs. For example, a conserved aspartic acid residue in the fourth cytoplasmic loop of MraY is thought to be a nucleophile initiating the reaction by forming a covalent phosphosugar-enzyme intermediate.[5] We observed

two additional conserved aspartic acid residues in our sequence alignments (D156 and D168), and mutation of either of these residues to alanine results in loss of activity.

Interestingly, all the mutants in which aspartic acid residues that were mutated to alanine resulted in misfolded or aggregated protein when analyzed by gel chromatography (Figure 3-5). The same was observed when these aspartic acid residues were mutated to asparagine residues (Figure 3-5). We hypothesized that mutation of these key residues might fundamentally alter the structure of PglC, destabilizing it in solution. Similar results were observed in mutagenesis studies of the C-terminal domain of WbaP, in which mutations of certain aspartic acid residues resulted in a shift in gel migration, suggesting an altered structure.[20] It is possible that isolating these proteins from the cell envelope fraction (CEF) into detergent micelles during purification might contribute to protein instability *in vitro*. In order to preserve the structure of these mutants as much as possible, they were additionally assayed for catalytic activity in their CEF forms. However, even in this more stable form, the aspartate mutants were inactive (Table 3-4).

Table 3-4 Summary of activity assays performed with PglC aspartate mutants CEFs. Activity is defined as follows: wild-type enzyme levels of activity (+++), mutant enzyme inactive with 100x CEF as compared to wild-type (-), estimated by gel analysis.

Mutation	**Predicted Role of Mutated Residue**	**Activity**
WT		+++
D92A (CEF)	Coordination of metal cofactor	-
D156A(CEF)	Coordination of metal cofactor/Catalytic residue	-
D168A (CEF)	Coordination of metal cofactor/Catalytic residue	-

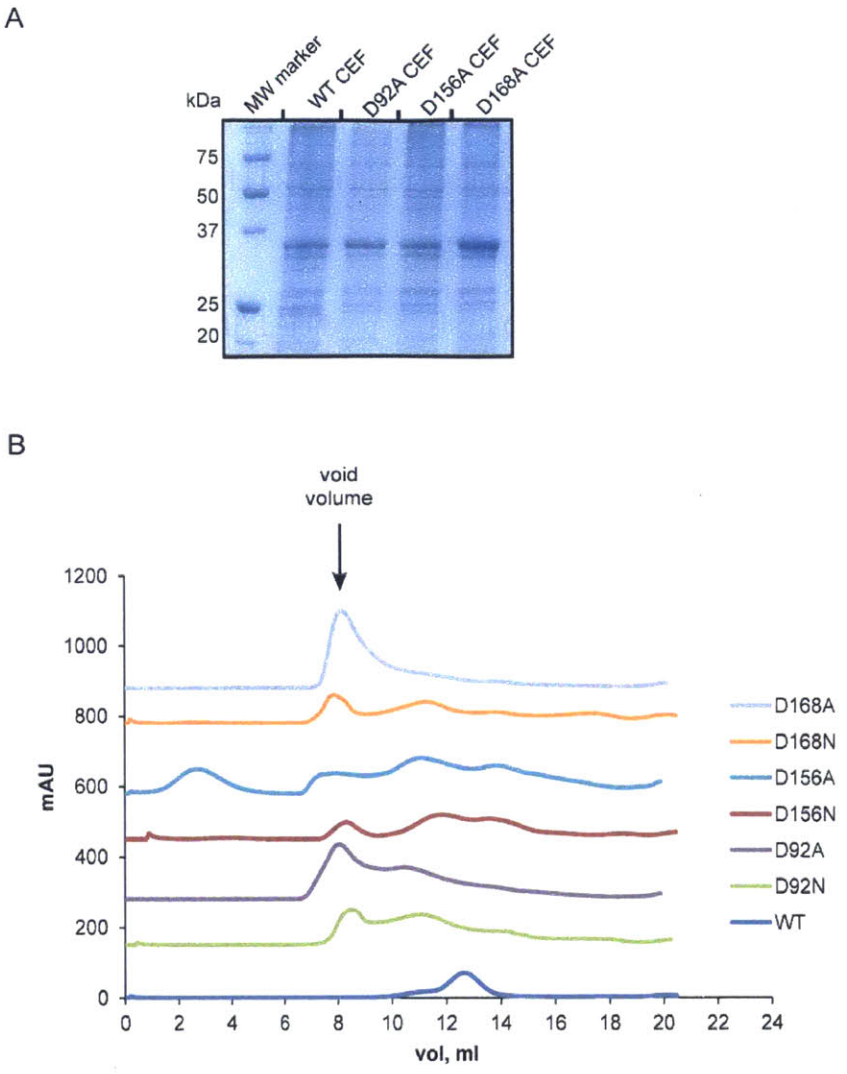


Figure 3-5 Analysis of PglC aspartate mutants. (A) SDS-PAGE analysis to assess the purity of the cell envelope fractions of each aspartate mutant. (B) Gel filtration analysis of the purified mutants.

A similar bioinformatics analysis of the C-terminal domain of WbaP from *Salmonella enterica* compared several hundreds of WbaP homologs to identify conserved residues.[20] The most conserved charged or polar amino acids were mutated to alanine residues and mutant proteins were evaluated using *in vivo* complementation assays to observe the formation of LPS O-antigen. Unsurprisingly, there is overlap in the identities of conserved residues discovered in those experiments and in this study. Both studies identified the adjacent Asp-Glu pair as being

critical for activity (D92/E93 in the *C. jejuni* PglC), as well as an additional conserved aspartate residue more C-terminal in the sequence (D168 in PglC, corresponding to D458 in WbaP). Both studies also revealed a conserved arginine residue (R111 in PglC, R401 in WbaP) that is essential for catalysis.

Based on the bioinformatics analysis, a predicted 3D structure was constructed for the soluble domain of PglC using EVfold.[21] This method uses multiple sequence alignments and a maximum entropy model of the protein sequence to obtain residue pair couplings. These couplings are used to calculate residue-residue proximity in folded protein structures in conjunction with distance constraints, secondary structure predictions and molecular dynamics simulations, to yield predicted structures. As a control for the prediction for PglC, two similarly sized proteins were modeled and the resulting predicted structures varied from the experimentally observed structures within 3-5 Å C_α-RMSD (data not shown).

Seven thousand homologs of the PglC-like domain were used to generate predicted structures of PglC, of which the top ranking structure is shown (Figure 3-6A). The sequences used to generate this model lacked adequate homology for the very last 20 residues; therefore they appear as an unstructured loop in the predicted model. The prediction algorithms folded the transmembrane helix into the structure of the rest of the protein, resulting in a structure which is not physiologically relevant. Thus, the transmembrane domain was not included, and the prediction was limited to the soluble domain of PglC. We were encouraged by the presence of many of the residues identified in our mutagenesis studies in proximity of each other, suggesting they might form the active site of the protein (Figure 3-6B). Additionally, the non-conserved residues used as controls in our mutagenesis studies were far away from the apparent active site of the protein.

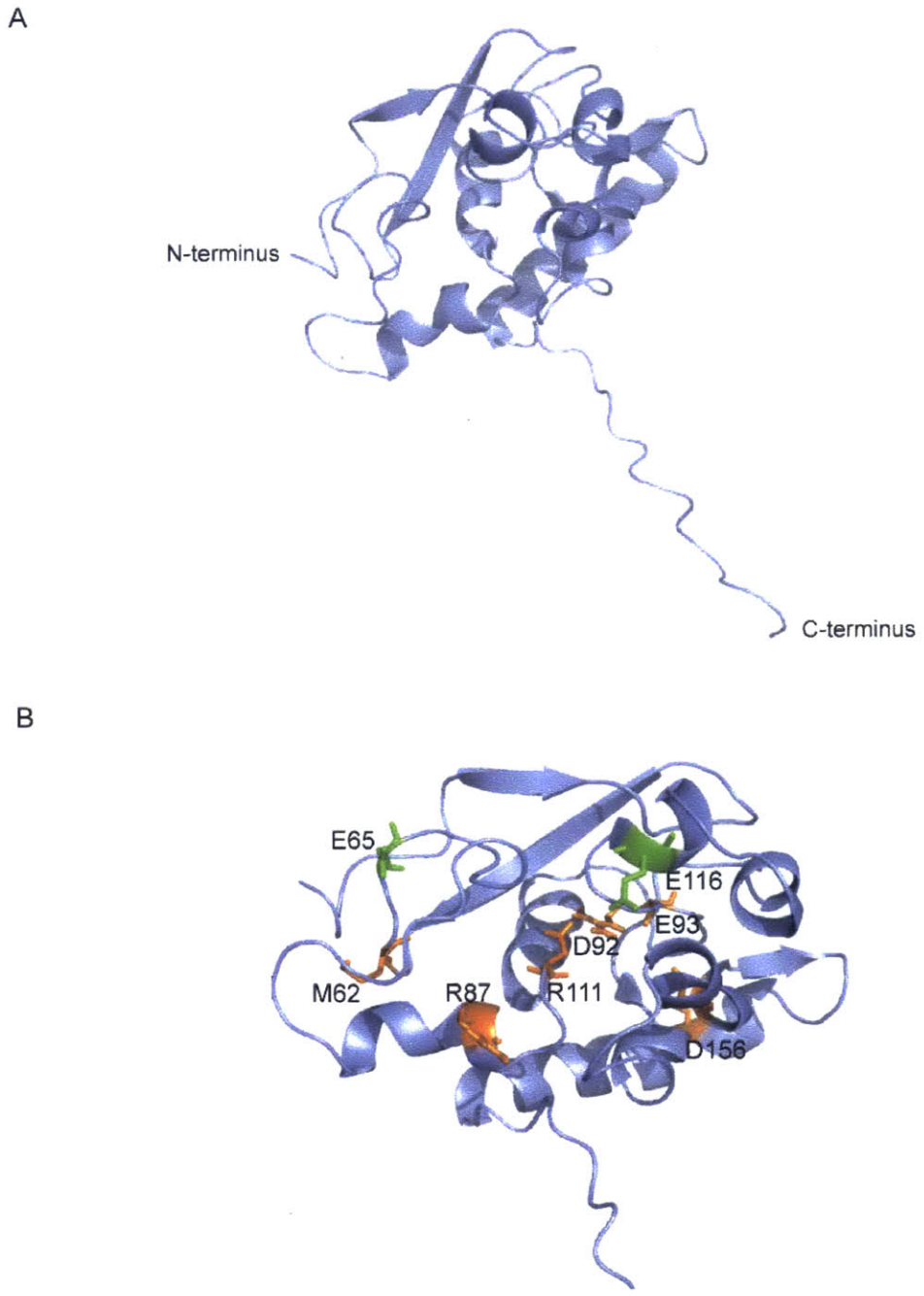


Figure 3-6 Predicted structure of the soluble domain of PglC generated by EVfold. (A) Highest ranked predicted structure. (B) Predicted structure showing residues mutated in this study. Conserved residues are depicted in orange, and non-conserved residues are shown in green.

We examined the domain between residues 160-180 in the PglC structure, which we had hypothesized formed an amphipathic helix as discussed in Chapter Two. Indeed, when compared to the helical wheel generated from the sequence, the predicted structure predicts an alpha helix for this domain, in which the hydrophobic residues line one face of the helix (Figure 3-6C,D).

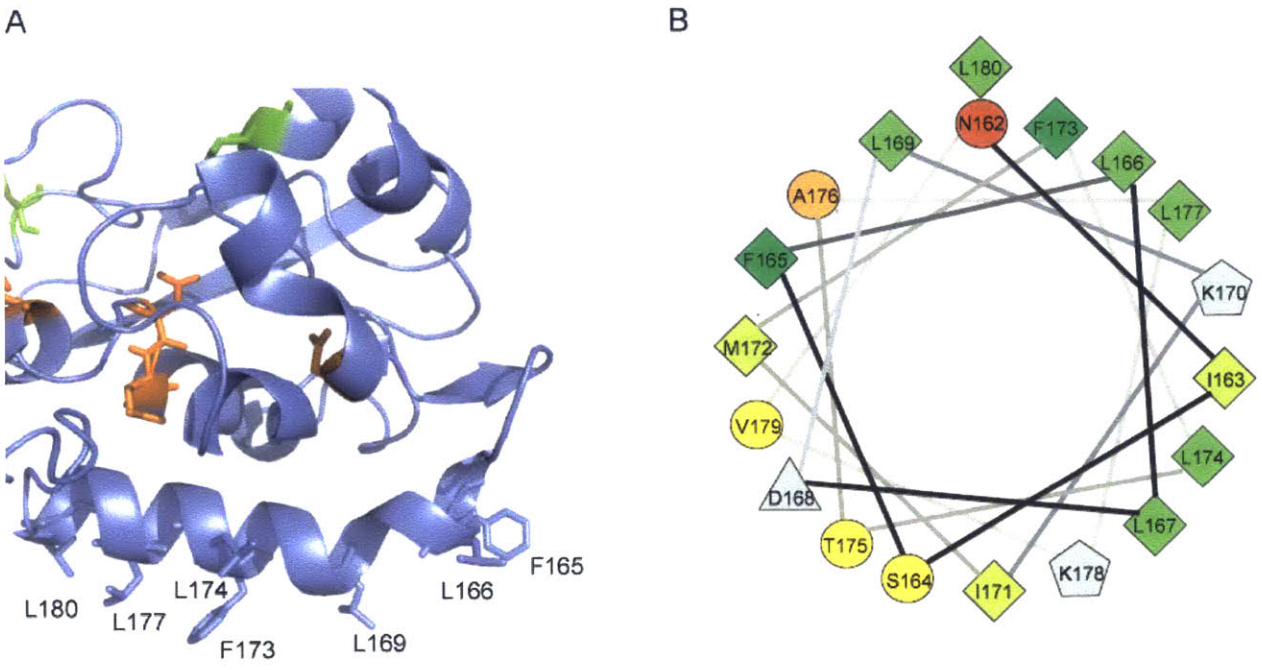


Figure 3-7 Structure of the predicted amphipathic helix in PglC (A) Predicted amphipathic helix of the highest ranked structure, with hydrophobic residues represented as sticks. (D) Helical wheel predicted for the amphipathic helix using the prediction software created by Don Armstrong and Raphael Zidovetzki: Version Id: wheel.pl,v 1.4 2009-10-20 21:23:36 don Exp. The most hydrophobic residue is shown in green, and the amount of green decreases proportionally to the hydrophobicity, with zero hydrophobicity coded as yellow. Hydrophilic residues are coded red with pure red being the most hydrophilic (uncharged) residue.

Conclusions

In the absence of a crystal structure, bioinformatics and sequence alignments are powerful methods that can provide information about the important amino acids in a protein. Our sequence analysis was greatly improved by including the PglC-like domains of the PglB (*Ng*)

and the WbaP families, as the alignment of all three families emphasized residues that may not have been revealed by analyzing the sequences of the PglC family alone. We were able to correlate residues identified from this analysis to residues in proximity of each other in the predicted model, suggesting that they participate in the activity of the enzyme.

Our previous efforts to improve the stability of PglC, described in Chapter Two, greatly aided these studies, as we were able to characterize the activity of almost all the mutants in their purified form. These results serve as a guide for future studies involving the labeling of PglC for biophysical analysis which are currently underway in our lab. The sequence alignments and subsequent mutagenesis provide valuable information about which residues can tolerate amino acid substitutions and which residues must be avoided for such purposes because they are essential for activity. Thus, the work described in this chapter sets the stage for future investigations of PglC, which will ultimately provide insight into the phosphoglycosyltransferase family of enzymes.

Acknowledgements

I am grateful to Lingqi Luo and Professor Karen Allen, for bringing their structural and computational expertise to this project. I would also like to thank Meredith Hartley and Monika Musiel-Siwek for reading this chapter and providing helpful comments.

Experimental Methods

Cloning of SUMO-PglC Mutants

PglC had previously been cloned into the pET SUMO vector as described in Chapter Two. Mutations were introduced into this sequence using the primers listed in Table 3-5, using the Quikchange protocol.

Table 3-5 Primers used to create PglC mutants.

Mutation	**Primer Sequence** (Altered codon in red)
P24A F	5′-TTTAGCTTTAGTGCTTTTAGTGCTTTTTTCTGCGGTGATTTTAATCA-3′
P24A R	5′-GTGATTAAAATCACCGCAGAAAAAAGCACTAAAAGCACTAAAGC-3′
P24W F	5′-GCTTTTAGTGCTTTTTTCTTGGGTGATTTTAATCACTGCT-3′
P24W R	5′-AGCAGTGATTAAAATCACCCAAGAAAAAAGCACTAAAAGC-3′
M62Q F	5′-TAAATTTAAAACCCAGAGCGATGAAAGAGATGAAAAGGGT-3′
M62Q R	5′-ACCCTTTTCATCTCTTTCATCGCTCTGGGTTTTAAATTTA-3′
M62I F	5′-TAAATTTAAAACCATAAGCGATGAAAGAGATGAAAAGGGT-3′
M62QI R	5′-ACCCTTTTCATCTCTTTCATCGCTTATGGTTTTAAATTTA-3′
R87Q F	5′-GCTTTTGGAAAAATCGTTCAAAGCTTAAGTTTGGATGAGC-3′
R87Q R	5′-GCTCATCCAAACTTAAGCTTTGAACGATTTTTCCAAAAGC-3′
D92N F	5′-AAATCGTTAGAAGCTTAAGTTTGAATGAGCTTTTGCAACTTTTTAATGTT-3′
D92N R	5′-AACATTAAAAAGTTGCAAAAGCTCATTCAAACTTAAGCTTCTAACGATTT-3′
D156N F	5′-TTCTTGGCAGAAAAAATTCGAACTTAATGTGTATTATGTG-3′
D156N R	5′-CACATAATACACATTAAGTTCGAATTTTTTCTGCCAAGAA-3′
D168N F	5′-GTGAAAAATATTTCTTTTTTGCTTAATTTAAAAATCATGTTTTTAACAG-3′
D168N R	5′-CTGTTAAAAACATGATTTTTAAATTAAGCAAAAAAGAAATATTTTTCAC-3′
D92A F	5′-AAATCGTTAGAAGCTTAAGTTTGGCGGAGCTTTTGCAACTTTTTAATGTT-3′
D92A R	5′-AACATTAAAAAGTTGCAAAAGCTCCGCCAAACTTAAGCTTCTAACGATTT-3′
D156A F	5′-TTCTTGGCAGAAAAAATTCGAACTTGCGGTGTATTATGTG-3′
D156A R	5′-CACATAATACACCGCAAGTTCGAATTTTTTCTGCCAAGAA-3′
D168A F	5′-GTGAAAAATATTTCTTTTTTGCTTGCGTTAAAAATCATGTTTTTAACAGCTTTAAAGG-3′
D168A R	5′-CCTTTAAAGCTGTTAAAAACATGATTTTTAACGCAAGCAAAAAAGAAATATTTTTCAC-3′
E93Q F	5′-AAATCGTTAGAAGCTTAAGTTTGGATCAGCTTTTGCAACTTTTTAATGTT-3′
E93Q R	5′-AACATTAAAAAGTTGCAAAAGCTGATCCAAACTTAAGCTTCTAACGATTT-3′
R111Q F	5′-ATATGAGTTTTGTTGGACCTCAACCTCTTTTGGTTGAGTATTTGCCTCTT-3′
R111Q R	5′-AAGAGGCAAATACTCAACCAAAAGAGGTTGAGGTCCAACAAAACTCATAT-3′
R144Q F	5′-ATGGGCGCAGGTAAATGGTCAAAATGCTATTTCTTGGCAG-3′

R144Q R	5′-CTGCCAAGAAATAGCATTTTGACCATTTACCTGCGCCCAT-3′
W149L F	5′-GTAAATGGTAGAAATGCTATTTCTCTGCAGAAAAAATTCGAAC TTGATGT-3′
W149L R	5′-ACATCAAGTTCGAATTTTTTCTGCAGAGAAATAGCATTTCTAC CATTTAC-3′
E65Q F	5′-AAAACCATGAGCGATCAAAGAGATGAAAAGG-3′
E65Q R	5′-CCTTTTCATCTCTTTGATCGCTCATGGTTTT-3′
E116Q F	5′-CCTAGACCTCTTTTGGTTCAGTATTTGCCTCTTTAC-3′
E116Q R	5′-GTAAAGAGGCAAATACTGAACCAAAAGAGGTCTAGG-3′

Overexpression of Wild-Type and SUMO-PglC Mutants

PglC was cloned into the pET-SUMO vector using P13 and P14 primers (Table 2-1) described in Chapter Two. The pET-SUMO-PglC plasmid was transformed into BL21-RIL cells (Agilent) for overexpression, using kanamycin and chloramphenicol for selection. Overexpression was performed using the Studier method.[22] In this method, 1 mL of an overnight cell culture was added to expression media containing 30 μg/mL kanamycin and 30 μg/mL chloramphenicol in 1 L of autoinduction media (0.1% (w/v) tryptone, 0.05% (w/v) yeast extract, 2 mM MgSO4, 0.05% (v/v) glycerol, 0.005% (w/v) glucose, 0.02% (w/v) α-lactose, 2.5 mM Na_2HPO_4, 2.5 mM KH_2PO_4, 5 mM NH_4Cl, 0.5 mM Na_2SO_4). Cells were allowed to grow with shaking for 3 h at 37 °C. After 3 h, the temperature was decreased to 16 °C, and the cells were incubated for 16 hours. Cells were harvested by centrifuging at 9000 x g, and cells were stored at -80 °C. Cell pellets were thawed in 10% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, 40 μL protease inhibitor cocktail (Calbiochem). The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65

minutes to pellet the CEF. If the CEF was to be used for activity assays, it was homogenized into 1% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl and stored at -80 °C. If protein was to be purified from the CEF, it was isolated and homogenized into 10% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl, 1% n-dodecyl beta-D-maltoside (DDM), using a glass homogenizer. This sample was incubated at 4 °C with gentle rocking for 16 hours, after which it was centrifuged (145,000 x g) to remove insoluble material. The resulting supernatant was incubated with 1 mL Ni-NTA resin for 16 hours. 20 µL protease inhibitor cocktail was added to prevent proteolysis The resin was washed with 30 mL Wash 1 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 20 mM imidazole), followed by a wash with 30 mL Wash 2 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 45 mM imidazole). PglC was eluted in 4 x 1 mL fractions of elution buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 300 mM imidazole). Gel filtration analysis was performed using a Superdex S200 10/300 column (GE Healthcare) equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM.

Activity Assays with Wild-Type PglC and PglC Mutants

Wild-type PglC and PglC mutants were assayed using a radioactive extraction-based assay. Assays contained 16 µM Und-P, 2.75% DMSO, 0.2% Triton X-100, 30 mM Tris pH 8, 5 mM $MgCl_2$, 7.5 µM [^{3}H]-UDP-diNAcBac, and 3 nM PglC in a final volume of 120 µL. Aliquots (10 µL) were taken at defined time points and quenched in 1 mL 2:1 $CHCl_3$:MeOH (v/v). The organic layer was washed three times with 400 µL PSUP (Pure Solvent Upper Phase, composed of 15 mL $CHCl_3$, 240 mL MeOH, 1.83 g KCl in 235 mL H_2O). The resulting aqueous layers were combined with 5 mL EcoLite (MP Biomedicals) liquid scintillation cocktail. Organic layers were combined with 5 mL OptiFluor (PerkinElmer). Both layers were analyzed on a Beckman

Coulter LS6500 scintillation counting system. When CEFs were used instead of pure protein, the CEFs were added at a final concentration of 8% (v/v).

Conservation Analysis

For each family containing a PglC-like domain, a stringent set of homologous sequences was parsed out from its network by selecting the original target and its neighboring nodes which have direct links with the target. We used the Clustal Omega[23] to respectively generate a high-quality multiple sequence alignment (MSA) for homologous sequences from each class and the set mixing all 3 families of PGTs. The MSAs were further edited to remove gaps, and subjected to sequence logo construction through WebLogo 3.[16]

Evolutionary Coupling Analysis and *De-novo* Structure Modeling

A state-of-art web server platform of evolutionary coupling analysis based *de-novo* structural modeling [21] was used to predict the structure for PglC non-transmembrane domain in *C. jejuni*. HHblits, a lightning-fast HMM-HMM alignment based iterative protein sequence searching method[23], was used to acquire a sufficient number of homologous sequences where an accurate alignment was built. After the coverage/gap filtration, the MSA was subjected to a global maximum entropy model based algorithm, EVCoupling[21] to achieve a set of 'direct' residue couplings, which highly likely represent spatial proximity. A subset of high scoring residue pairs was converted to "evolutionary inferred contacts" (EICs), which were then utilized by a distance-geometry and simulated annealing based approach, CNSsolve[25] to predict 3D structures.

References

(1) Valvano, M. A. Export of O-Specific Lipopolysaccharide. *Front. Biosci.* **2003**, *8*, s452–s471.

(2) Price, N. P.; Momany, F. A. Modeling Bacterial UDP-HexNAc: Polyprenol-P HexNAc-1-P Transferases. *Glycobiology* **2005**, *15* (9), 29R – 42R.

(3) Burda, P.; Aebi, M. The Dolichol Pathway of N-Linked Glycosylation. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1999**, *1426* (2), 239–257.

(4) Chung, B. C.; Zhao, J.; Gillespie, R. A.; Kwon, D.-Y.; Guan, Z.; Hong, J.; Zhou, P.; Lee, S.-Y. Crystal Structure of MraY, an Essential Membrane Enzyme for Bacterial Cell Wall Synthesis. *Science* **2013**, *341* (6149), 1012–1016.

(5) Lloyd, A. J.; Brandish, P. E.; Gilbey, A. M.; Bugg, T. D. H. Phospho-N-Acetyl-Muramyl-Pentapeptide Translocase from Escherichia Coli: Catalytic Role of Conserved Aspartic Acid Residues. *J. Bacteriol.* **2004**, *186* (6), 1747–1757.

(6) Glover, K. J.; Weerapana, E.; Chen, M. M.; Imperiali, B. Direct Biochemical Evidence for the Utilization of UDP-Bacillosamine by PglC, an Essential Glycosyl-1-Phosphate Transferase in the Campylobacter Jejuni N-Linked Glycosylation Pathway. *Biochemistry* **2006**, *45* (16), 5343–5350.

(7) Hartley, M. D.; Morrison, M. J.; Aas, F. E.; Børud, B.; Koomey, M.; Imperiali, B. Biochemical Characterization of the O-Linked Glycosylation Pathway in Neisseria Gonorrhoeae Responsible for Biosynthesis of Protein Glycans Containing N,N'-Diacetylbacillosamine. *Biochemistry* **2011**, *50* (22), 4936–4948.

(8) Saldías, M. S.; Patel, K.; Marolda, C. L.; Bittner, M.; Contreras, I.; Valvano, M. A. Distinct Functional Domains of the Salmonella Enterica WbaP Transferase That Is Involved in the Initiation Reaction for Synthesis of the O Antigen Subunit. *Microbiology* **2008**, *154* (2), 440–453.

(9) Wang, L.; Liu, D.; Reeves, P. R. C-Terminal Half of Salmonella Enterica WbaP (RfbP) Is the Galactosyl-1-Phosphate Transferase Domain Catalyzing the First Step of O-Antigen Synthesis. *J Bacteriol* **1996**, *178* (9), 2598–2604.

(10) Furlong, S. E.; Ford, A.; Albarnez-Rodriguez, L.; Valvano, M. A. Topological Analysis of the Escherichia Coli WcaJ Protein Reveals a New Conserved Configuration for the Polyisoprenyl-Phosphate Hexose-1-Phosphate Transferase Family. *Sci. Rep.* **2015**, *5*.

(11) Lehrer, J.; Vigeant, K. A.; Tatar, L. D.; Valvano, M. A. Functional Characterization and Membrane Topology of Escherichia Coli WecA, a Sugar-Phosphate Transferase Initiating the Biosynthesis of Enterobacterial Common Antigen and O-Antigen Lipopolysaccharide. *J. Bacteriol.* **2007**, *189* (7), 2618–2628.

(12) Anderson, M. S.; Eveland, S. S.; Price, N. P. Conserved Cytoplasmic Motifs That Distinguish Sub-Groups of the Polyprenol phosphate:N-Acetylhexosamine-1-Phosphate Transferase Family. *FEMS Microbiol. Lett.* **2000**, *191* (2), 169–175.

(13) Amer, A. O.; Valvano, M. A. Conserved Amino Acid Residues Found in a Predicted Cytosolic Domain of the Lipopolysaccharide Biosynthetic Protein WecA Are Implicated in the Recognition of UDP-N-Acetylglucosamine. *Microbiology* **2001**, *147* (11), 3015–3025.

