ELUCIDATING THE MOLECULAR BASIS OF INHIBITION OF ENZYMES ESSENTIAL FOR BATERIALVIRULENCE: THE ACTIVITY OF THREE GLYCOSYLTRANSFERASES AND A POTENTIAL DRUG TARGET, UNDECAPRENYL PYROPHOSPHATE SYNTHASE

by

Elleansar Naa Norley Okwei

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Chemistry

Charlotte

2016

Approved by:

_____
Dr. Jerry Troutman

_____
Dr. Donald Jacobs

_____
Dr. Juan Vivero-Escoto

_____
Dr. Kirill Afonin

ABSTRACT

ELLEANSAR NAA NORLEY OKWEI. Elucidating the molecular basis of inhibition of enzymes essential to bacterial virulence: the activity of three glycosyltransferases and a potential drug target, UPPS. (Under the direction of DR. JERRY TROUTMAN)

The fight against bacterial infection is one of the most important branches of modern medicine. Due to the emergence of strains of antibiotic-resistant bacteria, researchers have been presented with complex challenges in an effort to combat these stronger bacteria that pose a significant threat to our health. Bacteria possess certain distinct features that protect them from the outside environment, such as their cell wall, and, in the case of Gram-negative bacteria, the outer membrane layer of their cell envelope. The cell wall is absolutely essential for bacterial survival, such that removal or an alteration in any of its components, results in the death of bacteria. Thus, many research projects aim to tackle pathways leading to formation of the polysaccharide cell wall. In this work, we study the activity and inhibition of four proteins which contribute to bacterial virulence. The first two proteins, LgmA and LgmC, are thought to be part of the pathway that modifies Lipid A, a component of the outer membrane, in *Bordetella pertussis*. Both enzymes are studied alongside WecA, a well-known, already characterized hexosyltransferase from *E. coli*, to explore differences in activity and inhibition of different classes of glycosyltransferases. Lastly, we conduct structural and computational studies of undecaprenyl pyrophosphate synthase (UPPS), a protein that is indispensable for cell wall biosynthesis. Our results provide some insight into the functionality and inhibition of each enzyme, and consequently, may contribute to the improvement or design of novel inhibitors of bacterial enzymatic activity.

DEDICATION

To my amazing parents, Moses Nii Narku Okwei and Georgina Abla Okwei, to whom I owe everything. Their relentless hard work, and their support, encouragement, and unshakable love for their children is beyond my comprehension.

To my favourite brother and absolute partner in crime, Aaron Nii Nortey Okwei, for being my best friend.

To my favourite sisters and best friends, Sheba Naa Norkor, Miriam Naa Norkai, and Adelaide Agnes Korkor Okwei, to whom acting like a second mother has brought me some of the greatest joys in life, and, of course, some of the greatest pains. They make me a proud big sister.

And to the rest of my incredible family and friends who have been with me on my journey, I am very grateful.

ACKNOWLEDGEMENTS

I would like to thank the people whose contributions were crucial to the completion of this work. Dr. Fernandez for providing *Bordetella pertussis* cells. Neil Price for providing inhibitors, including the tunicamycin-derived inhibitors synthesized by the Biao Yu lab. To the members of the Troutman lab, including Joseph Hazel for cloning lgmA gene and wecA gene into DH5α cells, Katelyn Erickson for performing mass spectrometry on LgmA product, and Sunita Sharma for tips on working with membrane proteins. To the entire Biomolecular Physics Group, especially Brittany Smith and Chris Singer for their help with the DCM, and Fareeha Kanwal for helping generate DCM data. Huge thanks to the entire Burson staff and faculty; to Dr Thomas Walsh for the tuition assistance scholarship, and the NIH for research funding. I am very grateful to Dr. Dennis Livesay for his constructive criticism and helping me better understand the computational world from a chemist's perspective. And finally, to my committee members: Dr. Juan Vivero-Escoto and Dr. Kirill Afonin, who have generously given their time and advice; to Dr. Donald Jacobs for his patient tutelage, and for helping me navigate the world of computational biophysics; and to my advisor, Dr. Jerry Troutman, for his incredible insight and constant guidance throughout my projects.

TABLE OF CONTENTS

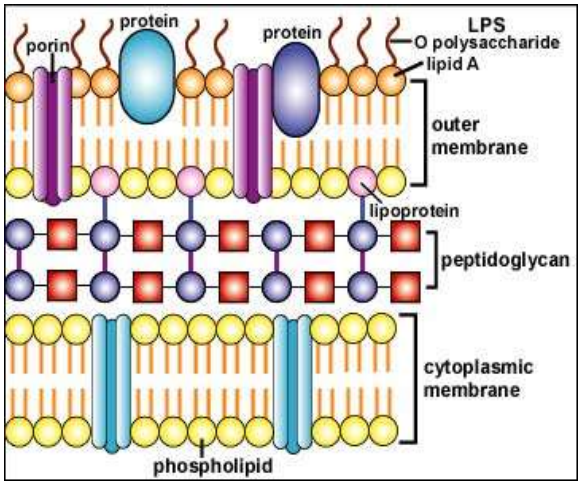
LIST OF ABBREVIATIONS

2CN-BP	2-cyano-bactoprenylphosphate
BLAST	Basic Local Alignment Search Tool
BP	Bactoprenyl phosphate
CC	Cooperativity- Correlation
DCM	Distance Constraint Model
GlcNAc	N-acetyl-glucosamine
GlcNAc-1-P	N-acetyl-glucosamine-1-phosphate
GT	Glycosyltransferase
FLU	Fluorescence
FPP	Farnesyl diphosphate
HPLC	High performance liquid chromatography
IPP	Isopentenyl diphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LGMA/B/C	Lipid A-Glucosamine Modification Protein A, B, or C
MES	2-(N-morpholino)ethanesulfonic acid
PSI-BLAST	Position-Specific Iterative Basic Local Alignment Search Tool
QSFR	Quantitative Stability/Flexibility Relationships
SDS-PAGE	Sodium dodecyl sulfate – Polyacrylamide gel electrophoresis
TRIS	Tris(hydroxymethyl)aminomethane
UDP	Uridine diphosphate
UP	Undecaprenylphosphate
UPPS	Undecaprenyl diphosphate synthase

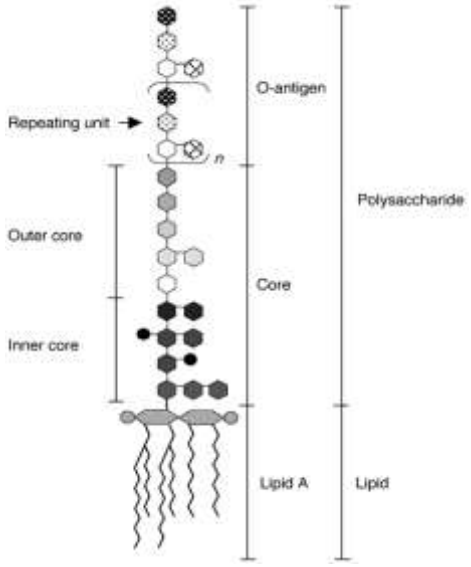
CHAPTER 1: INTRODUCTION

1.1 Lipopolysaccharides

The envelope of pathogenic Gram-negative bacteria, such as *Bordetella pertussis, Escherichia coli, Salmonella enterica*, and *Pseudomonas aeruginosa*, is composed of two cell membranes, a glycerophospholipid-based inner membrane, and a very distinctive outer membrane that consists of lipopolysaccharides (LPS) (Figure 1). Because they are crucial for physiological membrane functions and essential for bacterial growth and viability, they are the primary target of components of the host's immune response [3]. Endotoxin, to be distinguished from exotoxin (which is secreted by both Gram-positive and Gram-negative bacteria), is a composition of the bacterial body and needs not be secreted; as a result, LPS of dead bacteria provide the same amount of toxicity as live bacteria. LPS derived from different groups of Gram-negative bacteria have a common basic structure comprising of a hydrophobic region, lipid A, which anchors the molecule to the membrane and is also responsible for the toxic properties of the molecule, a hydrophilic core polysaccharide chain, and a repeating hydrophilic oligosaccharide side chain (the O-antigen) that is specific to the bacterial serotype [4].



Structure and Composition of the Gram-Negative Cell Wall. Gary Kaiser. [1, 2]



Ogawa et al. 2007

Figure 1: LPS is the major component of the outer cell membrane of Gram-negative bacteria. Left: Overall structure of the gram-negative bacterial cell membrane (Image by: Gary Kaiser [1]). Right: The general chemical structure of LPS. A hydrophobic lipid section, lipid A; a hydrophilic core polysaccharide chain; and a hydrophilic O-antigenic oligosaccharide. (Ogawa et al. 2007) [2].

1.2 Lipid A and Its Structure

Lipid A anchors the LPS within the membrane while the polysaccharide interacts with the external environment including the defenses of the animal or plant host species. It has been noted that the existence of lipid A-containing lipopolysaccharide in the most ancient and primitive Gram-negative bacteria demonstrates that it is absolutely required for their survival, shielding them from a variety of aggressive conditions [5].

Lipid A is a unique and distinctive phosphoglycolipid, the structure of which is highly conserved among species. The basic structure contains D-gluco-configured pyranosidic hexosamine residues, present as β(1→6)-linked dimers. The disaccharide phosphoryl groups in the 1 and 4’ positions, and (R)-3-hydroxy fatty acids at positions *O*-2, *O*-3, *O*-2' and *O*-3' in ester and amide linkages, of which two are usually further

acylated at their 3-hydroxyl group. Variations in the structure arise from the degree of phosphorylation, the presence of phosphate substituents, and most importantly in the nature, chain length, number, and position of the acyl groups [6]. In the lipid A of the most studied organism *Escherichia coli* (Figure **2**), the hydroxy fatty acids are C_{14} in chain length, and the hydroxy groups of the two (R)-3-hydroxy fatty acids of the distal GlcN-residue are acylated by non-hydroxy fatty acids (12:0 and 14:0)[5-7].

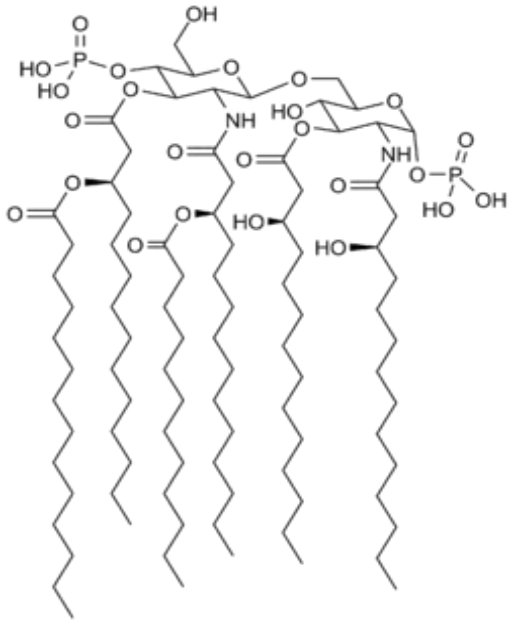


Figure 2: Chemical structure of Lipid A (*E. coli*).

1.3 Modification of Lipid A and Its Function.

Determination of the exact chemical structures of lipid A from various Gram-negative bacteria shows that the molecule can be further modified in response to environmental stimuli [6, 7]. These modifications have been implicated in virulence of pathogenic Gram-negative bacteria and are representative of one of the microbial surface remodeling mechanisms employed by bacteria to help evade the innate immune response [8].

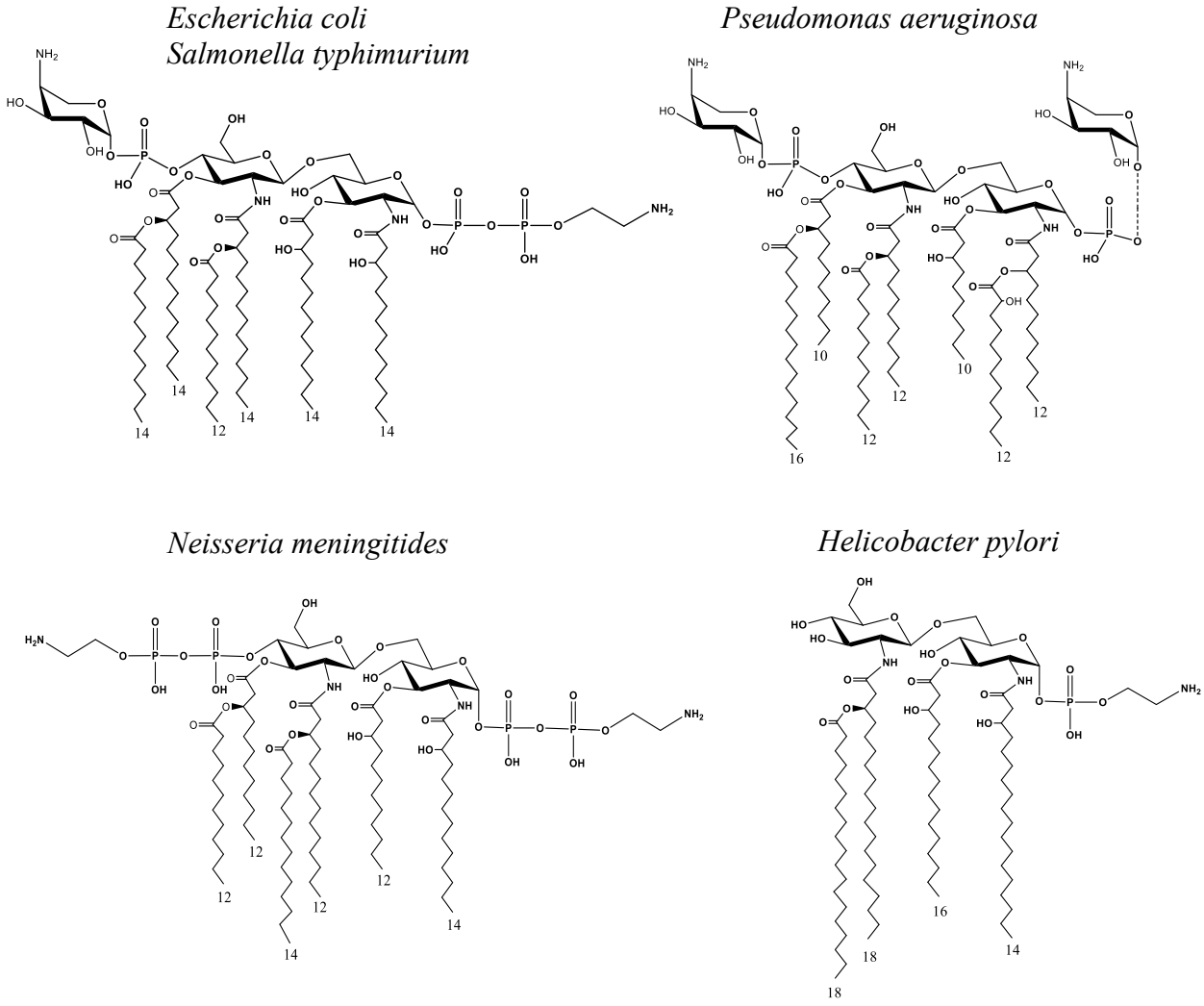


Figure 3: Lipid A structures of various pathogenic bacteria. In E. coli, *S. typhimurium, N. meningitidis*, and *P. aeruginosa*, each PO_4 group is substituted with L-4-aminoarabinose or a phospho-ethanolamine. In *H. pylori*, the 1' PO_4 is substituted with a phospo-ethanolamine while the 4' PO_4 is removed altogether.

As aforementioned, the variable length of the fatty acyl chains of lipid A dictates the extent of toxicity, with an increased number of chains resulting in higher toxicity. Similarly, the 1- and 4'-phosphates attached to the glucosamine backbone can be modified and/or removed. Gunn et al. (1998) found that a number of Gram-negative bacteria contain latent enzymes capable of modifying the lipid A phosphates [9]. Modified lipid A structures of some pathogenic bacteria are shown [9-12] (Figure **3**).

Substitution of lipid A with L-Ara4N and phosphoethanolamine residues are greatly elevated in polymyxin-resistant mutants of *S. typhimurium* and *E. coli* K-12 [13-16]. Although the function of these modifications are unknown, the current hypothesis is that masking of the lipid A phosphates with the cationic sugar reduces its net negative charge, thereby decreasing its vulnerability to cationic antimicrobial peptides (CAMPs).

1.4 Modification of Lipid A with Glucosamine in *B. pertussis*.

1.4.1 Proposed Pathway of Lipid A – Glucosamine Modification

The Fernandez lab (University of British Columbia, Canada) has shown that the modification of lipid A with glucosamine in *B. pertussis* results in increased resistance of the bacterium to numerous cationic antimicrobial peptides (CAMPs), as well as increased resistance to outer membrane perturbation [17, 18]. The focus of this project is to characterize the enzymes implicated in the pathway by which Lipid A from *B. pertussis* is proposed to be modified with glucosamine (GlcN).

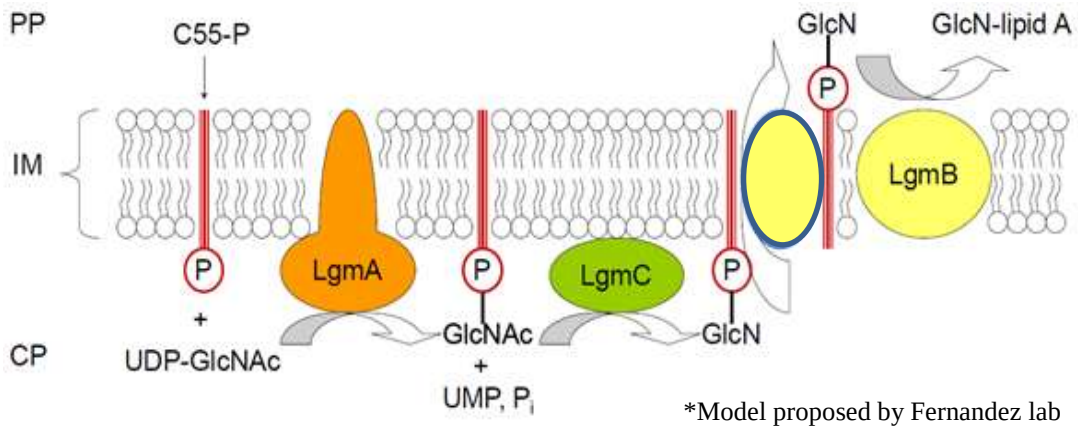


Figure 4: Lipid A modification pathway proposed. Putative functions have been assigned to each enzyme by the Fernandez lab (University of British Columbia). LgmA: glycosyl transferase, LgmC: deacetylase, LgmB: glycosyl transferase.

1.4.2 Identifying genes responsible for the production of the Lipid A-GlcN modification proteins

The genes encoding the proteins thought to be functional in the *B. pertussis* Lgm pathway have been identified, and putative functions assigned based on bioinformatics analyses and sequence comparison to homologs. LgmA and LgmB are reported homologs of ArnC and ArnT, respectively; ArnC and ArnT are characterized glycosyl transferases found in the pathway that modifies the phosphate of lipid A with aminoarabinose in species like *E. coli*, and *Salmonella typhimurium.* LgmC has predicted structural similarity to NaxD, a characterized enzyme shown to be important for lipid A modification in *Francisella tularensis,* and a member of the YdjC-like superfamily of deacetylation proteins [19, 20]. The final proposed pathway is shown (Figure 4). LgmA is proposed to act as a glycosyl transferase transferring GlcNAc to C55-P, LgmC then deacetylates the GlcNAc moiety leaving behind GlcN, which gets transferred to Lipid A by the other glycosyl transferase, LgmB.

1.5 Glycosyltransferases

1.5.1 Glycosyltransferase Reactions

Glycosyltransferase reactions are immensely important in many biological systems. The biosynthesis of disaccharides, oligosaccharides, polysaccharides, and glycoconjugates are governed by the action of a myriad of glycosyltransferase (GT) enzymes that catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules such as another sugar, a lipid, protein, or antibiotic. The few GTs which have resolved crystal structures are grouped into families based on fold type, but the vast majority of glycosyltransferases are generally classified based on their amino acid sequence similarity;

this makes precise functional predictions often unreliable or inaccurate. Although the number of distinct families based on sequence alone is vast, the same three-dimensional fold is expected to occur within each of the families. Previously, the GT families that received the most attention were those that play major roles in recognition or signaling events such as those adding terminal sugars in glycoconjugates, however, further study of GTs has exposed their roles in areas including mammalian development, bacterial toxicity, and human disease. Their essential roles in bacterial systems have also identified them as potential antibiotic and therapeutic targets.

The acceptor specificity exhibited by a particular glycosyltransferase is central to its functional role. Even though poly-specificity is common among GT families, it is known that GTs have a donor (usually a uridine diphosphate (UDP)-linked sugar), an acceptor (most commonly C_{55}–P in bacteria, or dolichol-P in eukaryotes), and linkage specificity; unfortunately, not much is known about the catalytic mechanism of these enzymes or the molecular basis that account for donor and acceptor specificity [21]. It has been observed that perhaps the greatest challenge in the field is functional characterization of GTs. As of early 2008, there have been over 33,000 open reading frames that encode this class of enzymes, yet the donor and acceptor specificity for the vast majority, which is over 95%, is not known [22]. cDNAs and genes encoding GTs have been cloned from several sources, but some of the major hurdles to obtaining functional characterization include expression and purification of these enzymes, and most importantly, efficient assay methods.

