CHARACTERIZATION OF THE BIOSYNTHETIC PATHWAY OF CAPSULAR POLYSACCHARIDE A IN *BACTEROIDES FRAGILIS*

by

Sunita Sharma

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Biology

Charlotte

2019

Approved by:

_____
Dr. Jerry M. Troutman

_____
Dr. Todd Steck

_____
Dr. Joanna Krueger

_____
Dr. Daniel Nelson

_____
Dr. Irina Nesmelova

ABSTRACT

SUNITA SHARMA. Characterization of the biosynthetic pathway of capsular polysaccharide A in *Bacteroides fragilis* (Under the direction of DR. JERRY M. TROUTMAN)

Capsular Polysaccharide A (CPSA) is a polymer of a tetrasaccharide unit found on the surface of the symbiotic gut bacteria *Bacteroides fragilis*. It has been suggested that CPSA is vital for maintaining a normal balance between the levels of Th1 and Th2 cells in a normal mammalian immune system. The gene locus responsible for CPSA biosynthesis has been identified. The operon has been proposed to encode one glycosyl-1-phosphate transferase (WcfS) and three glycosyltransferases (WcfN, -P and -Q), three sugar modifying enzymes (WcfM, WcfR and WcfO), a flippase (Wzx) and a polysaccharide polymerase (Wzy) based on bioinformatics tools. A route for the biosynthesis of CPSA is proposed. The initiating sugar transferase, WcfS has been previously identified and characterized. An *in vitro* method was used to enzymatically synthesize CPSA, which was assembled on a fluorescent analogue of the native bactoprenyl diphosphate anchor one sugar at a time. Utilizing reverse phase HPLC, functional characterization of the remaining glycosyltransferases, WcfN, WcfP and WcfQ involved in the CPSA biosynthesis was done. Also function of the hypothesized pyruvyltransferase WcfO was determined. This is the first study to characterize a pyruvyltransferase involved in polysaccharide biosynthesis from a prokaryote. The putative galactopyranose mutase WcfM was also characterized. The biosynthesis of the polysaccharide was also achieved in a single pot, compared to the multiple steps involved in chemical synthesis, displaying an enormous leap in the biosynthesis of complex molecules like CPSA.

ACKNOWLEDGMENT

Foremost, I would like to acknowledge the past members of the Troutman. Specifically, Christina Martinez and Vanessa Amabo for synthesizing the bactoprenyl analogues that I used in characterizing the various enzymes. Katelyn Erickson for characterizing the reaction products and helping me immensely with mass spectrometry. Your help has been invaluable. I also want to thank Dr. Todd L. Lowary for providing the lab with UDP-galactofuranose, which helped in characterizing the mutase enzyme. I would also like to acknowledge the funding received for the project from NIH to Dr. Troutman. The funding helped me to focus on the project full time. I would like to thank my advisor Dr. Troutman, for believing in me and taking me in his lab when I was just a novice student. You have helped me learn a lot. I would also like to thank my committee members, for moulding me to a better science person. Lastly, I would like to thank my husband who stuck by me through all times, and helped me see the light at the end of the tunnel, and my brothers whose pep talks helped me through tough times. All of your support has been invaluable.

DEDICATION

Dedicated to all the Women aspiring to be in science, You can do it, as I could do it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AADGal	2-N-acetylamido-4-amino-6-deoxy-galactopyranose
B. fragilis	*Bacteroides fragilis*
BPP	bactoprenyl diphosphate
BP	bactoprenyl phosphate
C. jejuni	*Campylobacter jejuni*
CAZY	Carbohydrate-Active Enzymes database
CE	capillary electrophoresis
CPS	capsular polysaccharide
CPSA	capsular polysaccharide A
E. coli.	*Escherichia coli*
EAE	experimental autoimmune encephalomyelitis
ESI-MS	electrospray ionization mass spectrometry
HPLC	high performance liquid chromatography
HLA-DM	human leukocyte antigen DM
HLA-DR	human leukocyte antigen DR
HR-MS	high resolution mass spectrometry
GAG	glucosaminoglycan
Galf	galactofuranose
GalNAc	*N*-acetylgalactosesamine
GlcNAc	*N*-acetylglucosamine
Galp	galactopyranose
iNOS	inducible nitric oxide synthase

IPTG	isopropylthiogalactopyranoside
LB	Luria Bertani
LC/MS	Liquid chromatography mass spectrometry
MALDI-MS	Matrix assisted laser desorption/ionization mass spectrometry
MHCII	major histocompatibility class II
NMR	nuclear magnetic resonance spectroscopy
PglF	a dehydratase
P. aeruginosa	*Pseudomonas aeruginosa*
PHYRE^2	Protein Homology/analogy Recognition Engine v 2.0
R.leguminosarum	*Rhizobium leguminosarum*
S. pombe	*Schizosaccharomyces pombe*
TACA	tumor-associated carbohydrate antigen
TCR	T-cell receptor
TLC	Thin layer chromatography
UDP	uridine diphosphate
V. vulnificus	*Vibrio vulnificus*
2AA-BP	2-amideaniline bactoprenyl monophosphate
2CNA-BP	2-nitrileaniline bactoprenyl monophosphate
4,6-pyr-Gal	4,6-pyruvate-galactose
4,6-pyr-Glc	4,6-pyruvate-glucose