(14) Furlong, S. E.; Valvano, M. A. Characterization of the Highly Conserved VFMGD Motif in a Bacterial Polyisoprenyl-Phosphate N-Acetylaminosugar-1-Phosphate Transferase. *Protein Science* **2012**, *21* (9), 1366–1375.

(15) Amer, A. O.; Valvano, M. A. Conserved Aspartic Acids Are Essential for the Enzymic Activity of the WecA Protein Initiating the Biosynthesis of O-Specific Lipopolysaccharide and Enterobacterial Common Antigen in Escherichia Coli. *Microbiology* **2002**, *148* (2), 571–582.

(16) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14* (6), 1188–1190.

(17) Patel, K. B.; Ciepichal, E.; Swiezewska, E.; Valvano, M. A. The C-Terminal Domain of the Salmonella Enterica WbaP (UDP-galactose:Und-P Galactose-1-Phosphate Transferase) Is Sufficient for Catalytic Activity and Specificity for Undecaprenyl Monophosphate. *Glycobiology* **2012**, *22* (1), 116–122.

(18) Albright, C. F.; Orlean, P.; Robbins, P. W. A 13-Amino Acid Peptide in Three Yeast Glycosyltransferases May Be Involved in Dolichol Recognition. *PNAS* **1989**, *86* (19), 7366–7369.

(19) Marrero, P. F.; Poulter, C. D.; Edwards, P. A. Effects of Site-Directed Mutagenesis of the Highly Conserved Aspartate Residues in Domain II of Farnesyl Diphosphate Synthase Activity. *J. Biol. Chem.* **1992**, *267* (30), 21873–21878.

(20) Patel, K. B.; Furlong, S. E.; Valvano, M. A. Functional Analysis of the C-Terminal Domain of the WbaP Protein That Mediates Initiation of O Antigen Synthesis in Salmonella Enterica. *Glycobiology* **2010**, *20* (11), 1389–1401.

(21) Marks, D. S.; Colwell, L. J.; Sheridan, R.; Hopf, T. A.; Pagnani, A.; Zecchina, R.; Sander, C. Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS ONE* **2011**, *6* (12), e28766.

(22) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr. Purif.* **2005**, *41* (1), 207–234.

(23) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.

(24) Buchan, D. W. A.; Minneci, F.; Nugent, T. C. O.; Bryson, K.; Jones, D. T. Scalable Web Services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **2013**, *41* (Web Server issue), W349–W357.

(25) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; et al. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54* (Pt 5), 905–921.

(26) Holm, L.; Rosenström, P. Dali Server: Conservation Mapping in 3D. *Nucleic Acids Res.* **2010**, *38* (Web Server issue), W545–W549.

(27) Garcia-Boronat, M.; Diez-Rivero, C. M.; Reinherz, E. L.; Reche, P. A. PVS: A Web Server for Protein Sequence Variability Analysis Tuned to Facilitate Conserved Epitope Discovery. *Nucleic Acids Res* **2008**, *36* (Web Server issue), W35–W41.

Chapter 4

Design and Evaluation of Inhibitors for a Bacterial Phosphoglycosyltransferase

The work described in this chapter was performed in collaboration with Dr. Maria Walvoort, a post-doctoral associate in the Imperiali Group. Dr. Walvoort synthesized all the PglC inhibitors, and performed UMP/CMP-Glo assays with these inhibitors. The work presented in this chapter is currently in preparation as a manuscript for publication.

Introduction

Bacterial cell surfaces display a variety of glycoconjugates, many of which are essential for survival or for host cell interactions. These include capsular polysaccharides, lipopolysaccharides, peptidoglycan, and glycoproteins. The biosynthesis of many of these glycoconjugates is initiated by phosphoglycosyltransferases (PGTs), a family of enzymes that transfer a C1′-phosphosugar from a nucleotide-activated precursor onto a polyprenol phosphate acceptor. PGTs are attractive antimicrobial targets due to their importance in the synthesis of cell-surface glycans. For example, as illustrated in Figure 4-1 MraY initiates the biosynthesis of peptidoglycan by catalyzing the transfer of 102-phospho-MurNAc-L-Ala-D-Glu-*m*-DAP-D-Ala-D-Ala to undecaprenol phosphate (Figure 4-1A). WecA transfers GlcNAc-1-phosphate to undecaprenol phosphate, in the first step of O-antigen synthesis (Figure 4-1B). MraY and WecA are polytopic integral membrane proteins, with 10 and 11 transmembrane domains respectively. Importantly, WecA is homologous to the GlcNAc-1-P transferase (GPT), which initiates the dolichol pathway for N-linked protein glycosylation in eukaryotes.

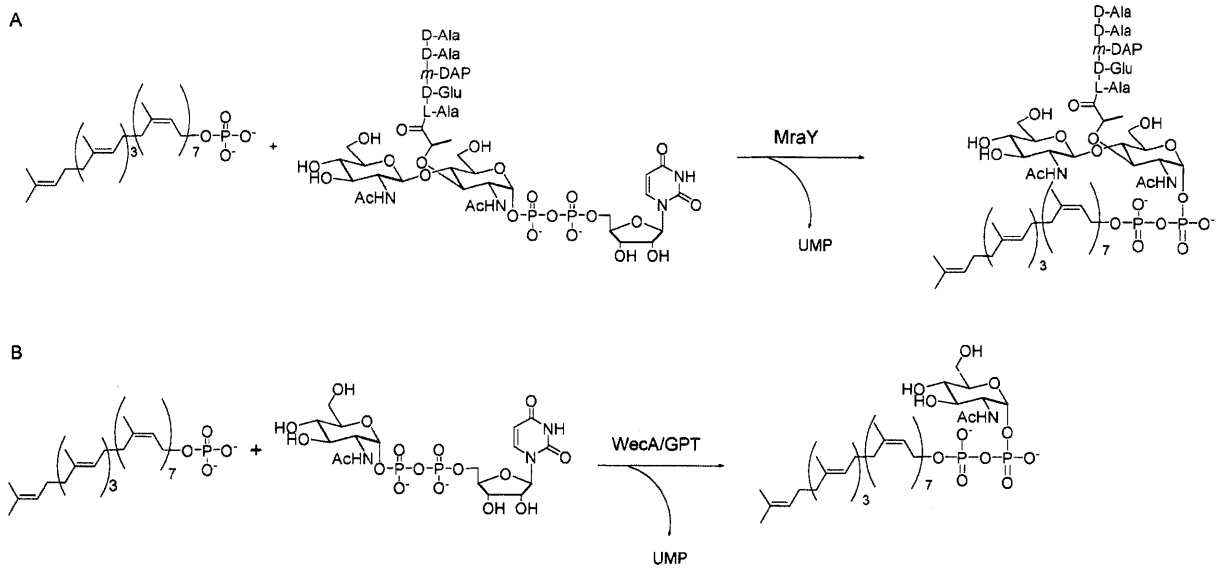


Figure 4-1 Reactions catalyzed by phosphoglycosyltranferases. (A) MraY and (B) WecA.

Nucleoside antibiotics are a family of natural products that exhibit a diverse range of biological activities ranging from antibacterial, antifungal, antitumor, antiviral, herbicidal, insecticidal, to immunosuppressive properties.[1] Nucleoside antibiotics target bacteria by inhibiting phosphoglycosyltransferases (PGTs). Tunicamycin, muriedomycin and liposidomycin are nucleoside antibiotics isolated from the fermentation broth of *Streptomyces* (Figure 4-2).[2–4] Tunicamycin (Figure 4-2B) contains a uracil moiety, a fatty acid, and two linked sugars (GlcNAc and tunicamine). The tunicamine hydroxyl groups are proposed to coordinate a divalent metal in the active site of metal-dependent PGTs, much like the diphosphate moiety in the UDP-donor sugar.[5] The GlcNAc moiety is believed to impart substrate specificity, as tunicamycin is an extremely potent inhibitor of WecA and its eukaryotic homolog, dolichol-P N-acetylglucosamine-1-P transferase (GPT), with IC_{50}s of 11 nM and 7 nM respectively.[6,7] The high potency of tunicamycin towards both these PGTs is attributed to its resemblance to the two substrates that associate with the enzyme in the ternary reaction complex. Mureidomycin (Figure 4-2C) is a peptidyl nucleoside natural product containing a 3′-deoxyuridine sugar attached by an enamide linkage to a peptide chain. The enamide was initially believed to play a key role in the inhibition of MraY, but this was later ruled out by testing analogs with enamide functional groups.[8] The amino terminus is essential for inhibition, suggesting that mureidomycins may use a metal ion displacing mechanism to bind to PGTs.[9] Liposidomycins (Figure 4-2D) are fatty acyl nucleoside antibiotics, which inhibit the formation of lipid intermediates in glycoconjugate biosynthesis. Liposidomycins are selective inhibitors of MraY and use a slow-binding mechanism to inhibit the reaction.[10] These natural products share key features with the substrates of PGT reactions, highlighted in Figure 4-2.

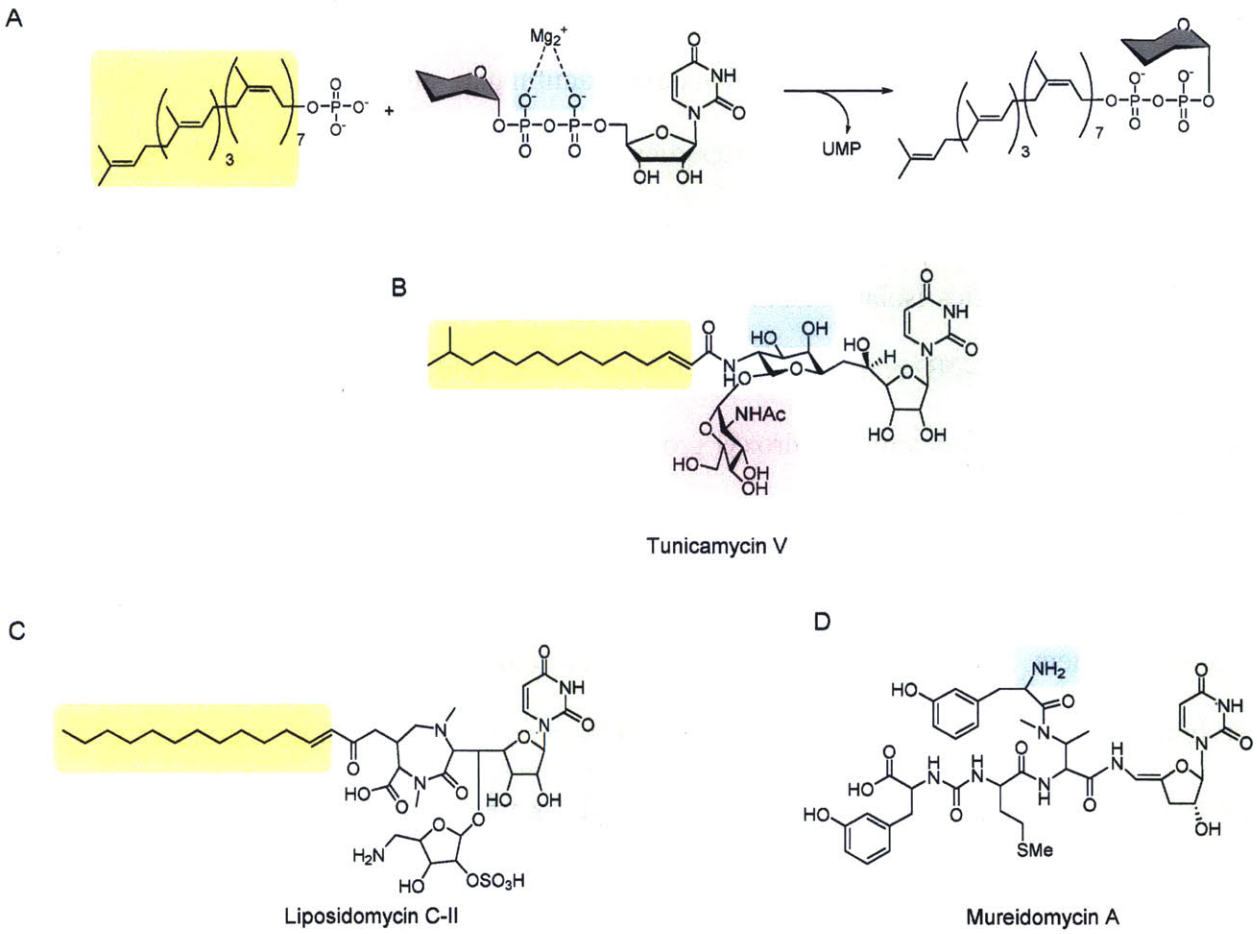


Figure 4-2 Structures of nucleoside antibiotics with key structural features highlighted. (A) A typical PGT reaction. (B) Structure of Tunicamycin V, (C) Structure of Liposidomycin C-II, and (D) Structure of Mureidomycin A. The uridine moiety is highlighted in gray, the metal-binding or metal-displacing groups are highlighted in blue, the carbohydrate is highlighted in pink, and the fatty acid or isoprene substrate is highlighted in yellow.

The inhibition of bacterial PGTs by natural products has been studied in detail due to their antibiotic potential. Kinetic analysis suggests that nucleoside antibiotics are differentially selective for PGT families. For example, liposidomycins are reported to inhibit MraY roughly 30-500-fold more effectively than tunicamycin.[11] In contract, tunicamycin inhibits the eukaryotic GPT 30-300-fold more effectively than liposidomycins.[11] The mode of inhibition of PGTs by nucleoside antibiotics also varies across PGT families. For example, tunicamycin reversibly

inhibits MraY, while liposidomycin B and mureidomycin A are both slow-binding inhibitors of this enzyme.[10] The results of inhibition studies with nucleoside antibiotics such as these have paved the way for the rational design of antibiotics targeted towards bacterial PGTs.[8,12,13] An important caveat to many of the biochemical studies is that many of them were performed with impure protein, in solubilized membrane fractions.

The work in this chapter focuses on PglC, a representative of the simplest family of PGTs, containing a single transmembrane helix and a globular C-terminal domain. PglC initiates the biosynthesis of the N-glycan in *Campylobacter jejuni*. PglC catalyzes the transfer of *N,N'*-diacetylbacillosamine-1-phosphate from UDP-*N,N'*-diacetylbacillosamine (UDP-diNAcBac) to undecaprenol phosphate, thereby forming the first membrane-associated intermediate (Figure 4-3B). PglC is essential for N-linked glycosylation in *C. jejuni*, a protein modification that is associated with the virulence of this human pathogen (Figure 4-3C). Eliminating N-linked protein glycosylation greatly reduces adhesion to and invasion of *C. jejuni* to host cells. Thus, inhibition of PglC would potentially attenuate the pathogenicity of *C. jejuni*. Our strategy to develop inhibitors for PglC is to synthesize molecules inspired by nucleoside antibiotics, incorporating structural features shared by these natural products and the substrates of the PglC reaction. Importantly, the presence of a "PglC-like domain" in other bacterial PGTs (Figure 4-3A) increases the scope of this project, with the potential to expand the number of protein targets of inhibitors developed in this body of work.

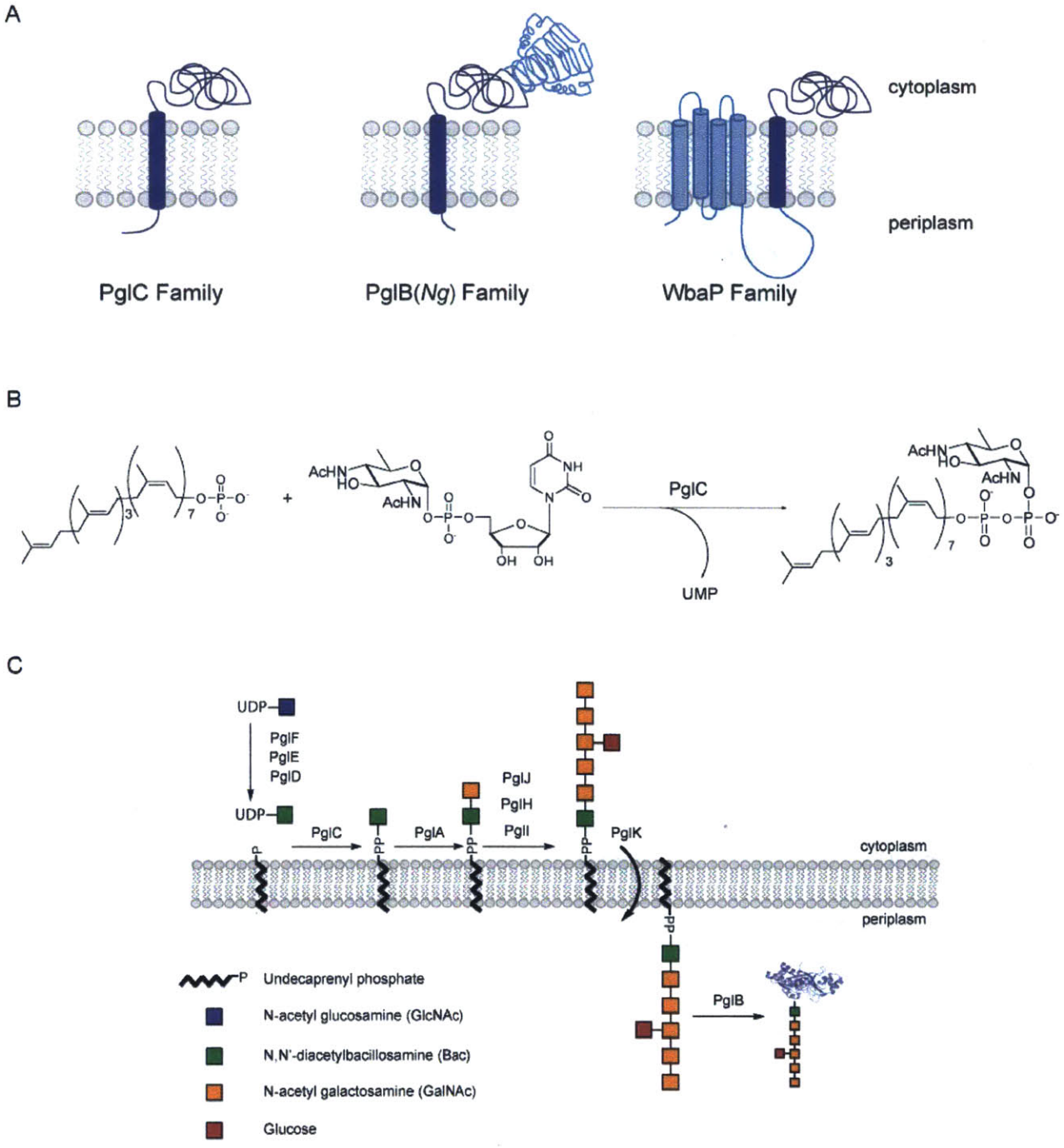


Figure 4-3 PglC initiates N-glycan biosynthesis in *C. jejuni*. (A) Bacterial PGT families containing the "PglC-like domain", shown in dark blue. (B) Reaction catalyzed by PglC. (C) N-linked glycosylation pathway of *C. jejuni*.

This chapter describes the development and optimization of inhibition assays suitable for use with PglC. The mechanism of the PglC reaction is investigated using kinetic analysis. Finally, the design and development of different families of inhibitors inspired by nucleoside antibiotics is discussed, with an outlook to further development of these scaffolds to include functional groups that will impart increased specificity to inhibitors.

Results and Discussion

Overexpression and Purification of PglC

PglC was expressed with a SUMO solubility tag. As discussed in Chapter Two, the SUMO tag greatly increases the yield of PglC during overexpression and purification, and helps maintain PglC in solution once purified, at concentrations less than 2 mg/mL. PglC was used as the SUMO fusion for inhibition studies, as the SUMO domain did not appear to interfere with the activity of PglC.

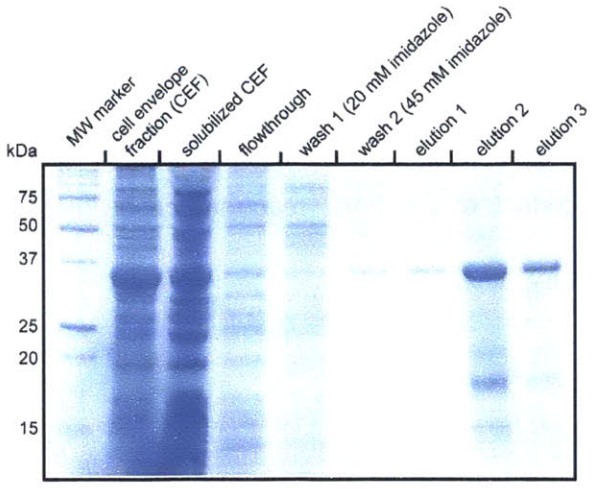


Figure 4-4 Purification of PglC with a SUMO solubility tag. The expected molecular weight of the product is ~35 kDa.

PglC Activity Assay

PglC was initially characterized using a radioactivity-based coupled assay with PglA, the next enzyme in the Pgl pathway (Figure 4-5A).[14] The function of PglA is to elaborate the Und-PP-diNAcBac product of the PglC reaction using UDP-GalNAc to form Und-PP-diNAcBac-GalNAc. To quantify substrate turnover, aliquots of the reaction are quenched at various time points and extracted to separate the water-soluble [^{3}H]-UDP-GalNAc starting material from the organic-soluble Und-PP-diNAcBac-[^{3}H]-GalNAc product, and liquid scintillation counting is performed to calculate product conversion. Although this assay is convenient due to easy commercial access to a radiolabeled substrate, it is not ideal for PglC inhibition assays because the uridine fragments that are the target of this inhibitor development initiative could potentially inhibit PglA, which also uses a UDP-sugar as a substrate, thereby giving a false positive inhibition result. In order to simplify the assay to include only PglC, radiolabeled UDP-diNAcBac needed to be synthesized. This was achieved using a previously described chemoenzymatic synthesis protocol described for UDP-diNAcBac,[15] modified to introduce [^{3}H]-acetyl-CoA at the last step (Figure 4-5B). With radiolabeled UDP-diNAcBac at hand, the formation of the tritiated product of the PglC reaction can be directly detected (Figure 4-5C) using an extraction-based assay.

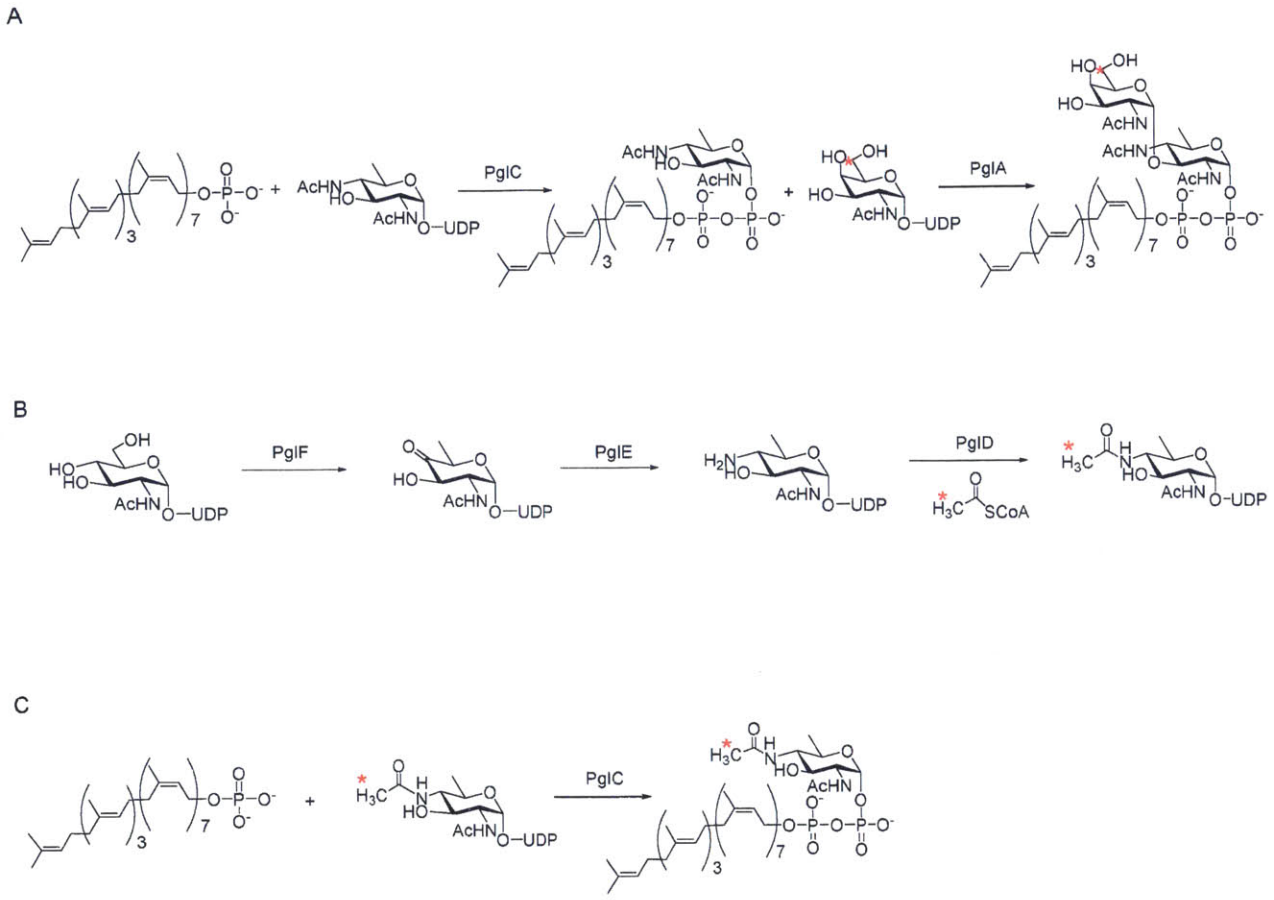


Figure 4-5 PglC radioactivity-based activity assays and synthesis of [^{3}H]-UDP-diNAcBac. (A) Coupled assay with PglA and [^{3}H]-labeled UDP-GalNAc. (B) Chemoenzymatic synthesis of [^{3}H]-labeled UDP-diNAcBac. Modified assay with [^{3}H]-labeled UDP-diNAcBac. The position of the tritium label is indicated by a red asterisk.

Reaction conditions were varied to determine optimal pH, and concentrations of Triton X-100. Additionally, the reaction was performed with 10 mM EDTA to investigate if the reaction is metal ion dependent, and concentrations of $MgCl_2$ were varied to determine optimal metal ion concentrations (Figure 4-6).