A significant hurdle to characterizing glycosyltransferases has been in developing an efficient method allowing quantifiable characterization. Progress has been made in these areas with the successful characterization of important glycosyltransferases such as MraY

[23], MurG [24], and WecA [25], using a radiolabeled technique. More recently, a FRET-based microplate reader assay was developed to detect bacterial membrane-bound enzyme MraY activity by BMG LABTECH.

We present a novel fluorescence-detecting high performance liquid chromatography (HPLC) assay to monitor the activity and inhibition of two glycosyltransferase enzymes. A fluorescent moiety, 2-cyanoaniline (2-CNA), is conjugated to a lipid substrate, undecaprenyl phosphate (C_{55}-P) synthesized in a reaction catalyzed by the enzyme, undecaprenyl pyrophosphate synthase (UPPS). The progress of the glycosyltransferase reaction is monitored by separation and detection at an excitation wavelength of 340 nm, and emission at 390 nm [26].

Here, we report the functional characterization of a phosphoglycosyltransferase, WecA, and a glycosyltransferase, LgmA, as well as explore the activity of a putative deacetylase, LgmC. The reason for using WecA is two-fold: WecA represents the superfamily under which most initiating hexosyltransferases fall, and also serves as a test case to demonstrate the efficiency and reliability of our experimental assay. LgmA, on the other hand, is a novel enzyme that belongs to a more unique family of GTs that is yet to be successfully characterized. Similarly, there are few other characterized N-acetylhexosamine deacetylases, making LgmC of particular interest. For the first time here, we report successful expression of the lipid A modification proteins, LgmA and LgmC, and implement the fluorescence-detecting HPLC technique to characterize this enzyme. To further understand the activity of GTs of different classes, we perform additional inhibition studies to investigate the differences in susceptibility to inhibitors.

1.5.2 WecA

WecA is an already characterized integral membrane protein, belonging to a family of polyisoprenyl phosphate *N*-acetylhexosamine-1-phosphate transferases. It is required for the biosynthesis of O-specific lipopolysaccharide and enterobacterial common antigen in *Escherichia coli* and other enteric bacteria [27]. It is one of a number of GTs thought to be very important in biological systems, such as MraY, WbcO, TagP, and WbpL which have been implicated in the biosynthesis of different bacterial cell envelope polymers (peptidoglycan, teichoic acids, or rhamnose-glucose polysaccharide) [28], as well as the eukaryotic GPT that transfers GlcNAc-1-P to dolichyl phosphate initiating the N-linked glycoprotein biosynthesis [29]. WecA transfers *N*-acetylglucosamine-1-phosphate (GlcNac-1-P) to C_{55}-P, to yield C_{55}-P-P-GlcNAc, the lipid intermediate involved in synthesis of various components of the bacterial cell wall. Although structural and functional characterization of proteins belonging to this superfamily have been greatly held back by problems encountered with their overexpression and purification, WecA is one of the few that have been successfully characterized [25], and therefore stands as a good representative of GlcNac-1-P transferase activity. Its membrane topology consists of 11 transmembrane segments, five cytoplasmic domains and five periplasmic domains [30]. Members of the HexNAc (N-acetyl-hexosaminyl) transferase family require the bound lipid carrier, C55-P, as an acceptor substrate, as well as a UDP-N-acetyl-hexosamine (UDP-linked sugars), but they tend to differ in their UDP-sugar substrate specificity, and also differ by their vulnerability to glycosyltransferase inhibitors.

1.5.3 LgmA

LgmA has been implicated in the Lipid A modification pathway of *Bordetella pertussis*, the Gram-negative bacterium that causes whooping cough. This modification in *B. pertussis* has been shown to result in increased resistance of the bacterium to numerous cationic antimicrobial peptides, as well as increased resistance to outer membrane perturbation [17, 19].

LgmA is thought to be the first protein in the pathway that transfers GlcNAc to C_{55}-P, upon which the GlcNAc moiety is deacetylated before the remaining glucosamine (GlcN) is transferred to Lipid A to complete the modification. This protein is thus of interest because it is a member of the Lipid A modification proteins which have been linked to virulence of pathogenic Gram-negative bacteria and are representative of one of the microbial surface remodeling mechanisms employed by bacteria to help evade the innate immune response [7, 8]. Even more interesting, unlike most other characterized initiating hexose-phosphate glycosyltransferases (such as WecA which transfers GlcNAc-1-P from UDP-GlcNAc), LgmA is thought to transfer only GlcNAc to yield C_{55}-P-GlcNAc.

Although both LgmA and WecA are a members of the large, diverse glycosyltransferase 2 family (which transfer sugar from UDP-glucose, UDP-N-aectylgalactosamine, guanosine diphosphate (GDP)-mannose, or cytidine diphosphate (CDP)-abequose, to a range of substrates including cellulose, dolichol phosphate, and teichoic acids), LgmA is found to have higher sequence similarity to the bacterial dolichol-phosphate mannose I (DPM1) synthase. DPM1 is the catalytic subunit of eukaryotic DPM synthase which is required for synthesis of the glycosylphosphatidylinositol (GPI) anchor, N-glycan precursor, protein O-mannose, and C-mannose. The enzyme has three subunits

DPM1, DPM2, and DPM3. Many bacterial DPM1-like enzymes have been identified based on sequence similarity to eukaryotic DPM1; although the mechanism of the eukaryotic enzyme is well studied, the mechanism of the bacterial enzymes is not well understood [31-33].

Moreover, there seems to be only one characterized glycosyltransferase in which the lipid undecaprenyl phosphate is an acceptor. MurG is a key enzyme at the borderline between the two stages of peptidoglycan synthesis and can be considered a potential target of novel antibacterials. MurG N-acetylglucosaminyl transferase from *E. coli* catalyzes the addition of GlcNAc to lipid I, yielding GlcNAc-MurNAc-(pentapeptide)-pyrophosphoryl undecaprenol (lipid II), following a reaction characterized by a transferase, in which phospho-N-acetyl-muramoyl-pentapeptide from the cytoplasmic precursor UDP-MurNAc-pentapeptide is transferred to the membrane acceptor undecaprenyl phosphate to yield lipid I [24]. Thus, LgmA activity is not only of interest because it is a putative glycosyltransferase, but also because it is a putative UDP-GlcNAc : undecaprenylphosphate glycosyltransferase.

1.5.4 LgmC

LgmC is predicted to be a deacetylase, and thought to act in the second step of the Lipid A modification pathway, in which C_{55}-P-GlcNAc is deacetylated to yield C55-P-GlcN. A search for its conserved domains in the National Center for Biotechnology Information (NCBI), shows LgmC to be related to the carbohydrate esterase 4 (CE4) superfamily, as well as the predicted glycoside hydrolase or deacteylase ChbG from the uncharacterized YdjC-like family proteins from bacteria. The CE4 superfamily mainly includes chitin deacetylases, bacterial peptidoglycan N-acetylglucosamine deacetylases, and acetylxylan esterases, and is remotely related to the YdjC-like family of proteins. The

YdjC-like subfamily contains many hypothetical proteins, and is represented by the uncharacterized protein YdjC (also known as CbhG) from *E. coli*. The molecular function of this family is said to be unclear [34, 35]. Fortunately, a characterized deacetylase has been recently reported. *naxD*, initially reported to encode a hypothetical protein was revealed as a member of the YdjC superfamily based on sequence similarity. NaxD has since been shown to be necessary for deacetylation of undecaprenyl phosphate-GalNAc in the lipid A modification (with galactosamine) pathway in *Francisella novicida* [20]. As NaxD is one of very few characterized deacetylases (and probably the only characterized deacetylase from a Lipid A modification pathway), it piques higher interest into the activity of LgmC. LgmC, in this work, is characterized alongside LgmA and WecA, to better understand the activity of this putative deacetylase.

1.6 Unexpected Revelations in the Proposed Pathway

In attempting to confirm the assigned function of the proteins, we were presented with some unexpected results from the preliminary data. We were able to confirm activity in LgmA, the first enzyme proposed to act in the pathway. The second enzyme, LgmC, did not show activity as a deacetylase under our reaction conditions; in fact, it behaved like a glycosyltransferase. Not only did this complicate the proposed pathway, but it began to raise questions about the actual role of each enzyme, especially LgmB, that could probably not be addressed immediately. As such, we became very interested in the activity of LgmC as a possible novel glycosyltransferase. This presented an opportunity to study a range of glycosyltransferases, the enzymes that have been shown to be very necessary for bacteria survival. Our interested shifted to the activity of

the bacterial glycosyltransferases responsible for bacterial virulence, either by way of bacterial cell wall formation or outer membrane modification.

In this study, we present results on activity, substrate specificity, metal-dependence, surfactant effect, and finally, inhibition susceptibility of WecA and LgmA. In understanding the differences in activity and inhibition of GTs of different classes, we get more insight into their molecular basis of activity. We are then able to contribute to the growing array of GT inhibitors to further the design and development of antibacterial agents, and add to our understanding of the activity of these enzymes that have proven very important in many biological systems.

Additionally, we probe further into the activity of LgmC, which we hypothesize is wrongly predicted as a deacteylase. We present data that indicates LgmC may be acting in the capacity of a glycosyltransferase; in fact, we propose it may be acting similar to WecA, as a GlcNAc-1-P transferase. Although it has little structural similarity to characterized glycosyltransferases, we explore its activity alongside LgmA and WecA, as a novel phosphoglycosyltransferase from *B. pertussis*.

1.7 Undecaprenyl Pyrophosphate Synthase (UPPS)

1.7.1 UPPS Is Essential For Bacterial Cell Wall Biosynthesis

Undecaprenyl pyrophosphate (C_{55}-P) is required for transferring complex polysaccharides as demonstrated in the lipid A modification pathway. It is essential for bacterial cell wall synthesis, without which bacteria will have no protection from the external environment. It acts as the lipid carrier found in only bacterial polysaccharide biosynthesis pathways, making it an attractive antibacterial target. C_{55}-P is synthesized by the enzyme undecaprenyl pyrophosphate synthase (UPPS). UPPS is a member of the cis-

isoprenyl pyrophosphate synthases, such as dehydrodolichyl-pyrophosphate synthase, which catalyze the formation of long chain products ranging from C_{55} to C_{100} ([36, 37]. As UPPS is specifically responsible for generating C_{55}-P, understanding the structure of UPPS is an intriguing research area for scientists, particularly for the purposes of developing new antibiotics. Inhibiting UPPS provides a method that promises to be an efficient strategy for treating bacterial infections.

1.7.2 Activity of UPPS

Many structurally diverse isoprenoids produced in nature are important components of cellular machinery, and serve as reproductive hormones, constituents of membranes, and signal transduction components, among others. Prenyltransferases catalyze the prenyl chain elongation of prenyl diphosphates, which are the common precursors for all isoprenoids. Prenyltransferases are grouped according to the *cis*- and *trans*-isomerism of the products they form (Figure 5) [38].

UPPS is a member of the cis-isoprenyl diphosphate synthase (cis-IPPS) superfamily. It catalyzes eight consecutive condensation reactions of farnesyl diphosphate (FPP) with isopentenyl diphosphate (IPP) to form undecaprenyl diphosphate, which subsequently gets dephosphorylated by undecaprenyl diphosphate phosphatase (UPPP) [39] , to yield the final C_{55}-P product.

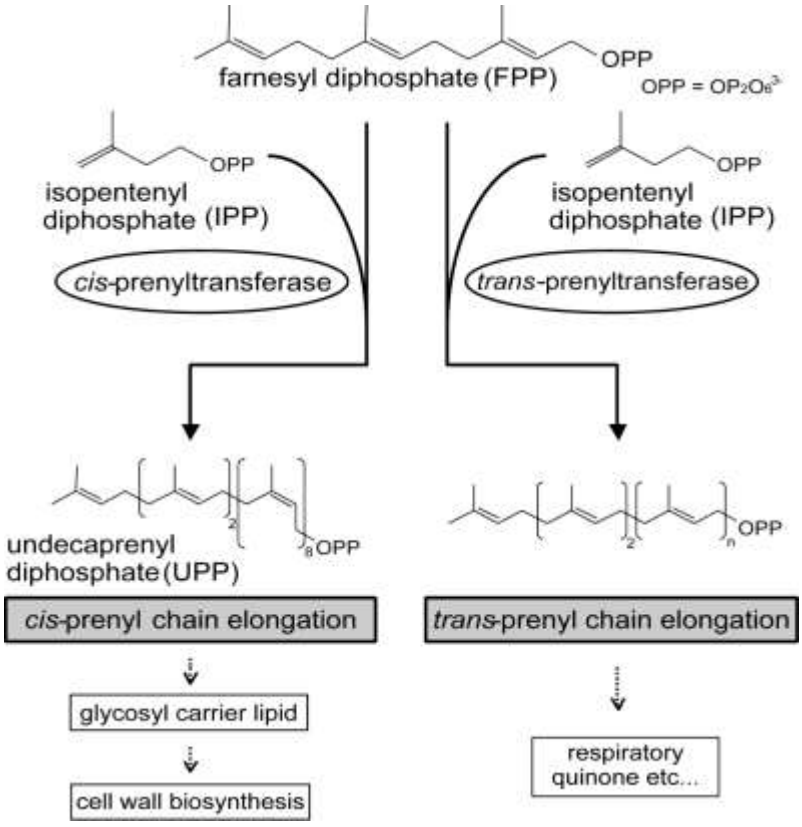


Figure 5: Schematic drawing of prenyl chain elongation. Fujihashi et al. 2001 ([38]).

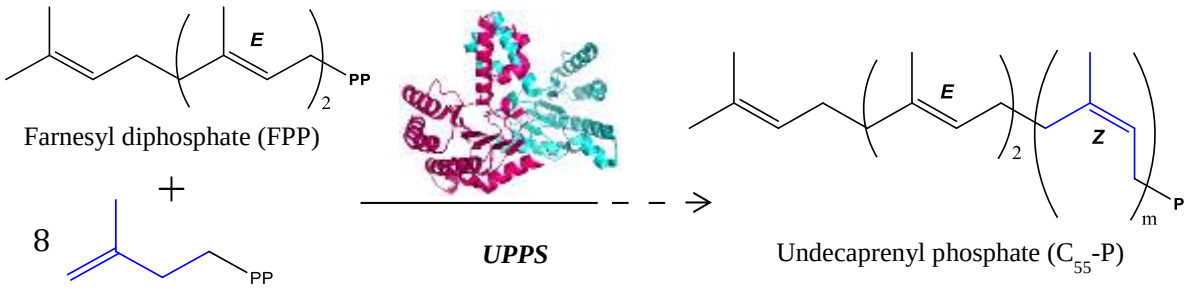


Figure 6: UPPS activity. UPPS catalyzes the condensation of 8 units of IPP and 1 unit of FPP to form undecaprenyl phosphate.

1.7.3 Overall Structure of UPPS

The overall structure of UPPS is conserved among species; there are 7 different species available in the Protein Data Bank (PDB). The enzyme is a functional homodimer, in which each monomer is crystallographically independent. The dimer, when viewed from the front, has the shape of a butterfly (Figure 7). Each subunit contains a catalytic domain, and a pairing domain. The contact interface of the dimer is about 15%. The topology seen from the front view of the dimer (Figure 8) shows six parallel β-strands (S1-6), and seven α-helices (H1, H2, H3 H5, H6, H8, and H10). The center is made up of a β-sheet core which is surrounded by five of the seven a-helices (H1, H2, H3, H5, and H10). Additionally, there are three short 3_{10}-helices (H4, H7, and H9) in each monomer [38, 40].

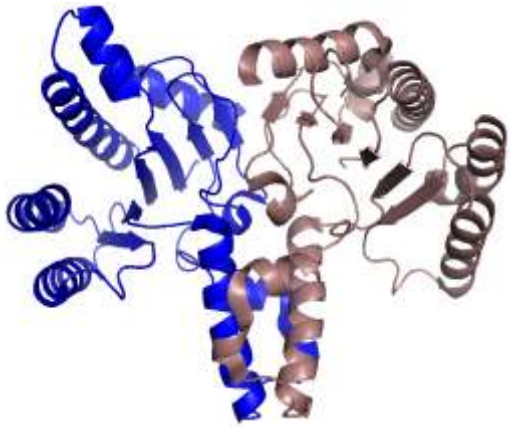


Figure 7: Overall structure of UPPS. The known crystal structures have a similar fold, and from the front view, have the shape of a butterfly. The enzyme is a functional homodimer. Each subunit is colored blue or violet. The structure shown here is from *E. coli* (PDB ID: 1JP3). Image rendered in PyMol.

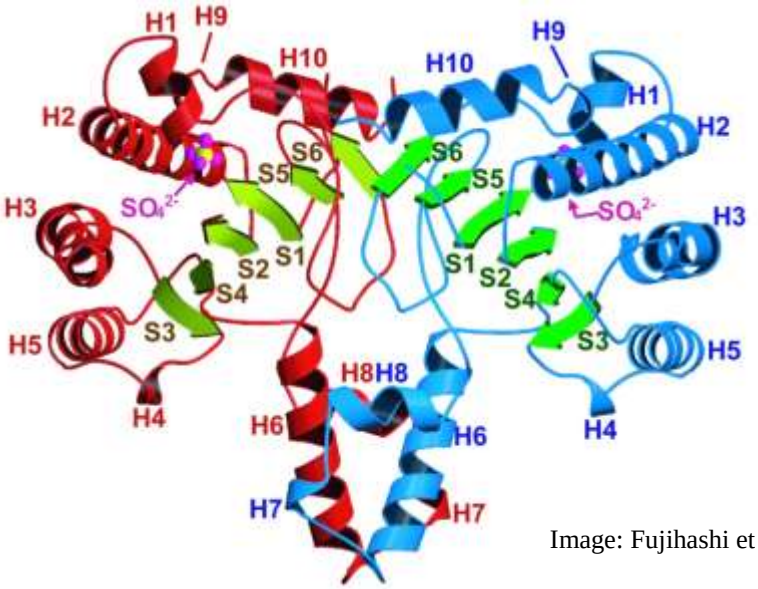


Image: Fujihashi et al, 2001.

Figure 8: Topology of overall structure of UPPS from M. luteus. The helices are labelled together with the sulfate ions found in the crystal structure.

The substrate binding site of UPPS was identified previously based on site-directed mutagenesis studies [41, 42]; the residues determined as important to FPP and IPP binding were: D-26, N-28, R-30, H-43, F-70, S-71, R-194, and E-198. A crystal structure of *E. coli* in complex with FPP (PDB ID: 1V7U), and another in complex with IPP and Mg (PDB ID: 1X07), confirm the importance of the residues identified. Furthermore, additional residues of importance have since been added to the substrate binding site. Figure 9 shows the location of the binding site in the protein. The two independent crystal structures superimposed on each other show that FPP and IPP both bind in the same binding site. The magnesium ion bound in the structure is thought to stabilize the negative phosphate charges of either FPP or IPP.

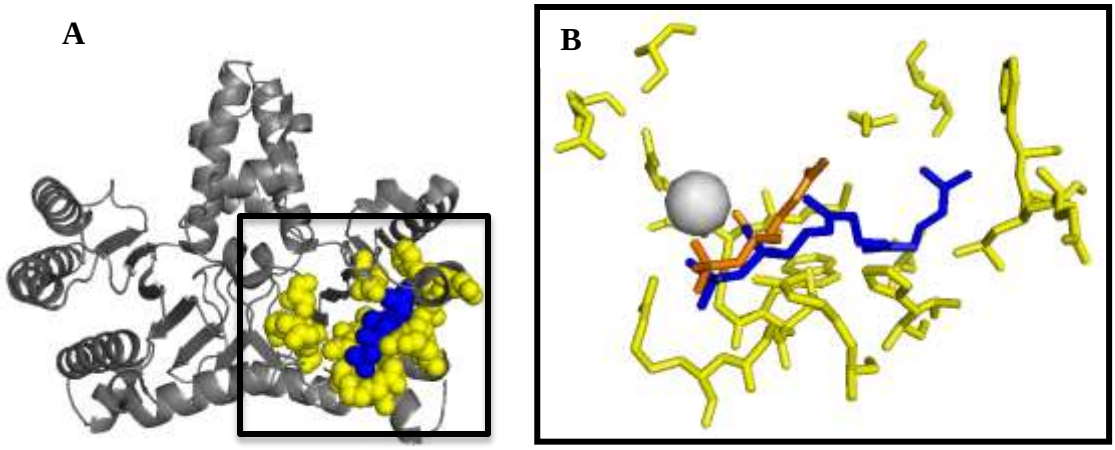


Figure 9: Crystal structure of E coli UPPS: A) complexed with FPP (PDB ID: 1V7U). The active site residues are represented as yellow spheres, and FPP is shown in a cozy perch in the active site as blue spheres. B) FPP from structure 1V7U superimposed onto *E. coli* UPPS complexed with IPP. FPP (blue sticks) (1V7U) and IPP (orange sticks) (1X07) shown in the active site (yellow sticks). Structure is shown bound with Mg (silvery-white sphere), which acts to stabilize the negative phosphate charges of FPP or IPP.