CHAPTER 1: INTRODUCTION

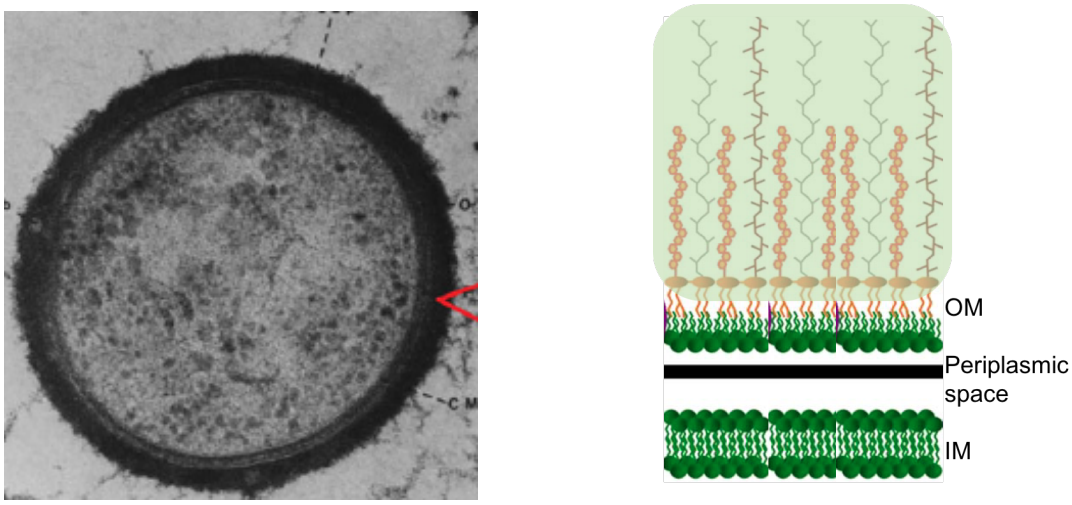
1.1 Bacterial Capsule Polysaccharides

A microbial cell surface is rich in various polysaccharides. They range from cell wall components such as mannan of yeast cells[1]; teichoic acid, lipopolysaccharides and peptidoglycan of bacterial cells[2]. Apart from the cell wall components, polysaccharides are found either associated with other cell surface molecules, or totally disassociated from the microbial cell, forming an amorphous layer of extracellular polysaccharide[3,4]. The extracellular polysaccharide in some bacteria exists in a well-organized layer, forming an outer enveloped layer around the bacteria, (Figure 1.1), also known as the capsule. It is found in both Gram-negative and Gram-positive bacteria[2].

Capsular polysaccharides are very diverse in their composition and structure[5]. They themselves as such do not contribute to the integrity of the microbial structure, but however form structures, that can be recognized. They are usually composed of repeating single units of common monosaccharides joined by glycosidic linkage[6,7], although most of the polysaccharides may contain one or more rare sugars[8,9]. The rare sugars may include L-hexoses, or furanose form of hexoses. The polysaccharide can be either a homo- or a heteropolymer. The sugar component of the polysaccharide varies also within the species of bacteria, which also determines their serological types[3,10].

Capsular polysaccharides play diverse roles in the bacterial physiology and ecology. They are beneficial to the bacteria in many ways. It protects the bacteria from

desiccation by forming a hydrated gel around the bacteria. For example, in *Escherichia coli*, desiccation initiates the expression of the genes encoding the enzymes for the biosynthesis of colanic acid[3]. Capsules also may help the organism to adhere to host cells or to each other, thereby facilitating and maintaining bacterial colonization of biological



a b

Figure 1.1: (a) Electron micrograph of a *Bacteroides fragilis* cell wall[11], showing capsular polysaccharides on its surface, marked by red arrow (The license to use the figure was granted by John Wiley and Sons; License number 4544010599387). (b) Cartoon representation of bacterial cell wall, showing the presence of capsular polysaccharide (light green) on the outer surface, though it is not sure how the polymer adheres to the cell surface.

and inanimate surfaces through formation of biofilms[12]. They serve as a barrier between the cell wall and the environment, and mediate host-pathogen interactions[13-15]. Cell surface polysaccharides are also responsible for virulence in bacteria, for example in *Vibrio vulnificus,* the capsule is responsible for its pathogenicity causing necrotizing wounds on the skin[16]. In some cases, they help in evading phagocytosis by migrating phagocytes and

tissue-fixed macrophages[17]. The more highly charged the capsular polysaccharide is, greater the degree of resistance to phagocytosis[17]. The presence of a capsule is actually thought to confer resistance to non-specific host defense mechanisms[18]. During invasive bacterial infections, interactions between the capsular polysaccharide and the host's immune system can decide the outcome of the infection[13,19]. In some cases, though, polysaccharides on these bacteria also may actually help in the development of the immune system in humans, as seen with the zwitterionic polysaccharides of type 1 capsular polysaccharide of *Streptococcus pneumoniae,* type 5 and 8 of *Staphylococcus aureus*, and capsular polysaccharide A (CPSA) of *Bacteroides fragilis*[20-23]. All of these have been shown to have some degree of immunomodulatory effects on the immune system, of which CPSA has been studied the most.