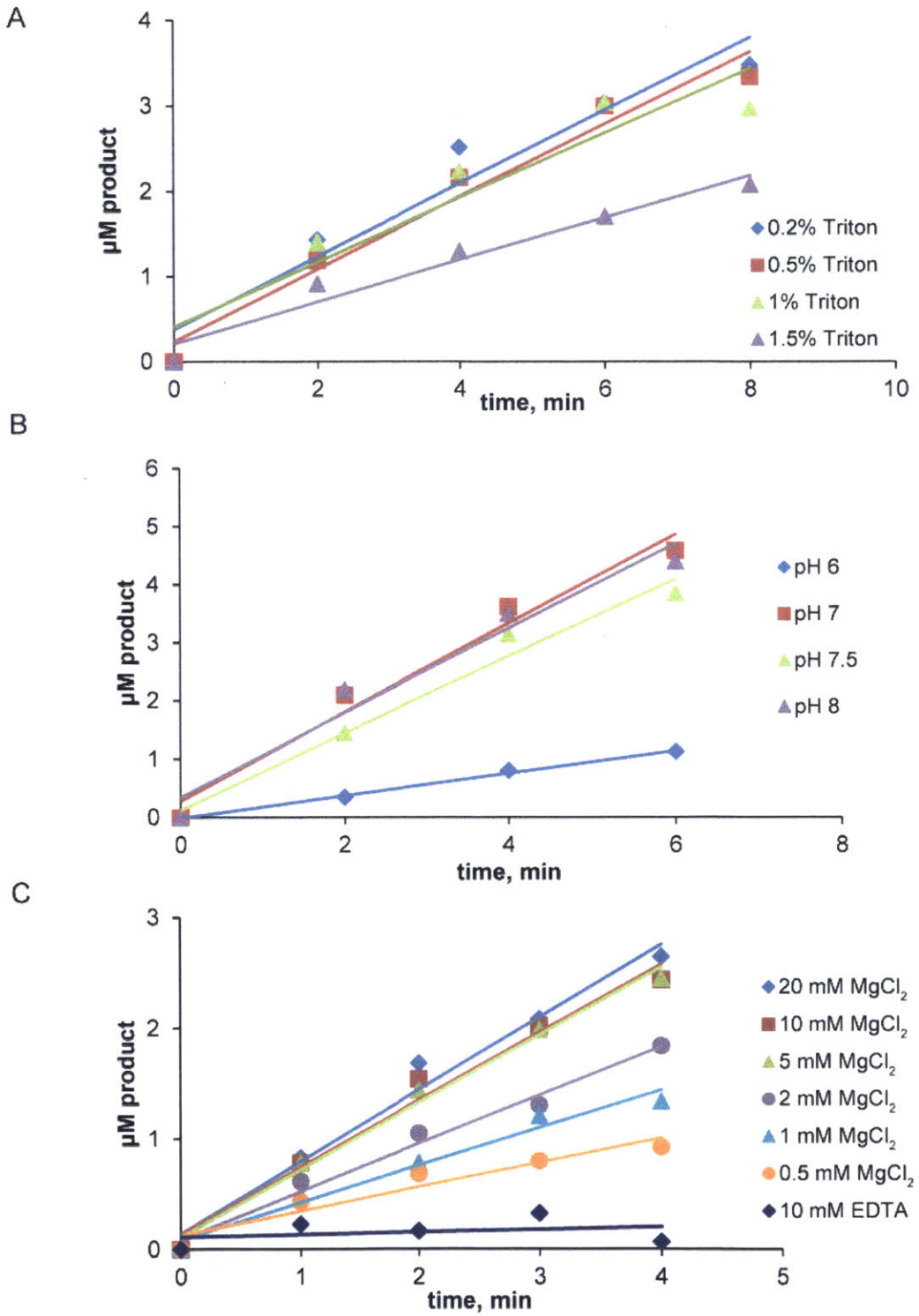


Figure 4-6 Optimization of the [^{3}H]-UDP-diNAcBac PglC activity assay. (A) Effect of Triton X-100 concentrations (B) Effect of pH of the reaction, and (C) Effect of concentration of $MgCl_2$.

The radioactivity-based assay was used to screen initial inhibitors. However, in order to test the increasing number of PglC inhibitors being developed over the course of this project, a more high-throughput assay was required. In collaboration with researchers at Promega, we began to test a prototype assay for PGTs, called the UMP/CMP-Glo assay. This is a coupled assay, that uses an enzyme to convert the UMP by-product of the PglC reaction to a luciferase signal, presumably via a UTP intermediate (Figure 4-7). Although the identity of the enzyme converting UMP to UTP was not disclosed, we hypothesize that it is a polyphosphate kinase.[16] This is a discontinuous assay that can be performed in a 96-well format, and greatly increased the speed at which PglC inhibitors could be tested. However, we observed that the smaller uridine fragments from early generations, tested at millimolar concentrations, inhibited the assay. The polyphosphate kinase converts UMP to UTP, and it is possible for many of the smaller inhibitors resembling the UMP substrate to bind to the enzyme, thereby inhibiting it. As our inhibitors were elaborated, we observed decreased inhibition of the UMP/CMP-Glo assay, presumably due to increased specificity for PglC. Thus, the extraction-based assay was used to screen initial generations of fragments, and the UMP/CMP-Glo assay was used to test later generations of inhibitors.

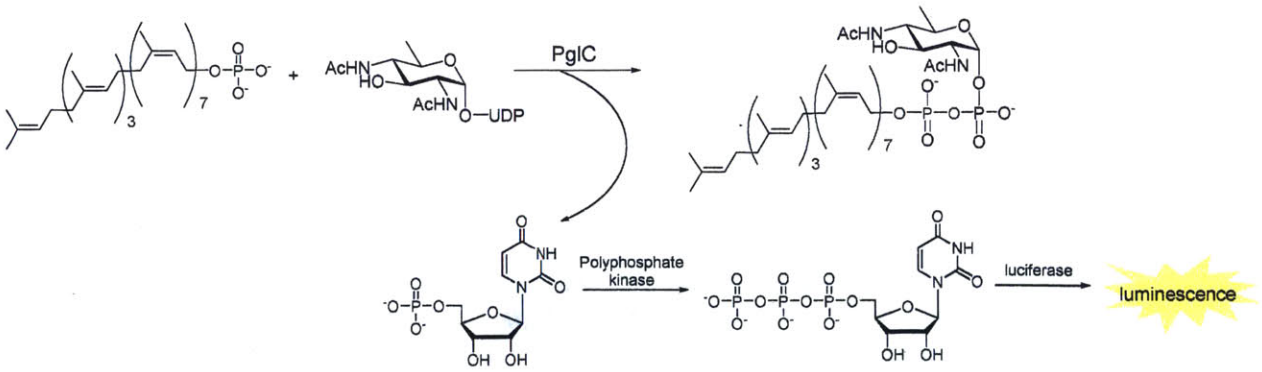


Figure 4-7 Scheme of the UMP/CMP-Glo assay used with PglC.

Determination of Kinetic Parameters of Substrates

Kinetic parameters were determined for each of the substrates of the PglC reaction. The radioactivity-based assay was used to determine the K_M of UDP-diNAcBac, at 20 μM Und-P, while the UMP/CMP-Glo assay was used to determine the K_M of Und-P, at 20 μM UDP-diNAcBac (Figure 4-8). An important caveat for the determination of kinetic parameters for Und-P is that it is difficult to predict how this substrate is distributed among detergent micelles which contain PglC, and it is not apparent what the effective concentration of Und-P is in the microenvironment of the micelles. Thus, the reported parameters for Und-P are apparent values.

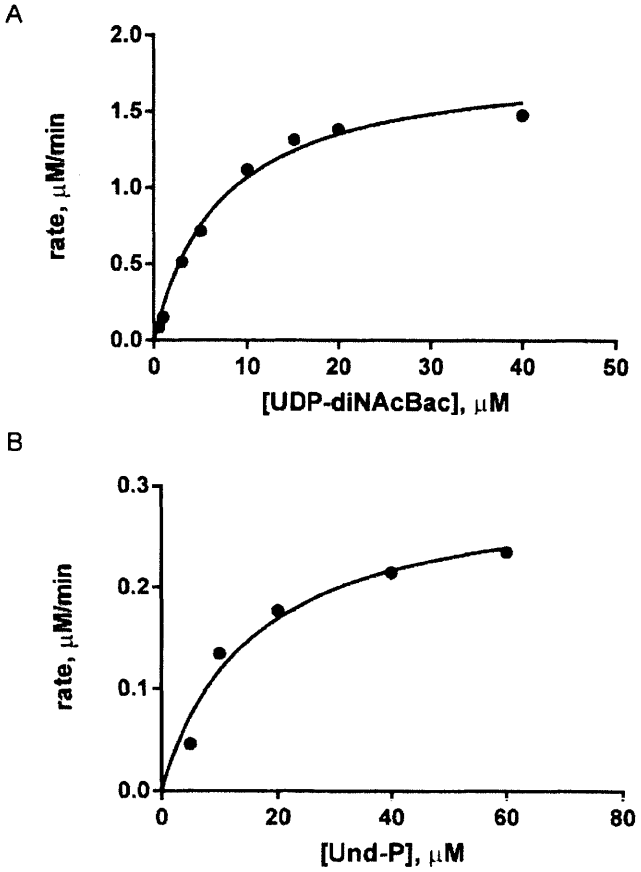


Figure 4-8 Michaelis-Menten plots for the substrates of the PglC reaction. (A) UDP-diNAcBac and (B) Und-P.

Table 4-1 Kinetic parameters determined for substrates of the PglC reaction.

	$K_{M\ app}$	$V_{max\ app}$	k_{cat}
UDP-diNAcBac	7.2 ± 1.1 μM	1.84 ± 0.09 μM/min	303 ± 90 min^{-1}
Und-P	15.6 ± 5.1 μM	0.3 ± 0.04 μM/min	460 ± 10 min^{-1}

Kinetic Analysis of the PglC Reaction

The overall phosphoglycosyltransferase reaction involves nucleophilic attack at the nucleotide-diphosphate-sugar donor, and transfer of a sugar-1-phosphate to polyprenol-phosphate. Two different mechanisms can be proposed for this bisubstrate reaction – a ping-pong mechanism, which involves the intermediacy of a covalent enzyme intermediate and the release of the first product, after which the second substrate binds and the final product is formed and released, or a sequential mechanism with formation of a ternary complex, in which both substrates bind simultaneously to the enzyme before product release (Figure 4-9). Kinetic analysis of the eukaryotic GPT isolated from a hen oviduct microsomal fraction indicates a sequential mechanism for this enzyme, with formation of a ternary complex.[7] In contrast, studies with a crude sample of MraY suggest that the transfer of substrate occurs via the formation of a covalent enzyme intermediate, which was reported to be trapped and observed by gel filtration, using radiolabeled UDP-MurNAc-pentapeptide.[17] However, the control experiment with the proposed nucleophilic Asp-267 residue mutated has not been performed. In order to provide insight into the mechanism of the PglC reaction, double-reciprocal plots were constructed with increasing substrate concentrations, for each of the substrates of PglC. In this analysis, parallel lines would indicate a Bi Bi Ping Pong mechanism, while intersecting lines indicate a Bi Bi random or Bi Bi ordered mechanism. Analysis of preliminary data with concentrations of each substrate varied produces intersecting lines, suggesting that the PglC reaction occurs through the formation of a ternary complex (Figures 4-10 and 4-11).

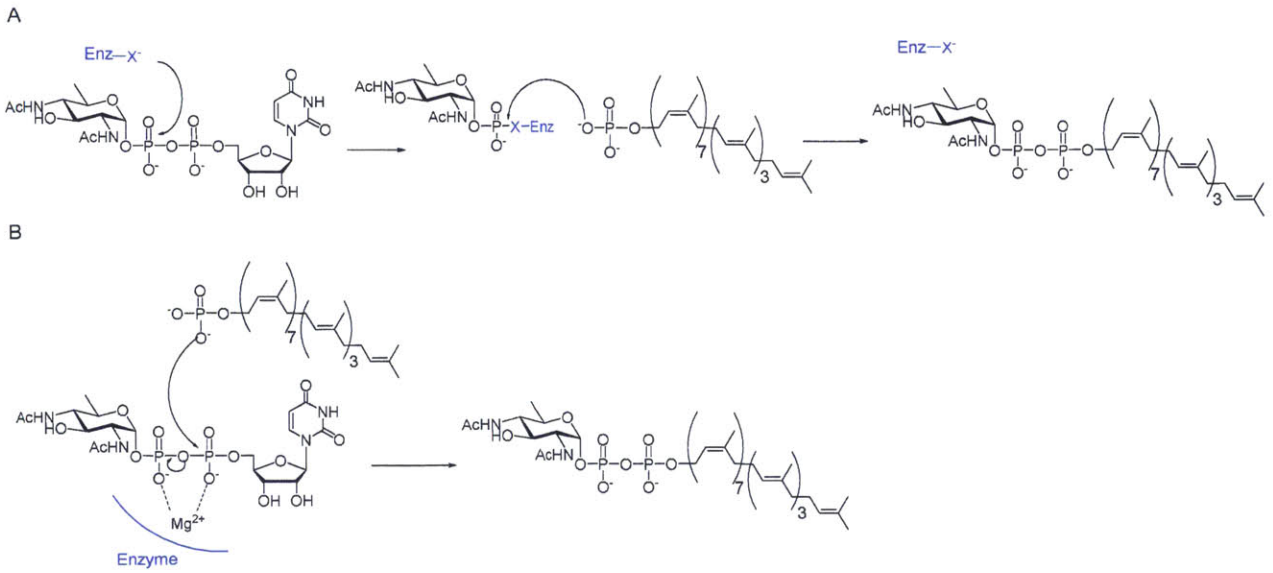


Figure 4-9 Possible reaction mechanisms for PglC. (A) Ping Pong mechanism, with formation of a covalent enzyme intermediate. (B) Bi Bi Random mechanism with formation of a ternary complex.

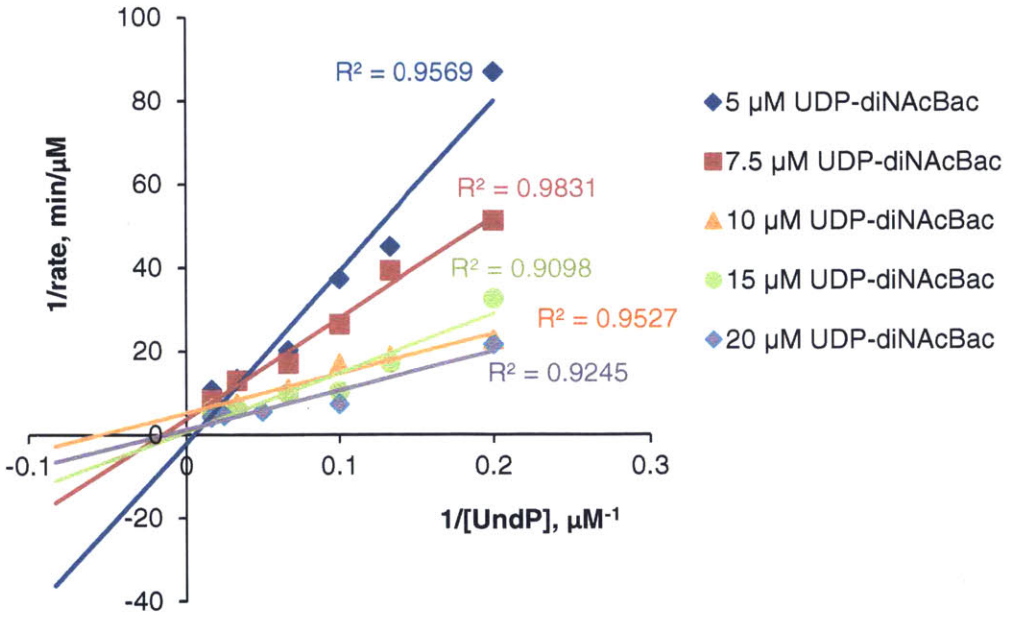


Figure 4-10 Lineweaver Burk plot of PglC reactions performed at multiple UDP-diNAcBac concentrations. Data used for the plot is provided in the appendix.

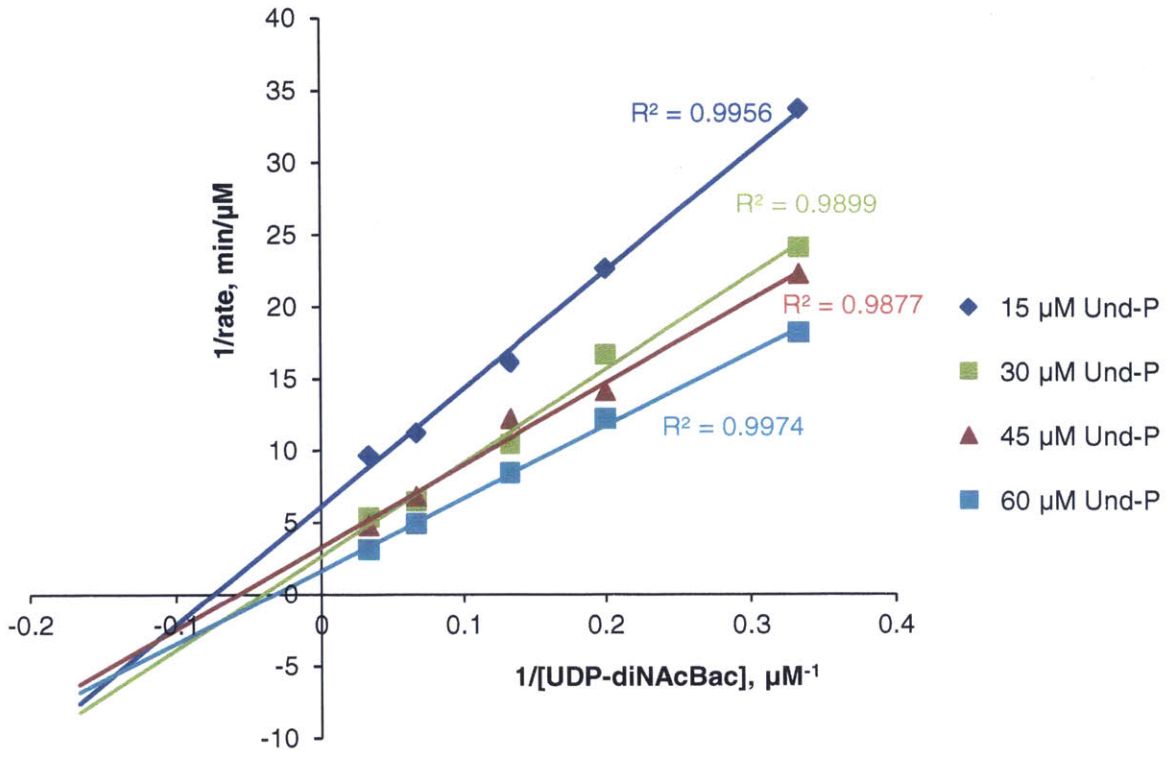


Figure 4-11 Lineweaver Burk plot of PglC reactions performed at multiple undecaprenol phosphate (Und-P) concentrations. Data used for the plot is provided in Table 4-2.

Table 4-2 Data used to generate Lineweaver Burk plots performed with varying concentrations of UDP-diNAcBac and Und-P (Figures 4-10 and 4-11).

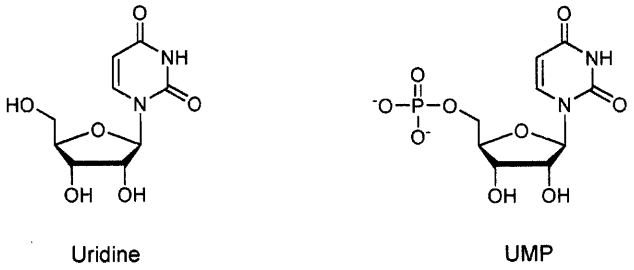
[UDP-diNAcBac], µM	[Und-P], µM	Rate, µM/min	R^2
5	5	0.0115	0.8984
5	7.5	0.0222	0.9504
5	10	0.0268	0.9328
5	15	0.0495	0.9893
5	30	0.0743	0.9965
5	60	0.092	0.9992
7.5	5	0.0195	0.9365
7.5	7.5	0.0254	0.8689
7.5	10	0.0376	0.9939
7.5	15	0.0581	0.9909
7.5	30	0.0762	0.9699
7.5	60	0.1177	0.9927

[UDP-diNAcBac], μM	[Und-P], μM	Rate, μM/min	R^2
10	5	0.0438	0.9867
10	7.5	0.0526	0.9955
10	10	0.0581	0.9909
10	15	0.0904	0.9744
10	30	0.1327	0.9901
10	60	0.1526	0.9722
15	7.5	0.0307	0.9872
15	10	0.0577	0.9956
15	15	0.095	0.9949
15	30	0.1554	0.9976
15	60	0.1687	0.9966
20	5	0.0462	0.9644
20	10	0.1348	0.9965
20	20	0.1772	0.9933
20	40	0.2146	0.9948
20	60	0.2351	0.9972
3	15	0.0297	0.9554
5	15	0.0442	0.9806
7.5	15	0.0621	0.9993
15	15	0.0892	0.9998
30	15	0.1038	0.9995
3	30	0.0416	0.9631
5	30	0.06	0.9679
7.5	30	0.096	0.9828
15	30	0.1548	0.9718
30	30	0.1876	0.986
3	45	0.0451	0.9572
5	45	0.0709	0.9952
7.5	45	0.0821	0.9977
15	45	0.1472	0.9985
30	45	0.2112	0.9863
3	60	0.0551	0.9896
5	60	0.082	0.9841
7.5	60	0.1187	0.9986
15	60	0.2046	0.9976
30	60	0.3238	0.9924

Evaluation of Inhibitor Precursors

Uridine and uridine monophosphate (UMP) were both evaluated for their ability to inhibit the PglC reaction (Figure 4-12). Interestingly, UMP, the by-product of the PglC reaction, appears to inhibit the reaction quite effectively at 1 mM, and at lower concentrations (data not shown). This finding supports our previous observations of the PglC reaction, which appears to saturate at approximately 30% product conversion, suggesting that the reaction is inhibited by the UMP by-product. In contrast, uridine does not appear to inhibit the PglC reaction at concentrations as high as 1 mM.

A



B

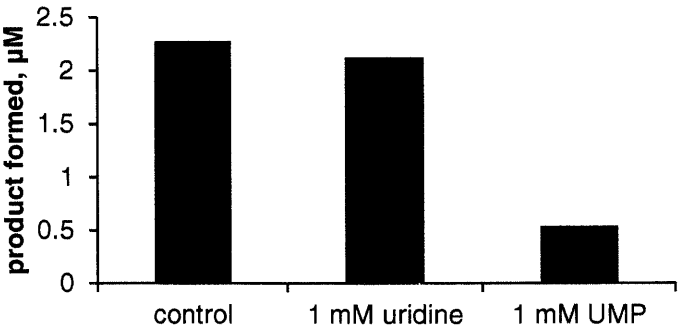


Figure 4-12 Inhibition of PglC by inhibitor precursors. (A) Structures of uridine and uridine monophosphate (UMP). (B) Inhibition of PglC by uridine and UMP at 1 mM.

Design and Evaluation of Metal Cofactor-Displacing Inhibitors

A series of PglC inhibitors was designed inspired by the proposed metal ion-displacing property of the amine group in mureidomycin. Uridine was modified to include an amine functional group, which could potentially displace the metal ion cofactor in the active site of PglC. The amines were linked to the uridine core with different alkyl chain linker lengths, as there is currently no structural information available to predict distances required to reach the metal ion-binding site. The first generation of compounds was tested using the radioactivity-based PglC assay high concentrations (5 and 1 mM), as they were expected to bind with only modest affinity. The amount of product formed after five minutes was compared at these concentrations of inhibitors (Figure 4-13).

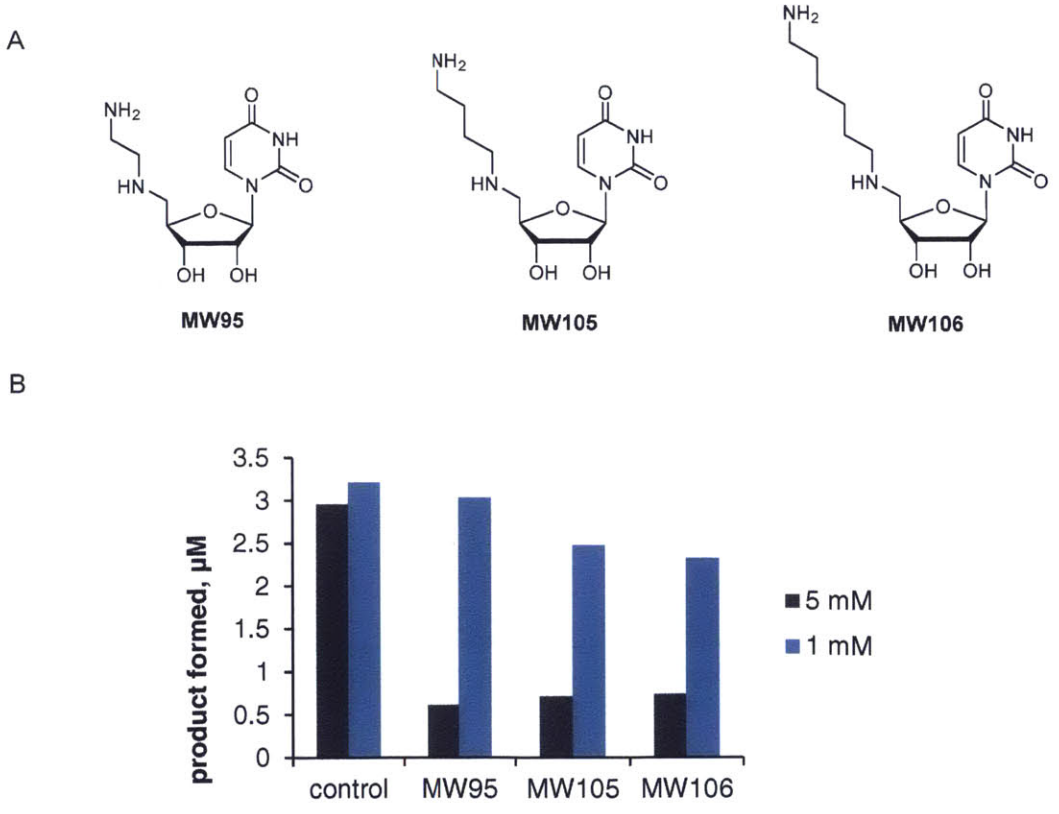


Figure 4-13 Inhibition of PglC by amine compounds. (A) Structures of amine compounds with varying linker lengths. (B) Inhibition of PglC by compounds at 5 mM and 1 mM.

The compound containing the longest amine (MW106) was then modified to include an azide group, which could be modified using copper-catalyzed azide-alkyne cycloaddition. A variety of alkynes were tested to vary the functional groups that might bind in the sugar-binding site of the enzyme (Figure 4-14). An intermediary concentration (3 mM) was selected to test these inhibitors, as previous experiments (Figure 4-13) demonstrated that 5 mM was too high of a concentration and 1 mM was too low of a concentration to observe trends in inhibition with varying amine lengths. The amount of product formed after five minutes was compared using the radioactivity-based PglC assay.

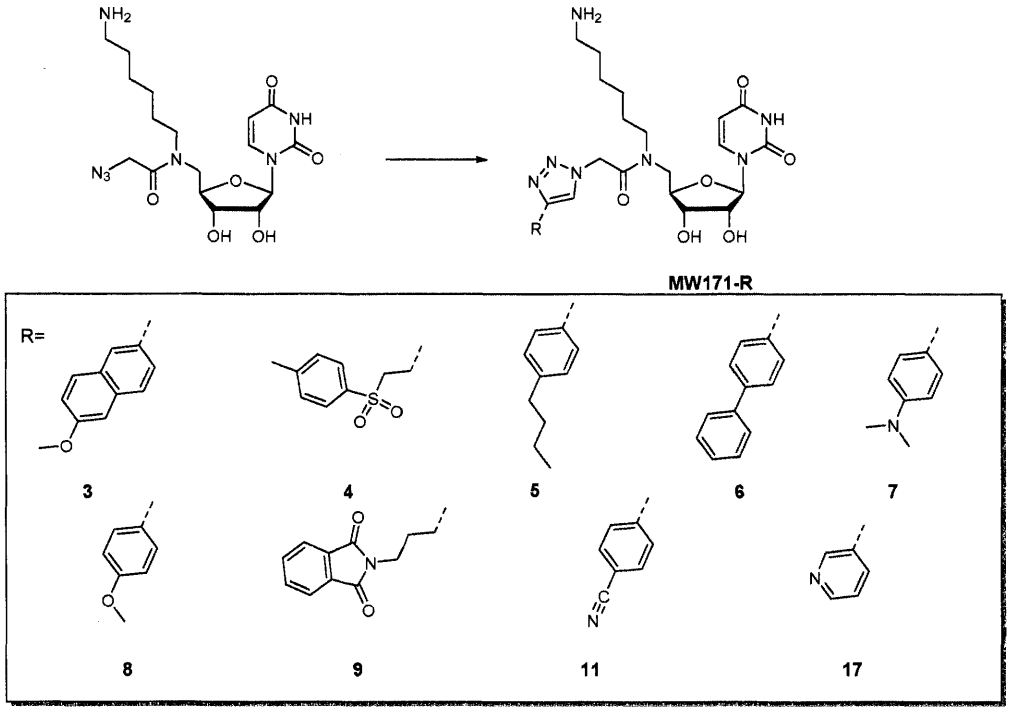


Figure 4-14 Library of modified amine compounds synthesized to test PglC inhibition, using copper catalyzed azide alkyne cycloaddition.