1.7.4 UPPS Inhibition

A lot of current work on UPPS is focused on developing inhibitors. Particularly because it is not produced in humans, targeting isoprenoid biosynthesis by inhibiting UPPS activity is a potentially important and effective route for antibacterials. Bisphosphonates that mimic the allylic diphosphate substrate have been found to effectively inhibit UPPS [43]. A recent publication by the Oldfield lab group identified antibacterial drug leads targeting isoprenoid biosynthesis, in which they reported X-ray crystallographic structures of 10 chemically diverse compounds (phosphonic, benzoic, and diketo acids, as well as bisamidine and a bisamine) that inhibit *E. coli* and *S. aureus* UPPS activity. Interestingly, it was also revealed that UPPS had four available binding sites for inhibitors (Figure 10) [44]. Other inhibitors have been identified through high

throughput screening such as tetramic and tetronic acids which inhibit *Streptococcus pneumoniae* UPPS [45].

Additional methods have been employed to aid in the search for UPPS inhibitors such as computer virtual screening [46], and molecular dynamics simulations. Through molecular dynamics simulations, multiple conformational states of UPPS have been shown to be recognized by different classes of inhibitor molecules [47], which some researchers have suggested may explain the difficulty in finding tight-binding inhibitors for UPPS. So far, only a limited number of UPPS inhibitors are available. Newer methods including radiolabeled assays, and fluorescent-detection methods ([26, 48], have been developed to monitor UPPS activity in real time in; this will enable more efficient ways to perform kinetics and inhibition studies in order to facilitate the inhibitor discovery process [49]. Due to the diverse range of inhibitors possible for inhibition, and the differences in susceptibility exhibited by various species, employing computational methods may contribute significantly to speeding up the drug discovery process.

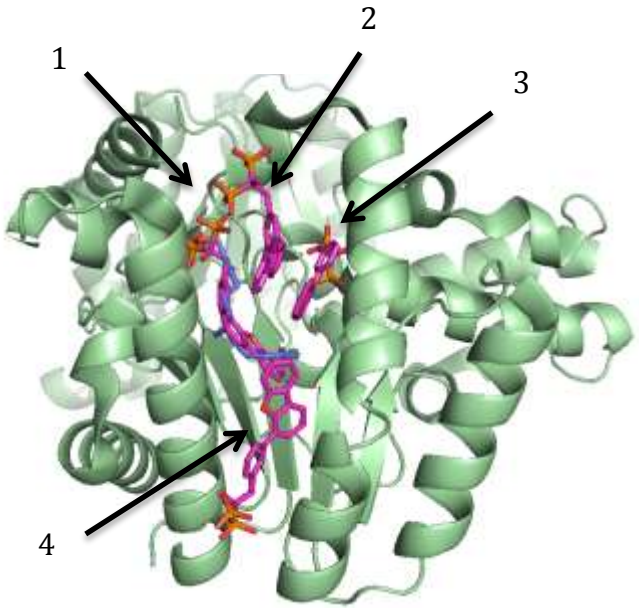


Figure 10: Structure of *E. coli* UPPS with bisphosphonate (BPH) inhibitor (pink) (PDB ID: 2E98). There are 4 inhibitor binding sites: site 1 shows one BPH superimposed with FPP (blue). The rest of the BPH molecules bind in sites 2, 3 and 4. Image rendered in PyMol.

In this work, we aim to integrate chemical tools and computational methods for data-driven drug discovery. The goal is to computationally demonstrate the hypothesis that structural and sequence differences across UPPS proteins lead to different mechanisms affecting compound selectivity and potency. We employ bioinformatics analyses to try to understand structural features of UPPS that may govern inhibitor selectivity. Further, we implement a Distance Constraint Model (DCM) (developed by the Jacobs group) [50] to explore the molecular basis for differences in activity across a wide array of UPPS families. We present a holistic approach that seeks to combine chemistry, bioinformatics, and computational methods to establish a comprehensive and detailed analysis of the factors influencing stability and flexibility across UPPS families. Our overarching aim is to differentiate molecular level details in sequence and structural variations to activity, to gain knowledge that can then be leveraged to create narrow spectrum antibiotics.

CHAPTER 2: EXPERIMENTAL

2.1 Protein Expression and Purification

The gene for each protein has been cloned into a pET30b expression vector (Fernandez Lab, UBC), with an amino-terminal protease-cleavable and purification tag that confers kanamycin resistance. This vector, encoding a hexahistidine sequence, was transformed into expression cell lines and selected for on Luria broth (LB) agar plates containing kanamycin (Kan). Glycerol stocks were prepared from small scale cultures of the transformed bacteria and stored at -80^{o}C to be used as needed.

A starter culture of 5 mL LB, 5 µL of 50 mg/mL Kan, and a micropipette tip poke of the glycerol stock was grown at 37^{o}C overnight, with shaking at 160 rpm. This small scale culture was then used to inoculate 1 L of sterile LB media (with 1 ml of 50 mg/mL kanamycin), and grown at 37^{o}C with shaking at 160 rpm. During log phase growth, when OD_{600} was between 0.6 and 0.8, gene expression of LgmA, LgmB, or LgmC was induced with 0.02% Isopropyl β-D-1-thiogalactopyranoside (IPTG) (for 4 hours at 37^{o}C or overnight at 25^{o}C). Cells were centrifuged for 15 mins at 5,000 x g, after which cell pellets were re-suspended in lysis buffer (50 mM Tris pH 8, 200 mM NaCl, 20 mM Imidazole) by shaking, and then sonicated; the cell lysate was obtained by centrifugation (for 75 mins at 30,000 rpm) to remove cell debris. The lysate containing soluble recombinant protein was purified by affinity chromatography, and membrane fractions were retained for LgmA, LgmB, and LgmC protein. Each was subsequently extracted

from the membrane fraction (in lysis buffer supplemented with 0.1% triton), and further purified on a Ni-NTA column.

The lysate containing soluble recombinant protein, or membrane-extracted protein was loaded on an immobilized metal affinity chromatography resin charged with Ni^{2+}. The resin was washed with 24 mL wash buffer (50 mM Tris, pH 8, 200 mM NaCl, 50 mM imidazole) followed by elution of bound protein with 10 mL of elution buffer (wash buffer supplemented to 500 mM imidazole). Purification of protein was followed by sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the cell lysate, flow through, wash, and elution fractions. Purified LgmA migrated as a ~38 kDa protein, in good agreement with the calculated molecular mass of 37,395 Da, LgmC migrated as ~40 kDa (predicted molecular weight 39,395 Da), and LgmB migrated as ~55 kDa (predicted molecular weight 55,000 Da). Purified protein was dialyzed at room temperature with buffer containing 50 mM Tris, pH 8 and 200 mM NaCl (3 x 1L, 4 hours each, with the third change going overnight). Protein concentration was determined by absorbance at 280 nm with the NanoDrop spectrophotometer. After dialysis, and protein quantification, glycerol stocks of aliquots were stored at -80^{o}C for future use. Membrane fractions for all 3 proteins were retained and used for activity assays. It should be noted here that no activity has been observed with extracted, Ni-NTA purified protein and thus all the data presented here are from activity assays performed with membrane fractions.

2.2 HPLC Activity Assays Utilizing Fluorescent Probes

Activity assays were performed using a fluorescent probe-based high performance liquid chromatography (HPLC) technique developed in the lab. Bactoprenyl phosphate

(BP), the starting material in the Lipid A modification pathway as well as the cell wall biosynthesis pathway, is synthesized from a single farnesyl diphosphate (FPP), and 8 units of isopentenyl diphosphate (IPP), in a reaction catalyzed by Undecaprenyl Pyrophosphate Synthase (UPPS). To monitor the activity of UPPS, fluorescent analogs of the precursors are utilized. 2-cyanoaline bactoprenyl phosphates (2CN-BP) (with variable chain length isoprenoids) have been synthesized, and are used as substrates in the activity assays [26]. Reversed-phase HPLC was performed with a C-18 column in which hydrophobic compounds are retained longer than polar compounds, allowing for tracking of the 2CN-BP (Ex: 340 nm, Em: 390 nm) based on retention times.

2.3 Replicating N-acetyl-glucosamine-1-phosphate (GlcNAc-1-P) Transferase Activity in WecA.

WecA has been shown to transfer GlcNAc-1-P to undecaprenyl phosphate. We were able to replicate this activity using our method in which the retention time of the starting material, 2CN-BPc5, decreased, indicating the formation of the more polar 2CN-BPc5-P-GlcNAC product. Activity was assayed in reactions containing 100 mM Bicine, 7.5 mM sodium cholate, 10 mM $MgCl_2$, 1.0 µM 2CN-BPc5, 1.0 mM UDP-GlcNAc, and 1 µL of 0.05 mg/mL WecA membrane fraction in a 100 µL reaction. Additional assays, including those used for kinetic studies and detergent effect on activity, were scaled down to 50 µL or 20 µL.

2.4 Confirming N-acetyl-glucosamine (GlcNAc) Transferase Activity in LgmA.

LgmA is predicted to act as a glycosyl transferase, and is hypothesized to transfer GlcNAc to undecaprenyl phosphate. In reaction conditions containing 50 mM MES

buffer pH 6, 10 mM $MnCl_2$, 0.1% Triton, 1.0 mM UDP-GlcNAc, 1.0 µM 2CN-BPc5, and LgmA membrane fraction, progress of the reaction was monitored.

2.5 Testing LgmC Protein for Deacetylation Activity

Two assays were employed for this method - the HPLC-based assay or a Capillary Electrophoresis (CE) assay. In the HPLC assay, LgmC enzyme was added to the LgmA reaction after formation of product, or was independently tested with isolated LgmA product. Different conditions were tested including soluble and membrane fractions of protein, different pHs, different metals, and different detergent types. The CE assay was used for testing LgmA as a UDP-GlcNAc deacetylase. Retention times based on charge and size of were monitored by absorbance at 260 nm (where uridine diphosphate strongly absorbs).

2.6 Activity Optimization, Enzyme Characterization, Kinetics, and Inhibition

To determine optimal conditions for enzymatic activity and to characterize the enzyme the following studies were performed: substrate specificity studies, metal-dependence, pH-dependence, and effect of surfactant concentration. Additionally kinetic studies and inhibition studies were also performed. In kinetic studies, reaction conditions were kept the same, while varying concentration of UDP-GlcNAc from 0.0 mM to 3.0 mM. Kinetic parameters were obtained by fitting data to Michaelis-Menten equation, and Hanes-woolf equation. For inhibition studies, a range of 13 other tunicamycin-derived inhibitors were also tested at 0.001 mg/mL inhibitor concentration per reaction. To determine the dissociation constant of tunicamycin, concentration was varied from 0.001 µg/mL to 0.009 µg/mL, to monitor inhibition kinetics. For both kinetic and inhibition

studies, reactions were quenched in n-propanol at specific time points to monitor rate of reaction.

2.7 Analysis of Structural Homologues

For bioinformatics analysis of WecA, the crystal structure of its homolog, MraY, in complex with the inhibitor, muraymycin D2, was analyzed. Inhibitor binding sites and other possible sites of enzyme-substrate / enzyme-inhibitor interactions were assessed to determine possible important interactions required for an effective inhibitor.

2.8 UPPS Structural Studies and Quantitative Stability/Flexibility Relationships Analysis

Crystal structures of UPPS from 7 bacterial species were obtained from the Protein Data Bank (PDB): *Micrococcus luteus* (1F75), *E. coli* (1JP3), *Helicobacter pylori* (2D2R), *Mycobacterium tuberculosis* (2VG4), *Campylobacter jejuni* (3UGS), *Streptococcus pneumoniae* (4Q9O), and *Staphylococcus aureus* (4U82). There are a range of other structures with a bound substrate of inhibitor, but, these were the only species available. To study the distribution of Quantitative Stability and Flexibility relationships (QSFR) in UPPS proteins, a maximum likelihood family tree was generated using multiple sequence alignment results obtained from PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool). The best scoring hits produced by PSI-BLAST were filtered using PISCES, and multiply aligned in MUSCLE. The resulting alignments were then improved using TrimAl for better analysis of phylogeny. The evolutionary tree was used as a guideline for the analysis of QSFR between proteins through minimal Distance Constraint Model (mDCM), an all-atom statistical mechanical model used to predict the thermodynamic and mechanical properties of a protein by

accounting for the non-additivity between free energy components through enthalpy-entropy compensation [50].

2.9 A Brief Overview of the Distance Constraint Model (DCM)

The DCM is used for simultaneously calculating thermodynamic and mechanical properties of proteins. It is based on a free energy decomposition scheme combined with constraint theory, allowing for microscopic interactions in the protein to be represented as mechanical distance constraints. Each distance constraint has an enthalpic and entropic cost [51, 52]. The microscopic interactions within the minimal DCM include covalent bonds, hydrogen bonds, and torsional forces. Overall, mechanical and thermodynamic concepts are integrated in a way that allow accurate flexibility characteristics of a given protein to be calculated over an ensemble of possible constraint topologies that are appropriately thermodynamically weighed [53].

2.10 UPPS Structural Differences across Species

Using the species obtained, we studied the active site of each structure based on results obtained from an NCBI (National Center Biotechnology Information) conserved domain search. Active sites were then aligned in the Pymol molecular visualization software to better study the 3D structure model. Quantitative Stability / Flexibility index parameters obtained from DCM were concurrently compared to determine if backbone flexibility or rigidity across the proteins could provide any insight into differences in activity between the various UPPS species.

CHAPTER 3: RESULTS AND DISCUSSION

3. Protein Expression

3.1 Successful Expression of Proteins Reveals Some New Information

LgmA, LgmB, and LgmC are the three proteins thought to act together to modify lipid A from *B. pertussis* with glucosamine. The gene thought to encode each protein had previously been identified and cloned into expression vectors. We report successful expression and isolation, for the first time, of each protein (Figure 11).

LgmA, a 352 amino acid protein with a predicted molecular weight of approximately 39 kDa, was resolved at approximately 38 kDA by SDS-PAGE gel, and Western blot analysis (Figure 11A). Additionally, finding that it was mostly retained in the membrane fraction agreed with its prediction as a membrane protein; it had been shown to have 3 transmembrane helices, by the TMHMM server [54].

3.1.1 LgmC Expression Reveals New Information about Its Size

LgmC, a 278 amino acid protein, was resolved at approximately 40 kDa; this was far from its predicted molecular weight of 31 kDa. It was also interesting to find that LgmC protein, which had been predicted to have no transmembrane helices, by the TMHMM server, was not found in the soluble fractions; instead, most of the protein was retained in the membrane fraction, as resolved by both SDS-PAGE and Western blot analysis (Figure 11B). Although this was not expected, none of these proteins have ever

been expressed, and thus there was no significant information on which to base or comparatively assess these initial findings.

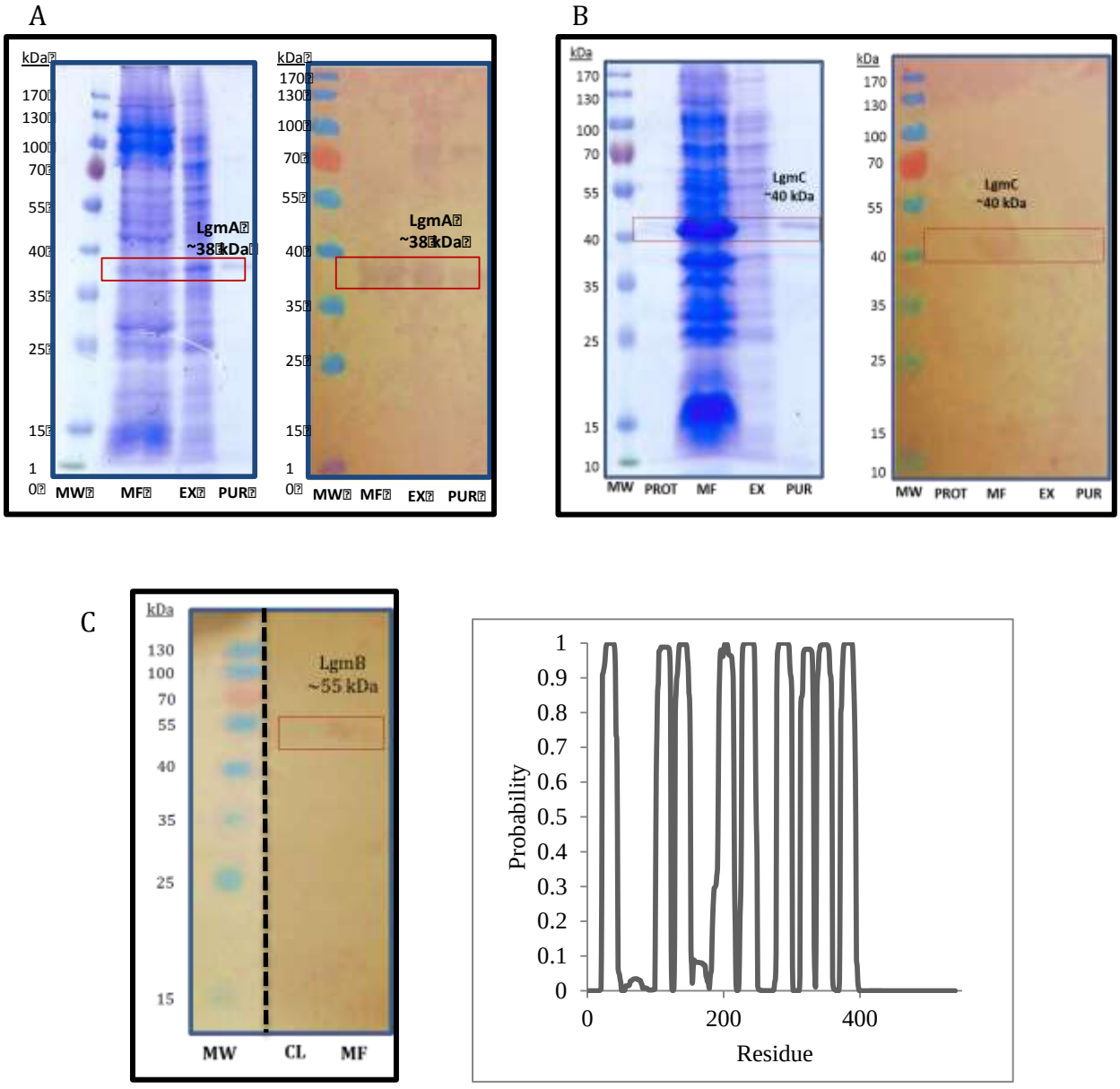


Figure 11: A and B) SDS-PAGE (left) and corresponding Western blot (right) show successful expression and purification of LgmA and LgmC. LgmA) Lanes: Molecular weight marker (MW), membrane fraction (MF), extracted protein from membrane fraction (EX), and Ni-NTA purified protein (PUR). Prominent bands for LgmA are visible at ~38 kDa. LgmC) Same as above, with dialyzed protein from the soluble lysate labeled PROT. Prominent bands for LgmC are visible at ~40 kDa. C) LgmB could only be expressed and resolved successfully as a membrane fraction. It is predicted to have 9 transmembrane helices by the TMHMM server. (Same labels as above).

3.1.2 LgmB Is a Predicted Membrane Protein with 9 Transmembrane Helices

LgmB was significantly more difficult to express. Armed with the knowledge that it was probably bound tightly to the membrane (having been predicted to have 9 transmembrane domains), we attempted various expression methods that have been shown to work for membrane proteins. Successful LgmB expression (in RP cell line and by auto-induction), is shown (Figure 11C) along with the transmembrane topology prediction. We were not successful in extracting it from the membrane under the same detergent conditions that had worked for LgmA and LgmC. Thus, LgmB protein could only be resolved by Western blot, and not SDS-PAGE (not shown). The estimated size of 55 kDa agrees with the prediction of its 528 amino acid sequence at 59 kDa.

3.1.3 Proteins Extracted From Membrane Did Not Show Activity

As previously observed, LgmA and LgmC were retained mostly in the membrane fractions. We attempted to extract these proteins from the membrane to purify further for activity assays. Preliminary assays showed that the extracted protein had no activity. As such, all activity assays reported herein are with membrane fractions. Recent literature demonstrated activity in a solubilized WecA enzyme [25]; unfortunately, we were unable to replicate this and only found activity in membrane fractions. Furthermore, WecA was not successfully resolved by Western blot; however, protein expression was confirmed as activity was present in membrane fractions of WecA.

3.2 Activity Assays

3.2.1 WecA is a GlcNAc-1-P transferase, and LgmA Acts a GlcNAc Transferase

Upon successful expression of the enzymes, we performed activity assay to determine enzymatic activity of each enzyme. WecA is an already characterized

GlcNAC-1-P transferase, for which we aimed to successfully replicate this activity. For LgmA, we aimed to firstly, establish activity, and secondly, confirm predicted function. Using the fluorescence-detecting HPLC method (Figure 12), in which fluorescent analogs of the bactoprenyl phosphate substrate can be tracked through separation on a hydrophobic column, activity of WecA and LgmA were monitored. As each enzyme was predicted to form a different product (Figure 13), we expected the individual products would be successfully separated on the hydrophobic column, as WecA product would be more polar.

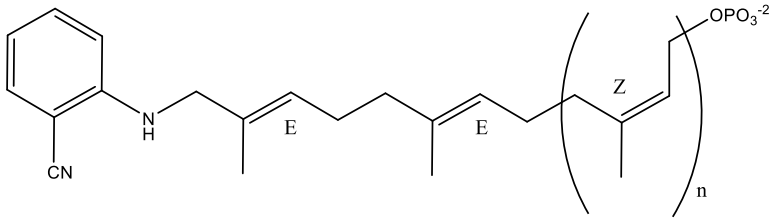


Figure 12**:** Variable chain length BP substrates. 2CN-BP(2-cyanoaniline bactoprenyl phosphate) substrates are synthesized in a UPPS-catalyzed reaction. The isoprenoid chain length (n) varies from 0 to 8 cis additions [BPc0 – BPc8]. This fluorescent analog of the BP can be tracked by HPLC with excitation at 340 nm, and emission at 390 nm.