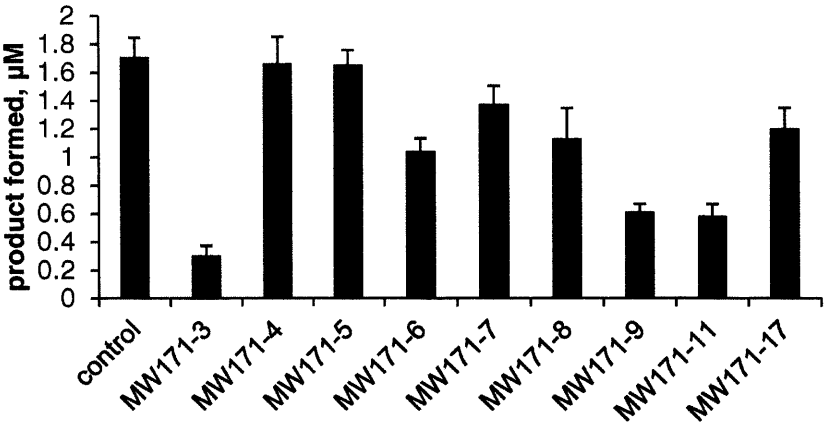


Figure 4-15 Inhibition of PglC by the library of modified amine compounds at a concentration of 3 mM.

The compounds showed varying degrees of inhibition of PglC. The most promising inhibitor, MW171-3, appeared to inhibit PglC completely, showing only background levels of signal (Figure 4-15). This compound is currently being elaborated further to include a fatty acid substituent with the goal of having this additional moiety bind in the isoprene-binding site of PglC. Control experiments to investigate the metal ion cofactor-displacing mechanism are also underway.

Design and Evaluation of Metal Cofactor-Coordinating Inhibitors

A series of PglC inhibitors was designed inspired by the proposed metal ion-coordination property of the sugar hydroxyl groups in tunicamycin. Uridine was elaborated to include carboxylic acids which could potentially coordinate the metal cofactor in the active site of PglC. The acids were linked to the uridine core with saturated and unsaturated linkers to vary the flexibility of the coordinating functional groups (Figure 4-16). The amount of product formed after five minutes was compared at these concentrations of inhibitors using the radioactivity-based PglC assay.

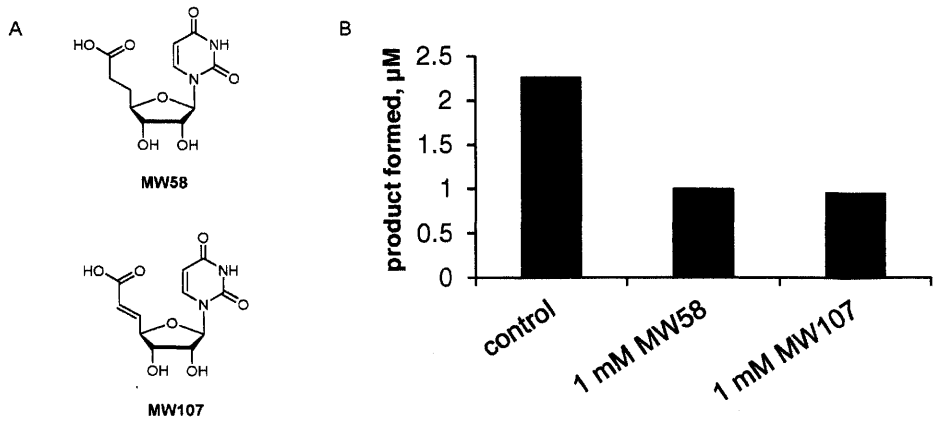


Figure 4-16 Inhibition of PglC by acid compounds. (A) Structures of acid tested and (B) Inhibition of PglC by compounds at 1 mM.

Both acids appeared to have inhibitory effects on the activity of PglC at concentrations of 1 mM. However, the structures do not contain branching points to easily elaborate the molecules to include functional groups that would impart specificity for PGTs. Thus, an alternative strategy was devised using amino acid side chains (aspartic acid and glutamic acid) to provide the required metal coordinating groups, as well as an azide amino acid that could be coupled to a fatty acid as well as conjugated to a sugar-like moiety via a triazole (Figure 4-17).

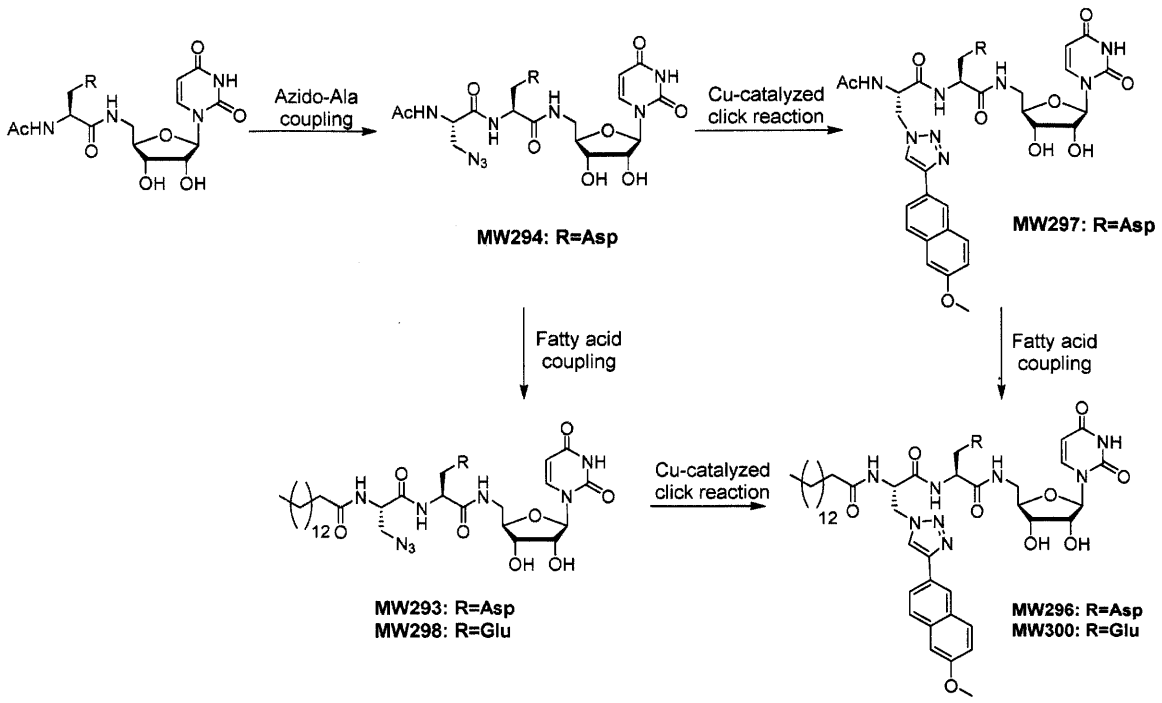


Figure 4-17 Scheme for the synthesis of inhibitors using amino acids. Aspartate and glutamate side chains provide the metal ion coordinating groups. An azide amino acid is used to provide two branching points at which functional groups can be attached using both peptide coupling as well as copper catalyzed azide alkyne cycloaddition.

Inhibitors synthesized using this strategy were tested using the UMP/CMP-Glo assay. These elaborated molecules could be tested at lower concentrations (100 μM) than earlier generations due to increased specificity for PglC, and showed much less inhibition of the assay than earlier generations at these concentrations. In order to control for the amount of assay inhibition by these compounds, a reading of a standard amount of UMP (2 μM) was measured in the presence of the inhibitors, and was compared to the luminescence signal in the absence of inhibitors. PglC assays were performed with the same concentrations of inhibitor. As Figure 4-18 shows, at 100 μM of inhibitor there is minimal inhibition of the assay. Importantly, the most elaborate compound with the aspartate side chain (MW296) which contains the myristic acid as

well as the aromatic sugar mimic appears to inhibit PglC more potently than the compounds with either of these functionalities alone (MW293 and MW297). Compounds with the glutamate side chains modified with the myristic acid only or with the acid and the aromatic naphthyl group also appear to inhibit PglC quite effectively at 100 μM.

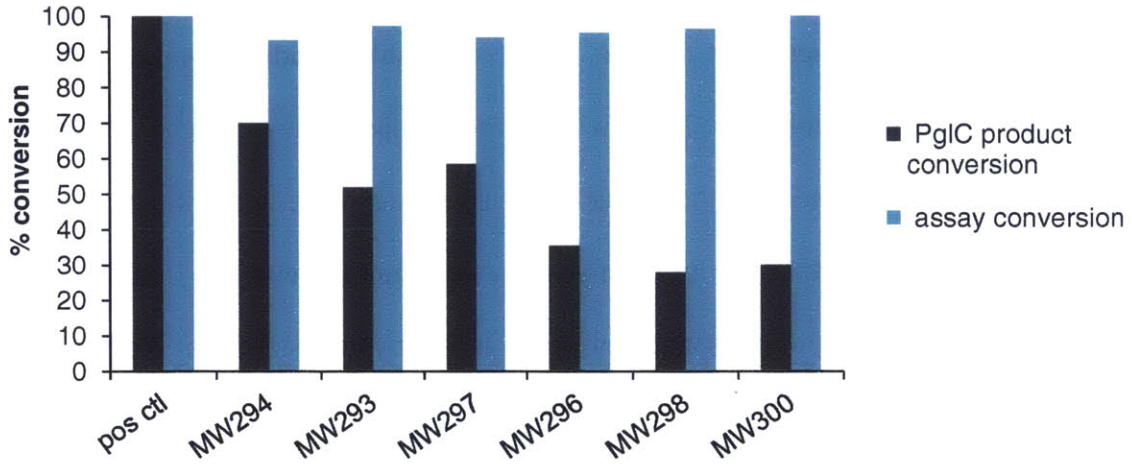


Figure 4-18 Inhibition of PglC by compounds synthesized using the amino acid strategy. Compounds were tested at concentrations of 100 μM. Normalized values of assay conversion and PglC product conversion are shown.

Further studies with these compounds have yielded preliminary IC_{50} values which can be used to compare inhibitors. These values were obtained using the UMP/CMP-Glo assay, at inhibitor concentrations no higher than 250 μM. Inhibition of the assay was tested using inhibitors at 250 μM to ensure that they did not inhibit the assay by more than 20%. The preliminary IC_{50} value for tunicamycin calculated using this assay is also reported (Table 4-2).

Table 4-3 IC_{50} values of inhibitors calculated for PglC

Inhibitor	IC_{50}, μM
Tunicamycin	100
MW296	64
MW298	60
MW300	53

Experiments are underway to investigate the mode of binding of these inhibitors to PglC. In particular, kinetic analysis with increasing inhibitor concentrations will be used establish whether the compounds are competitive for both substrates, thereby providing information about the binding mechanism. It should also be noted that it is possible that the compounds may not be competitive for the undecaprenol phosphate substrate, and that the myristic acid may serve to localize the inhibitor to the detergent micelle. This would still be advantageous, as it would increase accessibility of the inhibitor to PglC, which is embedded in the membrane and catalyzes a reaction at the membrane interface. A similar observation was made in the kinetic analysis of the mode of inhibition of MraY by tunicamycin, where tunicamycin was found not to be competitive for the dodecaprenol phosphate substrate. It was hypothesized that the hydrophobic "tail" of tunicamycin directed the molecule to detergent micelles in the assay rather than binding in the active site of the enzyme.[10]

Purification of PglC from *Helicobacter pullorum*

Helicobacter pullorum was recently discovered to have an N-linked protein glycosylation system.[18] The PGT in this pathway, PglC(*Hp*), is predicted to use a "HexNAc" substrate, based on mass spectrometry analysis. PglC(*Hp*) was expressed with a SUMO tag to increase solubility, and the protein was purified to homogeneity (Figure 4-19). Preliminary activity assays support UDP-GlcNAc as the carbohydrate substrate. Further characterization of this enzyme is underway, as it represents an additional target for the inhibitors designed in this study.

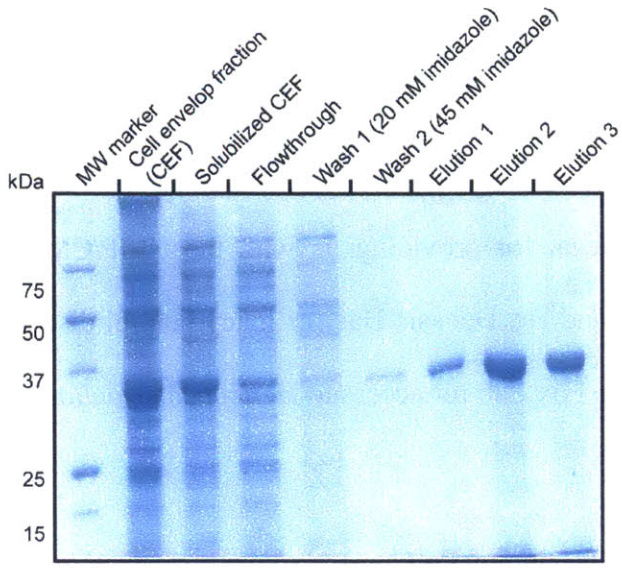


Figure 4-19 SDS-PAGE analysis of the purification of PglC from *H. pullorum* with a SUMO solubility tag. The expected molecular weight of the product is ~35 kDa.

Conclusions

PglC represents the family of simplest structures of PGTs, and therefore represents a tractable model for which inhibitors can be designed and evaluated. Our success in obtaining pure, stable protein greatly assisted the kinetic characterization of PglC. In addition, the development and optimization of PglC assays allowed a thorough evaluation of inhibitors designed for PglC.

The design of inhibitors presented in this chapter introduces two scaffolds which can potentially be expanded upon to impart specificity to particular enzymes. The strategies described here provide a modular approach to the synthesis of inhibitors for PGTs, with clear branching points at which variability and specificity can be introduced in the form of amino acids side chains, fatty acids, as well as alkynes conjugated to sugar mimics.

Acknowledgements

I am grateful to Maria (Marthe) Walvoort for being a wonderful collaborator and for synthesizing all the inhibitors for this project. We are thankful to Hicham Zegzouti, a senior research scientist at Promega, for providing us with the UMP/CMP-Glo assay. We are also thankful to Sonya Entova and Dr. Debasis Das for many helpful discussions about kinetics. We would also like to thank the Fox Lab for access to *Helicobacter pullorum* DNA.

Experimental Methods

Overexpression and Purification of PglC

PglC was cloned into the pET-SUMO vector using P13 and P14 primers (Table 2-1) described in Chapter Two. The pET-SUMO-PglC plasmid was transformed into BL21-RIL cells (Agilent) for overexpression, using kanamycin and chloramphenicol for selection. Overexpression was performed using the Studier method.[19] In this method, 1 mL of an overnight cell culture was added to expression media containing 30 μg/mL kanamycin and 30 μg/mL chloramphenicol in 1 L of autoinduction media (0.1% (w/v) tryptone, 0.05% (w/v) yeast extract, 2 mM $MgSO_4$, 0.05% (v/v) glycerol, 0.005% (w/v) glucose, 0.02% (w/v) α-lactose, 2.5 mM Na_2HPO_4, 2.5 mM KH_2PO_4, 5 mM NH_4Cl, 0.5 mM Na_2SO_4). Cells were allowed to grow with shaking for 3 h at 37 °C. After 3 h, the temperature was decreased to 16 °C, and the cells were incubated for 16 hours. Cells were harvested by centrifuging at 9000 x g, and cells were stored at -80 °C. Cell pellets were thawed in 10% of the original culture volume in 50 mM Tris pH 8.0, 150 mM NaCl, 40 μL protease inhibitor cocktail (Calbiochem). The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. The cells were incubated on ice for ten minutes between rounds of sonication. Cellular debris was removed by centrifugation at 9000 x *g* for 45 minutes. The resulting supernatant was

transferred to a clean centrifuge tube and subjected to centrifugation at 142,000 x g for 65 minutes to pellet the CEF. If the CEF was to be used for activity assays, it was homogenized into 1% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl and stored at -80 °C. If protein was to be purified from the CEF, it was isolated and homogenized into 10% of the original culture volume in 50 mM HEPES pH 7.5, 100 mM NaCl, 1% n-dodecyl beta-D-maltoside (DDM), using a glass homogenizer. This sample was incubated at 4 °C with gentle rocking for 16 hours, after which it was centrifuged (145,000 x g) to remove insoluble material. The resulting supernatant was incubated with 1 mL Ni-NTA resin for 16 hours. 20 µL protease inhibitor cocktail was added to prevent proteolysis The resin was washed with 30 mL Wash 1 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 20 mM imidazole), followed by a wash with 30 mL Wash 2 buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 45 mM imidazole). PglC was eluted in 4 x 1 mL fractions of elution buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM, 300 mM imidazole). Gel filtration analysis was performed using a Superdex S200 10/300 column (GE Healthcare) equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl, 0.03% DDM

Synthesis of [^{3}H]-UDP-diNAcBac

[^{3}H]-UDP-diNAcBac was prepared by previously described chemoenzymatic methods, with a modification at the final step to incorporate the tritium radiolabel.[15] UDP-GlcNAc was converted to UDP-4-amino-NAcBac using PglF and PglE. After purification, 5 mM UDP-4-amino-NAcBac was incubated with [^{3}H]-Acetyl coenzyme A (20 Ci/mmol, American Radiolabeled Chemicals), 50 mM HEPES pH 7.5, 50 mM NaCl, and 260 μM PglD, followed by a chase with 5 mM unlabeled acetyl coenzyme A for 2 hours. The radiolabeled product was purified on a Phenomenex Synergi C18 reversed-phase HPLC column, to yield [^{3}H]-UDP-diNAcBac (5.4 mCi/mmol).

Radioactivity-Based Activity Assays with PglC

PglC was assayed using a radioactive extraction-based assay. Assays contained 20 μM Und-P, 10% DMSO, 0.2% Triton X-100, 50 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 20 μM [^{3}H]-UDP-diNAcBac (5.4 mCi/mmol), and 2.5 nM PglC in a final volume of 60 μL. Inhibitors were added in DMSO, in a volume such that the total concentration of DMSO in the reaction was equal to 10% (v/v). PglC was pre-incubated in the reaction mixture lacking [^{3}H]-UDP-diNAcBac for ten minutes. After initiation of the reaction with [^{3}H]-UDP-diNAcBac, aliquots (20 μL) were taken at five minute time points and quenched in 1 mL $CHCl_3$:MeOH. The organic layer was washed three times with 400 μL PSUP (Pure Solvent Upper Phase, composed of 15 mL $CHCl_3$, 240 mL MeOH, 1.83 g KCl in 235 mL H_2O). The resulting aqueous layers were combined with 5 mL EcoLite (MP Biomedicals) liquid scintillation cocktail. Organic layers were combined with 5 mL OptiFluor (PerkinElmer). Both layers were analyzed on a Beckman Coulter LS6500 scintillation counting system.

UMP/CMP-Glo Activity Assays with PglC for Kinetic Analysis

PglC activity assays contained 20 μM Und-P, 10% DMSO, 0.1% Triton X-100, 50 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 20 μM UDP-diNAcBac, and 1 nM PglC in a final volume of 100 μL. Inhibitors were added to the reaction in DMSO, in a volume such that the total concentration of DMSO in the reaction was equal to 10% (v/v). PglC was pre-incubated in the reaction mixture lacking UDP-diNAcBac for five minutes. After initiation of the reaction with UDP-diNAcBac, aliquots (15 μL) were taken at defined time points (two, four and six minutes) and were quenched in 15 μL of the UMP/CMP-Glo quench mixture. The quenched samples were transferred to 96 Well Half Area Microplates (Greiner Bio-One). The plate was shaken gently for 16 minutes and incubated for 44 minutes at 25 °C after which the luminescence signal was recorded. Data used to generate the Lineweaver Burk plots are available in the Appendix.

UMP/CMP-Glo Inhibition Assays with PglC

PglC activity assays contained 20 μM Und-P, 10% DMSO, 0.1% Triton X-100, 50 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 20 μM UDP-diNAcBac, and 1 nM PglC in a final volume of 25 μL. Inhibitors were added to the reaction in DMSO, in a volume such that the total concentration of DMSO in the reaction was equal to 10% (v/v). PglC was pre-incubated in the reaction mixture lacking UDP-diNAcBac for five minutes. After initiation of the reaction with UDP-diNAcBac, the reaction was allowed to proceed for 20 minutes after which it was quenched with 25 μL of the UMP/CMP-Glo quench mixture. Quenched samples were transferred to 96 Well Half Area Microplates (Greiner Bio-One). The plate was shaken gently for 16 minutes and incubated for 44 minutes at 25 °C after which the luminescence signal was recorded. Inhibition of the assay was measured by quenching a 25 μL solution of 2 μM UMP in the presence of the

inhibitor, and measuring the corresponding luminescence reading. The percent conversion of the assay was calculated by comparing this value to the luminescence reading of a 2 μM solution of UMP in the absence of inhibitor.

Cloning PglC(*Hp*) into SUMO vector

PglC(*Hp*) was amplified from genomic DNA of *Helicobacter pullorum* strain MIT 09-6635, using primers shown in Table 4-4. PglC(*Hp*) was overexpressed in BL-21 RIL cells by autoinduction, using the same procedure as described above for PglC from *C. jejuni*.

Table 4-4 Primers used to clone PglC(*Hp*) into the SUMO vector.

PglC(*Hp*)-F (**BsaI**)	5′-CGCC**GGTCTCCAAGGT**ATGTATAAAAACCTGATTAA-3′
PglC(*Hp*)-R (**XhoI**)	5′-ATCG**CTCGAG**TTATTCGCTTTTATTGCCTGTGAATTTTTCC-3′

References

(1) Isono, K. Nucleoside Antibiotics: Structure, Biological Activity, and Biosynthesis. *J. Antibiot.* **1988**, *41* (12), 1711–1739.

(2) Takatsuki, A.; Arima, K.; Tamura, G. Tunicamycin, a New Antibiotic. I. Isolation and Characterization of Tunicamycin. *J. Antibiot.* **1971**, *24* (4), 215–223.

(3) Isono, F.; Katayama, T.; Inukai, M.; Haneishi, T. Mureidomycins A-D, Novel Peptidylnucleoside Antibiotics with Spheroplast Forming Activity. III. Biological Properties. *J. Antibiot.* **1989**, *42* (5), 674–679.

(4) Isono, K.; Uramoto, M.; Kusakabe, H.; Kimura, K.; Isaki, K.; Nelson, C. C.; McCloskey, J. A. Liposidomycins: Novel Nucleoside Antibiotics Which Inhibit Bacterial Peptidoglycan Synthesis. *J. Antibiot.* **1985**, *38* (11), 1617–1621.

(5) Xu, L.; Appell, M.; Kennedy, S.; Momany, F. A.; Price, N. P. J. Conformational Analysis of Chirally Deuterated Tunicamycin as an Active Site Probe of UDP-N-Acetylhexosamine:polyprenol-P N-Acetylhexosamine-1-P Translocases. *Biochemistry* **2004**, *43* (42), 13248–13255.

(6) Al-Dabbagh, B.; Mengin-Lecreulx, D.; Bouhss, A. Purification and Characterization of the Bacterial UDP-GlcNAc:Undecaprenyl-Phosphate GlcNAc-1-Phosphate Transferase WecA. *J Bacteriol* **2008**, *190* (21), 7141–7146.

(7) Keller, R. K.; Boon, D. Y.; Crum, F. C. N-Acetylglucosamine-1-Phosphate Transferase from Hen Oviduct: Solubilization, Characterization, and Inhibition by Tunicamycin. *Biochemistry* **1979**, *18* (18), 3946–3952.

(8) Gentle, C. A.; Bugg, T. D. H. Role of the Enamide Linkage of Nucleoside Antibiotic Mureidomycin A: Synthesis and Reactivity of Enamide-Containing Analogues. *J. Chem. Soc., Perkin Trans. 1* **1999**, No. 10, 1279–1286.

(9) Nigel I Howard, T. D. H. B. Synthesis and Activity of 5′-Uridinyl Dipeptide Analogues Mimicking the Amino Terminal Peptide Chain of Nucleoside Antibiotic Mureidomycin A. *Bioorganic & medicinal chemistry* **2003**, *11* (14), 3083–3099.

(10) Brandish, P. E.; Kimura, K. I.; Inukai, M.; Southgate, R.; Lonsdale, J. T.; Bugg, T. D. Modes of Action of Tunicamycin, Liposidomycin B, and Mureidomycin A: Inhibition of Phospho-N-Acetylmuramyl-Pentapeptide Translocase from Escherichia Coli. *Antimicrob. Agents Chemother.* **1996**, *40* (7), 1640–1644.

(11) Kimura, K.; Ikeda, Y.; Kagami, S.; Yoshihama, M.; Suzuki, K.; Osada, H.; Isono, K. Selective Inhibition of the Bacterial Peptidoglycan Biosynthesis by the New Types of Liposidomycins. *J. Antibiot.* **1998**, *51* (12), 1099–1104.

(12) Tanino, T.; Al-Dabbagh, B.; Mengin-Lecreulx, D.; Bouhss, A.; Oyama, H.; Ichikawa, S.; Matsuda, A. Mechanistic Analysis of Muraymycin Analogues: A Guide to the Design of MraY Inhibitors. *J. Med. Chem.* **2011**, *54* (24), 8421–8439.

(13) Mihalyi, A.; Jamshidi, S.; Slikas, J.; Bugg, T. D. H. Identification of Novel Inhibitors of Phospho-MurNAc-Pentapeptide Translocase MraY from Library Screening: Isoquinoline Alkaloid Michellamine B and Xanthene Dye Phloxine B. *Bioorganic & Medicinal Chemistry* **2014**, *22* (17), 4566–4571.

(14) Glover, K. J.; Weerapana, E.; Chen, M. M.; Imperiali, B. Direct Biochemical Evidence for the Utilization of UDP-Bacillosamine by PglC, an Essential Glycosyl-1-Phosphate Transferase in the Campylobacter Jejuni N-Linked Glycosylation Pathway. *Biochemistry* **2006**, *45* (16), 5343–5350.

(15) Olivier, N. B.; Chen, M. M.; Behr, J. R.; Imperiali, B. In Vitro Biosynthesis of UDP-N,N′-Diacetylbacillosamine by Enzymes of the Campylobacter Jejuni General Protein Glycosylation System. *Biochemistry* **2006**, *45* (45), 13659–13669.

(16) Motomura, K.; Hirota, R.; Okada, M.; Ikeda, T.; Ishida, T.; Kuroda, A. A New Subfamily of Polyphosphate Kinase 2 (Class III PPK2) Catalyzes Both Nucleoside Monophosphate and Nucleoside Diphosphate Phosphorylation. *Appl. Environ. Microbiol.* **2014**, AEM.03971–13.

(17) Lloyd, A. J.; Brandish, P. E.; Gilbey, A. M.; Bugg, T. D. H. Phospho-N-Acetyl-Muramyl-Pentapeptide Translocase from Escherichia Coli: Catalytic Role of Conserved Aspartic Acid Residues. *J. Bacteriol.* **2004**, *186* (6), 1747–1757.

(18) Jervis, A. J.; Langdon, R.; Hitchen, P.; Lawson, A. J.; Wood, A.; Fothergill, J. L.; Morris, H. R.; Dell, A.; Wren, B.; Linton, D. Characterization of N-Linked Protein Glycosylation in Helicobacter Pullorum. *J. Bacteriol.* **2010**, *192* (19), 5228–5236.

(19) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr. Purif.* **2005**, *41* (1), 207–234.

Chapter 5

Chemoenzymatic Assembly of Bacterial Glycoconjugates for Site-Specific Bioorthogonal Labeling

The work described in this chapter was performed in collaboration with Dr. Garrett Whitworth, a postdoctoral associate in the Imperiali group, and Dr. Ziqiang Guan from the Department of Biochemistry at Duke University. In addition to performing capillary electrophoresis analysis of PglC reactions, Dr. Whitworth synthesized undecaprenol-phosphate, UDP-GalNAz, UDP-4-Azido-diNAcBac, and the acetylated carbohydrates used for metabolic labeling studies. Dr. Guan performed mass spectrometry analysis of all the polyprenol-linked glycans. The work presented in this chapter is currently in preparation as a manuscript for publication.

Introduction

Campylobacter jejuni is the principal microbial cause of human gastroenteritis world-wide. *C. jejuni* is a Gram-negative bacterium that is commensal in livestock, but is a human pathogen. The main source of transmission to humans is through the consumption of contaminated poultry. In 1999, *C. jejuni* was the first bacterium discovered to have an N-linked protein glycosylation system.[1] Furthermore, the sequencing of the complete genome of *C. jejuni* revealed several additional gene clusters that are important for the biosynthesis and transport of glycoconjugates to the cell surface. [2–4]

The *C. jejuni* glycome comprises four components that play important roles in the biology and pathogenesis of the organism: the lipooligosaccharide (LOS), the capsular polysaccharide (CPS), Ser/Thr-linked (O-linked) glycans, and N-linked glycans (Figure 5-1). The LOS and CPS are structurally variable and are thought to play a role in the adherence and invasion of epithelial cells in the gastrointestinal tract of host organisms.[5,6] *C. jejuni* flagella require O-linked glycosylation for colonization and virulence in host organisms.[7] FlaA, a major flagellin structural protein, can be modified with pseudaminic acid and derivatives of this sugar at up to 19 different sites.[8] The N-linked protein glycosylation (Pgl) system modifies periplasmic and outer membrane proteins, and is involved in protein protection from proteolytic cleavage and increasing protein solubility. Additionally, N-linked protein glycosylation is intimately associated with colonization and host cell adherence and invasion.[9] These collective observations implicate glycan modifications in *C. jejuni* pathogenesis and highlight them as potential virulence targets that may be of therapeutic significance.