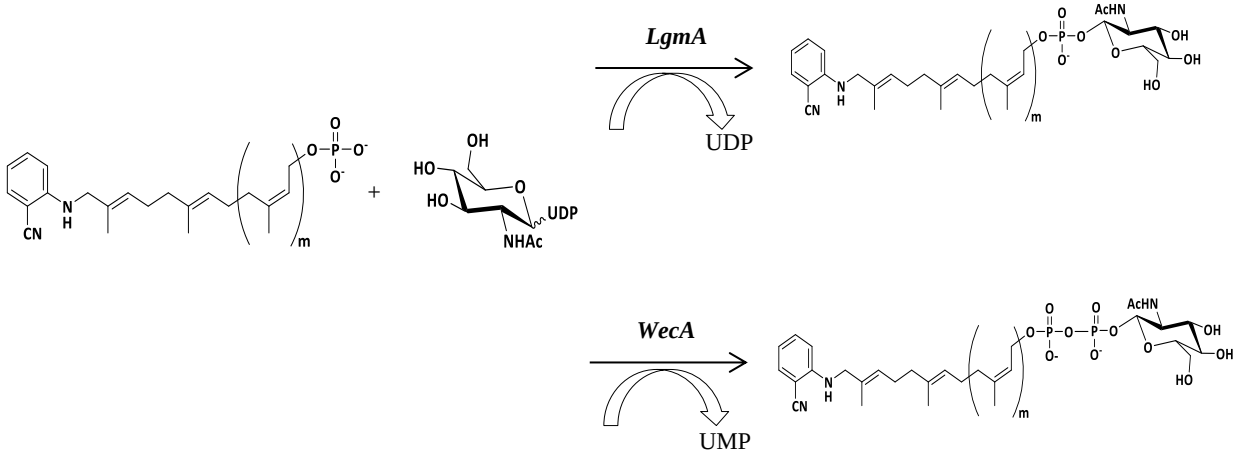


Figure 13: The reaction schemes show the difference in activity between LgmA GlcNAc transferase, and WecA GlcNAc-1-P transferase. LgmA transfers only the GlcNac moiety from UDP-GlcNAc to undecaprenyl phosphate toy yield Und-P-GlcNac while WecA, like most other glycosyltransferases, transfers the GlcNAc-1-P moiety.

WecA activity was replicated successfully; GlcNAc-1-P transferase activity was observed in membrane fractions of the enzyme (Figure 14), indicated by the formation of a more polar product which has less retention time than the starting substrate. Control assays containing no substrate, and/or membrane fraction without WecA gene showed no activity.

Similarly, LgmA membrane fractions were shown to have activity. The resulting product (Figure 14) had a retention time greater than that of 2CN-BP-P-GlcNAc, but less than that of the starting material. This is what would be expected for 2CNBP-GlcNAc which would be more polar than 2CN-BP due to the added sugar molecule, but less polar than 2CN-BP-P-GlcNAc, with one less phosphate. Mass spectral analysis of LgmA product (Figure 14) confirmed formation of the 2CN-BP-GlcNAc product, with a mass-to-charge ratio (892.34 m/z) in agreement with the calculated molecular weight (892.00 g/mol).

No activity was observed in any of the Ni-NTA purified lysates, or protein extracted from the membrane fraction in detergent. As such, only membrane fractions of LgmA protein were retained and used for assays. Control assays were run with no substrate, and/or membrane fraction without LgmA gene to ensure nothing in the membrane fraction would interfere with monitoring activity successfully. No activity was observed in any of the control assays (Appendix A). This demonstrated that LgmA had been successfully expressed, activity had been established, and its prediction as a GlcNAc transferase was confirmed. This makes LgmA one of the few lipid A modification proteins to be successfully expressed and shown to have activity, as well as one of the few UDP-GlcNAc:C55-P glycosyltransferase to be characterized.

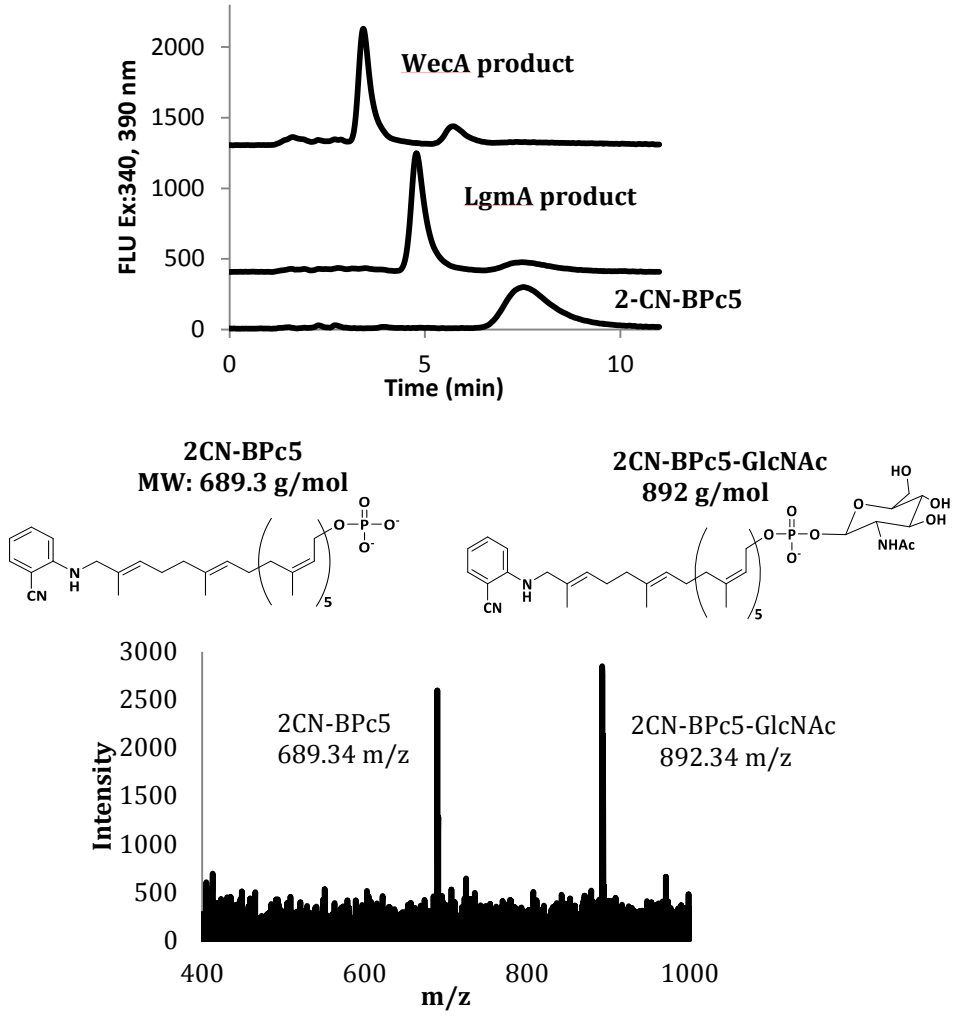


Figure 14: LgmA vs WecA activity. Above: WecA transfers GlcNAc-1-P from UDP-GlcNAc to 2-CN-BPc5 (2-cyanoaniline-(cis-5)bactoprenyl phosphate), while LgmA transfers GlcNAc from UDP-GlcNAc to the 2CN-BPc5. Product of the WecA reaction is seen to have a shorter retention time for the more polar 2CN-(c5)BP-P-GlcNAc compared to the 2CN-(c5)BP-GlcNAc product of LgmA (above). FLU on y-axis nidicates fluorescence intensity. Below: Mass spectrometry analysis of LgmA product (below): a m/z ratio of LgmA product is in agreement with the calculated molecular weight of 2CN-(c5)BP-GlcNAc, confirming the proposed function of LgmA.

3.3 Effect of Metal, pH, And Surfactant Concentration on LgmA Enzymatic Activity

Having confirmed the proposed function of LgmA, we sought to characterize the best conditions for this enzyme, and to compare it to WecA. WecA has been reported to require detergent, as well as a metal for activity. To understand if similar conditions

would be required for GlcNAc transferase activity, we tested metal-dependence, pH-dependence, and effect of surfactant concentration on LgmA activity.

3.3.1 pH Changes Have No Effect on LgmA Activity

To determine the optimum pH for LgmA activity, we tested activity at three pHs: 6.0, 7.5, or 8.0. As seen in Figure 15, pH changes within our tested range had no noticeable effect on activity. It seems in this range, there is no significant effect on amino acid ionization that would affect the protein's 3D-level structure, to affect enzymatic activity under our reaction conditions. We retained pH 6.0 for subsequent LgmA assays.

3.3.2 LgmA, Like WecA, Requires a Divalent Metal for Activity

Many of the characterized GTs have been shown to require a metal. The structural representative of the GT family, SpsA, has been shown to require metal for activity, and has been successfully crystallized in complex with both Mn and Mg [55]. To assess if the GlcNAc transferase would also require a metal, we explored metal-dependence of LgmA activity. We found that LgmA also required the presence of a divalent metal cation for activity, which has already been shown in WecA to be Mg^{2+} or Mn^{2+} as confirmed in our studies [25, 27]. LgmA had similar levels of activity in both metals (Figure **16**); however, we usually observed a slightly higher level of activity (in terms of turnover, and reaction time) for Mn. As such Mn was retained for LgmA reaction conditions.

3.3.3 Product Turnover Is Affected By Detergent Type And Concentration In LgmA.

In order to evaluate the activity of LgmA, it was necessary to identify a detergent that effectively solubilized the hydrophobic polyisoprenyl substrates, without inhibiting enzymatic activity. Efficient activity with membrane-bound LgmA had been found in

Triton X-100, and thus had been retained in all LgmA assays. However, triton was found to interfere significantly with mass spectral analysis. As such, we explored the efficiency of cholate in LgmA activity. In attempting to optimize cholate (which seems to be optimal for WecA activity as well), we observed that not only did detergent concentration affect LgmA activity, but detergent type also had an effect. Cholate was not as effective as triton for LgmA activity. Additionally, higher cholate concentrations were required for the different substrate lengths (

Figure **16**), while a fixed concentration of triton had worked in all substrates. The final conditions chosen for LgmA activity were pH 6.0, 0.1% Triton, and10 mM $MnCl_2$ in which the highest turnover was usually observed.

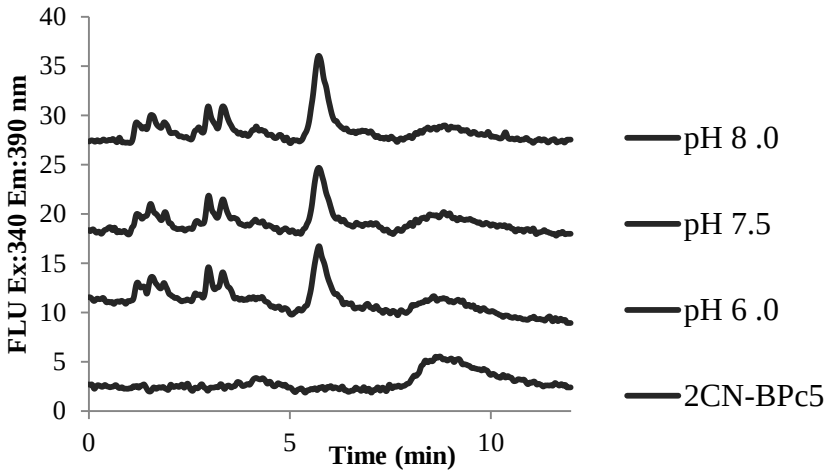


Figure 15: LgmA pH-dependence. LgmA activity was assayed at pH 6.0, 7.5, and 8.0. LgmA was functional at each pH, with no significant differences in activity observed.

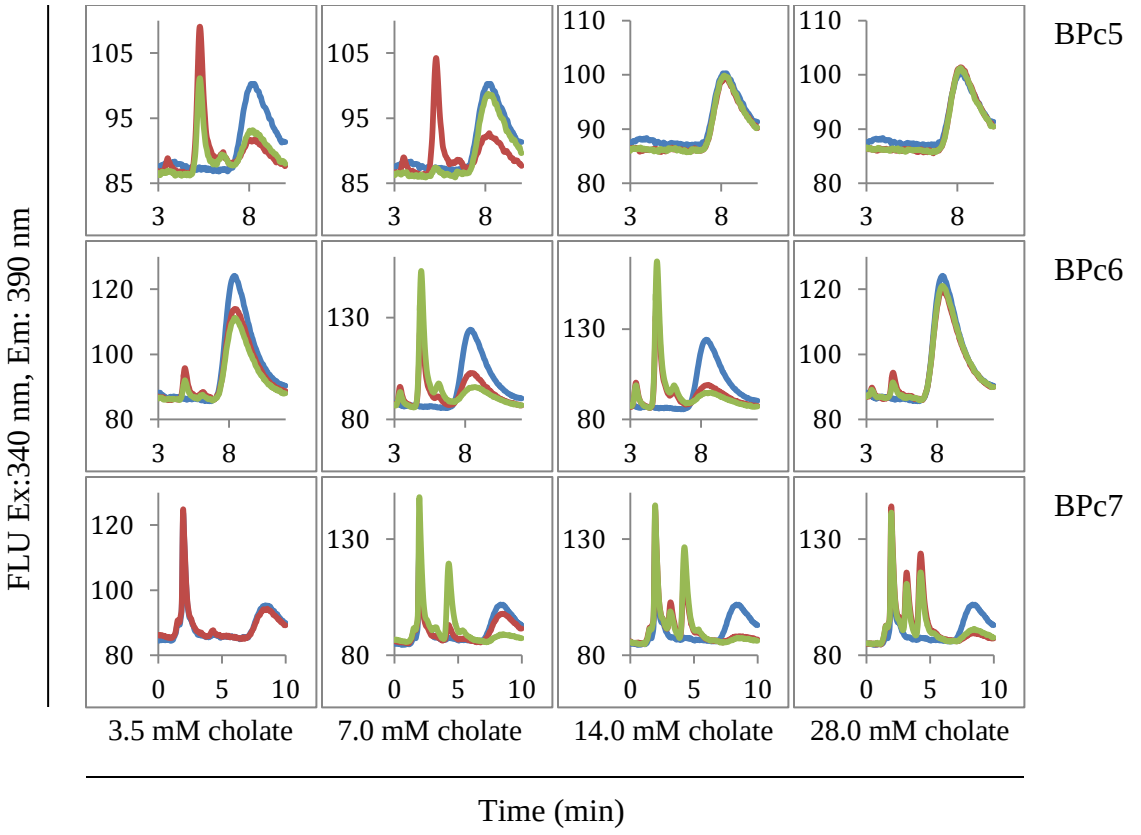


Figure 16: Metal-dependence and effect of detergent concentration. LgmA was found to have activity in cholate, although not as efficient as in triton. It was found that concentration of surfactant also had an effect on activity. Metal effect also showed that LgmA had activity in the presence of both Mn (red graph) and Mg (green graph). Control reactions are in blue.

3.4 LgmA Shows a High Preference for UDP-GlcNAc

We were interested in the range of substrates LgmA could work with. In reactions testing alternative sugar substrates (UDP-Glucose, UDP-Galactose, and UDP-GalNAc), we found LgmA had a high preference for GlcNAc (Figure 17). Each sugar was tested in place of UDP-GlcNAc at a concentration of 1 mM. Activity was only observed in the presence of UDP-GlcNAc, indicating a very high preference of LgmA for using GlcNAc as a substrate. Not only does this confirm the accuracy of its functional prediction, but it also supports its possible role in *B. pertussis* lipid A-GlcN modification, since the

initiating hexosyltransferase in the suggested pathway would most likely require a selective GlcNAc transferase (as LgmA has been shown to be). Activity in the presence of UDP-GlcNAc was usually seen at submillimolar concentrations within about 10 minutes. On the other hand, no significant activity was observed for any of the other substrates, although a small amount of product formation was observed at concentrations greater than 10 mM and reaction times beyond an hour. LgmA was therefore determined to be highly specific for UDP-GlcNAc, similar to WecA.

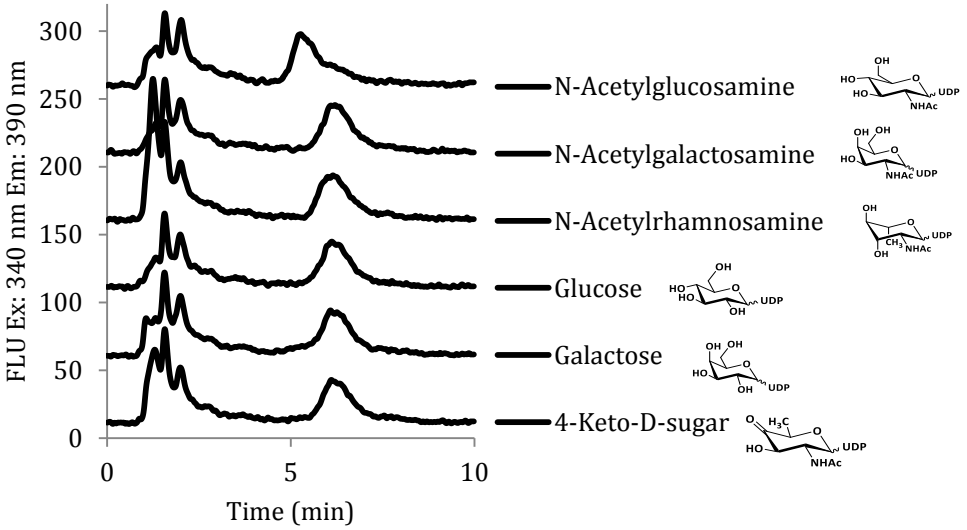


Figure 17: Substrate specificity studies show LgmA has a strong preference for GlcNAc. Reactions were performed in the presence of 1.0 mM substrate.

3.5 LgmA Does Not Discriminate Between Isoprenoid Chain Lengths

As we have found LgmA to have high specificity for its sugar donor substrate, we wanted to find out how selective it would be for the lipid-phosphate acceptor. To determine whether isoprenoid chain length plays a role in LgmA activity, assays were performed using variable length bactoprenyl phosphate (BP) substrates, with length ranging from 0 cis additions to 8 cis additions (BPc0 – BPc8). Results show activity for LgmA with 2CN-BPc1 through 2CN-BPc8, while no activity is observed for 2CN-BPc0.

2CN-BPc2 and 2CN-BPc3 also have activity, although peaks are not very well-resolved (Figure **18**).

3.6 Triton Acts As a Better Surfactant for LgmA Activity than Cholate

Concentration of surfactant was assessed with varying bactoprenyl phosphate chain lengths. Activity was present at all lengths in 0.1% triton. However, differences in activity at varying lengths were observed with sodium cholate, as observed in Figure **16**. Thus, we performed additional assays to compare LgmA activity in triton to LgmA activity in different concentrations cholate. Results are summarized in Table **1**, in which better activity is seen with higher isoprenoid chain lengths at higher concentrations of sodium cholate (critical micelle concentration: 14 mM). No activity was observed with BPc1 or BPc8 in cholate; seeing as the shorter length substrates had better activity in lower concentration of cholate, we can speculate that further decreasing the concentration of cholate may help with BPc1 product formation.

3.7 Comparing Kinetics and Inhibition of LgmA and WecA

3.7.1 LgmA Demonstrates Higher Affinity for GlcNAc than WecA

Kinetic studies were performed to understand the affinity for UDP-GlcNAc between WecA (GlaNAc-1-P transferase) and LgmA (GlcNAc transferase). Using Michaelis-Menten (MM) plots, and Hanes-woolf (HW) plots (Figure 19), LgmA was determined to have higher affinity for UDP-GlcNAc with a K_m of 0.013 mM +/- 0.00067 at 0.5 mg/mL total protein, while WecA has a K_m of 0.342 mM +/- 0.06 at 0.42 mg/mL total protein.

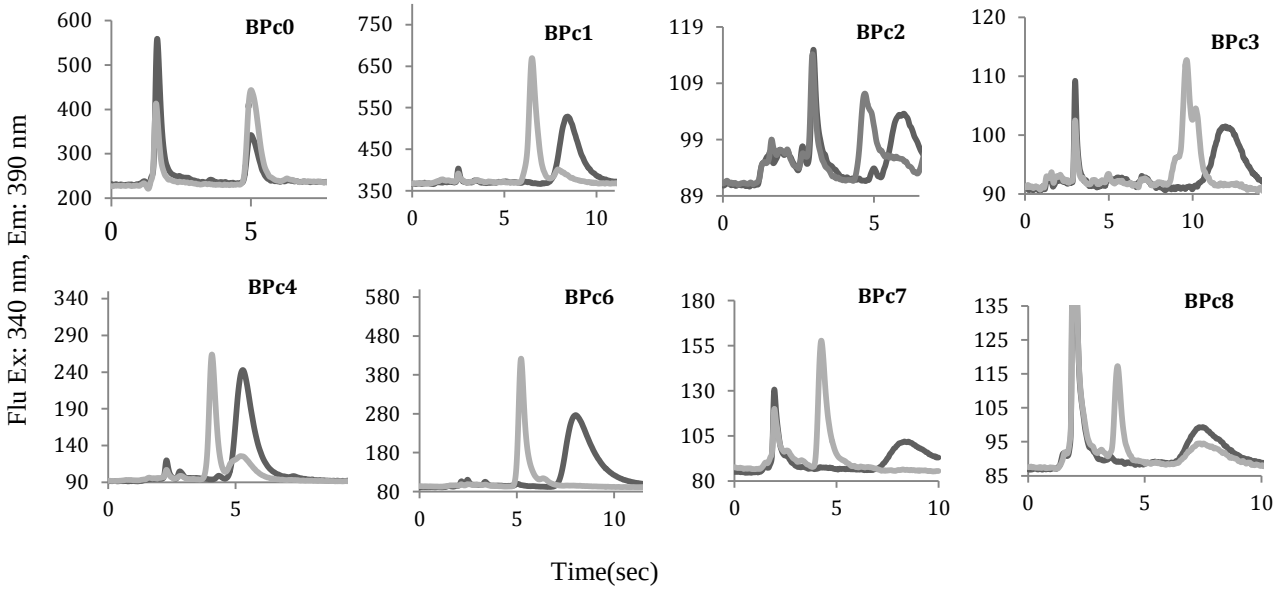


Figure 18: LgmA does not discriminate much between BP chain lengths. Black graphs represent control reactions (with no LgmA) and grey graphs show reactions containing LgmA. Successful transfer of GlcNAc to 2CN-BP is indicated by lower retention times in the LgmA-containing (grey) reactions. Successful transfer of GlcNAc is seen in 2CN-BPc1 through 2CN-BPc8, but not in 2CN-BPc0.

Table 1: Activity of LgmA with varying bactoprenyl phosphate chain lengths was assessed in the presence of varying concentrations of cholate and compared to activity in Triton. Differences in activity are observed with sodium cholate: higher length isoprenoids seem to have better activity at higher cholate concentrations than the shorter length isoprenoids.

	Triton	Cholate			
BPc(X)	0.1 %	3.5 mM	7.0 mM	14.0 mM	28.0 mM
0	-	-	-	-	-
1	-	-	-	-	-
2	69.1	70.7	68.6	69.3	16.6
3	72.9	41.1	62.5	60.9	76.4
4	46.4	30.5	41.2	23.4	8.9
5	19.6	33.5	52.6	3.9	0.8
6	23.2	5.7	19.9	17.3	5.9
7	39.7	0.2	0.6	3.1	1.6
8	51.9	-	-	-	-

3.7.2 WecA is Inhibited More Effectively by Tunicamycin than LgmA

Not many glycosyltransferases have been assessed against a range of inhibitors, so, inhibition tests were performed to gauge the effect of various tunicamycin-derived inhibitors on these two enzymes. In addition to tunicamycin, we tested ten tunicamycin-derivatives T-0 through T-10 (which retain the glucosamine moiety, but have different terminal moieties) (Figure 20), and three novel derivatives: reduced tunicamycin, double-reduced tunicamycin, and quinovosamycin (structure not available due to intellectual property concerns).

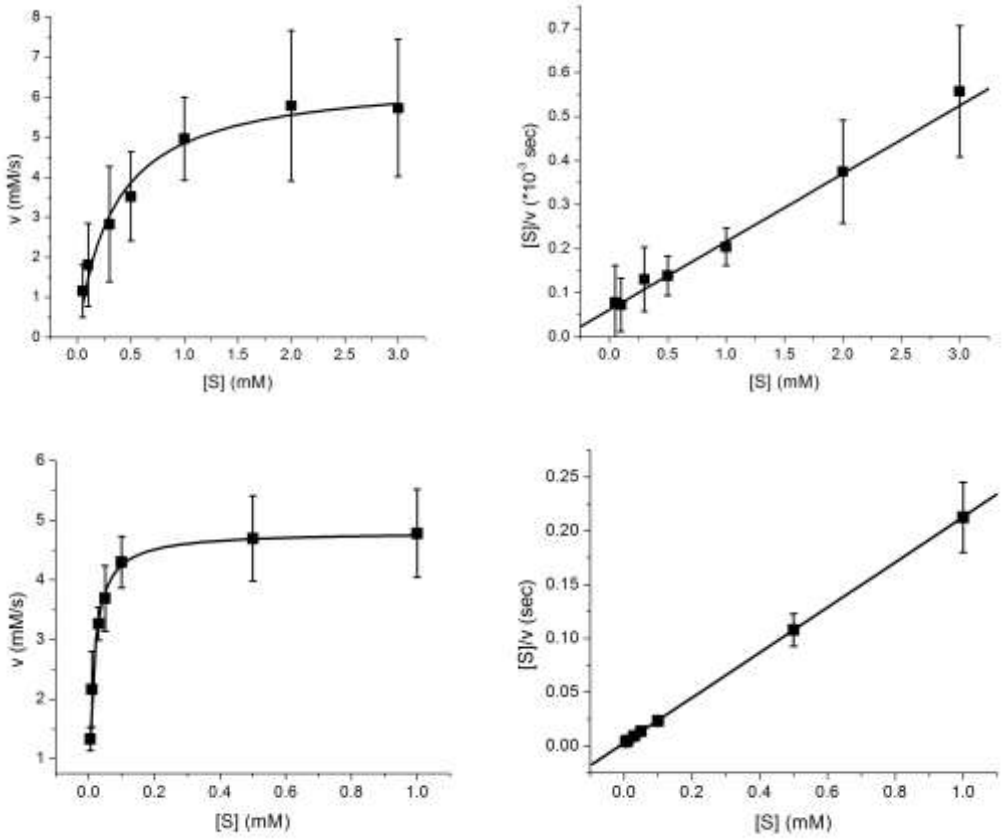


Figure 19: Kinetic studies. WecA and LgmA both have a high specificity for GlcNAc. LgmA displays a higher K_m of 0.013 mM +/- 0.00067 at 0.5 mg/mL total protein [above: MM plot (left) and corresponding HW plot (right)], while WecA has a K_m of 0.342 mM +/- 0.06 at 0.42 mg/mL total protein [below: MM plot (left) and corresponding HW plot (right)].

Tunicamycin has a structure that mimics the N-acetyl-D-hexoasamine-1-phosphate translocases/transferases, and therefore acts an effective inhibitor in bacterial cell wall synthesis and the N-glycosylations of eukaryotic proteins; it has been shown to inhibit WecA at submicromolar concentrations [25], as well as the eukaryotic UDP-GlcNAc:dolichyl-phosphate GlcNAc-1-phosphate transferase (GPT), which was inhibited by 50% or greater at antibiotic concentrations of 0.05-0.1 µg/mL. To probe the specificity of tunicamycin and to promote the design of other antibiotics that may have a similar mechanism of action, we assessed enzymatic activity of WecA and LgmA against ten tunicamycin-derived inhibitors synthesized in the Biao Yu lab [56], as well as a novel derivative of tunicamycin containing an N-acetyl-quinivosamine moiety instead of a GlcNAc moiety [57].

We expected to observe inhibition with all the derivatives since interactions with GlcNAc seemed to be the main determinant of inhibition route. Table 2 and Figure 21 are a summary of inhibition data with all the derivatives. We calculated percent of product formation to determine the presence or absence of inhibition. We found that the enzymes were not inhibited to the same extent. WecA was more effectively inhibited at very low concentrations of tunicamycin (than LgmA) at an estimated K_i of 0.04µg/mL as seen in the Dixon plot (Figure 22). Although a *Ki* value was not determined for LgmA in the presence of tunicamycin, it was determined that it took higher concentrations of tunicamycin to inhibit activity (Figure 23). Our results suggest that bacterial GlcNAc-1-P transferase, and the GlcNAc transferase, may not be inhibited in the same manner, as seen with tunicamycin and tunicamycin-derivatives.

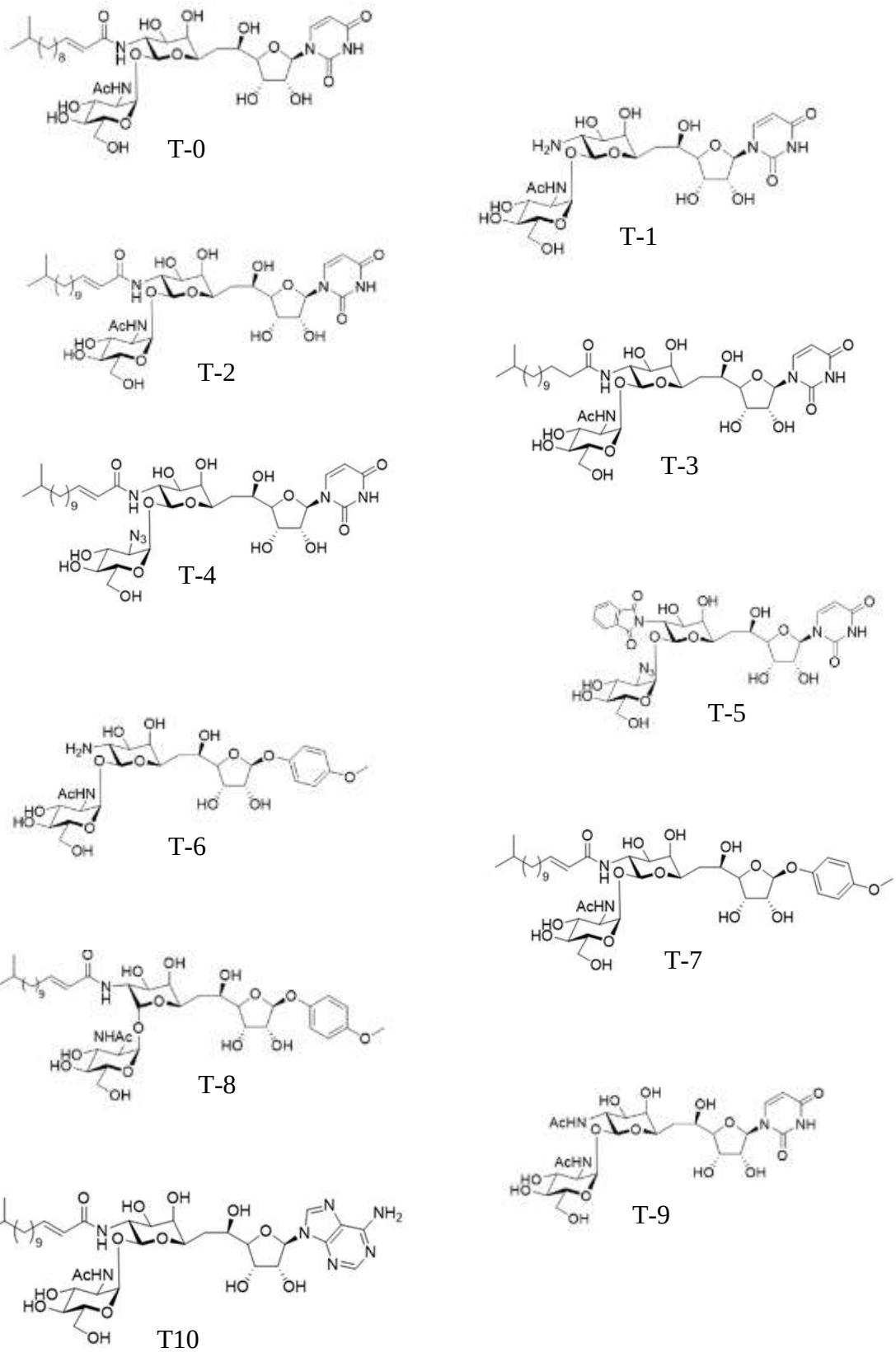


Figure 20: Tunicamycin-derived derivatives synthesized in the Biao Yu lab.

Table 2: WecA activity and LgmA activity were assessed against a range of tunicamycin-derived inhibitors. Results are analyzed based on product peak areas, using the uninhibited WecA reaction as a baseline to determine extent of inhbition. Reactions were performed in assays containing 0.001 mg/mL of inhibitor per reaction.

	Product turnover (%)	
Inhibitor	**WecA**	**LgmA**
None	44.331	25.937
TUN	0	15.697
T-0	0	11.502
T-1	0	16.906
T-2	0	10.492
T-3	0	11.923
T-4	0	13.778
T-5	28.763	15.469
T-6	58.587	18.796
T-7	56.085	25.123
T-8	56.791	23.821
T-9	14.498	20.848
T-10	24.509	29.079
double-reduced TUN	0	25.364
reduced TUN	0	26.325
QVM	0	19.930

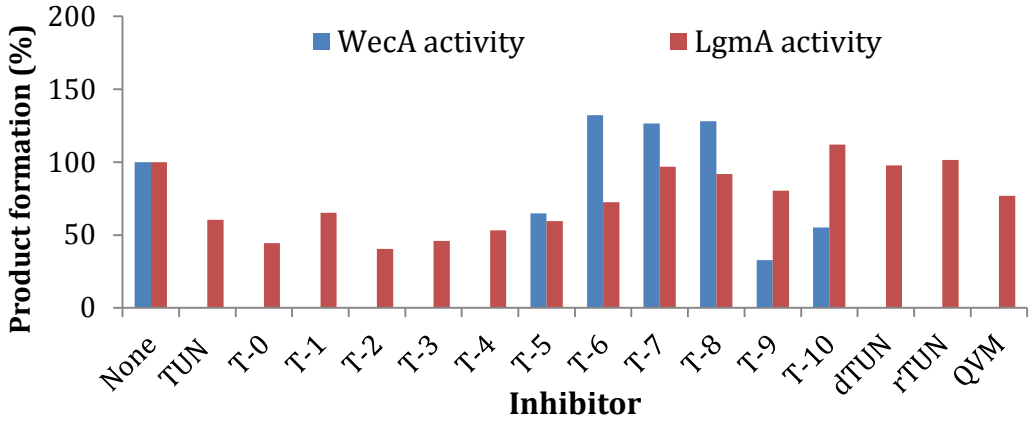


Figure 21: LgmA activity was not efficiently inhibited by tunicamycin. At the same concentration that inhibited WecA activity (blue graphs), LgmA activity was not affected (red graphs). Extent of inhibition was assessed against activity of enzyme with no inhibitor. A threshold of 0% product formation(in the presence of inhibitor) was set to indicate effective inhibition; less than 50% product formation (in the presence of inhibitor) to indicate some inhibition, and above 50% product formation (in the presence of inhibitor) to indicate no inhibition. *These results were found to be consistent across three different sets of experiments; however, error bars are not shown as only this set was performed under the same conditions (0.001 mg/mL inhibitor, 0.5 mM UDP-GlcNAc; HPLC conditions: 50% propanol, 50% ammonium bicarbonate).

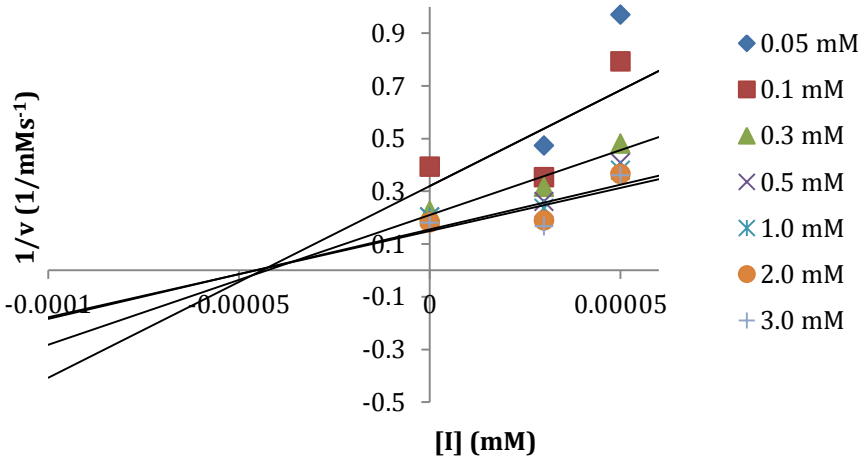


Figure 22: Dixon plot of WecA inhibition with Tunicamycin. Inhibition was tested with GlcNAc concentrations ranging from 0.05 mM to 3.0 mM, and tunicamycin concentrations ranging from 0.001 µg/mL to 0.05 µg/mL. Results from the Dixon plot showed a K_i at about 0.04 µg/mL.

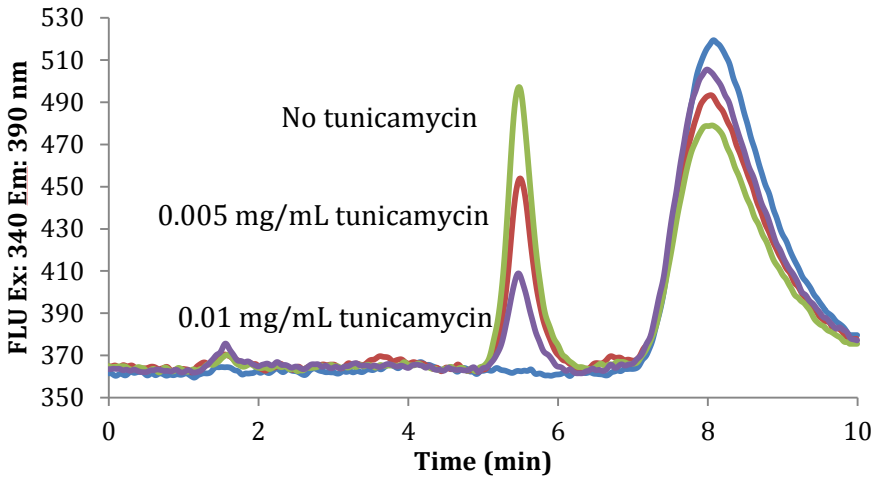


Figure 23: LgmA activity was monitored for 10 minutes in the presence of tunicamycin. The green graph has no inhibitor, while the red and purple graphs contain 0.005 mg/mL and 0.01 mg/mL, respectively.

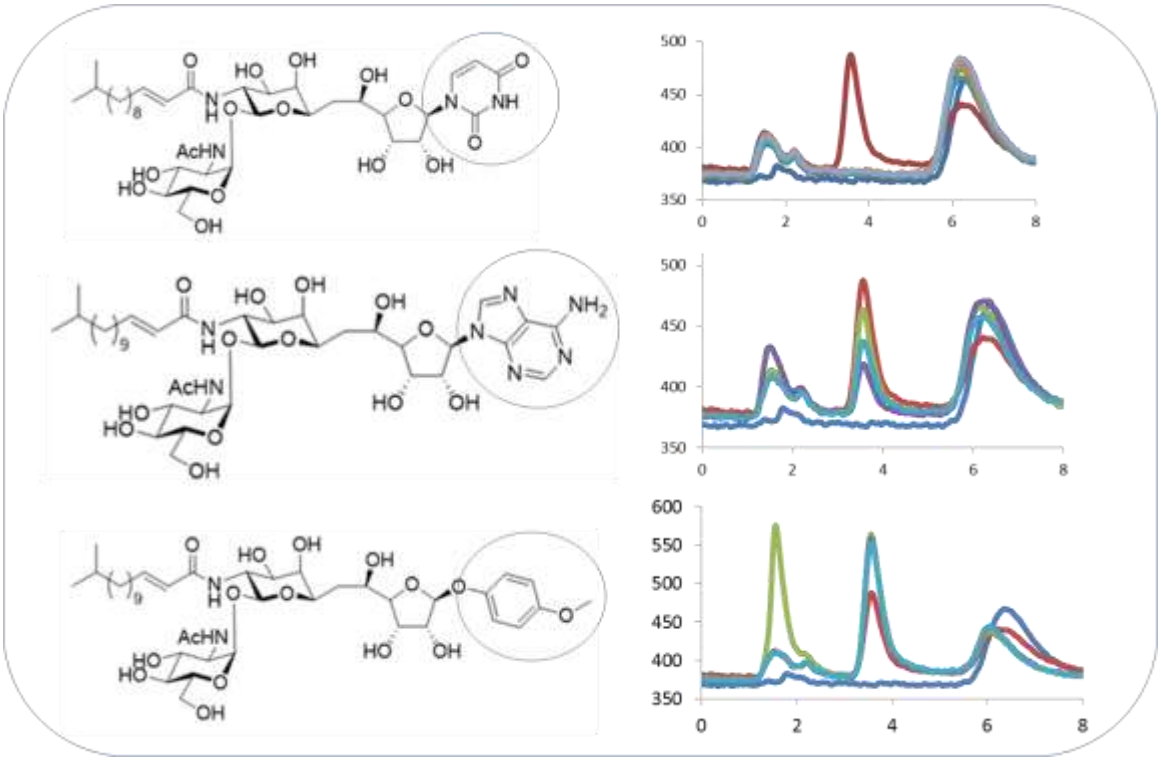


Figure 24: Summary of WecA inhibition with tunicamycin-derivatives. Compounds with a uracil moiety show more effective inhibition, while compounds with an anisole moiety show no inhibition at all. Compounds with another nucleobase moiety (e.g. adenine), have some inhibition. Red graphs on each plot represents WecA activity with no inhibitor. Other colors represent WecA activity in the presence of an inhibitor. In the first graph (on top), complete inhibition is observed for all the inhibitors containing the uracil moiety; the second graph shows some inhibition, indicated by 50% or less product formation compared to the red graph. The last graph (bottom) showed activity even higher than that with no inhibitor, indicating no inhibition.

3.8 Structural Analysis of WecA Paralogue, MraY, Shows that the Uracil Moiety Is Required For Effective WecA Inhibition

A pattern was observed with WecA inhibition data as seen in Figure 24. The inhibitors containing a uracil moiety showed the most effective inhibition. Structural studies were performed using MraY, a paralogue of WecA, to understand binding interaction between the enzyme and inhibitor. MraY, like WecA, has been shown to be susceptible to tunicamycin and muraymycin.