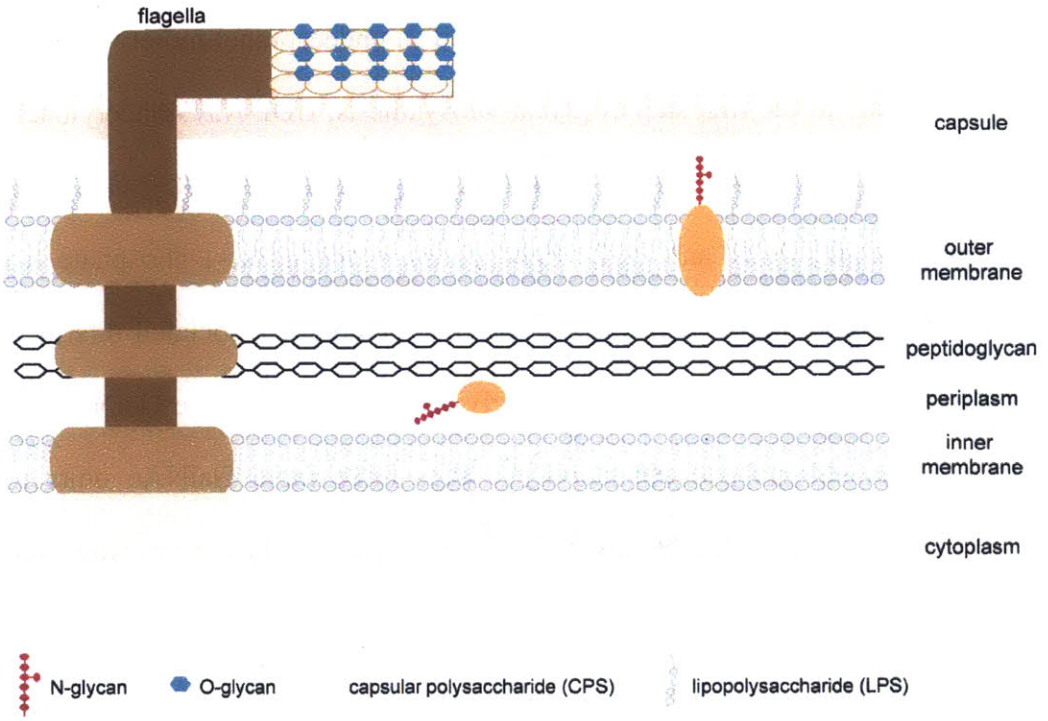


Figure 5-1 Schematic representation of the *C. jejuni* glycome and cell-surface glycoconjugates. Adapted from Young et al. [10]

Bacterial glycoproteins are increasingly being recognized as important virulence factors. As methods to characterize and identify these diverse and unique glycans improve, their roles in infection have been revealed in more detail. For example, O-glycosylation of the flagellar proteins of *Helicobacter pylori* is essential for flagella formation, motility and virulence.[11,12] O-glycosylation of the type IV pilin of *Neisseria gonorrhoeae* is required for binding to receptors in cervical epithelial cells, thereby causing cervical infections.[13] *Acinetobacter baumannii* is an emerging "superbug" with a characterized O-linked protein glycosylation system. Disruption of glycosylation in *A. baumannii* results in a diminished ability to form biofilms as well as reduced virulence.[14]

Since the discovery of the N-linked Pgl system in *C. jejuni*, the Imperiali Lab has biochemically characterized the enzymes in this pathway (Figure 5-2). The glycan, a branched

heptasaccharide, is assembled in a stepwise fashion, on an undecaprenol moiety embedded in the cytoplasmic membrane. In the first step of glycan biosynthesis, UDP-*N*,*N'*-diacetylbacillosamine (UDP-diNAcBac) is synthesized from UDP-*N*-acetylglucosamine (UDP-GlcNAc) by the sequential action of PglF, PglE, and PglD. PglC transfers diNAcBac-1-phosphate from UDP-diNAcBac to undecaprenol phosphate, forming the first membrane-associated intermediate. The glycosyltransferases PglA and PglJ elaborate the glycan by adding a single GalNAc each. PglH, a glycosyltransferase with polymerase activity[15], adds three more GalNAc units to form a hexasaccharide. PglI completes the heptasaccharide by adding a single branching glucose. The glycan is then translocated across the inner membrane by the flippase PglK and transferred by the oligosaccharyl transferase, PglB, onto an asparagine side chain within a D/E-X_1-N-X_2-S/T sequon (where X_1 and X_2 are not proline) of a substrate protein. The deletion of any *pgl* gene except *pglI* disrupts formation of N-linked glycoproteins.[16]

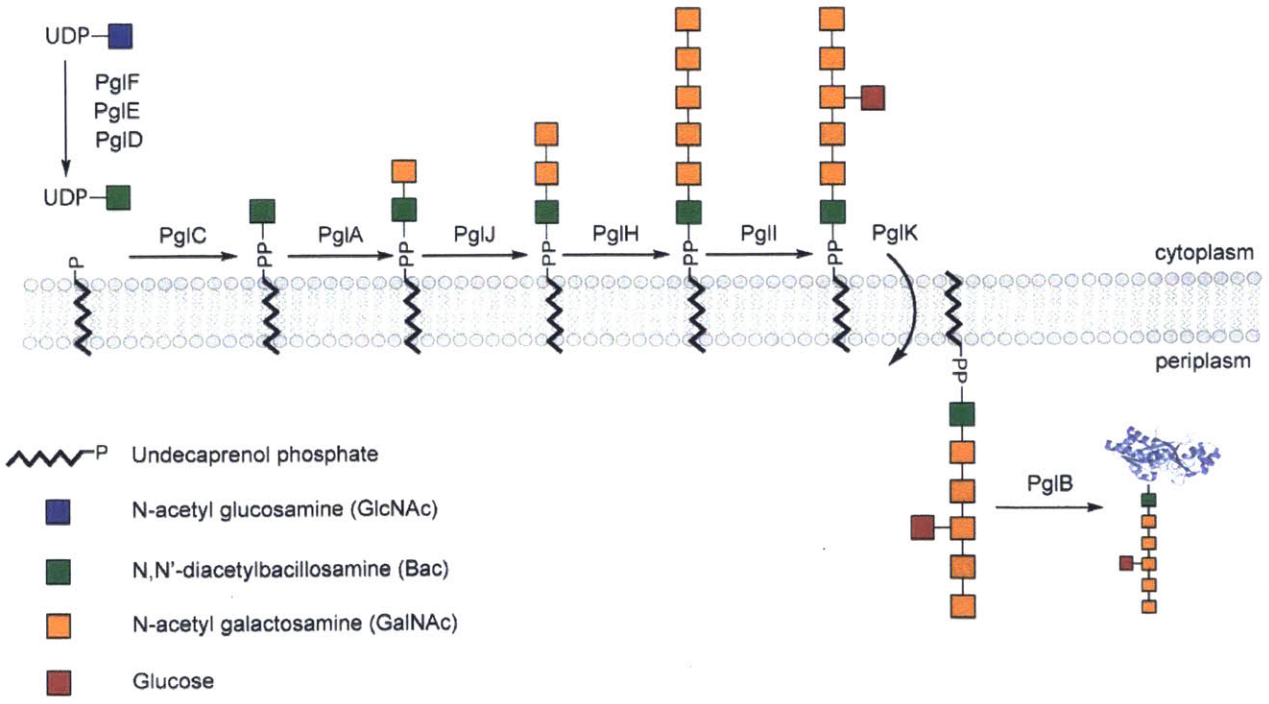


Figure 5-2 The N-linked protein glycosylation pathway in *C. jejuni*

The exact role of N-linked protein glycosylation in the virulence of *C. jejuni* has yet to be elucidated. Gene knockout and mutation studies point to a role in host cell adhesion and invasion. For example, *pglB* and *pglE* knockout strains of *C. jejuni* are unable to adhere to and invade human epithelial cells, and are less effective at colonizing mice intestinal tracts than the wild type strain.[16] In separate studies, *pglH*, *pglB* and *pglK* mutants were shown to be impaired in their ability to colonize chick ceacum.[17] These results suggest the presence of host cell surface receptors that bind the glycan, facilitating adherence and subsequent invasion. To date, there has been only one report of such a receptor, suggesting that the *C. jejuni* heptasaccharide interacts with human host cells via the human macrophage galactose-type lectin (MGL) receptor.[18] The MGL receptor is predominantly expressed on human dendritic cells and transmits signals that modulate immune cell function. Immunoblot experiments and infection assays suggest that MGL recognizes the terminal GalNAc residue in *C. jejuni* glycoproteins. Further studies are required to determine whether there are additional receptors for the glycan, and to characterize the nature of the interactions between the N-glycan and the surface of host cells.

Our approach to understanding the role of N-linked glycans in *C. jejuni* pathogenicity involves the use of sugars with bioorthogonally reactive functionality. Specifically, we aim to metabolically label *C. jejuni* with sugars that include a reactive "handle" and to use copper-catalyzed azide-alkyne cycloaddition to investigate the role of these glycans on the cell surface. Potential applications include imaging studies to probe the distribution of the glycan on the cell surface, as well as the identification of receptors on the human gastrointestinal epithelial layer *in vivo*. For these purposes, our goal is to incorporate azide-modified sugars into cell surface N-linked glycoproteins, which can then be reacted with bioorthogonal probes for subsequent imaging or isolation. This method was pioneered by the Bertozzi group in studying eukaryotic

glycosylation, and has been used to label glycans in mammalian cells[19], live zebrafish[20] and mice.[21] Metabolic labeling using unnatural carbohydrates has also been successful in studying the glycome of bacteria such as *Helicobacter pylori* [22]and to image lipopolysaccharides in *E. coli* and *Legionella pneumophila*.[23] At the outset of these studies the tolerance of the Pgl enzymes for incorporation of azide functionality, had not been demonstrated, although it has been documented that synthetically prepared substrates including truncated (C_{20}) polyprenol diphosphate derivatives of 6-azido-GlcNAc and 2-azidoGlcNAc (GlcNAz) failed to serve as substrates in peptide glycosylation by PglB.[24]

Results and Discussion

In order to determine whether azide-modified sugars would be successfully incorporated into a native undecaprenol-diphosphate-linked heptasaccharide, we first performed *in vitro* experiments to examine whether the Pgl enzymes would accept these substrate analogs. Traditionally, azides have been tolerated by glycosyltransferases when incorporated into the N-acetyl functional group. Of the carbohydrates that comprise the *C. jejuni* heptasaccharide, six contain N-acetyl functional groups that can be modified with an azide moiety (Figure 5-3). The first three enzymes in the Pgl pathway, PglC, PglA and PglJ could potentially deliver an azide-modified carbohydrate into a single, specific position in the glycan. In contrast, the polymerase activity of PglH could potentially result in incorporation of an azide-modified carbohydrate into the three terminal GalNAc sites of the glycan. The C-2 position of UDP-diNAcBac was not considered for azide modification because the presence of an N-acetyl group at this position is known to be important for substrate recognition by the oligosaccharyltransferase.[25]

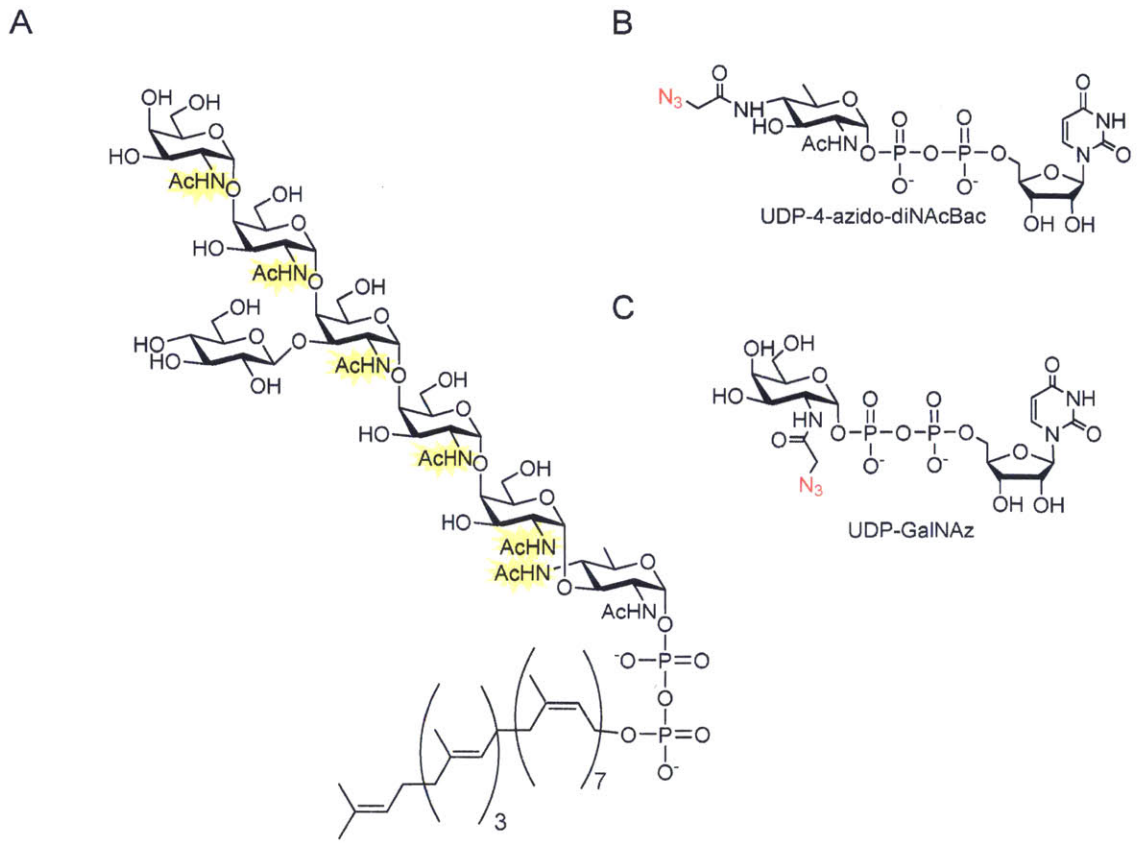


Figure 5-3 (A) Structure of the *C. jejuni* heptasaccharide. Functional groups highlighted in yellow indicate positions at which an azide can be introduced. (B) Structure of UDP-4-azido-diNAcBac, an azido analog of UDP-diNAcBac (C) Structure of UDP-GalNAz, an azido analog of UDP-GalNAc.

Synthesis of Substrates and Enzymes for *in vitro* Assays

In order to assemble these complex glycoconjugates and study the tolerance of Pgl enzymes for unnatural azide-modified sugars, multiple substrates and purified enzymes needed to be prepared. A combination of chemical and enzymatic synthesis was used to synthesize the carbohydrate and isoprene substrates.

Synthesis of Undecaprenol Phosphate

A C_{50}-C_{60} polyprenol mixture was extracted from *Rhus typhina* leaves, using previously described methods,[26] followed by phosphorylation with phosphoramidite $(FmO)_2PN{i}Pr_2$ (Figure 5-4). Subsequent oxidation and deprotection led to the desired product. It is important to note that the stereochemistry of the bacterial undecaprenol double bonds are 2-*trans* to 8-*cis* rather than the 3-*trans* to 7-*cis* shown for the plant-derived undecaprenol used here. Although the polyprenol phosphate synthesized was a mixture of the C_{50}-C_{55} isoprenes, the product and derivatives of the product will be referred to as Und-P for simplicity.

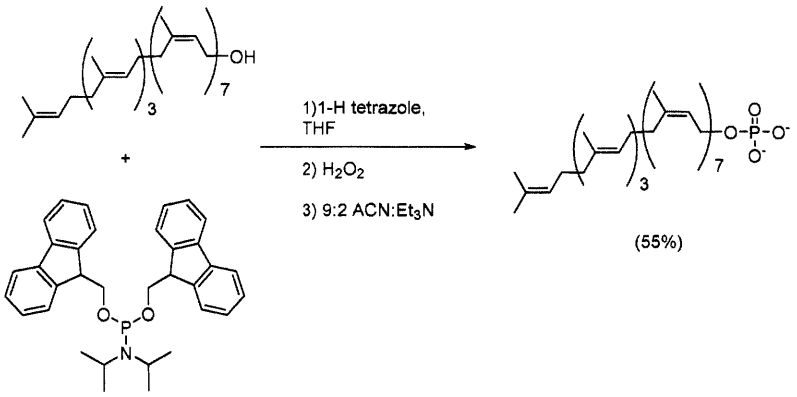


Figure 5-4 Synthesis of polyprenol phosphate from polyprenols isolated from *Rhus typhina* leaves

Chemoenzymatic Synthesis of Carbohydrate Substrates

UDP-diNAcBac was generated enzymatically from UDP-GlcNAc using previously described methods (Figure 5-5A).[27] UDP-4-azido-diNAcBac, a novel unnatural azide-modified sugar, was chemically synthesized from UDP-4-amino-4,6-dideoxy-GlcNAc (an intermediate in the UDP-diNAcBac biosynthetic pathway) through selective chemical chloroacetylation followed by azide substitution (Figure 5-5B). UDP-GalNAz was synthesized chemoenzymatically starting with N-acetylgalactosamine, as described previously (Figure 5-5C).[28]

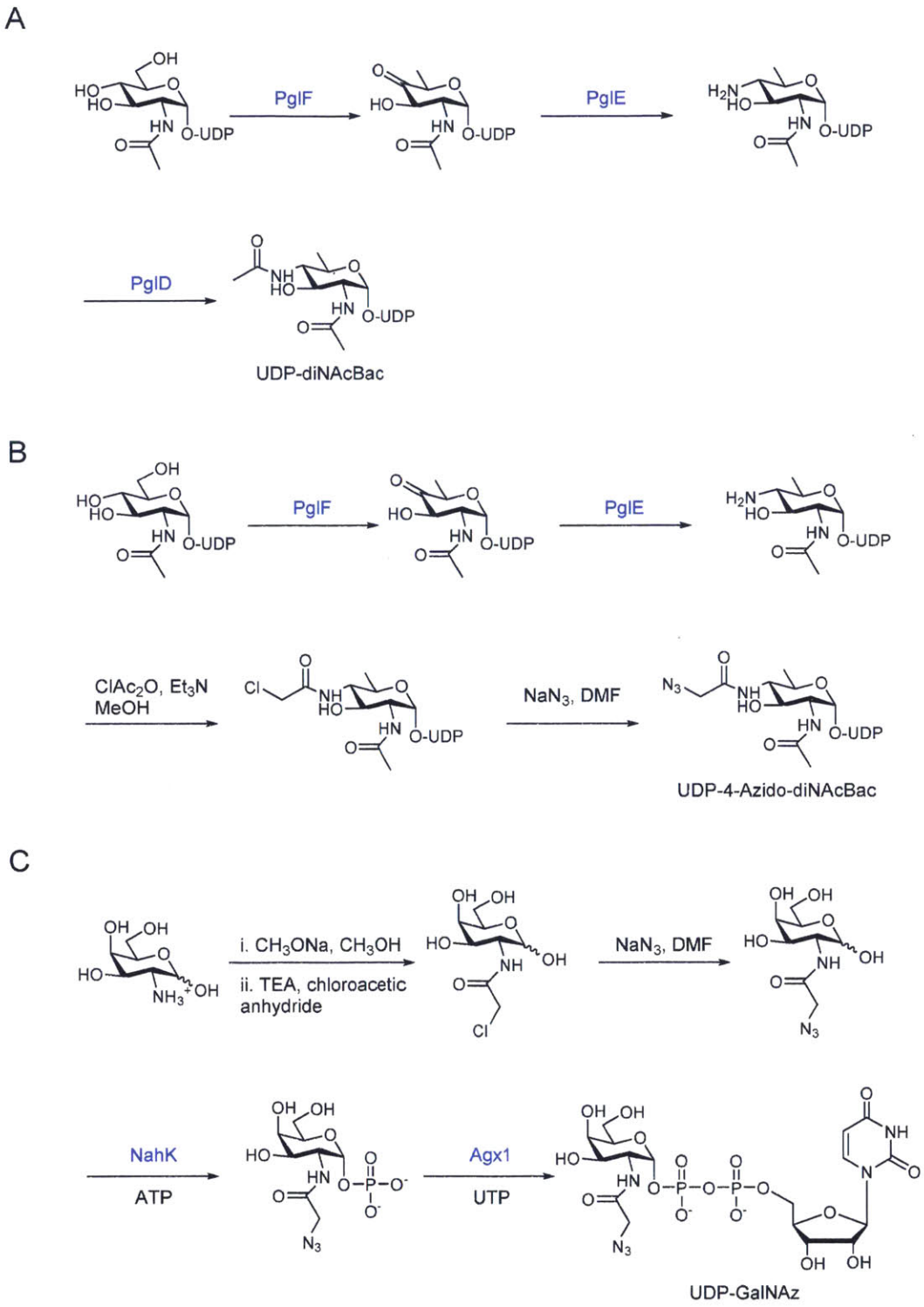


Figure 5-5 Chemoenzymatic synthesis of UDP-sugars used in this study: (A) UDP-diNAcBac (B) UDP-4-Azido-diNAcBac (C) UDP-GalNAz. Enzymes are highlighted in blue.

Overexpression and Purification of Pgl enzymes

Enzymes from the Pgl pathway were expressed heterologously in *E. coli*, and purified to homogeneity using Ni-NTA resin, with the exception of PglI, which was used as a cell envelope fraction (CEF) (Figure 5-6). PglC was used as purified protein for kinetic studies, and as a CEF for synthesis of glycan intermediates because it is more stable and active in the CEF form. All five of these glycosyltransferases are either integral membrane proteins or membrane associated proteins, and were prepared accordingly.

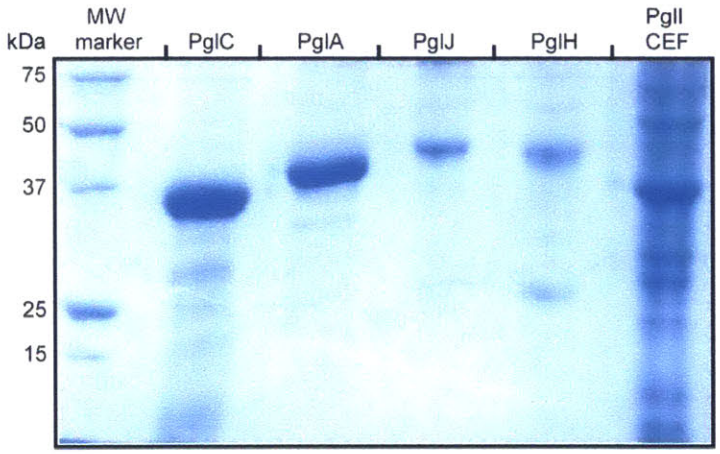


Figure 5-6 SDS-PAGE analysis of the Pgl enzymes used in this study. PglI was used as an impure CEF; the major band at ~35 kDa corresponds to the molecular weight of PglI.

Glycosyltransferase Assays with Azide-modified Carbohydrates

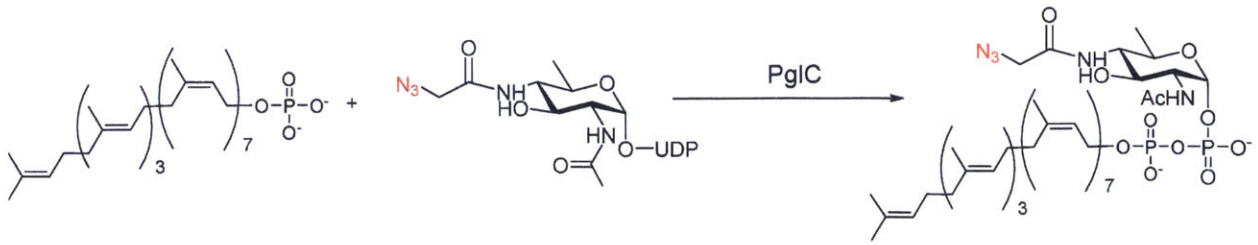
With these enzymes and substrates in hand, the tolerance of each of the glycosyltransferases was tested by comparing the amount of product formed between the natural carbohydrate substrate and the corresponding azide-derivative. Different analytical methods had to be used to evaluate each enzyme, as the separation between starting material and products varied greatly depending on the method of analysis. For example, radiolabeled starting materials

were used in the PglJ reaction, and the products could be clearly separated from reactants by normal-phase HPLC. However, using similar conditions for the PglA reaction did not result in adequate separation between the starting material and products, suggesting that the addition of a single sugar to the monosaccharide starting material did not sufficiently change the migration properties of the product on a normal-phase chromatography column. Reactions with PglA were therefore analyzed by cleaving the glycan from the polyprenol diphosphate, followed by normal-phase HPLC analysis. In the case of PglC, the addition of a sugar to undecaprenol phosphate would alter the migration time of the product. Using [^{32}P] or [^{33}P]-radiolabeled polyprenol phosphates as starting material would not be suitable for the time-frame of the synthesis and subsequent analysis due to the short half-lives of these isotopes. Thus, the PglC reaction was analyzed using capillary electrophoresis.

PglC Activity with UDP-4-Azido-diNAcBac

In order to compare the tolerance of PglC for UDP-diNAcBac and UDP-4-azido-diNAcBac, substrates were incubated with PglC and Und-P for 60 min, followed by quenching of the reaction and cleavage of the monosaccharide from the Und-PP lipid carrier. The free monosaccharide was fluorescently labelled with 8-aminopyrene-1,3,6-trisulfonate (APTS) for analysis by capillary electrophoresis (CE). To standardize peak area and electrophoretic mobility, maltose was used as an internal standard. The azide derivative was well tolerated with product formation corresponding to ~60-65% of the natural substrates rate at high carbohydrate concentrations (Figure 5-7).

A



B

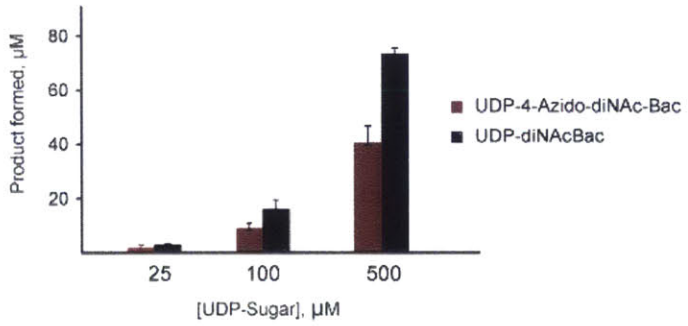
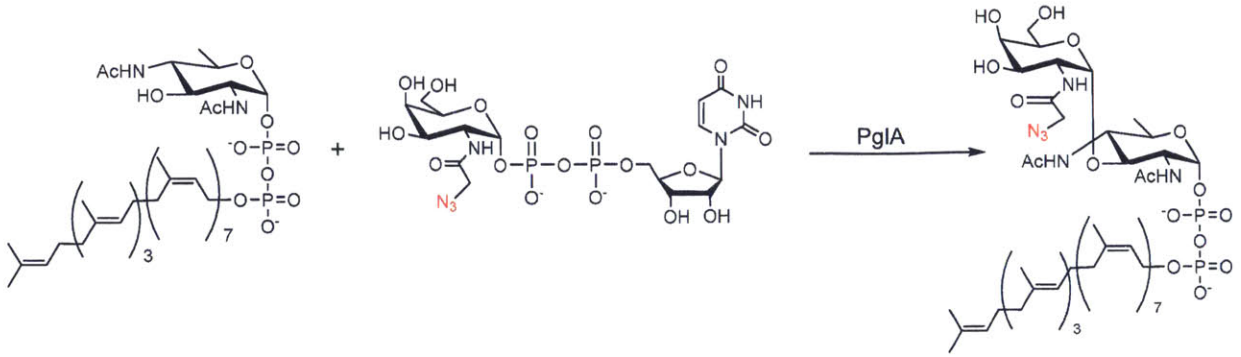


Figure 5-7 Analysis of PglC reactions with UDP-4-azido-diNAcBac and UDP-diNAcBac. (A) Reaction of carbohydrates with PglC. (B) Comparison of product formed with the two carbohydrate substrates.