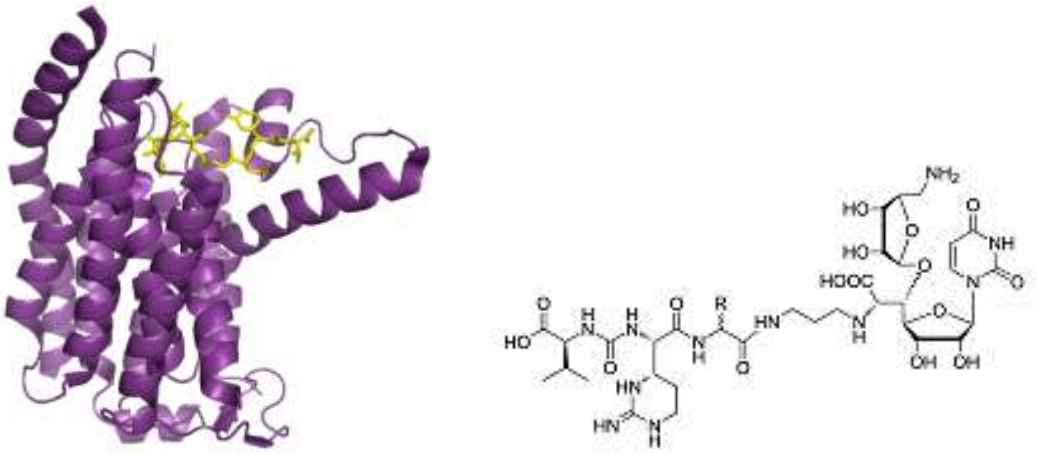


Figure 25: Crystal structure of MraY, a WecA paralogue, crystallized with inhibitor, muraymycin. Structure of muraymycin is shown on the right.

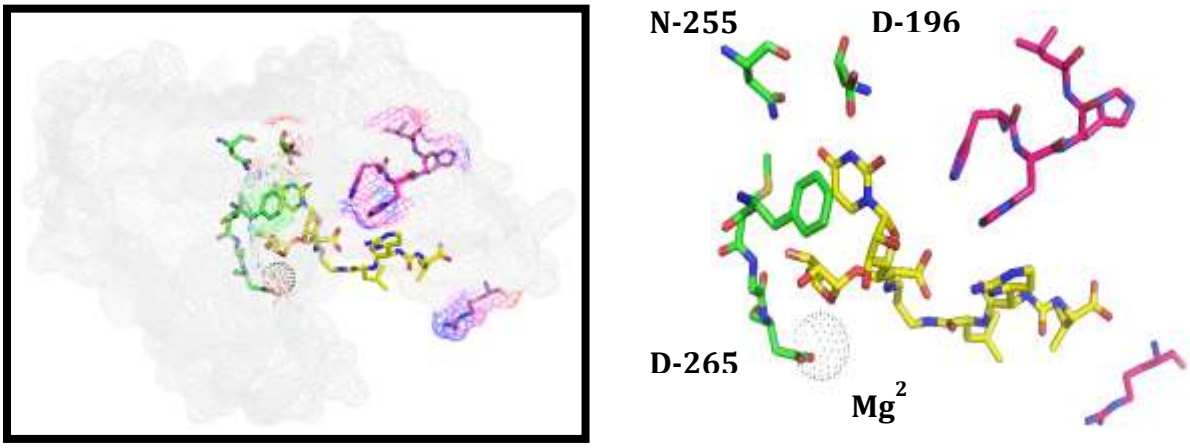


Figure 26: (Left) Structure of MraY shown as a mesh with active site residues highlighted in green and pink; muraymycin is colored yellow. (Right) Zoom-in of active site; the uracil moiety of muraymycin seems to be interacting with N-255 and D-196 via hydrogen bonds. The active site is stabilized by Mg (via interaction with D-265).

Figure 25 shows the crystal structure of MraY (PDB ID: 5CKR) shown with the inhibitor muraymycin (colored yellow). The structure of muraymycin, as seen on the right, contains the uracil moiety (as in tunicamycin). The important residues in the active site are highlighted in the mesh structure (Figure 26). The green residues represent the active site, and the pink residues are the predicted substrate (UDP-sugar) binding site; the

muraymycin inhibitor is colored yellow. On the right is a zoomed-in view of the active site. From the atom-colored structure (nitrogen = blue, oxygen = red), we see hydrogen bonding interactions between the active site residues: the positively charged end of N-255 is interacting with the ketone coming off the fourth carbon in the uracil ring. The negatively charged D-196 interacts with the nitrogen in the 3-position of the uracil ring. These two seem to the main stabilizing interactions between the active site and the uracil moiety of the inhibitor.

These interactions align with our observations from the inhibition activity in WecA. We predict that since the active sites of WecA and MraY are highly conserved (including the D-196, N-255, D-265 residues) [sequence comparison shown above], WecA structure is very likely similar to MraY, and is most likely inhibited in the same manner. From our observations, the inhibitors with the uracil moiety have the best inhibition activity due to stabilized hydrogen bond interactions with the active site, facilitating more effective binding. The inhibitors with the anisole moiety, in place of the uracil ring, have no inhibition, most likely due to destabilization of strong H-bond interactions. Thus, if H-bond interaction is important for effective binding, a compound that will be able to form hydrogen bonds with D-196 and N-255 would bind better; this is seen with the adenine substitution. Some hydrogen bonding is present, and there is some inhibition observed; however, binding is not as tight as with uracil (possibly due to

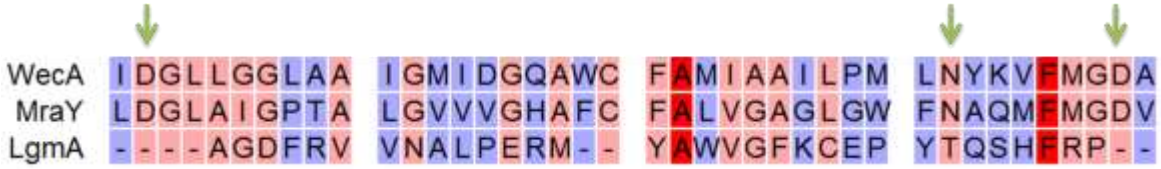
positioning of the molecules, or size), and thus inhibition is not as effective as seen with uracil.

These results give us a significantly better insight into the inhibition of WecA, and other similar enzymes that have been shown to be susceptive to tunicamycin and tunicamycin-derived inhibitors. It is known that certain features of the structure may be important for effective inhibition: the sugar moiety (that may mimic the sugar substrate), the uridine moiety (that may bind where uridine diphosphate (UDP) from the UDP-linked sugar binds), and the isoprene chain. We have been able to confirm here the importance of the uridine moiety; most importantly, that the uracil ring is important for binding. We were not able to conclude the role of the sugar substrate or the role of the isoprene in inhibitor binding, at least in terms of active site interactions analyzed with the paralogue. Further studies will require a closer look into how the sugar moieties or the isoprene chain/length play a role in inhibitor binding. It may be possible that while they may not be directly binding to the active site, they may provide some structural characteristics to the inhibitor that facilitate better binding. Our results may contribute to the design of more effective inhibitors for this class of enzymes.

3.8.1 LgmA Structure and Fold May Differ From Other GTs of Its Superfamily

Based on our results and structural analysis, we can further posit that the structure and fold of LgmA (and possibly other GlcNAc transferases), is not very similar to GlcNAc-P transferases, particularly in the active or substrate-binding site. From sequence comparison, we can make additional assumptions: although we have shown that LgmA, like WecA and MraY, also requires a divalent metal, we note that its active site is different from that of WecA. The D-265 required for Mg binding in WecA and MraY, is

not found in LgmA. Additionally, the D-196 and N-255 required for uracil binding in WecA and MraY is not conserved in LgmA: D-196 is not present, and N-255 is replaced by a threonine in LgmA. This makes sense with our data, and we can conclude that due to differences in make-up of active site and substrate-binding site residues, LgmA (and similar GlcNAc transferases) will not be inhibited in the same manner as WecA (and other similar GlcNA-1-P transferases).



3.9 LgmC Activity

3.9.1 LgmC Does Not Have Deacetylase Activity

LgmC was predicted to be a deacetylase based on its sequence; it has been proposed that it removes the acetyl group from the GlcNAc moiety after it has been transferred by LgmA to C_{55}-P. All assays testing LgmC as a deacetylase did not yield any activity under our reaction conditions. Various conditions were tested including different buffers and pHs, surfactant type and concentration, metals, protein from soluble and membrane fractions, as well as protein expressed in different cell lines. No deactylation activity was observed. Sample chromatograms are shown (Figure 27).

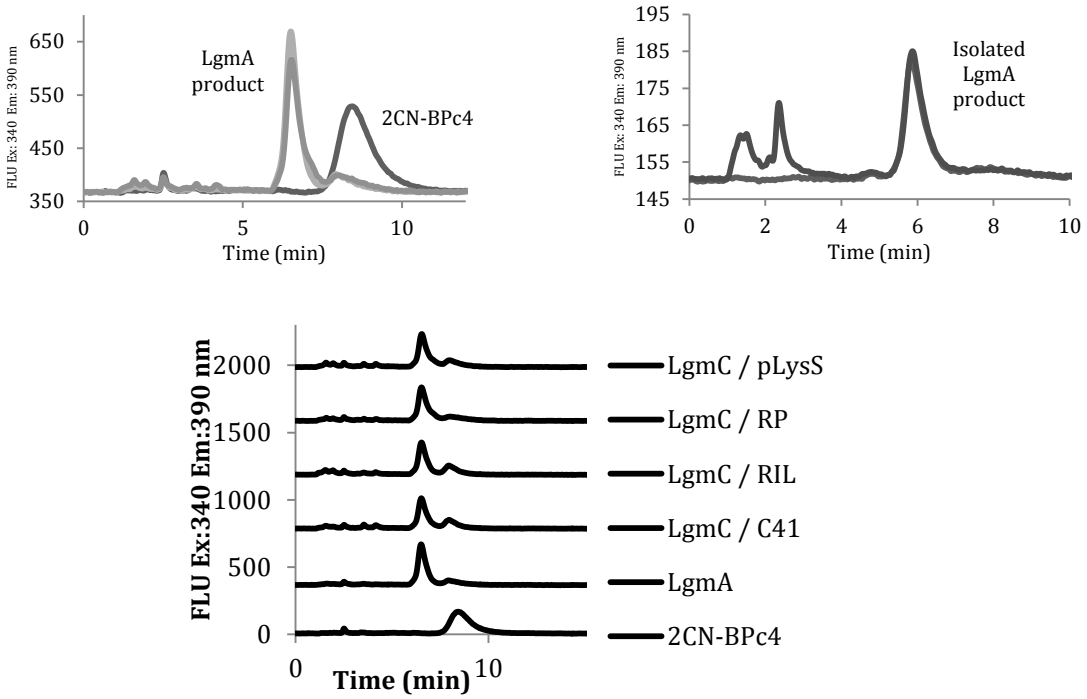


Figure 27: All assays in which LgmC was tested as a deacetylase showed no activity. In the data shown here, LgmC was added to the LgmA reaction (top left), or was tested independently with isolated LgmA product (top right). LgmC protein was expressed in various cell lines and tested with LgmA product.

3.9.2 LgmC Shows Activity – But Not as a Deacetylase

Upon subjecting the protein to a plethora of conditions, we were surprised to find that the protein actually had activity. In reaction conditions containing 50 mM MES buffer pH 6, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 28 mM sodium cholate, 1 µM 2CN-BPc7, 1 mM UDP-GlcNAc, LgmA membrane fraction, and LgmC membrane fraction, activity was observed (Figure 28). As LgmA had been added to the reaction, it initially seemed LgmC could have deacetylase activity after all. It was interesting to observe, although, that while the expected LgmC product (2CN-(c7)BP-GlcN) would have a different retention time than LgmA product (2CN-(c7)BP-GlcNAc), the difference was larger than expected.

To ensure activity was indeed being observed with LgmC, and not other particulates of the membrane fraction, control assays were run (Figure 29).

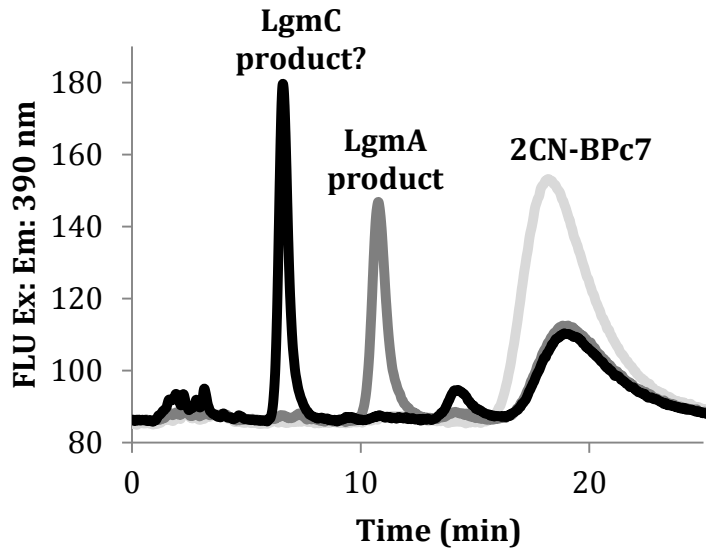


Figure 28: LgmC has activity. In reaction conditions containing 50 mM MES buffer pH 6, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 28 mM sodium cholate, 1 µM 2CN-BPc7, 1 mM UDP-GlcNAc and LgmC membrane fraction, activity was observed. The light grey peak represents the starting material 2CN-BPc7, with the black peak showing what may possibly be LgmC product. An LgmA reaction was run concurrently to show the difference in retention time for both products, shown in dark grey.

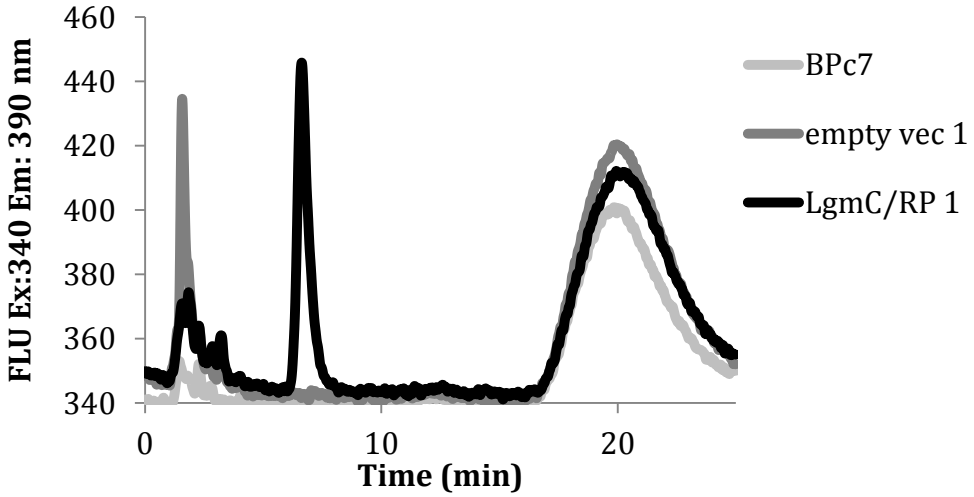


Figure 29: LgmC activity control assays. An empty vector (without LgmC gene) was expressed and its membrane fraction used as a negative control. The results show that the membrane fraction from the empty vector (dark grey) has no activity under the reaction conditions, while LgmC does (black).

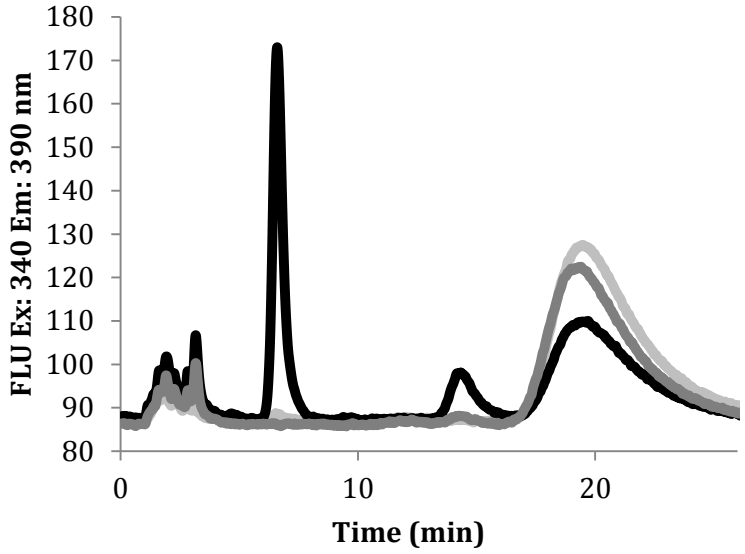


Figure 30: LgmC does not require LgmA for activity. Additional controls were run to identify what substrate LgmC required. LgmA reaction (light grey peak), did not show activity under conditions required for LgmC activity. LgmC did not have activity in the absence of UDP-GlcNAc (dark grey). LgmC showed activity in the presence of UDP-GlcNAc and LgmA (black).

3.9.3 LgmC Does Not Require LgmA for Activity

Additional control assays for LgmC revealed that even though LgmC required UDP-GlcNAc for activity, it did not actually require LgmA or LgmA product to function (Figure 30). In a reaction using the new conditions determined for LgmC function, LgmA was found to have no activity, even though LgmC product was still observed. This raised questions as to what exactly LgmC was doing, since it was not acting on LgmA product. New assay conditions were employed for LgmC in which no LgmA was added – LgmC function was not affected. Reaction progress was monitored in which formation of product was observed over time, as reactant decreased (Figure **31**). It was thus established that LgmC was not acting as a deacetylase (of LgmA product). We also noted that its product retention time was akin to what we usually see for glycosyltransferases.

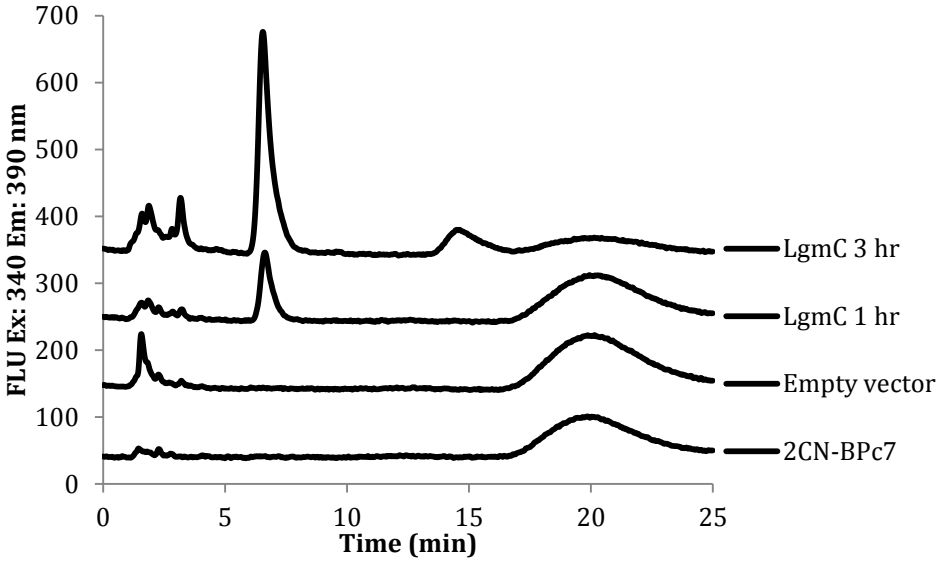


Figure 31: LgmC acts on UDP-GlcNAc. The empty vector control has no activity under LgmC reaction conditions. LgmC product formation is observed after one hour, and product formation is increased after three hours, while 2CN-BPc7 starting material decreases, indicating its conversion to product.

3.9.4 LgmC Seems To Be a GlcNAc-1-P Transferase

We suggest LgmC has glycosyltransferase activity. The first probability, and more intriguing prospect, is that it performs a combination of deacetylation and glycosyltransferase activity: while this falls in line with the Lipid A modification pathway proposed, it renders LgmA, which is supposed to initiate the pathway by transferring only GlcNAc, redundant. The second probability of LgmC function is that it is acting as a GlcNAc-1-P transferase, the family under which most GTs fall. To test this assumption, we assessed LgmC activity against WecA, the characterized GlcNAc-1-P transferase.

LgmC was re-expressed along with WecA. LgmC was mostly retained in the membrane fraction, as resolved by the Western blot (Figure 32). Although WecA was

successfully expressed, as seen in activity assays, we are still unable to resolve WecA by Western blot. Activity assays show LgmC product has similar retention time to WecA product (Figure 33). To further explore LgmC function, its activity was assessed against tunicamycin, which inhibits GlcNAc-1-P transferases. We found that LgmA activity was effectively inhibited by tunicamycin (Figure 34).

Our results support LgmC activity as a GlcNAc-1-P transferase. This is an interesting find, and incites future work into understanding more about LgmC activity. It is possible LgmC is a novel glycosyltransferase with dual activity, by performing both deacetylation and transfer of GlcNAc to C_{55}-P.

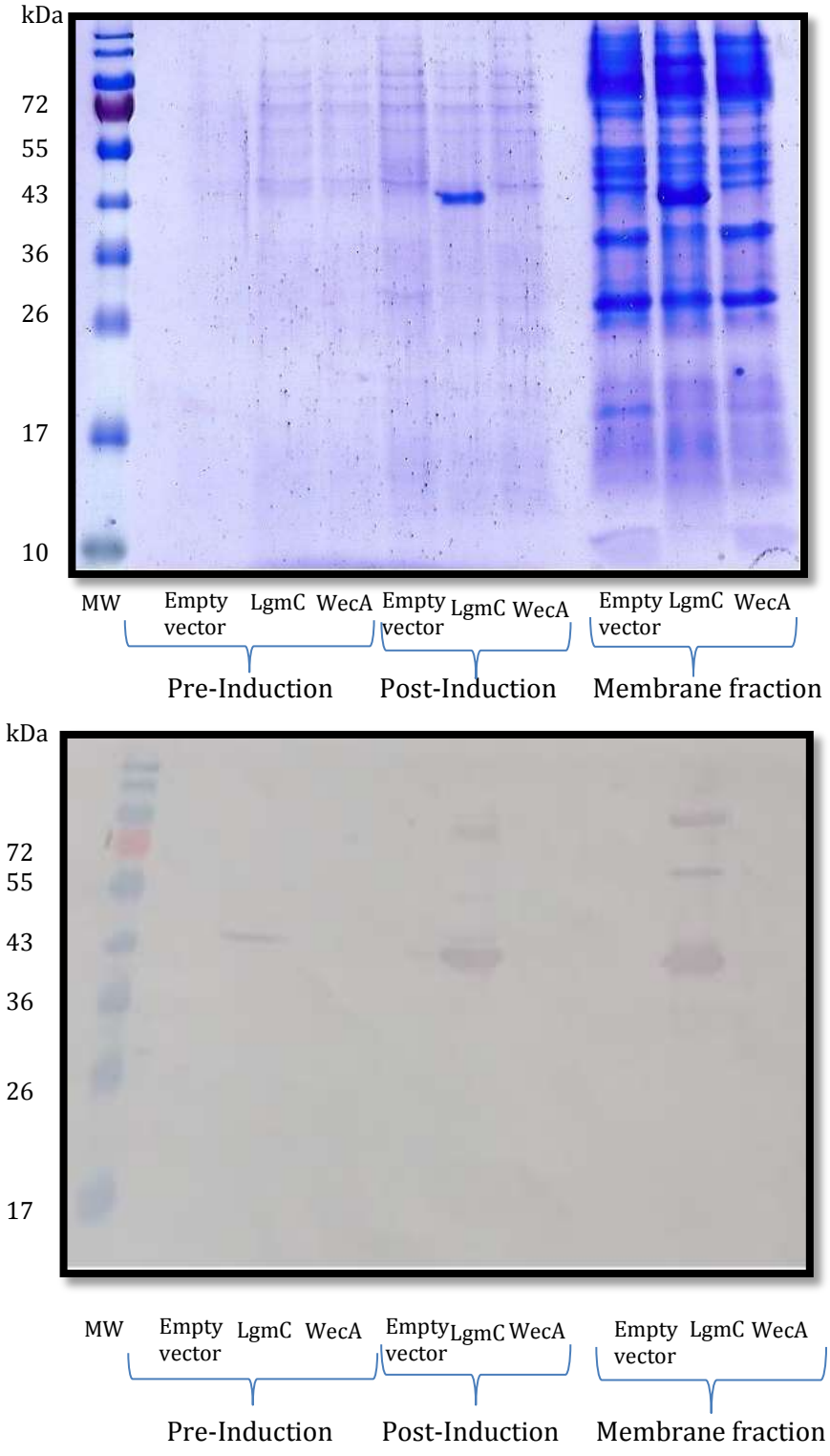


Figure 32: LgmC and WecA expression. LgmC was expressed to compare alongside WecA. However, WecA bands could not be resolved by Western blot.

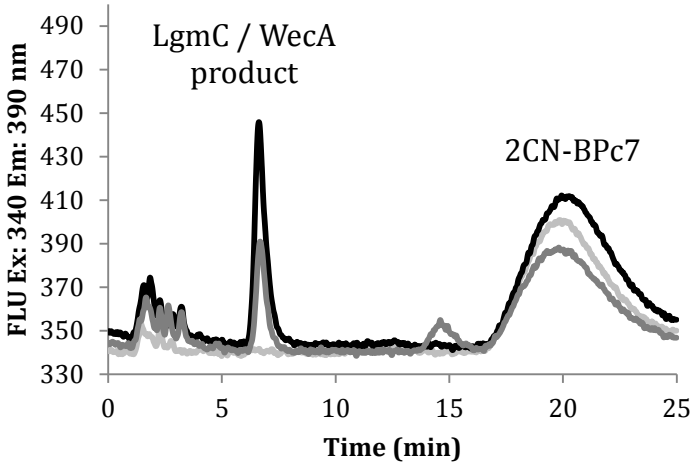


Figure 33: LgmC and WecA product have the same retention time, suggesting that both enzymes are forming the same product.

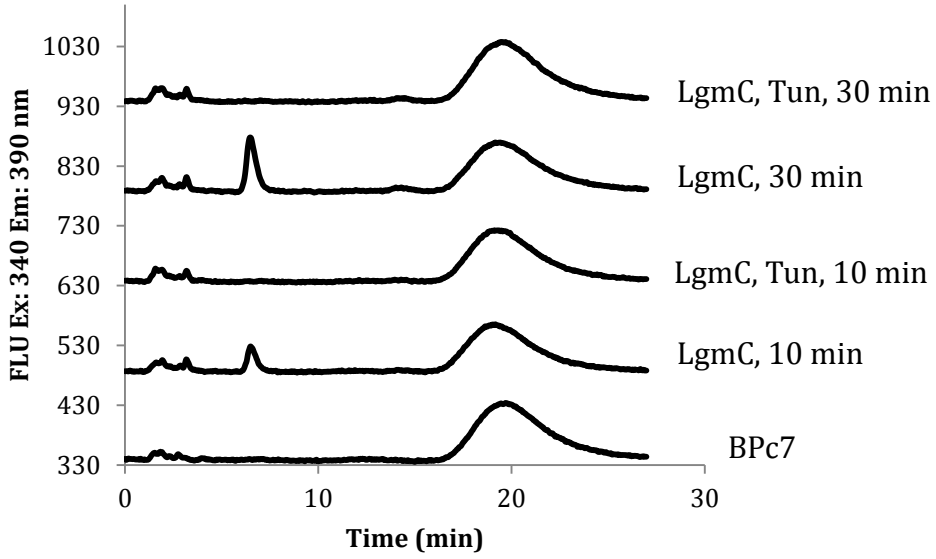


Figure 34: LgmC is inhibited by tunicamycin. In reactions that were allowed to go for 10 or 30 minutes before quenching with propanol, LgmC product formation is not observed.

3.10 Structural and computational studies of UPPS

Undecaprenyl diphosphate synthase (UPPS), a very promising antibacterial target, has been shown to have slight differences in activity and inhibition. As mentioned previously, recent studies have reported on inhibitors that affect some UPPS species, while others are unaffected. This knowledge prompted the study into attempting to understand the molecular bases on which UPPS proteins from different organisms exhibit differences in activity; these differences can be exploited in selective targeting of various UPPs enzymes. Herein, we summarize some computational and bioinformatics methods we implemented in the structural and computational studies of UPPS, and discuss the potential directions in which our preliminary data may lead.

3.10.1 UPPS Computational Studies Endeavor 1: Apply Distance Constraint Model (DCM) on conformational ensembles of UPPS to remove irrelevant conformations.

The structures of proteins fluctuate in various timescales and with various amplitudes [58]. Since these fluctuations are important for function, it is beneficial to complement the structural information, by accounting for their fluctuations. Conformational ensembles are powerful tools in representing the range of conformations that can be sampled by proteins. They have been used in studying fundamental properties of proteins such as the mechanism of molecular recognition, and the early stages of protein folding [59]. Although these ensembles do not generally report on protein dynamics, the can report on the amplitude of the dynamics, which has been found to be useful in analyzing protein behavior. To effectively study dynamics of UPPS activity, we turned to the use of conformational ensembles.

Various approaches have been proposed in the literature to generate conformational ensembles for proteins. For the purposes of studying the dynamics of UPPS with various possible inhibitors, we intended to do the following: 1) Start with known crystal structures. 2) Employ homology modeling and docking to generate putative all-atom structures. This was an important step because, of the hundreds of thousands of sequences available for UPPS proteins, there are only 36 crystal structures available, currently. For an in-depth study, 36 crystal structures would not be sufficient. Homology modeling would enable us to study those proteins of interest which had no structures available. Furthermore, to better visualize enzyme-substrate or enzyme-inhibitor interactions, we would employ docking techniques. The various conformations in combination with docking of various inhibitors would result in a plethora of conformational ensembles on which dynamic simulations would be based. Of course, not all the conformations will be relevant for studies. In the final step, we would apply the computational distance constraint model (DCM), which reliably predicts thermodynamic and mechanical properties, on multiple conformations to remove the irrelevant conformations for more efficient analysis.

To generate these conformations, we explored the efficiency of the Rosetta modeling software for our purposes [60]. We found it very capable, as it had a wide array of capabilities; we successfully carried out processes such as modeling mutations, and docking (Figure 35) as well as implementing the ‘backrub’ feature, in which a flexible backbone is employed as opposed to the more commonly used rigid backbone. We successfully generated 2,000 conformations for each of the 36 UPPS crystal structures

available (Figure 36). We did so for 3 temperatures: 25°C, 37°C, and 65°C, giving us 6,000 total conformations for each UPPS structure.

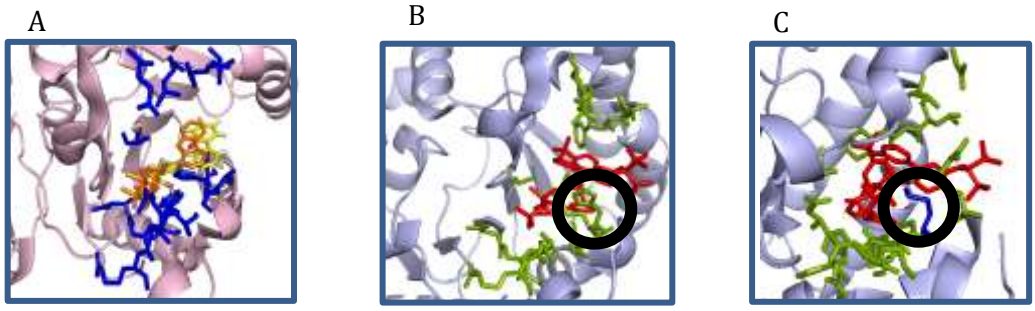


Figure 35: A) *E. coli* UPPS was successfully docked with an inhibitor (BPH-629) (yellow), and was found to be in the same location seen for the inhibitor (orange) in the original PDB structure (2E98). B) His43 (circled in black) was successfully mutated to Met (circled in black) (C) in *C. jejuni* UPPS.

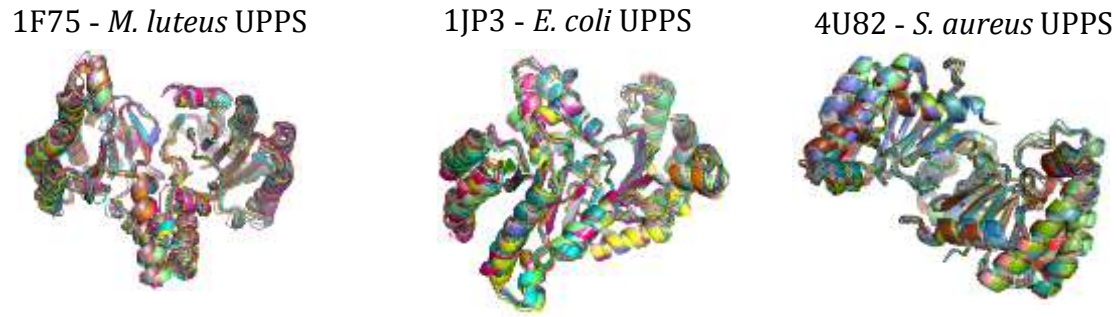


Figure 36: Sample of conformational ensembles. Each color represents a different conformation.

Having successfully generated conformations for future analysis, our focus shifted to deciding the most appropriate computational study for preliminary analysis. Conveniently, of the 36 UPPS crystal structures available in the PDB, 7 were from different species. This presented a good opportunity to study structure and activity of UPPS from different species without an immediate requirement for homology modeling.

3.10.2 UPPS Computational Studies Endeavor 2: Comparison of Quantitative Stability / Flexibility Relationships (QSFR) across various UPPS species

3.10.2.1 The dataset

In this study, we compare evolutionary conservation of QSFR properties of UPPS from five bacterial species specifically *E. coli* (1JP3), *M. luteus* (1F75), *H. pylori* (2D2R), *C. jejuni* (3UGS), and *S. pneumoniae* (4Q9O). As UPPS is known to be a functional dimer, the two other available structures were not included; *M. tuberculosis* has been reported (and is crystallized) as a functional tetramer [61], and *S. aureus*, although reported as a functional dimer, was present in the crystal structure as a monomer. It can be dimerized to be employed in futures studies, but we do not include it this time to maintain a fair playground for comparison. *M. tuberculosis* UPPS, on the other hand, is a unique enzyme that probably warrants an independent study. A superposition of the five representative structures is presented in Figure 37, highlighting the structural similarity; the adjacent diagram provides the CLC Sequence Viewer neighbor joining tree to reveal their sequence relationships.

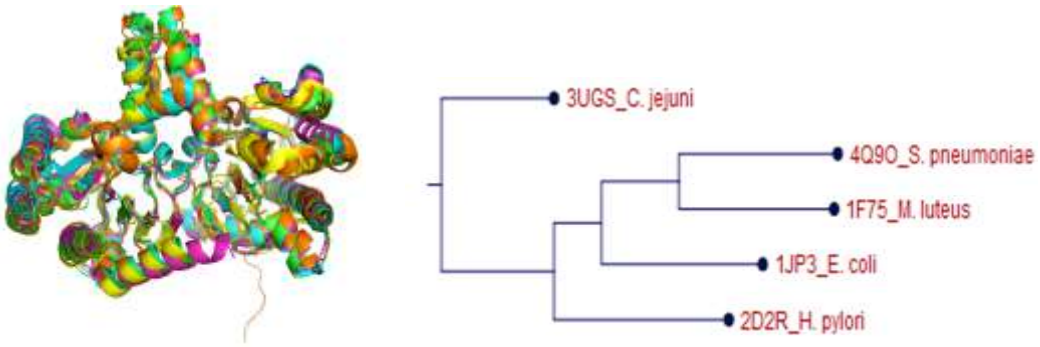


Figure 37: The five representative UPPS enzymes. (Left) Superposition of the E. coli (blue), M. luteus (green), H. pylori (pink), and C. jejuni (yellow), and S. pneumoniae (orange) structures. (Right) Neighbor joining tree of the five considered enzymes.

To further identify where each representative protein lies in the phylogeny of the cis-IPPS (cis-isoprenoid pyrophosphate synthase) superfamily, a detailed phylogenetic tree was generated (Figure 38). This evolutionary tree is used as a guideline for the analysis of quantitative stability/flexibility relationships (QSFR) between proteins through minimal Distance Constraint Model (mDCM).

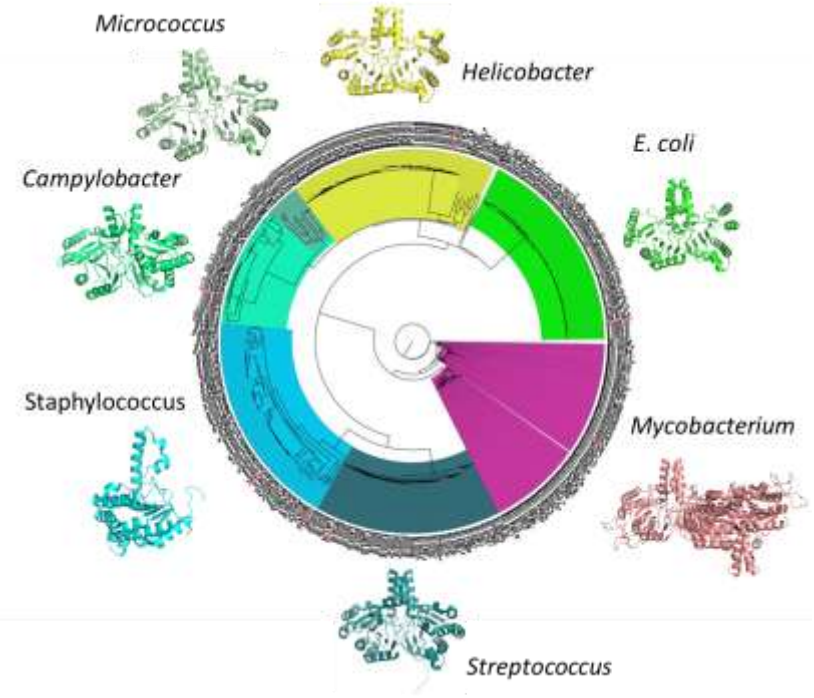


Figure 38: UPPS phylogenetic tree. A maximum likelihood family tree was generated using multiple sequence alignment results obtained from PSI-BLAST. The best scoring hits produced by PSI-BLAST were filtered using PISCES, and multiply aligned in MUSCLE. The resulting alignments were then improved using TrimAl for better analysis of phylogeny.

3.11 Slight differences in active site of UPPS from various species do not reveal much about slight differences in activity.

Active site residues were compared between the species. Unsurprisingly, most of the residues were conserved, especially in the substrate, farnesyl diphosphate (FPP),

binding region. There were a few differences noted; however, these did not give much insight into active site interaction with other inhibitors. Table 3 shows the alignment of active site residues. There are only a few residues that are not entirely conserved; even the side chain character is still retained. For instance, in column 10, *M. luteus* and *E. coli* have a valine, while *H. pylori*, *C. jejuni*, and *S. aureus* have a leucine, methionine, and isoleucine, respectively; even though the specific amino acid is not the same, the hydrophobicity of the side chain is retained. Thus, active site analysis alone does not allow for in-depth insight into slight differences in activity between the species.

Table 3: Alignment of active site residues from different UPPS species. Differences among species are shown in bold.

UPPS PDB ID	ACTIVE SITE RESIDUES													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1F75 *M. luteus*	D	G	N	G	R	R	H	G	**M**	**V**	A	**L**	**P**	**W**
1JP3 *E. coli*	D	G	N	G	R	R	H	G	**A**	**V**	A	**L**	**F**	**W**
2D2R *H. pylori*	D	G	N	G	R	R	H	G	**V**	**L**	A	**M**	**L**	**W**
3UGS *C. jejuni*	D	G	N	G	R	R	H	G	**V**	**M**	A	**K**	**R**	**F**
4Q9O *S. pneumoniae*	D	G	N	G	R	R	H	G	**M**	**L**	A	**L**	**P**	**W**
4U82 *S. aureus*	D	G	N	G	R	R	H	G	**M**	**I**	A	**N**	**W**	**W**

3.12 DCM calculates cooperativity correlation and flexibility index

In addition to the thermodynamic quantities, the mDCM calculates a number of mechanical properties from the set of Quantitative Stability/Flexibility Relationships (QSFR) metrics that are particularly useful: the Flexibility Index (FI), and Cooperativity

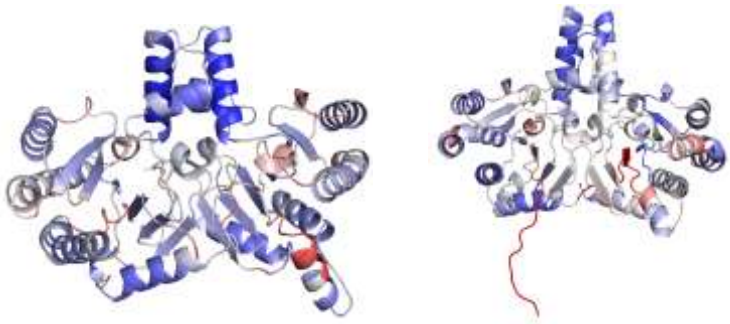
Correlation (CC). The flexibility index describes backbone flexibility - positive flexibility index values quantify the number of degrees of freedom within a local region, while negative FI values quantify the number of redundant constraints. The backbone is said to be *isostatically* rigid when FI = 0. Cooperativity Correlation (CC) depicts the complete set of residue-to-residue couplings. CC is described by an N x N matrix, where N is the number of residues in the protein. Each pixel on the CC plot is colored based on the correlations between residues. When it is colored blue, it indicates co-rigidity, whereas red means they are flexibly correlated. White indicates no correlation, although it does not necessarily imply rigidity or flexibility [62].

3.12.1 Differences in cooperativity correlation and flexibility index comparison across UPPS species

Although many differences were not observed in sequence and crystal structure analysis, substantial differences were noted in the flexibility dynamics results from DCM (Figure 39). Interestingly, flexibility seems to be conserved in the active site regions for all the proteins. Most of the alpha-helices seem to be rigid (colored blue), while the beta-sheets in the core of the protein have more flexibility (colored red). What this may suggest is that, in addition to the conserved active site residues, dynamics in and around the active site play a particular role in dictating which molecules the enzyme will successfully bind. As such, with further analysis of the residues in the flexible regions, we may find clues in dynamics that translate to slight differences in activity across species. We further observe in the results that UPPS from *M. luteus* and *E. coli* have more flexibility (at estimated melting temperature – Tm). *C. jejuni* and *H. pylori* have more rigidity in the overall structure. These results support that the flexibility in some of

these proteins may allow interaction with a more diverse range of inhibitors; while we cannot confirm this for *M. luteus*, this has been observed in *E. coli* UPPS, which has already been crystallized with a range of inhibitors. . We may even guess that due to the similarities in QSFR properties, *E. coli* and *M. luteus* may exhibit similar trends in activity, at least compared to the other species shown.