PglA Activity with UDP-GalNAz

The tolerance of PglA for UDP-GalNAc and UDP-GalNAz was compared using a coupled assay with PglC to make the monosaccharide starting material. Reactions were performed with an excess of the UDP-sugars, and were quenched at 10 and 60 minutes, to isolate the disaccharide products The resulting disaccharide was cleaved from the Und-PP-linked carrier, and labeled with 2-aminobenzamide (2-AB) for analysis by normal-phase HPLC with fluorescence detection. The azide-modified substrate appears to be well tolerated by PglA with product formation comparable to the native substrate after 60 minutes (Figure 5-8).

A



B

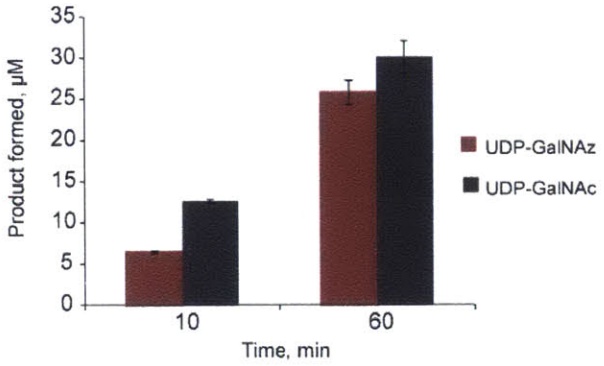


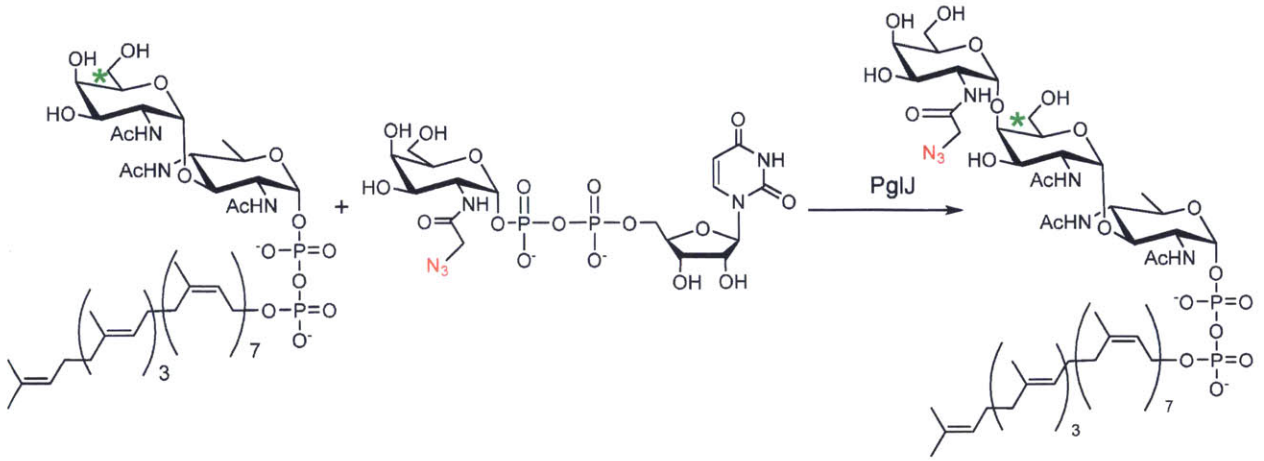
Figure 5-8 Analysis of PglA reactions with UDP-GalNAc and UDP-GalNAz. (A) Reaction of carbohydrates with PglA. (B) Comparison of product formed with the two UDP-sugar substrates.

PglJ Activity with UDP-GalNAz

The tolerance of PglJ for UDP-GalNAc and UDP-GalNAz were tested by incubating PglJ with [^{3}H]-labeled Und-PP-diNAcBac-GalNAc dissacharide and the two carbohydrate substrates (200 µM). Reactions were quenched at either a 10 or 60 min time point and purified by normal-phase HPLC. Levels of radioactivity were determined for each fraction to quantify product formation. The amount of product formed with UDP-GalNAz was 50% less than the product formed with UDP-GalNAc at the 10 min time point. However, both substrates showed > 90%

conversion after 60 min (Figure 5-9), thus indicating that PglJ can efficiently incorporate an azide into the Und-PP-glycan.

A



B

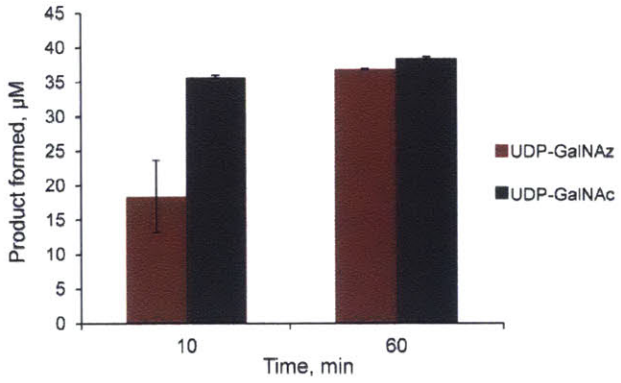


Figure 5-9 Analysis of PglJ reactions with UDP-GalNAc and UDP-GalNAz. (A) Reaction of carbohydrates with PglJ. (B) Comparison of product formed with the two carbohydrate substrates.

PglH Activity with UDP-GalNAz

The tolerance of PglH for UDP-GalNAz was tested using radiolabeled substrates. PglH was incubated with UDP-GalNAc and UDP-GalNAz and [^{3}H]-labeled Und-PP-diNAcBac-

GalNAc-GalNAc trisaccharide. Reactions were quenched after four hours, and analyzed by normal-phase HPLC. The radioactivity associated with each fraction was quantified to determine the amount of product formed. No azide-containing product was observed with PglH, even with a ten-fold excess of UDP-GalNAz (Figure 5-10). Reactions were also performed with 0.5 equivalents of UDP-GalNAz, as intermediate tetra-, penta- and hexasaccharides had previously been observed under these conditions[15] (data not shown), but no product was observed in this case either. Thus, under these conditions, PglH does not appear to accept UDP-GalNAz as a substrate.

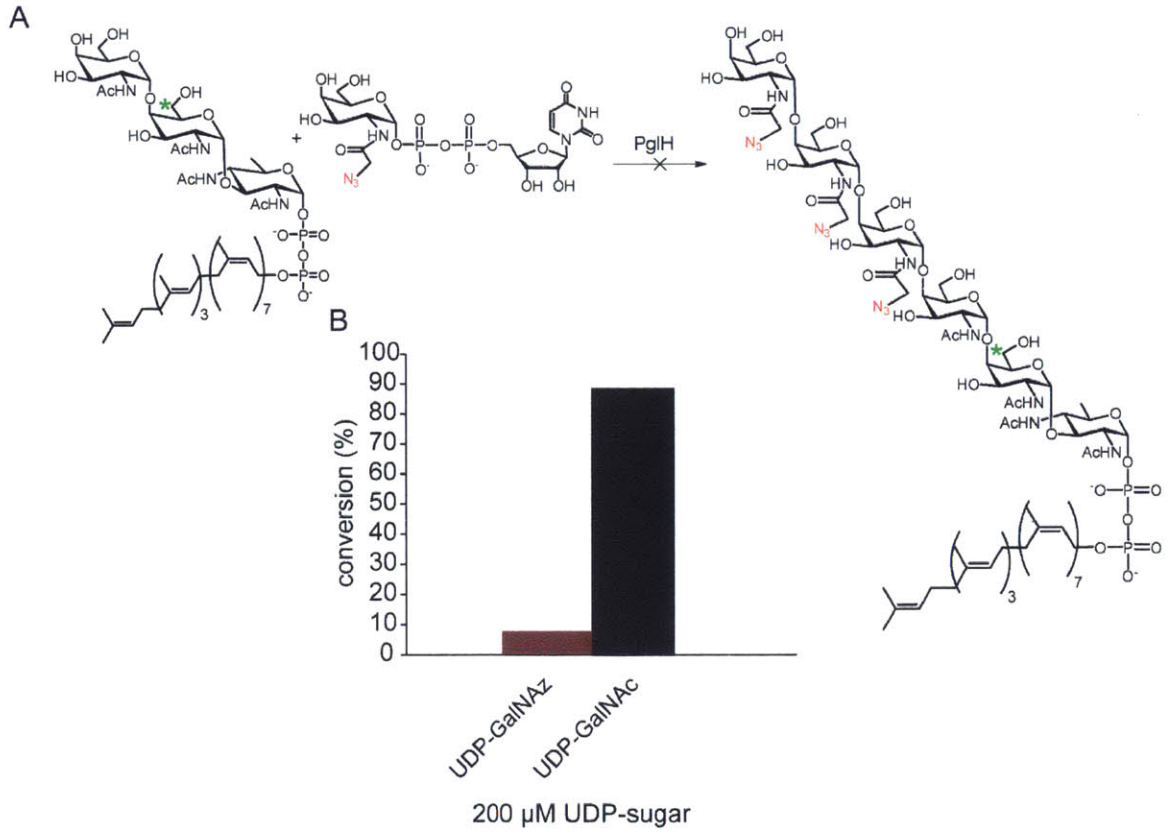


Figure 5-10 Analysis of PglH reactions with UDP-GalNAc and UDP-GalNAz. (A) Reaction of carbohydrates with PglH. The position of the [^{3}H] radiolabel is indicated with a green asterisk. (B) Comparison of product formed with the two UDP-sugar substrates.

Biosynthesis of Azide-Modified Glycoconjugates

The glycan was built up to the trisaccharide intermediate enzymatically, with an azide-modified carbohydrate inserted at each of the three positions (Figure 5-11). Quenching and washing steps were performed after the synthesis of each disaccharide to ensure that the azide-modified carbohydrate was added only at a single position in the trisaccharide. Synthesis of these products was confirmed using HPLC and mass spectrometry (Figure 5-11).

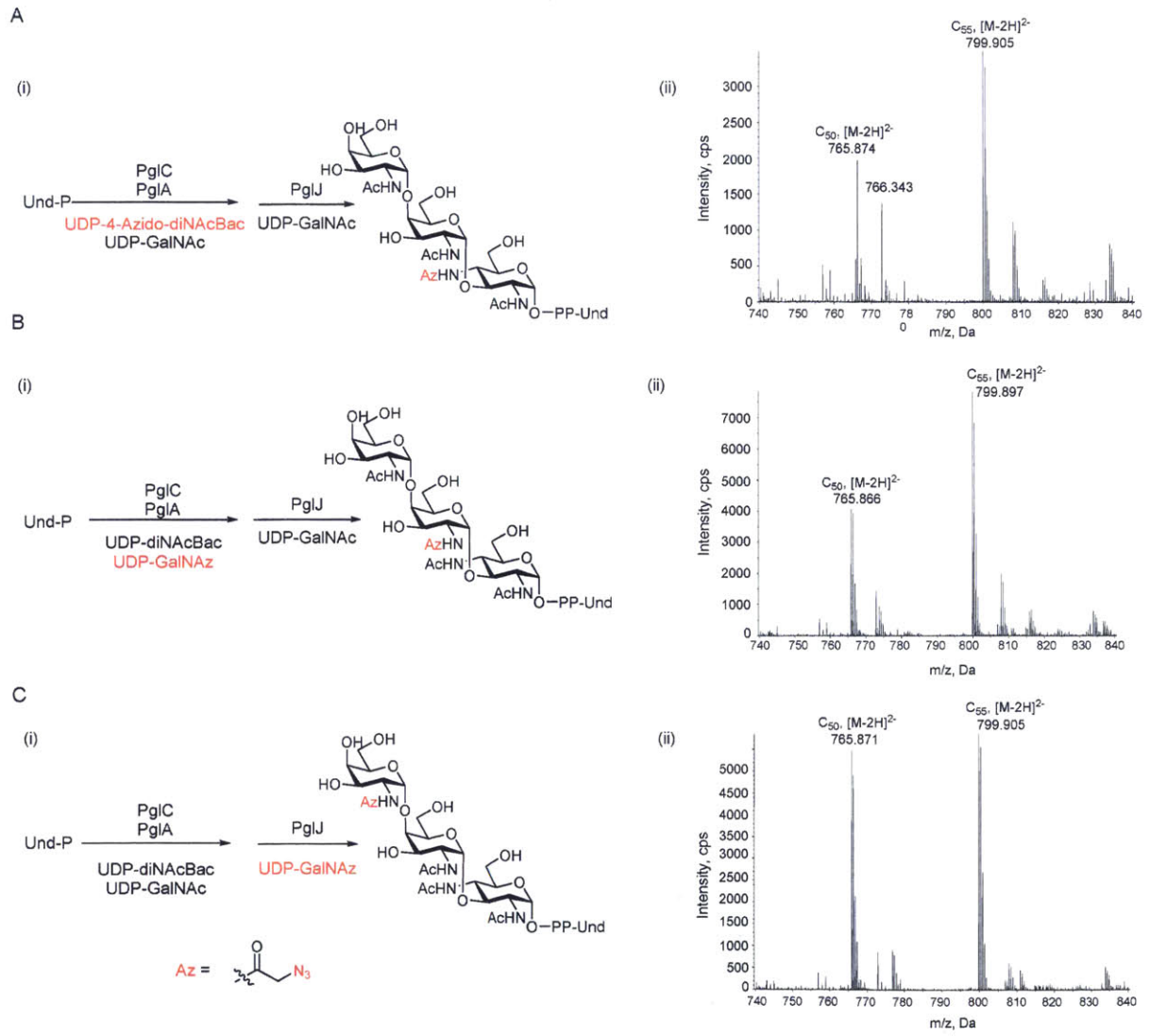


Figure 5-11 Synthesis and mass spectrometry analysis of of trisaccharides with azide-modified carbohydrates in each position. (A) Und-PP-4-azido-diNAcBac-$(GalNAc)_2$. (B) Und-PP-diNAcBac-GalNAz-GalNAc. (C) Und-PP-diNAcBac-GalNAc-GalNAz. Mass spectrometry traces show products with the C_{50} and C_{55} isoprenols.

Successful synthesis of trisaccharides with azide-modified carbohydrates in the first two positions demonstrates the ability of downstream glycosyltransferases to elaborate an azide-modified glycan precursor. To confirm that this is also true for a glycan with GalNAz at the third position, synthesis of the heptasaccharide was completed using PglH and PglI (Figure 5-12).

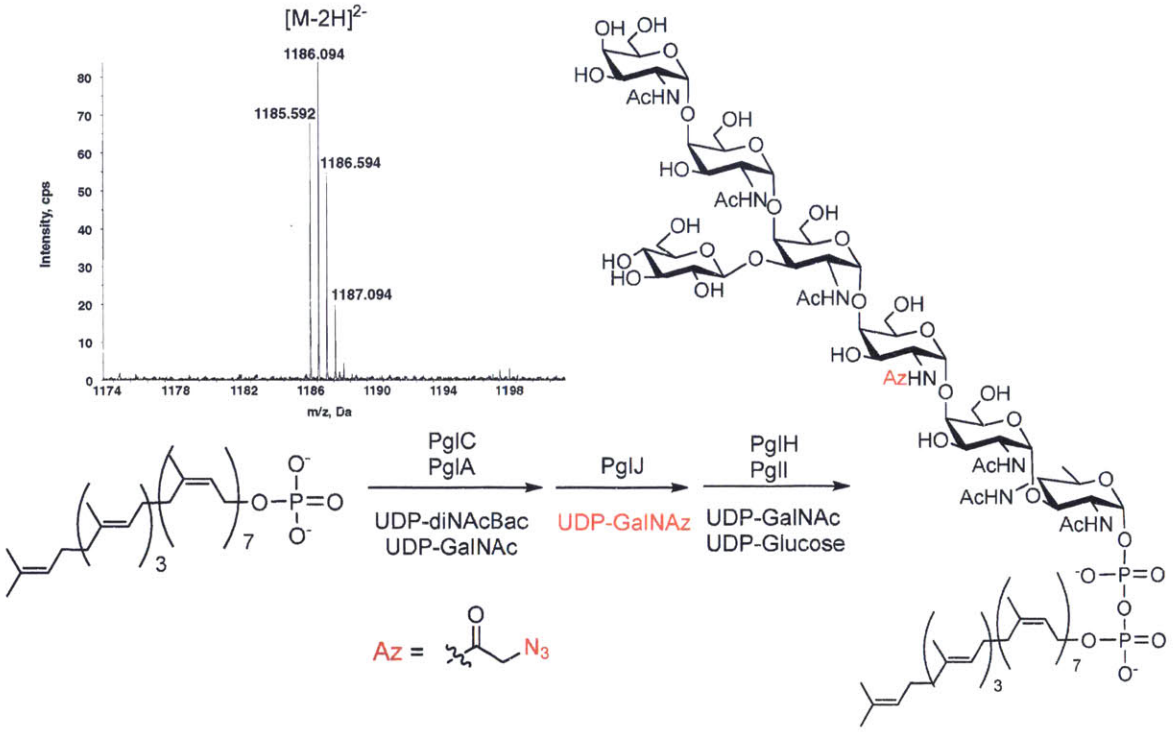


Figure 5-12 Synthetic scheme for the enzymatic synthesis of the full heptasaccharide containing an azide-modified carbohydrate in the third position. Inset: Mass spectrometry analysis of the azide-modified heptasaccharide product.

Oligosaccharyltransferase Assays with Azide-modified Glycoconjugates

In order for this system to be applicable to the biosynthesis of glycoproteins, the tolerance of PglB for the azide-modified glycans also needed to be verified. The ability of PglB to transfer each of these azide-modified trisaccharides to a peptide substrate was investigated using [^{3}H]-labeled substrates (Figure 5-13).

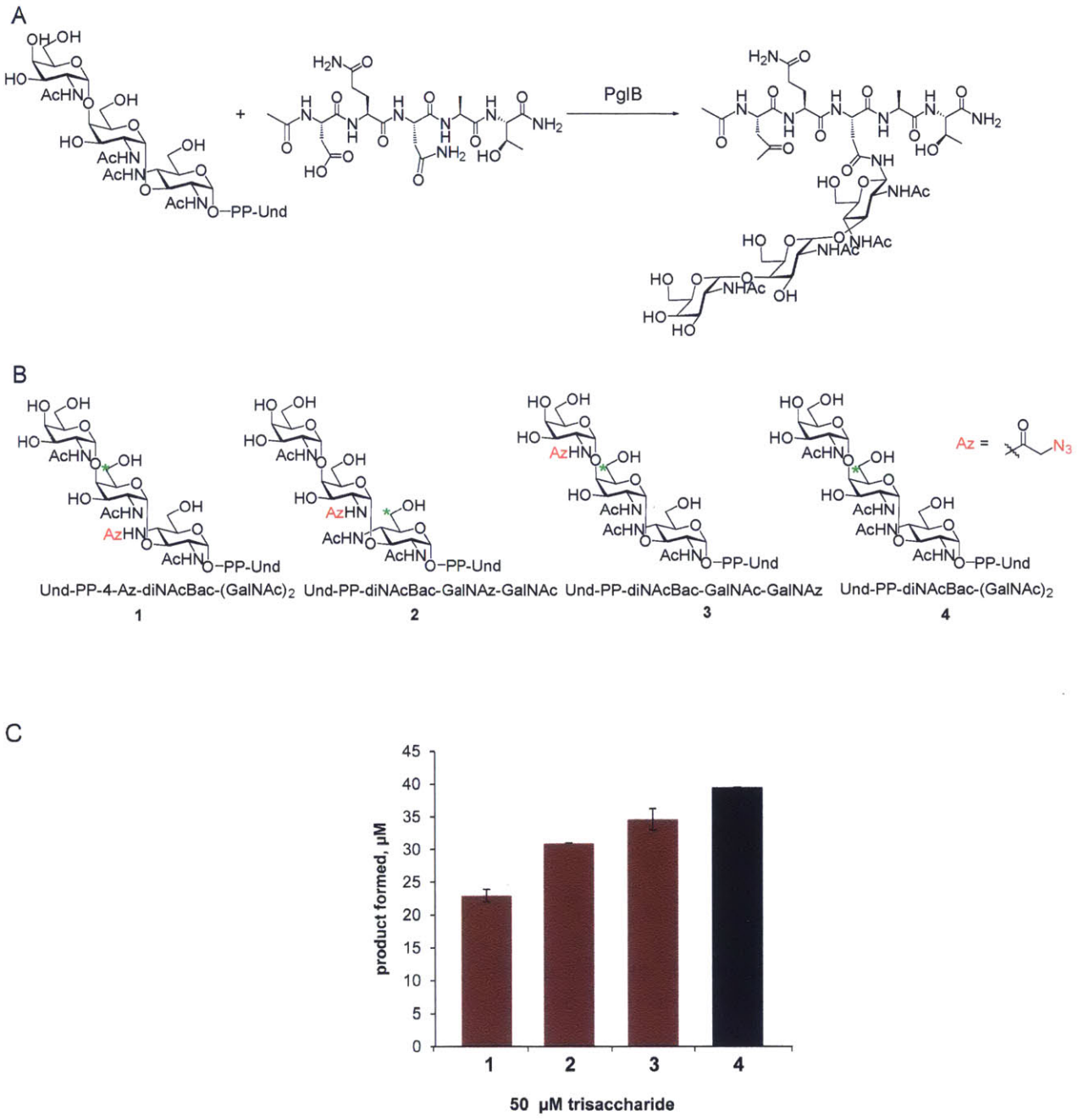


Figure 5-13 Analysis of PglB reactions with azide-modified trisaccharide substrates. (A) PglB-catalyzed transfer of a glycan to a peptide substrate. (B) Structures of native and azide-modified trisaccharide substrates. Radiolabeled positions are indicated with a green asterisk. (C) Comparison of products formed with all four trisaccharide precursors.

Click Reactions with Glycoconjugates

The utility of these glycoconjugates as tools for subsequent biorthogonal labeling was tested at the polyprenol diphosphate-linked glycan level as well as in the glycopeptide form. An azide-modified polyprenol diphosphate-linked trisaccharide was reacted with Acetylene-545 fluorophore using copper-catalyzed click chemistry. Copper-catalyzed click reactions with an azide-modified glycopeptide and alkyne-conjugated-biotin were also successful (Figure 5-14). The resulting fluorophore-linked glycoconjugates have potential for use in mechanistic studies of downstream glycosyltransferases in the Pgl pathway.

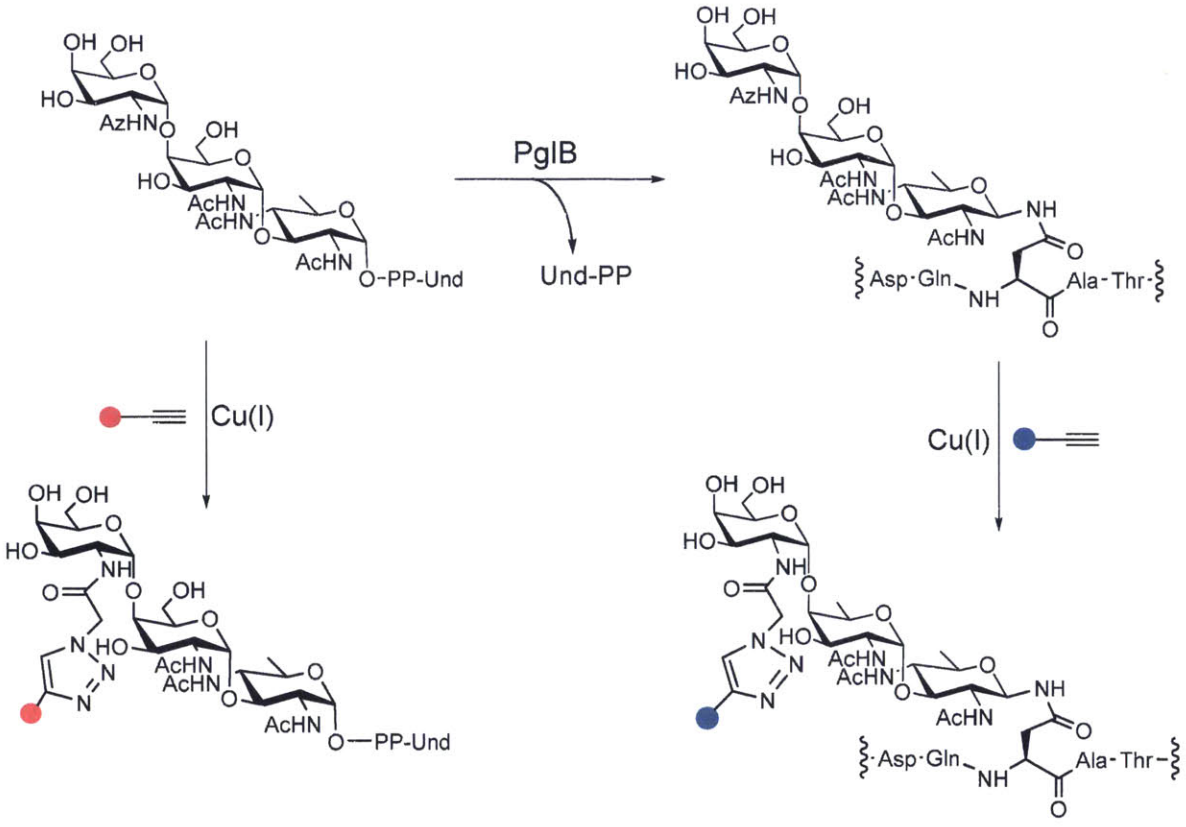


Figure 5-14 Scheme for copper-catalyzed azide-alkyne cycloaddition with azide-modified glycoconjugates. The trisaccharide can be modified at the polyprenol-glycan level as well as in glycopeptide form.

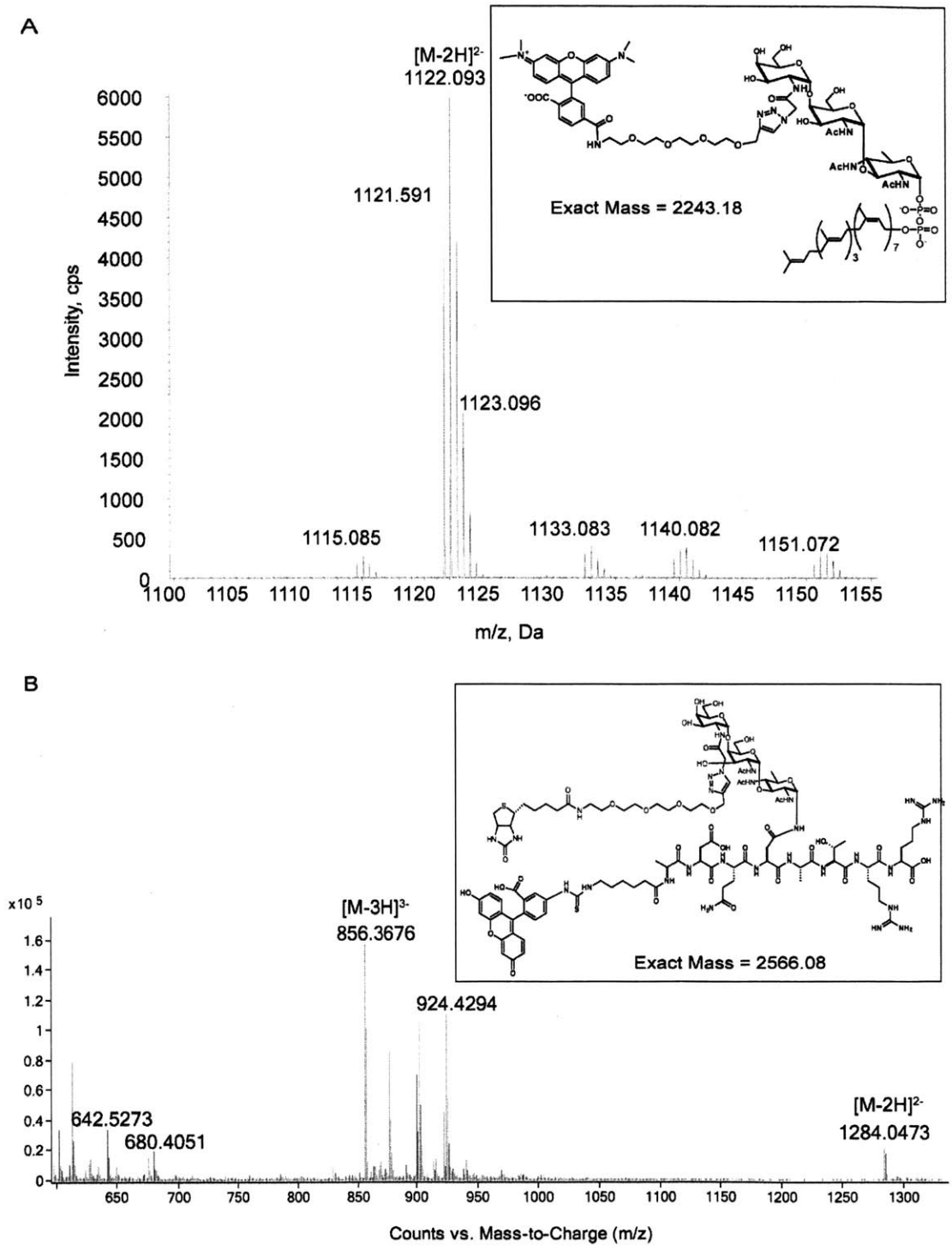


Figure 5-15 Mass spectrometry analysis of the azide-alkyne clicked glycoconjugate products. (A) Und-PP-diNAcBac-GalNAc-GalNAz clicked with Acetylene-545 fluorophore. Inset: chemical structure of the product. (B) Azide-modified glycopeptide (with the diNAcBac-GalNAc-GalNAz trisaccharide) clicked with alkyne-conjugated biotin. Inset: chemical structure of the product.