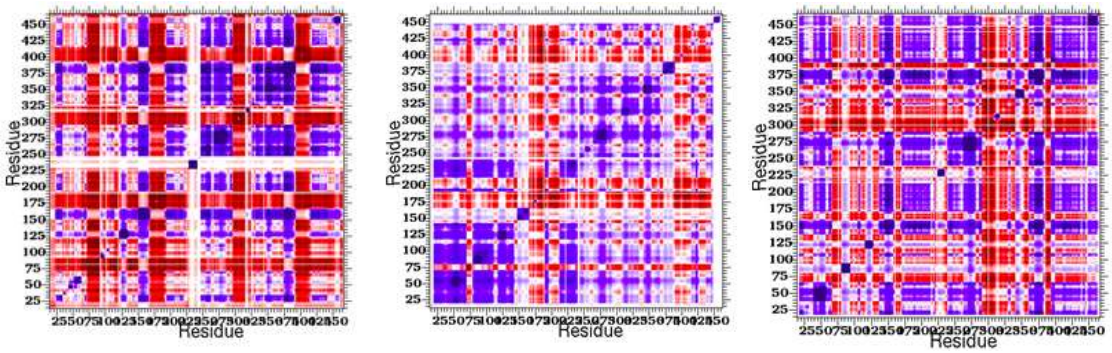
1F75 – *M. luteus* 1JP3 – *E. coli* 2D2R – *H. pylori*



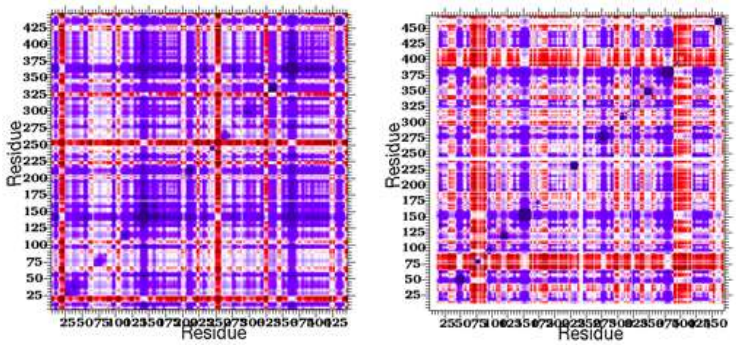
3UGS – *C. jejuni* 4Q9O – *S. pneumoniae*

Figure 39: Protein flexibility at estimated melting temperature mapped onto crystal structure. Blue indicates rigid areas and red indicates flexible areas.

Further differences are observed in the co-rigidity / co-flexibility CC plots (Figure 40), where differences among species are very pronounced. In *M. luteus*, we see large areas of red, indicating a high number of co-flexibility between various residues. Another interesting observation we can make is, even though *E. coli* appears to have more backbone flexibility than *H. pylori* (as seen in the flexibility index), it seems to have more regions of correlated rigidity than *H. pylori*. Additionally, *C. jejuni* retains much correlated rigidity, as seen in the flexibility index as well.



1F75 – *M. luteus* 1JP3 – *E. coli* 2D2R – *H. pylori*



3UGS – *C. jejuni* 4Q9O – *S. pneumoniae*

Figure 40: Cooperativity Correlation (CC) plots depicting the correlations between residues. Blue indicates co-rigidity, red means they are flexibly correlated. White indicates no correlation, although it does not necessarily imply rigidity or flexibility.

Going forward, these experiments will need to be repeated to ensure QSFR results are reproducible for each structure. Also, it will be important to determine the effect of mechanical perturbations on protein dynamics; e.g. how disruption of the correlated flexibility/rigidity relationships impact overall dynamics. Furthermore, employing molecular dynamics simulations and assessment of activity in various conformations will supplement structural studies; this will likely prove powerful tools in better understanding UPPS activity, inhibition, and dynamics, and ultimately speed up the drug discovery process.

CHAPTER 4: CONCLUSIONS AND FUTURE WORK

4.1 LgmA and WecA

LgmA was confirmed to be a GlcNAc transferase as predicted. This sets it apart from the more characterized initiating hexosyltransferases that transfer GlcNAc-1-P, like WecA. Unlike other GTs which have been found to exhibit poly-specificity, LgmA has been shown to have a high specificity for UDP-GlcNAc (with a K_m of 0.013 mM against WecA, with a K_m of 0.3 mM). Taken together, our results support the implication of LgmA in the Lipid A modification pathway with glucosamine in *B. pertussis.* With such high specificity for UDP-GlcNAc and high affinity for the substrate, it makes sense that it would be found in such a pathway.

We further determined that although both enzymes require GlcNAc for function, they are not both inhibited in the same manner. Based on sequence comparisons of the predicted active sites, the D-265 required for Mg binding in WecA and MraY, is not found in LgmA. Additionally, the D-196 and N-255 required for uracil binding in WecA and MraY is not conserved in LgmA: D-196 is not present, and N-255 is replaced by a threonine in LgmA. Thus, we can conclude that due to differences in make-up of active site and substrate-binding site residues, LgmA (and similar GlcNAc transferases) will not be inhibited in the same manner as WecA (and other similar GlcNA-1-P transferases). Seeing as the presence of these residues were critical for effective tunicamycin inhibition in WecA, we reason that, for GlcNAc-1-P inhibition, the uracil moiety (in tunicamycin) is very important to binding activity; these results may be useful in subsequent design of

phosphoglycosyltransferase (or glycosyltransferase) inhibitors. Additionally, as there were not obvious interactions in the substrate-binding site that indicated how the enzymes discriminated between different kinds of sugars, future inhibition studies might explore the effect of other non-GlcNAc moieties on these enzymes.

4.2 LgmC

LgmC was predicted to be a deacetylase, but was found to have some GlcNAc-1-P transferase activity instead. As a potential novel phosphoglycosyltrnasferase, its activity warrants further investigation. It is not entirely impossible that it is both deacetylating and transferring UDP-GlcNAc at the same time. Its prospects as a dual-activity protein is indeed intriguing and can be further investigated. Elucidating its function will also address the unanswered questions of the Lipid A modification pathway in *B. pertussis*. We have showed that LgmA is likely a member of the pathway; nonetheless, the function observed in LgmC renders LgmA activity redundant.

4.3 UPPS

UPPS from various bacterial species have been shown to have slight differences in activity. With the goal of improving our understanding of UPPS protein dynamics to facilitate discovery and design of antibacterials, we compared the dynamical properties of five UPPS proteins. It was interesting to note that there were significant differences in backbone flexibility, as well as correlated flexibility/rigidity across UPPS from different species. These are results that would otherwise not be obvious from simple sequence and structural studies. This preliminary data, therefore, reveals many directions in which computational UPPS studies could be directed for drug design purposes. There is a number of possible future work for UPPS studies. First, and more directly related to this

study, is a more in-depth analysis of dynamics across families; as we have shown that there are major differences in flexibility, further comparisons need to be made to identify regions of the enzymes most affected, and how this impacts overall activity. Another future direction would be to compare dynamics between the monomer and dimer – UPPS has been shown to be functional in the dimeric form, but not the monomeric form. Differences in the dynamics between the two might indicate how the dimer is active over the monomer. Finally, as UPPS has been further shown to have both a bound and unbound conformations, understanding the dynamics in this area will be extremely important for understanding conformational changes upon binding of an inhibitor. Understanding how this differs from species to species may be one of the most valuable pieces of information that can be gleaned from this study. Additionally, employing molecular dynamics simulations may prove to be a powerful tool in combination with the aforementioned studies. The preliminary data obtained from this work may serve as a platform on which to base future analysis of various other UPPS species.

REFERENCES

1. Kaiser, G. *The Gram-Negative Cell Wall. The Prokaryotic Cell: Bacteria*.

2. Ogawa, T.A., Y. Makimura, Y. Tamai, R., *Chemical structure and immunobiological activity of Porphyromonas gingivalis lipid A.* Frontiers in Bioscience, 2007(12): p. 3795-812.

3. Wiese, A., et al., *The dual role of lipopolysaccharide as effector and target molecule.* Biol Chem, 1999. **380**(7-8): p. 767-84.

4. Rietschel, E.T., et al., *Bacterial endotoxin: molecular relationships of structure to activity and function.* FASEB J, 1994. **8**(2): p. 217-25.

5. Christie, W.W. *Lipid A and bacterial polysaccharides*. The American Oil Chemists' Society (AOCS) Lipid Library; Available from: lipidlibrary.aocs.org

6. Kabanov, D.S. and I.R. Prokhorenko, *Structural analysis of lipopolysaccharides from Gram-negative bacteria.* Biochemistry (Mosc), 2010. **75**(4): p. 383-404.

7. Raetz, C.R., et al., *Discovery of new biosynthetic pathways: the lipid A story.* J Lipid Res, 2009. **50 Suppl**: p. S103-8.

8. Trent, M.S., *Biosynthesis, transport, and modification of lipid A.* Biochem Cell Biol, 2004. **82**(1): p. 71-86.

9. Gunn, J.S., et al., *PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance.* Mol Microbiol, 1998. **27**(6): p. 1171-82.

10. Ernst, R.K., et al., *Specific lipopolysaccharide found in cystic fibrosis airway Pseudomonas aeruginosa.* Science, 1999. **286**(5444): p. 1561-5.

11. Guo, L., et al., *Regulation of lipid A modifications by Salmonella typhimurium virulence genes phoP-phoQ.* Science, 1997. **276**(5310): p. 250-3.

12. Guo, L., et al., *Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides.* Cell, 1998. **95**(2): p. 189-98.

13. Vaara, M., et al., *Characterization of the lipopolysaccharide from the polymyxin-resistant pmrA mutants of Salmonella typhimurium.* FEBS Lett, 1981. **129**(1): p. 145-9.

14. Vaara, M., *Agents that increase the permeability of the outer membrane.* Microbiol Rev, 1992. **56**(3): p. 395-411.

15. Helander, I.M., I. Kilpelainen, and M. Vaara, *Increased substitution of phosphate groups in lipopolysaccharides and lipid A of the polymyxin-resistant pmrA mutants of Salmonella typhimurium: a 31P-NMR study.* Mol Microbiol, 1994. **11**(3): p. 481-7.

16. Nummila, K., et al., *Lipopolysaccharides of polymyxin B-resistant mutants of Escherichia coli are extensively substituted by 2-aminoethyl pyrophosphate and contain aminoarabinose in lipid A.* Mol Microbiol, 1995. **16**(2): p. 271-8.

17. Marr, N., et al., *Substitution of the Bordetella pertussis lipid A phosphate groups with glucosamine is required for robust NF-kappaB activation and release of proinflammatory cytokines in cells expressing human but not murine Toll-like receptor 4-MD-2-CD14.* Infect Immun, 2010. **78**(5): p. 2060-9.

18. Shah, N.R., R.E. Hancock, and R.C. Fernandez, *Bordetella pertussis lipid A glucosamine modification confers resistance to cationic antimicrobial peptides and increases resistance to outer membrane perturbation.* Antimicrob Agents Chemother, 2014. **58**(8): p. 4931-4.

19. Shah, N.R., et al., *Minor modifications to the phosphate groups and the C3' acyl chain length of lipid A in two Bordetella pertussis strains, BP338 and 18-323, independently affect Toll-like receptor 4 protein activation.* J Biol Chem, 2013. **288**(17): p. 11751-60.

20. Llewellyn, A.C., et al., *NaxD is a deacetylase required for lipid A modification and Francisella pathogenesis.* Mol Microbiol, 2012. **86**(3): p. 611-27.

21. Breton, C. and A. Imberty, *Structure/function studies of glycosyltransferases.* Curr Opin Struct Biol, 1999. **9**(5): p. 563-71.

22. Lairson, L.L., et al., *Glycosyltransferases: structures, functions, and mechanisms.* Annu Rev Biochem, 2008. **77**: p. 521-55.

23. Henrich, E., et al., *Lipid Requirements for the Enzymatic Activity of MraY Translocases and in Vitro Reconstitution of the Lipid II Synthesis Pathway.* J Biol Chem, 2016. **291**(5): p. 2535-46.

24. Crouvoisier, M., D. Mengin-Lecreulx, and J. van Heijenoort, *UDP-N-acetylglucosamine:N-acetylmuramoyl-(pentapeptide) pyrophosphoryl undecaprenol N-acetylglucosamine transferase from Escherichia coli: overproduction, solubilization, and purification.* FEBS Lett, 1999. **449**(2-3): p. 289-92.

25. Al-Dabbagh, B., D. Mengin-Lecreulx, and A. Bouhss, *Purification and characterization of the bacterial UDP-GlcNAc:undecaprenyl-phosphate GlcNAc-1-phosphate transferase WecA.* J Bacteriol, 2008. **190**(21): p. 7141-6.

26. Mostafavi, A.Z., et al., *Fluorescent probes for investigation of isoprenoid configuration and size discrimination by bactoprenol-utilizing enzymes.* Bioorg Med Chem, 2013. **21**(17): p. 5428-35.

27. Amer, A.O. and M.A. Valvano, *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**(Pt 11): p. 3015-25.

28. Chung, B.C., et al., *Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis.* Science, 2013. **341**(6149): p. 1012-6.

29. Lehrman, M.A., *A family of UDP-GlcNAc/MurNAc: polyisoprenol-P GlcNAc/MurNAc-1-P transferases.* Glycobiology, 1994. **4**(6): p. 768-71.

30. Lehrer, J., et al., *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): p. 2618-28.

31. Azzouz, N., et al., *Regulation of Paramecium primaurelia glycosylphosphatidyl-inositol biosynthesis via dolichol phosphate mannose synthesis.* Biochimie, 2001. **83**(8): p. 801-9.

32. Oriol, R., et al., *Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate.* Mol Biol Evol, 2002. **19**(9): p. 1451-63.

33. Kapitonov, D. and R.K. Yu, *Conserved domains of glycosyltransferases.* Glycobiology, 1999. **9**(10): p. 961-78.

34. Naumoff, D.G., *Hierarchical classification of glycoside hydrolases.* Biochemistry (Mosc), 2011. **76**(6): p. 622-35.

35. Marchler-Bauer, A., et al., *CDD: NCBI's conserved domain database.* Nucleic Acids Res, 2015. **43**(Database issue): p. D222-6.

36. Rip, J.W., et al., *Distribution, metabolism and function of dolichol and polyprenols.* Prog Lipid Res, 1985. **24**(4): p. 269-309.

37. Quellhorst, G.J., Jr., et al., *Identification of Schizosaccharomyces pombe prenol as dolichol-16,17.* Biochem Biophys Res Commun, 1998. **244**(2): p. 546-50.

38. Fujihashi, M., et al., *Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase.* Proc Natl Acad Sci U S A, 2001. **98**(8): p. 4337-42.

39. El Ghachi, M., et al., *The bacA gene of Escherichia coli encodes an undecaprenyl pyrophosphate phosphatase activity.* J Biol Chem, 2004. **279**(29): p. 30106-13.

40. Ko, T.P., et al., *Mechanism of product chain length determination and the role of a flexible loop in Escherichia coli undecaprenyl-pyrophosphate synthase catalysis.* J Biol Chem, 2001. **276**(50): p. 47474-82.

41. Pan, J.J., L.W. Yang, and P.H. Liang, *Effect of site-directed mutagenesis of the conserved aspartate and glutamate on E. coli undecaprenyl pyrophosphate synthase catalysis.* Biochemistry, 2000. **39**(45): p. 13856-61.

42. Kharel, Y., et al., *Identification of Significant residues for homoallylic substrate binding of Micrococcus luteus B-P 26 undecaprenyl diphosphate synthase.* J Biol Chem, 2001. **276**(30): p. 28459-64.

43. Guo, R.T., et al., *Bisphosphonates target multiple sites in both cis- and trans-prenyltransferases.* Proc Natl Acad Sci U S A, 2007. **104**(24): p. 10022-7.

44. Zhu, W., et al., *Antibacterial drug leads targeting isoprenoid biosynthesis.* Proc Natl Acad Sci U S A, 2013. **110**(1): p. 123-8.

45. Peukert, S., et al., *Design and structure-activity relationships of potent and selective inhibitors of undecaprenyl pyrophosphate synthase (UPPS): tetramic, tetronic acids and dihydropyridin-2-ones.* Bioorg Med Chem Lett, 2008. **18**(6): p. 1840-4.

46. Kuo, C.J., et al., *Structure-based inhibitors exhibit differential activities against Helicobacter pylori and Escherichia coli undecaprenyl pyrophosphate synthases.* J Biomed Biotechnol, 2008. **2008**: p. 841312.

47. Sinko, W., et al., *Applying molecular dynamics simulations to identify rarely sampled ligand-bound conformational states of undecaprenyl pyrophosphate synthase, an antibacterial target.* Chem Biol Drug Des, 2011. **77**(6): p. 412-20.

48. Teng, K.H., et al., *Fluorescent substrate analog for monitoring chain elongation by undecaprenyl pyrophosphate synthase in real time.* Anal Biochem, 2011. **417**(1): p. 136-41.

49. Teng, K.L., P., *Undecaprenyl diphosphate synthase, a cis-prenyltransferase synthesizing lipid carrier for bacterial cell wall biosynthesis.* Molecular Membrane Biology, 2012. **29**(7): p. 267-273.

50. Jacobs, D.J., et al., *Elucidating quantitative stability/flexibility relationships within thioredoxin and its fragments using a distance constraint model.* J Mol Biol, 2006. **358**(3): p. 882-904.

51. Jacobs, D.J. and S. Dallakyan, *Elucidating protein thermodynamics from the three-dimensional structure of the native state using network rigidity.* Biophys J, 2005. **88**(2): p. 903-15.

52. Livesay, D.R., et al., *A flexible approach for understanding protein stability.* FEBS Lett, 2004. **576**(3): p. 468-76.

53. Verma, D., et al., *Towards comprehensive analysis of protein family quantitative stability-flexibility relationships using homology models.* Methods Mol Biol, 2014. **1084**: p. 239-54.

54. *TMHMM Server v. 2.0. Center for Biological sequence analysis. Technical University of Denmark.*

55. Tarbouriech, N., S.J. Charnock, and G.J. Davies, *Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases.* J Mol Biol, 2001. **314**(4): p. 655-61.

56. Li, J. and B. Yu, *A modular approach to the total synthesis of tunicamycins.* Angew Chem Int Ed Engl, 2015. **54**(22): p. 6618-21.

57. Price, N.P. and B. Tsvetanova, *Biosynthesis of the tunicamycins: a review.* J Antibiot (Tokyo), 2007. **60**(8): p. 485-91.

58. Karplus, M. and J. Kuriyan, *Molecular dynamics and protein function.* Proc Natl Acad Sci U S A, 2005. **102**(19): p. 6679-85.

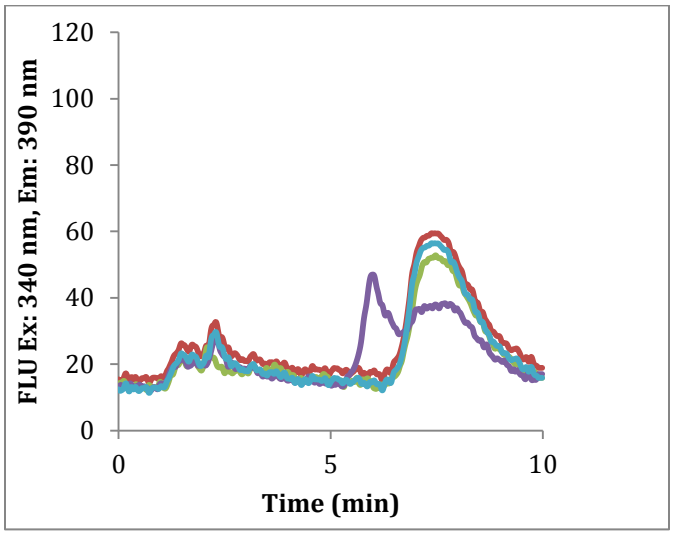
59. Salvatella, X., *Understanding protein dynamics using conformational ensembles*, in *Advances in Experimental Medicine and Biology*. 2013. p. 67-85.

60. Rohl, C.A., et al., *Protein structure prediction using Rosetta.* Methods Enzymol, 2004. **383**: p. 66-93.

61. Wang, W., et al., *The structural basis of chain length control in Rv1086.* J Mol Biol, 2008. **381**(1): p. 129-40.

62. Brown, M.C., et al., *A case study comparing quantitative stability-flexibility relationships across five metallo-beta-lactamases highlighting differences within NDM-1.* Methods Mol Biol, 2014. **1084**: p. 227-38.

APPENDIX A: CONTROL ASSAYS FOR LGMA ACTIVITY



To ensure that activity being observed was in fact due to LgmA enzymatic activity, control assays were run under reaction conditions with: 1) no LgmA membrane fraction; 2) no UDP-GlcNAc; 3) blank membrane fraction (without LgmA gene). No activity was monitored under any of the control reactions. Only the reaction containing the required substrates and LgmA membrane fraction resulted in a peak shift indicating product turnover.