Metabolic Labeling of *C. jejuni* with Azide-Modified Carbohydrates

Many steps are required for effective incorporation of azide-modified glycans into N-linked glycoproteins in an *in vivo* system (Figure 5-16A). First, carbohydrates are delivered to cells as the corresponding peracetyl derivatives to increase cell membrane permeability and enter the cell by passive diffusion (Figure 5-16a). Once in the cytoplasm, promiscuous endogenous esterases must de-O-acetylate the carbohydrate (Figure 5-16b). This step is known to be inefficient in bacteria, due to slow esterase activity.[29] The azide-modified carbohydrate then needs to be phosphorylated at the C-1 position, followed by uridinylation to yield the UDP-activated substrate, ready for incorporation into the Pgl pathway (Figure 5-16c,d). In order for this step to be successful, enzymes responsible for biosynthesis of UDP-sugars in the cell must be able to tolerate the azide-modified substrate to a reasonable extent. If azide-modified sugars are converted to the corresponding UDP-activated forms in the cell, our studies strongly suggest that Pgl glycosyltransferases can incorporate them into glycoproteins (Figure 5-16e). Upon reaching the cell-surface, azide-modified glycoproteins can react with biorthogonal probes such as alkyne-modified fluorophores or purification tags such as biotin (Figure 5-16h). An alternate biosynthesis pathway to UDP-GalNAc, via the GalE enzyme that catalyzes the epimerization between UDP-GlcNAc and UDP-GalNAc, exists in *C. jejuni* (Figure 5-16j).[30] If Ac_4-GlcNAz was incorporated into the cell and activated to UDP-GlcNAz, and if GalE could epimerize this sugar to UDP-GalNAz, this would provide an alternate way to incorporate azide-modified sugars into the N-glycan. The success of this strategy would require GlmU to uridinylate GlcNAz-1-phosphate (Figure 5-16i), and GalE to be promiscuous enough to epimerize an azide-modified substrate.

Our studies demonstrate that multiple *C. jejuni* glycosyltransferases can tolerate azido sugar substrates. However, attempts to metabolically label *C. jejuni* glycoproteins *in vivo* have proven unsuccessful, presumably due to the absence of an efficient carbohydrate salvage pathway which would activate a carbohydrate to the UDP-linked form that could be used by glycosyltransferases. We attempted metabolic labeling of *C. jejuni* with many protected azide-modified carbohydrates (Figure 5-16B). Our first attempts used acetylated GalNAz and GlcNAz, and relied on the efficiency of esterases as well as biosynthetic enzymes to activate these sugars to their UDP-conjugates. *C. jejuni* cells were grown in the presence of azide-modified sugars, lysed and then incubated with alkyne-conjugated biotin, followed by western blot analysis using an α-streptavidin antibody. The absence of α-streptavadin-reactive bands indicated that metabolic labeling with these sugar derivatives was unsuccessful (data not shown). We hypothesized that the phosphorylation step in the biosynthesis of UDP-sugars may be the limiting step; therefore this step was bypassed by using phospho-sugars with peracetylated hydroxyl groups and protected at the phosphate moiety with S-acetyl-2-thioethyl (SATE) groups used for metabolic labeling. The SATE protecting group has been used previously for similar purposes in the metabolic labeling of O-glycans and to introduce bioorthogonally reactive lipids into eukaryotic cells.[31,32] However, when analyzed by western blot using a an α-streptavidin antibody, we were unable to detect bands corresponding to azide-modified glycoproteins, suggesting that azide-modified sugars were still not being incorporated into glycoproteins. These results suggest that native UDP-sugar biosynthetic enzymes are unable to accept azide-modified sugars as substrates. One strategy to circumvent this issue could be to engineer salvage pathway in *C. jejuni* to include enzymes known to convert the azide-modified substrates into their UDP-activated forms, such as NahK and Agx1 (enzymes used in the chemoenzymatic synthesis of

these sugars, Figure 5-5). We are currently investigating derivatives that could be efficiently incorporated into the glycosylation pathway. Additionally, we have developed a method to label the N-glycan using GalO, an enzyme that selectively modified terminal GalNAc residues.[33]

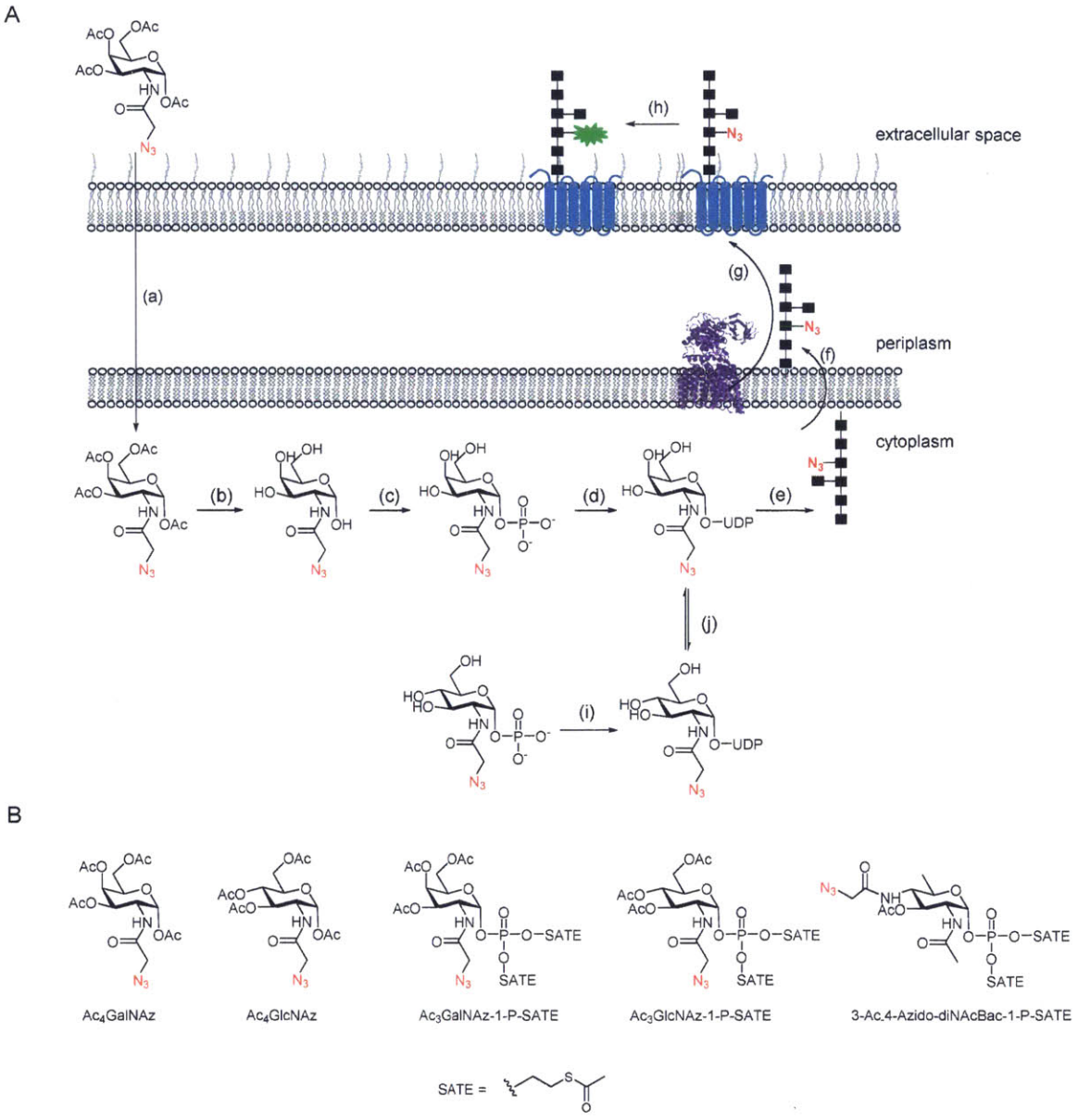


Figure 5-16 (A) Scheme for the metabolic labeling of *C. jejuni*: (a) Passive diffusion across membranes into the cytosol, (b) De-acetylation by esterases, (c) Phosphorylation by a hexose kinase, (d) Uridinylation by GlmU, (e) Incorporation into the N-glycan by the Pgl pathway enzymes, (f) Transfer to the periplasm by the flippase, PglK, (g) Synthesis of glycoproteins by PglB, (h) Click chemistry to append an extrinsic tag such as a fluorophore, (i) Uridinylation of GlcNaz by GlmU, (j) Epimerization of UDP-GlcNAz to UDP-GalNAz by GalE. (B) Protected azide-modified sugar derivates used for metabolic labeling studies.

Methods for efficient preparation and modification of complex glycans with defined modification sites are critical for analyzing the molecular basis for interactions between bacteria and host cells. Bacterial glycans present particular challenges for study due to the prevalence of highly modified "non-standard" carbohydrates, exacerbating the task of chemical synthesis. We have previously demonstrated that the *C. jejuni* N-glycan can be prepared on an analytical scale, in a one-pot reaction, using enzymes in the Pgl pathway.[34] The chemical synthesis of the *C. jejuni* N-linked glycan is extremely labor-intensive[35] and requires significant repurposing for the assembly of variants that include uniquely modified carbohydrates. Therefore we set out to establish the practicality and limitations of a general chemoenzymatic approach using native and modified nucleotide sugar donors.

Chemoenzymatic methods provide an important complement to chemical synthesis enabling access to chemically defined materials for biological studies including the generation of glycan arrays, vaccines, and molecular imaging probes.[36] Here we present a systematic approach for the production of chemically defined glycans with bioorthogonal conjugation handles representing intermediates and products in a well-characterized bacterial pathway for protein glycosylation derived from *Campylobacter jejuni*. Since the bacterial gene clusters encoding enzymes in the biosynthesis and utilization of polyprenol-diphosphate linked glycans in N- and O-linked bacterial protein glycosylation can now be readily identified using bioinformatics approaches,[37,38] we anticipate that this work will provide a useful precedent for the application of parallel approaches for the preparation of glycan targets from other physiologically clinically relevant pathogenic bacteria.

Conclusions

The *in vitro* studies presented in this chapter underscore the importance of defining substrate tolerances of enzymes prior to embarking on metabolic labeling studies. To our knowledge, this is the first report demonstrating the tolerance of the Pgl glycosyltransferases towards azide-modified carbohydrates. The glycosyltransferases PglC, PglA and PglJ are each capable of introducing azide-modified carbohydrates into specific positions of the *C. jejuni* N-glycan, while PglH appears to be unable to accept azide-modified substrates. The oligosaccharyltransferase, PglB, can transfer azide-modified glycans to a peptide substrate. The resulting glycoconjugates are valuable chemical biology tools, allowing the specific labeling of substrates and products of glycosyltransferases in order to mechanistically characterize enzymes in bacterial N-glycosylation pathways, and to further investigate this complex multienzyme process.

Acknowledgements

Dr. Garrett Whitworth was an invaluable collaborator, bringing his excellent carbohydrate synthesis skills to this project. Dr Ziqiang Guan's mass spectrometry analysis was extremely helpful for this project. I am thankful to Professor George Wang for sending us plasmids encoding genes for NahK and Agx1. Monika Musiel-Siwek very kindly provided us with precious, pure PglB for glycopeptide experiments. Cristina Zamora's patience in fixing the fluorescence detector and HPLC is greatly appreciated. I am also grateful to Amy Rabideau and the Pentelute Lab for access to their mass spectrometers for glycopeptide analysis. Finally, I am thankful to Julie Silverman and Joris De Schutter for reading this chapter and providing helpful comments.

Experimental Methods

Synthesis of polyprenol monophosphate (C50-60)

Polyprenol (C50-60) was extracted from the leaves of *Rhus typhina* (staghorn sumac) as previously described.[26] Bis(9*H*-fluoren-9-ylmethyl)-diisopropylamidophosphite was prepared from diisopropylamidophosphodichloridite (Sigma-Aldrich) as described by Watanabe *et al.* [39] Bis(9*H*-fluoren-9-ylmethyl)-diisopropylamidophosphite (90 mg, 0.20 mmol) was dissolved in THF (3 mL, dry) and added dropwise to a solution of polyprenol (50 mg, 0.07 mmol) in THF (3 mL, dry) at 0 °C. 1-H tetrazole (14 mg, 0.20 mmol) was added and the reaction was allowed to stir overnight at room temperature. Once the reaction was judged to be complete by TLC (Toluene:EtOAc, 99.5:0.5) it was cooled to -78 °C and H_2O_2 (2 drops of 30% H_2O_2, BDH) was added. The reaction was allowed to come up to room temperature and stirred for an additional 3 h. Upon completion, the reaction was quenched with 10% sodium sulfite (10 mL) and the compound was extracted with EtOAc (3 x 40 mL). The organic layers were washed with 1 M KH_2PO_4 (10 mL), saturated $NaHCO_3$ (10 mL), and brine (10 mL). The organic layer was dried ($MgSO_4$) and concentrated, the resulting oil was dissolved in ACN:Et_3N (9:2, 6 mL, dry) and stirred at room temperature, overnight. The reaction was dry loaded onto a flash chromatography column and the product was eluted with a gradient of EtOAc:MeOH to EtOAc:MeOH:H_2O:Et_3N (9:2 to 5:2:0.5:0.01). The desired polyprenol monophosphate (C50-60) was isolated as a colorless oil (55% yield over 3 steps) with ^1H and ^{31}P NMR corresponding to published values.[40]

Synthesis of UDP-4-azido-diNAcBac

The starting material, UDP-4-amino-4,6-dideoxy-GlcNAc was prepared enzymatically from UDP-GlcNAc as described by Olivier and coworkers.[27] Chloroacetic anhydride (87 mg, 510 μmol) and triethylamine (71 μL, 510 μmol) were added to a solution of UDP-4-amino-6-

deoxy-GlcNAc (15 mg, 26 μmol) in anhydrous MeOH (3 mL). The solution was stirred for 3 days at room temperature, concentrated by rotary evaporation and purified on a HiTrap Q HP anion exchange column (5 mL, GE Healthcare Life Sciences) using an AKTA prime plus FPLC. A flow rate of 2 mL/min and a gradient of 1% to 50% Buffer B (Buffer A = H_2O, Buffer B = 0.9 M NH_4HCO_3 in H_2O) over 120 mL was used to elute the desired product. Fractions containing the desired product were lyophilized and redissolved in H_2O 3 times to remove NH_4HCO_3. The isolated chloroacetyl-intermediate was dissolved in anhydrous DMF (5 mL). Sodium azide (17 mg, 260 μmol) was added and the suspension was stirred at room temperature for 3 days. DMF was removed by rotary evaporation, the resulting residue was dissolved in H_2O and purified on the HiTrap Q HP anion exchange column as previously described. Fractions containing UDP-4-azido-diNAcBac were lyophilized and redissolved in H_2O 3 times to remove NH_4HCO_3. UDP-4-azido-diNAcBac (12 mg, 70% over two steps) was isolated as a white solid and characterized by NMR. ^{1}H NMR (400 MHz, D_2O) δ 1.35-1.40 (d, 3H, *J* = 6.2 Hz, H-6"), 2.08 (s, 3H, NHAc-**CH_3**), 2.97-3.03 (dd, 1H), 3.10-3.20 (m, 2H), 3.23-3.27 (m, 1H), 3.89-3.92 (dd, 1H, *J* = 5.4 Hz), 3.93-3.97 (dd, 1H, *J* = 10.1 Hz), 4.10-4.15 (ddd, 1H, *J* = 3.0, 6.1, 10.1 Hz), 4.17 (s, 2H, C-**CH_2**-N_3), 4.19-4.35 (m, 5H), 4.36-4.38 (dd, 1H, *J* = 3.0 Hz),5.50-5.54 (dd, 1H, *J* = 3.2, 7.0 Hz, H-1"), 5.94-5.97 (d, 1H, *J* = 8.1 Hz, H-5), 5.97-6.00 (d, 1H, *J* = 4.6 Hz, H-1'), 7.94-7.97 (d, 1H, *J* = 8.1 Hz, H-6).

Synthesis of UDP-GalNAz

UDP-GalNAz was prepared chemoenzymatically from N-acetyl-galactosamine using previously described methods. [28]

Protein expression and purification

PglB was cloned and purified as described previously.[41] PglA, PglJ, PglH, and PglI were cloned as described previously.[34] PglC was prepared as a SUMO fusion, using the pET-SUMO vector. Plasmids were transformed into *E. coli* BL21-CodonPlus (DE3)-RIL cells (Agilent) using kanamycin and chloramphenicol for selection. Protein was expressed using a modified autoinduction method described by Studier.[42] In this method, 1 mL of an overnight cell culture was added to expression media containing 30 μg/mL kanamycin and 30 μg/mL chloramphenicol in 1 L of autoinduction media (0.1% (w/v) tryptone, 0.05% (w/v) yeast extract, 2 mM MgSO4, 0.05% (v/v) glycerol, 0.005% (w/v) glucose, 0.02% (w/v) α-lactose, 2.5 mM Na_2HPO_4, 2.5 mM KH_2PO_4, 5 mM NH_4Cl, 0.5 mM Na_2SO_4). Cells were allowed to grow with shaking for 3 h at 37 °C. After 3 h, the temperature was decreased to 16 °C, and the cells were incubated with shaking overnight. Cells were harvested at 9000*g*, and cell pellets were stored at -80 °C.

Purification of PglA, PglJ, PglH

Cell pellets were resuspended in 10% of the original culture volume in buffer A (50 mM Tris pH 8, 100 mM NaCl) plus 1% Triton and 50 μL protease inhibitor cocktail (Calbiochem). The cells were lysed by two rounds of sonication for 90 seconds each, at an amplitude of 50% with one-second on/off pulses. Cell debris was removed by ultracentrifugation at 142,000*g*, and the supernatant was incubated with 5 mL of Ni-NTA resin pre-equilibrated in buffer A. The sample was left on a rotator for 1 h at 4 °C and then poured into a column. The resin was washed (40 mL) with buffer B (50 mM Tris pH 8, 100 mM NaCl, 20 mM imidazole), and then washed (80 mL) with buffer C (50 mM Tris pH 8, 100 mM NaCl, 45 mM imidazole). The proteins were then eluted with buffer D (50 mM Tris pH 8, 100 mM NaCl, 300 mM imidazole). Elutions were dialyzed overnight in 4 L of buffer A. All purification steps were performed at 4 °C.

Purification of PglC, PglI

In the case of PglC and PglI, cell pellets were resuspended in 10% of the original culture volume in buffer A (50 mM HEPES pH 7.5, 100 mM NaCl) and a protease inhibitor cocktail (Calbiochem) and then were subjected to sonication. To prepare cell membrane fractions, the cell lysate was centrifuged to remove cellular debris (6,000*g*, 45 min), followed by pelleting of the membranes (142,000*g*, 1 h). The cell envelope fraction (CEF) was then resuspended in 0.25% of the original culture volume in buffer A (50 mM HEPES, pH 7.5, 150 mM NaCl) and stored at −80 °C. The CEF was either used directly for synthesis of isoprene substrates, or was purified further to obtain homogenous protein. To purify protein, the CEF was homogenized in 1% *n*-dodecyl-β-D-maltoside (DDM) and incubated on a rotator overnight at 4 °C after the addition of a protease inhibitor cocktail. The homogenized CEF was centrifuged (142,000*g*, 60 min), and the resulting supernatant was incubated with Ni-NTA resin for 3 h. The resin was washed (40 mL) with buffer B (50 mM HEPES pH 7.5, 100 mM NaCl 20 mM imidazole, 0.03% DDM), and then washed (40 mL) with buffer C (50 mM Tris pH 8, 100 mM NaCl, 45 mM imidazole, 0.03% DDM). The proteins were then eluted with buffer D (50 mM Tris pH 8, 100 mM NaCl, 300 mM imidazole, 0.03% DDM). Elutions were dialyzed overnight in 4 L of buffer A. All purification steps were performed at 4 °C.

Reactions with PglC, Und-P and UDP-diNAcBac/UDP-4-azido-diNAcBac

Undecaprenol-phosphate, at a final concentration of 500 μM was resuspended in 1% Triton X-100 and 3% DMSO by vortexing and sonication. To this was added 5 mM $MgCl_2$, 30 mM Tris pH 8.0, 10 nM PglC, and 50 μM, 200 μM or 500 μM UDP-diNAcBac or UDP-4-azido-diNAcBac. Reactions were quenched in 900 μL 2:1 $CHCl_3$:MeOH and washed three times with 400 μL PSUP. The organic layer was dried and resuspended in 500 μL 1:1 H_20:isopropanol + 20

μL 2M Trifluoroacetic acid, and incubated in a sand bath at 50 °C for 20 minutes to cleave the glycan from the isoprene. Following cleavage the dried, reducing monosaccharide samples and internal standards were subject to reductive-amination with 8-aminopyrene-1,3,6-trisulfonate (APTS, Beckman-Coulter). Reductive-amination was carried out with minor modifications to the Beckman-Coulter protocol. Samples were dissolved in 200 mM APTS in 7.5% acetic acid (1 μL), 18Ω deionized H_2O (1 μL), and 1.0 M $NaBH_3CN$ in THF (2 μL, Sigma-Aldrich) and incubated in a 60 °C sand bath for 90 min. After labeling, samples were diluted with 196 μL of 18Ω deionized H_2O to bring the total volume up to 200 μL, and analyzed with a ProteomeLab PA800 (Beckman-Coulter) equipped with a laser-induced fluorescence (LIF) detector and a 488 nm argon-ion laser. An N-CHO capillary (Beckman-Coulter) with an i.d. of 50 μm and a length of 50 cm was used to determine the electrophoretic mobility of each sample.

Reactions with PglA, Und-PP-diNAcBac and UDP-GalNAc/UDP-GalNAz

The tolerance of PglA for the two substrates was compared using a coupled PglC/PglA assay. Undecaprenol-phosphate, at a final concentration of 40 μM was resuspended in 1% Triton X-100 and 3% DMSO by vortexing and sonication. Solubilized undecaprenol-phosphate was incubated with 5 mM $MgCl_2$, 30 mM Tris pH 8.0, 1 μM PglC, 200 μM UDP-diNAcBac, 10 nM PglA, 200 μM UDP-GalNAc. Reactions were quenched in 500 μL 2:1 $CHCl_3$:MeOH and washed three times with 300 μL PSUP. The organic layer was dried and resuspended in 500 μL 1:1 H_2O:isopropanol + 20 μL 2M Trifluoroacetic acid, and incubated in a sand bath at 50 °C for 20 minutes to cleave the glycan from the isoprenol-diphosphate. Following cleavage the dried, reducing monosaccharide samples and internal standards were reductively-aminated with 2-aminobenzamide (2-AB, Sigma). A solution of 2-AB (5 mg) was prepared in 100 μl of a 30% (v/v) acetic acid:DMSO mixture. This dye solution was added to 6 mg of sodium

cyanoborohydride, and aliquots of 10 µl of this reagent were added to dried samples of glycans for labeling, and heated on a sand bath to 60°C for 4 h. After labeling, excess fluorophore was removed using Glycoclean cartridges (ProZyme), following the recommended protocol. Purified samples analyzed by normal-phase HPLC using a GlykosepN column (ProZyme, 250 x 4.6 mm), and detected with an excitation wavelength of 330 nm and an emission wavelength of 420 nm. Peak areas were quantified using a 2-AB labeled standard curve of GlcNAc, which was purified in the same manner.

Reactions with PglJ, Und-PP-diNAcBac-GalNAc and UDP-GalNAc/UDP-GalNAz

Tritiated Und-PP-diNAcBac-GalNAc (0.016 mCi/µmol), at a final concentration of 40 µM was resuspended in 0.08% Triton X-100 and 10% DMSO by vortexing and sonication. To this was added 50 mM $MnCl_2$, 50 mM Bicine pH 8.5, 90 nM PglJ, and 100 µM UDP-GalNAc or UDP-GalNAz. Reactions were quenched in 500 µL 2:1 $CHCl_3$:MeOH and dried down. Products were purified using a normal phase column (YMC-PVA-SIL-NP, 250 × 4.6 mm) and separated over 8-14% Buffer B gradient where Buffer A is 4:1 $CHCl_3$:MeOH and Buffer B is 10:10:3 $CHCl_3$:MeOH:2M NH_4HCO_3. Purified fractions (1 mL) were combined with 5 mL OptiFluor (PerkinElmer) and were analyzed on a Beckman Coulter LS6500 scintillation counting system.

Azide/Alkyne Copper-catalkyzed Click Chemistry

Und-PP-diNAcBac-GalNAc-GalNAz (0.25 mg, 156 nmol) was dissolved in *tert*-butanol (1 mL) and Acetylene-Fluor 545 (30 µL of a 20 mM solution in DMSO, 600 nmol, Click Chemistry Tools) was added. Deionized H_2O (0.94 mL) was added to the reaction mixture along with $CuSO_4$ (30 µL of a 20 mM solution in deionized H_2O) and sodium ascorbate (30 µL of a 20 mM solution in deionized H_2O). The reaction vessel was sealed, covered in aluminum foil and shaken at room temperature for 16 hours. Silica was added and the reaction was concentrated and

dry loaded onto a column for purification (EtOAc:MeOH:H_2O:Et_3N, 5:2:0.5:0.01). The Und-PP-diNAcBac-GalNAc-GalNAz-triazole-PEG4-TAMRA product was isolated as a pink syrup (Yield 40%): (EtOAc:MeOH:H_2O:Et_3N, 5:2:0.5:0.01) R_f = 0.30; HRMS (*m/z*): $[M-2H]^{2-}$ calcd. for $C_{117}H_{172}N_{10}O_{29}P_2^{2-}$ = 1121.589, found = 1121.591.

Azide-modified glycopeptide, glycosylated with diNAcBac-GalNAc-GalNAz, (5 μg, 2 nmol) was dissolved in water at a final concentration of 40 μM. Acetylene-PEG4-Biotin (Click Chemistry Tools) was added at a final concentration of 1 mM. The reaction was supplemented with 100 μM CuSO4, 500 μM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) and 200 μM sodium ascorbate, and shaken at room temperature for 16 hours. Formation of the clicked glycopeptide was confirmed by LC-MS. $[M-2H]^{2-}$ calcd. for $C_{109}H_{160}N_{28}O_{40}S_2$ = 1282.54, found = 1283.04.

Synthesis of Und-PP-4-Azido-diNAcBac-GalNAc-GalNAc

Und-P (0.76 mg), at a final concentration of 500 μM, was resuspended in 2% Triton X-100 and 3% DMSO by vortexing. To this mixture, 2 mM $MgCl_2$, 30 mM Tris pH 8.0, 2 mM UDP-4-azido-diNAcBac, 500 μM [^{3}H]-UDP-GalNAc (5.7 mCi/mmol), 35 μM PglA and 70 μl PglC CEF were added to synthesize the disaccharide, Und-PP-4-azido-diNAcBac-GalNAc. The reaction was incubated with shaking at room temperature overnight, quenched in 3 mL 2:1 $CHCl_3$:MeOH, and washed three times with 2 mL PSUP (Pure Solvent Upper Phase, composed of 15 mL $CHCl_3$, 240 mL MeOH, 1.83 g KCl in 235 mL H_2O). The disaccharide product was dried down, and resuspended in 5% DMSO and 0.05 % Triton X-100. 50 mM Bicine pH 8.5, 10 mM $MnCl_2$, 640 μM UDP-GalNAc, and 400 μM PglJ were added and the reaction was incubated with shaking at room temperature overnight. The reaction was dried down and purified using

normal phase HPLC methods. The trisaccharide product (52% yield over 2 steps, 0.75 mg) was further characterized using high-resolution mass spectrometry.

Synthesis of Und-PP-diNAcBac-GalNAz-GalNAc

Und-P, at a final concentration of 500 μM (0.8 mg) was resuspended in 2% Triton X-100 and 3% DMSO by vortexing. 2 mM $MgCl_2$, 600 μM [^{3}H]-UDP-diNAcBac (5.2 mCi/mmol), 30 mM Tris pH 8.0, 2.4 mM UDP-GalNAz, 35 μM PglA and 35% (v/v) PglC as a CEF isolate were added to synthesize the disaccharide, Und-PP-diNAcBac-GalNAz. The reaction was incubated with shaking at room temperature overnight, and quenched in 3 mL 2:1 $CHCl_3$:MeOH, and washed three times with 2 mL PSUP. The disaccharide product was dried down, and resuspended in 5% DMSO and 0.05 % Triton X-100. 50 mM Bicine pH 8.5, 10 mM $MnCl_2$, 300 μM UDP-GalNAc, and 140 μM PglJ were added and the reaction was incubated with shaking at room temperature overnight. The reaction was dried down and purified using normal phase HPLC methods. The trisaccharide product (35% over 2 steps, 1 mg) was further characterized using high-resolution mass spectrometry.

Synthesis of Und-PP-diNAcBac-GalNAc-GalNAz

Und-P, at a final concentration of 500 μM (0.6 mg) was resuspended in 2% Triton X-100 and 3% DMSO by vortexing. To this mixture, 2 mM $MgCl_2$, 30 mM Tris pH 8.0, 600 μM UDP-diNAcBac, 500 μM [^{3}H]-UDP-GalNAc (3.3 mCi/mmol), 90 μM PglA and 35% (v/v) PglC as a CEF isolate were added to synthesize the disaccharide, Und-PP-diNAcBac-GalNAc. The reaction was incubated with shaking at room temperature overnight, and quenched in 3 mL 2:1 $CHCl_3$:MeOH, and washed three times with 2 mL PSUP. The reaction yield (78%) was determined by quantifying the radioactivity transferred from the aqueous-soluble substrate to the organic-soluble product. The disaccharide product was dried down, and resuspended in 5%

DMSO and 0.05 % Triton X-100. 50 mM Bicine pH 8.5, 10 mM $MnCl_2$, 1.7 mM UDP-GalNAz, and 180 μM PglJ were added and the reaction was incubated with shaking at room temperature overnight. The reaction was dried down and purified using normal phase HPLC methods. The trisaccharide product (35% over 2 steps, 0.4 mg) was further characterized using high-resolution mass spectrometry

Synthesis of Und-PP-diNAcBac-GalNAc-GalNAz-$(GalNAc)_3$-Glc

Und-P, at a final concentration of 500 μM was resuspended in 2% Triton X-100 and 3% DMSO by vortexing. 2 mM $MgCl_2$, 30 mM Tris pH 8.0, 600 μM UDP- diNAcBac, 500 μM [^{3}H]-UDP-GalNAc (16.3 mCi/mmol), 30 μM PglA and 35% (v/v) PglC CEF were added to synthesize the disaccharide, Und-PP- diNAcBac-GalNAc. The reaction was incubated with shaking at room temperature overnight, and quenched in 3 mL 2:1 $CHCl_3$:MeOH, and washed three times with 2 mL PSUP. The reaction yield was determined by quantifying the radioactivity transferred from the aqueous-soluble substrate to the organic-soluble product. The disaccharide product was dried down, and resuspended in 5% DMSO and 0.05 % Triton X-100. 50 mM Bicine pH 8.5, 10 mM $MnCl_2$, 1.5 mM UDP-GalNAz, and 25 μM PglJ were added and the reaction was incubated with shaking at room temperature overnight. The reaction was dried down and purified using normal phase HPLC methods. The reaction yield was calculated by quantifying the amount of tritium associated with the elution peaks corresponding to the disaccharide starting material and trisaccharide product. The trisaccharide product was dried down, and resuspended in 5% DMSO and 0.05 % Triton X-100. 50 mM Bicine pH 8.5, 10 mM $MnCl_2$, 3 mM UDP-GalNAc, 500 μM UDP-Glucose, 35 μM PglH, 30% (v/v) PglI CEF were added and the reaction was incubated with shaking at room temperature overnight. The product were purified using a normal phase column (YMC-PVA-SIL-NP, 250 × 4.6 mm) and separated

over a 20-45% Buffer B gradient where Buffer A is 4:1 $CHCl_3$:MeOH and Buffer B is 10:10:3 $CHCl_3$:MeOH:2M NH_4HCO_3. The reaction yield (26%) was calculated using the tritium signal associated with the elution peak corresponding to the final heptasaccharide product. The product was characterized using RPLC-ESI/MS.

PglB Assays with Azide-Modified Trisaccharides

Azide-modified trisaccharides, at a final concentration of 50 μM were resuspended in 10% DMSO by vortexing. 10 mM $MnCl_2$, 50 mM HEPES, pH 7.5, 140 mM sucrose, 1.2 % Triton X-100, 250 μM of the peptide (FITC-DQNATRR, synthesized by Boston Open Labs), and 2 μM PglB were added. Aliquots (20 μL) were taken at defined time points and quenched in 1 mL 3:2:1 $CHCl_3$:MeOH: 4 mM $MgCl_2$, and washed three times with 2 mL PSUP. The resulting aqueous layers were combined with 5 mL EcoLite (MP Biomedicals) liquid scintillation cocktail. Organic layers were combined with 5 mL OptiFluor (PerkinElmer). Both layers were analyzed on a Beckman Coulter LS6500 scintillation counting system.

Synthesis of Glycopeptides

Und-PP-trisaccharides, at a final concentration of 50 μM were resuspended in 10% DMSO by vortexing. 10 mM $MnCl_2$, 50 mM HEPES, pH 7.5, 140 mM sucrose, 1.2 % Triton X-100, 250 μM of the peptide (FITC-DQNATRR, synthesized by Boston Open Labs), and 2 μM purified PglB were added. Reactions were incubated at room temperature with shaking for 3 hours, and quenched in 3 mL 3:2:1 $CHCl_3$:MeOH: 4 mM $MgCl_2$, and washed three times with 2 mL PSUP. The aqueous layer containing the glycopeptide was dried down and purified using a C18 reverse phase column (YMC-Pack ODS-A, 250 × 4.6 mm) and separated over 20-40% Buffer B gradient where Buffer A H_2O/0.1% TFA and buffer B is CH_3CN/0.1% TFA. The glycopeptide product was further characterized by high-resolution ESI-MS.

Reverse phase liquid chromatography-electrospray ionization mass spectrometry (RPLC-ESI/MS)

Reverse phase LC-ESI/MS was performed using a Shimadzu LC system (comprising a solvent degasser, two LC-10A pumps and a SCL-10A system controller) coupled to a high resolution TripleTOF5600 mass spectrometer (Applied Biosystems, Foster City, CA). LC was operated at a flow rate of 200 µl/min with a linear gradient as follows: 100% of mobile phase A was held isocratically for 2 min and then linearly increased to 100% mobile phase B over 14 min and held at 100% B for 4 min. Mobile phase A consisted of methanol/acetonitrile/aqueous 1 mM ammonium acetate (60/20/20, v/v/v). Mobile phase B consisted of 100% ethanol containing 1 mM ammonium acetate. A Zorbax SB-C8 reversed-phase column (5 µm, 2.1 x 50 mm) was obtained from Agilent (Palo Alto, CA). The MS settings were as follows: Ion spray voltage (IS) = -4500 V; Curtain gas (CUR) = 20 psi; Ion source gas 1 (GS1) = 20 psi; De-clustering potential (DP) = -55 V; Focusing potential (FP) = -150 V. Data acquisition and analysis were performed using the Analyst TF1.5 software (Applied Biosystem, Foster City, CA).

References

(1) Szymanski, C. M.; Yao, R.; Ewing, C. P.; Trust, T. J.; Guerry, P. Evidence for a System of General Protein Glycosylation in Campylobacter Jejuni. *Mol. Microbiol.* **1999**, *32* (5), 1022–1030.

(2) Karlyshev, A. V.; Linton, D.; Gregson, N. A.; Lastovica, A. J.; Wren, B. W. Genetic and Biochemical Evidence of a Campylobacter Jejuni Capsular Polysaccharide That Accounts for Penner Serotype Specificity. *Mol. Microbiol.* **2000**, *35* (3), 529–541.

(3) Linton, D.; Karlyshev, A. V.; Hitchen, P. G.; Morris, H. R.; Dell, A.; Gregson, N. A.; Wren, B. W. Multiple N-Acetyl Neuraminic Acid Synthetase (neuB) Genes in Campylobacter Jejuni: Identification and Characterization of the Gene Involved in Sialylation of Lipo-Oligosaccharide. *Mol. Microbiol.* **2000**, *35* (5), 1120–1134.

(4) Wood, A. C.; Oldfield, N. J.; O'Dwyer, C. A.; Ketley, J. M. Cloning, Mutation and Distribution of a Putative Lipopolysaccharide Biosynthesis Locus in Campylobacter Jejuni. *Microbiology (Reading, Engl.)* **1999**, *145 (Pt 2)*, 379–388.

(5) Bacon, D. J.; Szymanski, C. M.; Burr, D. H.; Silver, R. P.; Alm, R. A.; Guerry, P. A Phase-Variable Capsule Is Involved in Virulence of Campylobacter Jejuni 81-176. *Molecular Microbiology* **2001**, *40* (3), 769–777.

(6) Guerry, P.; Szymanski, C. M.; Prendergast, M. M.; Hickey, T. E.; Ewing, C. P.; Pattarini, D. L.; Moran, A. P. Phase Variation of Campylobacter Jejuni 81-176 Lipooligosaccharide Affects Ganglioside Mimicry and Invasiveness In Vitro. *Infect Immun* **2002**, *70* (2), 787–793.

(7) Guerry, P.; Ewing, C. P.; Schirm, M.; Lorenzo, M.; Kelly, J.; Pattarini, D.; Majam, G.; Thibault, P.; Logan, S. Changes in Flagellin Glycosylation Affect Campylobacter Autoagglutination and Virulence. *Molecular Microbiology* **2006**, *60* (2), 299–311.

(8) Thibault, P.; Logan, S. M.; Kelly, J. F.; Brisson, J. R.; Ewing, C. P.; Trust, T. J.; Guerry, P. Identification of the Carbohydrate Moieties and Glycosylation Motifs in Campylobacter Jejuni Flagellin. *J. Biol. Chem.* **2001**, *276* (37), 34862–34870.

(9) Szymanski, C. M.; Burr, D. H.; Guerry, P. Campylobacter Protein Glycosylation Affects Host Cell Interactions. *Infect Immun* **2002**, *70* (4), 2242–2244.

(10) Young, K. T.; Davis, L. M.; DiRita, V. J. Campylobacter Jejuni: Molecular Biology and Pathogenesis. *Nat Rev Micro* **2007**, *5* (9), 665–679.

(11) Schirm, M.; Soo, E. C.; Aubry, A. J.; Austin, J.; Thibault, P.; Logan, S. M. Structural, Genetic and Functional Characterization of the Flagellin Glycosylation Process in Helicobacter Pylori. *Molecular Microbiology* **2003**, *48* (6), 1579–1592.

(12) Josenhans, C.; Vossebein, L.; Friedrich, S.; Suerbaum, S. The neuA/flmD Gene Cluster of Helicobacter Pylori Is Involved in Flagellar Biosynthesis and Flagellin Glycosylation. *FEMS Microbiology Letters* **2002**, *210* (2), 165–172.

(13) Jennings, M. P.; Jen, F. E.-C.; Roddam, L. F.; Apicella, M. A.; Edwards, J. L. Neisseria Gonorrhoeae Pilin Glycan Contributes to CR3 Activation during Challenge of Primary Cervical Epithelial Cells. *Cellular Microbiology* **2011**, *13* (6), 885–896.

(14) Iwashkiw, J. A.; Seper, A.; Weber, B. S.; Scott, N. E.; Vinogradov, E.; Stratilo, C.; Reiz, B.; Cordwell, S. J.; Whittal, R.; Schild, S.; et al. Identification of a General O-Linked Protein Glycosylation System in Acinetobacter Baumannii and Its Role in Virulence and Biofilm Formation. *PLoS Pathog* **2012**, *8* (6), e1002758.

(15) Troutman, J. M.; Imperiali, B. Campylobacter Jejuni PglH Is a Single Active Site Processive Polymerase That Utilizes Product Inhibition to Limit Sequential Glycosyl Transfer Reactions. *Biochemistry* **2009**, *48* (12), 2807–2816.

(16) Kelly, J.; Jarrell, H.; Millar, L.; Tessier, L.; Fiori, L. M.; Lau, P. C.; Allan, B.; Szymanski, C. M. Biosynthesis of the N-Linked Glycan in Campylobacter Jejuni and Addition onto Protein through Block Transfer. *J. Bacteriol.* **2006**, *188* (7), 2427–2434.

(17) Karlyshev, A. V.; Everest, P.; Linton, D.; Cawthraw, S.; Newell, D. G.; Wren, B. W. The Campylobacter Jejuni General Glycosylation System Is Important for Attachment to Human Epithelial Cells and in the Colonization of Chicks. *Microbiology* **2004**, *150* (6), 1957–1964.

(18) Van Sorge, N. M.; Bleumink, N. M. C.; van Vliet, S. J.; Saeland, E.; van der Pol, W.-L.; van Kooyk, Y.; van Putten, J. P. M. N-Glycosylated Proteins and Distinct Lipooligosaccharide Glycoforms of Campylobacter Jejuni Target the Human C-Type Lectin Receptor MGL. *Cell. Microbiol.* **2009**, *11* (12), 1768–1781.

(19) Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. Copper-Free Click Chemistry for Dynamic in Vivo Imaging. *PNAS* **2007**, *104* (43), 16793–16797.

(20) Laughlin, S. T.; Baskin, J. M.; Amacher, S. L.; Bertozzi, C. R. In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science* **2008**, *320* (5876), 664–667.

(21) Chang, P. V.; Prescher, J. A.; Sletten, E. M.; Baskin, J. M.; Miller, I. A.; Agard, N. J.; Lo, A.; Bertozzi, C. R. Copper-Free Click Chemistry in Living Animals. *PNAS* **2010**, *107* (5), 1821–1826.

(22) Koenigs, M. B.; Richardson, E. A.; Dube, D. H. Metabolic Profiling of Helicobacter Pyloriglycosylation. *Mol. BioSyst.* **2009**, *5* (9), 909–912.

(23) Dumont, A.; Malleron, A.; Awwad, M.; Dukan, S.; Vauzeilles, B. Click-Mediated Labeling of Bacterial Membranes through Metabolic Modification of the Lipopolysaccharide Inner Core. *Angew. Chem. Int. Ed.* **2012**, *51* (13), 3143–3146.

(24) Liu, F.; Vijayakrishnan, B.; Faridmoayer, A.; Taylor, T. A.; Parsons, T. B.; Bernardes, G. J. L.; Kowarik, M.; Davis, B. G. Rationally Designed Short Polyisoprenol-Linked PglB Substrates for Engineered Polypeptide and Protein N-Glycosylation. *J. Am. Chem. Soc.* **2014**, *136* (2), 566–569.

(25) Wacker, M.; Feldman, M. F.; Callewaert, N.; Kowarik, M.; Clarke, B. R.; Pohl, N. L.; Hernandez, M.; Vines, E. D.; Valvano, M. A.; Whitfield, C.; et al. Substrate Specificity of Bacterial Oligosaccharyltransferase Suggests a Common Transfer Mechanism for the Bacterial and Eukaryotic Systems. *PNAS* **2006**, *103* (18), 7088–7093.

(26) Swiezewska, E.; Sasak, W.; Mańkowski, T.; Jankowski, W.; Vogtman, T.; Krajewska, I.; Hertel, J.; Skoczylas, E.; Chojnacki, T. The Search for Plant Polyprenols. *Acta Biochim. Pol.* **1994**, *41* (3), 221–260.

(27) Olivier, N. B.; Chen, M. M.; Behr, J. R.; Imperiali, B. In Vitro Biosynthesis of UDP-N,N'-Diacetylbacillosamine by Enzymes of the Campylobacter Jejuni General Protein Glycosylation System. *Biochemistry* **2006**, *45* (45), 13659–13669.

(28) Guan, W.; Cai, L.; Wang, P. G. Highly Efficient Synthesis of UDP-GalNAc/GlcNAc Analogues with Promiscuous Recombinant Human UDP-GalNAc Pyrophosphorylase AGX1. *Chem. Eur. J.* **2010**, *16* (45), 13343–13345.

(29) Antonczak, A. K.; Simova, Z.; Tippmann, E. M. A Critical Examination of Escherichia Coli Esterase Activity. *J Biol Chem* **2009**, *284* (42), 28795–28800.

(30) Fry, B. N.; Feng, S.; Chen, Y.-Y.; Newell, D. G.; Coloe, P. J.; Korolik, V. The galE Gene of Campylobacter Jejuni Is Involved in Lipopolysaccharide Synthesis and Virulence. *Infect Immun* **2000**, *68* (5), 2594–2601.

(31) Yu, S.-H.; Boyce, M.; Wands, A. M.; Bond, M. R.; Bertozzi, C. R.; Kohler, J. J. Metabolic Labeling Enables Selective Photocrosslinking of O-GlcNAc-Modified Proteins to Their Binding Partners. *PNAS* **2012**, *109* (13), 4834–4839.

(32) Neef, A. B.; Schultz, C. Selective Fluorescence Labeling of Lipids in Living Cells. *Angewandte Chemie International Edition* **2009**, *48* (8), 1498–1500.

(33) Whitworth, G. E.; Imperiali, B. Selective Biochemical Labeling of Campylobacter Jejuni Cell-Surface Glycoconjugates. *Glycobiology* **2015**, cwv016.

(34) Glover, K. J.; Weerapana, E.; Imperiali, B. In Vitro Assembly of the Undecaprenylpyrophosphate-Linked Heptasaccharide for Prokaryotic N-Linked Glycosylation. *Proc Natl Acad Sci U S A* **2005**, *102* (40), 14255–14259.

(35) Amin, M. N.; Ishiwata, A.; Ito, Y. Synthesis of N-Linked Glycan Derived from Gram-Negative Bacterium, Campylobacter Jejuni. *Tetrahedron* **2007**, *63* (34), 8181–8198.

(36) Cummings, R. D.; Pierce, J. M. The Challenge and Promise of Glycomics. *Chemistry & Biology* **2014**, *21* (1), 1–15.

(37) Kumar, M.; Balaji, P. V. Comparative Genomics Analysis of Completely Sequenced Microbial Genomes Reveals the Ubiquity of N-Linked Glycosylation in Prokaryotes. *Mol. BioSyst.* **2011**, *7* (5), 1629–1645.

(38) Iwashkiw, J. A.; Vozza, N. F.; Kinsella, R. L.; Feldman, M. F. Pour Some Sugar on It: The Expanding World of Bacterial Protein O-Linked Glycosylation. *Mol. Microbiol.* **2013**, *89* (1), 14–28.

(39) Watanabe, Y.; Nakamura, T.; Mitsumoto, H. Protection of Phosphate with the 9-Fluorenylmethyl Group. Synthesis of Unsaturated-Acyl Phosphatidylinositol 4,5-Bisphosphate. *Tetrahedron Letters* **1997**, *38* (42), 7407–7410.

(40) Branch, C. L.; Burton, G.; Moss, S. F. An Expedient Synthesis of Allylic Polyprenyl Phosphates. *Synthetic Communications* **1999**, *29* (15), 2639–2644.

(41) Jaffee, M. B.; Imperiali, B. Optimized Protocol for Expression and Purification of Membrane-Bound PglB, a Bacterial Oligosaccharyl Transferase. *Protein Expr Purif* **2013**, *89* (2), 241–250.

(42) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr. Purif.* **2005**, *41* (1), 207–234.

Appendix

Table A-1 Attempts towards the purification of PglC for X-ray crystallography (** indicates experiments performed by Dr. Meredith Hartley)

Date	Detergent (solubilization %, elution %)	[Total protein] during solubilization	Yield	Comments
PglC-pET24a**				
4-28-08	Triton X-100 (1%, 0.5%)	1-2 mg/mL	0.1 mg/L	PglC in low speed spin pellet
5-5-08	Detergent screen: CHAPS (2%) DDM (0.2%) NP-40 (0.5%) Empigen (0.4%)	1-2 mg/mL	-	20-40% solubilization of CEF with DDM, NP-40 and Empigen Little solubilization with CHAPS
5-8-08	NP-40 (0.5%, 0.5%)	1-2 mg/mL	0.1 mg/L	Poor solubilization and poor binding to Ni-NTA
5-12-08	Empigen (0.5%, 0.5%)	3-5 mg/mL	0.1 mg/L	Poor solubilization and poor binding to Ni-NTA
PglC-pET-NO**				
5-30-08	NP-40 (0.5%, 0.5%)	1-3 mg/mL	Very little	Poor solubilization and binding
6-25-08	Triton X-100 Screened detergent concentrations at 0.2%, 0.4%, 0.6%, 0.8%, and 1.0%	0.6-1 mg/mL	-	All Triton X-100 concentrations showed poor solubilization
PglC-pGEX**				
6-30-08	Triton X-100 (1%, 0.5%)	3-5 mg/mL	Very little	Poor solubilization of CEF
PglC-pGBH-MH**				
9-19-08	Detergent screen Triton X-100 (1%) NP-40 (1%) Empigen (1%) DDM (1%)	0.8-1.3 mg/mL	-	Triton X-100 is the best at solubilization
9-22-08	Trition X-100 (1%, 0.1%)	3-5 mg/mL	0.3 mg/L	Solubilization and binding to Ni-NTA is okay, but yield after lysis is poor.
9-30-08	Trition X-100 (1%, 0.1%)	3-5 mg/mL	1.5 mg/L	PglC is lysed with 1 mg/mL lysozyme
10-31-08	Triton X-100 (1%, 0.1%)	3-5 mg/mL	1.5 mg/L	Works well
3-30-09	DDM (1%, 0.03%)	1-2 mg/mL	0.3 mg/L	Lysing was poor – malfuntioning sonicator

4-30-09	DDM (1%, 0.03%)	1-2 mg/mL	0.1 mg/L	" "
5-7-09	DDM (1%, 0.02%)	1-2 mg/mL	1 mg/L	None
5-25-09	LDAO (1%, 0.03%)	1-2 mg/mL	Very low	??
6-9-09	LDAO (2.5%, 0.03%)	1-2 mg/mL	Very low	Binding to Ni-NTA
6-16-09	CHAPSO (2.5%, 1%)	3-5 mg/mL	1.5 mg/L	50% monodispersed
7-8-09	CHAPSO (2.5%, 1%)	3-5 mg/mL	1.5 mg/L	All aggregate in SEC
7-28-09	CHAPSO (1%, 1%)	3-5 mg/mL	1.5 mg/L	Extra high salt wash
8-27-09	CHAPSO (1%, 1%)	7-9 mg/mL	1.5 mg/L	70% monodispersed by SEC (however, <5% of original yield was recovered after SEC)
9-14-09	CHAPSO (1%, 1%)	3-5 mg/mL	1.5 mg/L	Did not pass through 100kDa filter
9-23-09	LDAO (0.05%, 0.05%)	3-5 mg/mL	Very low	Does not bind Ni-NTA resin
10-1-09	β-OG (1.1%, 1.1%)	3-5 mg/mL	--	β-OG similar to CHAPSO
PglC-pGBH-MH-FLAG**				
11-4-09	LDAO (0.05%, 0.05%)	3-5 mg/mL	Low	Poor binding to resin
PglC-pGBH-MH**				
12-14-09	LDAO (0.05%, 0.05%)	3-5 mg/mL	Low	Used TALON resin – poor binding
1-5-09	DDM (0.03%), LDAO (0.05%)	3-5 mg/mL	0.1 mg/L	Poor solubilization
1-11-09	DDM (0.03%), LDAO (0.05%)	3-5 mg/mL	Low	Used French press to lyse
1-11-09	DDM (0.03%), LDAO (0.05%)	3-5 mg/mL	Low	Solubilized with overnight incubation
1-26-10	DDM (1%, 0.03%)	3-5 mg/mL	1.5 mg/L	Higher detergent percentage improved solubilization
2-15-10	DDM (1%) β-OG (1.4%)	3-5 mg/mL	1.5 mg/L	With and without PglD, PglC precipitated after elution

sPglC-pET-NO (TM helix ends at Ile-35, made $sPglC_{36}$, $sPglC_{39}$ and $sPglC_{42}$ constructs)

4-28-11	Inclusion body prep with 6 M urea, no detergent	-	2 mg/L	Aggregated when analyzed by size exclusion
PglC-pET24a inclusion body prep				
4-17-12	Triton X-100 (1%) during solublization with 6M Urea	1-2 mg/mL	< 1 mg/L	Refolded on Ni resin. Protein was not active.
PglC-pET-NO inclusion body prep				
4-22-12	Triton X-100 (1%) during solubilization with 6M urea	1-2 mg/mL	< 1 mg/L	Refolded by dialysis and on Ni resin. Protein was not active (Ni resin) and aggregated (dialysis)
SUMO-sPglC (TM helix ends at Ile-35, made $sPglC_{36}$, $sPglC_{39}$ and $sPglC_{42}$ constructs)				
6-18-12	-	-	~1 mg/L	Protein present in soluble fraction, SUMO cleavage not efficient, sPglC is inactive
SUMO-PglC (*C. jejuni*), grown in LB Media				
6-11-12	Triton X-100 (1%), reduced Triton (0.05%)	2-3 mg/mL	1 mg/L	SUMO protease cleavage not efficient
7-1-12	Triton X-100 (1%), DDM (0.03%)	2-3 mg/mL	1 mg/L	Protein mostly aggregated, shows a small monodisperse peak by SEC analysis
9-26-12	Lauryl Maltose Neopentyl Glycol (LMNG) (1%), LMNG (0.003%)	2-3 mg/mL	< 1 mg/L	Protein aggregated, elutes in void volume when analyzed by SEC
10-3-12	Octyl glucose neopentyl glycol (OGNG) (66 mM), OGNG (2.55 mM)	2-3 mg/mL	< 1 mg/L	OGNG not efficient at solubilization of protein out of CEF
PglC-pET-NO (*C. concisus*)				
10-26-12	Triton X-100 (1%), DDM (0.03%)	1-2 mg/mL	<1 mg/L	Protein elutes in void volume when analyzed by SEC
10-26-12	DDM (1%), DDM (0.03%)	1-2 mg/mL	<1 mg/L	Protein mostly elutes in void volume when analyzed by SEC, small shoulder visible at elution time for protein
PglC-pET-NO (*C. lari*)				

11-2-12	Triton X-100 (1%), DDM (0.03%)	1-2 mg/mL	<1 mg/L	Protein elutes in void volume when analyzed by SEC
11-2-12	DDM (1%), DDM (0.03%)	1-2 mg/mL	<1 mg/L	Protein mostly elutes in void volume when analyzed by SEC, small shoulder visible at elution time for protein
SUMO-PglC (*C. lari, C.concisus*), grown in LB media				
12-21-12	DDM (1%), DDM (0.03%)	2-3 mg/mL	~1 mg/L	Protein not aggregated when analyzed by SEC
1-8-13	LMNG (1%), LMNG (0.003%)	2-3 mg/mL	~1 mg/L	Protein not aggregated when analyzed by SEC but many peaks visible on SEC trace
sPglC-pET-NO (*C. lari, C.concisus*)				
5-13-13	Protein present in inclusion bodies	-	2-3 mg/L (*C. concisus*) 1-2 mg/L (*C. lari*)	Inclusion bodies analyzed by Drew Lynch (Allen Lab)
SUMO-PglC (*C. jejuni, C. lari, C.concisus*), overexpressed by autoinduction				
12-3-13	DDM (1%), DDM (0.03%)	3-5 mg/mL	6-8 mg/L	