1.2 Immunomodulatory effects of CPSA

B. fragilis is an obligate anaerobic bacterium which colonizes the intestinal tract of the human gut, and essentially all other mammals. It is an integral component of the normal gastrointestinal flora[24,25]. It is classified as a Gram-negative, non-spore forming and anaerobic bacilli. This mammalian symbiont and opportunistic pathogen depends on its capsular layer for virulence as well as for symbiosis in the mammalian gut[26]. Eight capsule polysaccharides can be expressed on its surface, depending on the environmental niche of the organism, designated as CPSA through CPSH[27-29]. Capsular polysaccharide A is one of the eight polysaccharides found on the surface of *B. fragilis,* and is the most abundant. CPSA plays a role in abscess formation when the bacterium localizes outside of its normal niche in the gastrointestinal tract or during surgical procedures[30]. However, this view has been challenged when it was found that treating the animal with the CPSA and then

introducing the abscess-inducing bacteria resulted in the immune system of the animal protecting itself against the production of abscesses. Furthermore, few studies have also claimed that the abscess formation by *B. fragilis* actually prevents infection in the wound by other pathogenic bacteria[14,31,32].

CPSA is a unique polymer. It has both negatively and positively charged motifs present on each repeating monomer, making it a zwitterionic molecule[33,34] (Figure 1.2). The presence of this zwitterionic character has been attributed to the novel immunologic activity displayed by CPSA. The zwitterionic character has been shown to modulate the mammalian immune system by interacting with the adaptive immune system[35]. Elimination of either charge group in CPSA results in a lack of *in vivo* activation of the T-cells[36].

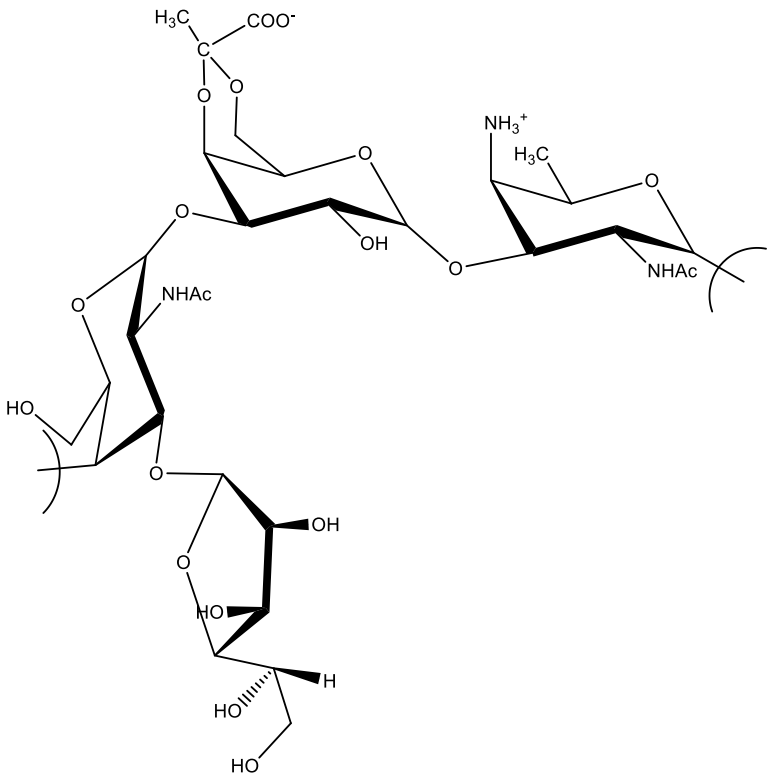


Figure 1.2: Tetrameric repeat unit of the CPSA found on *B. fragilis*. It consists of an acetamido-4-amino-6-deoxygalactopyranose (AADGal), 4,6-pyruvate galactose (4,6-pyr-Gal), N-acetylgalactosamine (GalNAc), and a galactofuranose (Gal*f*) sugar.

CPSA modulates the immune system by its stimulation of a T-cell dependent form of immunity that provides protection against the formation of the intraabdominal abscesses. At the molecular level, CPSA interacts with the MHCII pathway similar to traditional protein antigens[37]. The first step is endocytosis of CPSA by the antigen-presenting cells like dendritic cells. Once in the endosome, CPSA is depolymerized based on the chemical reaction, deaminative cleavage[38]. This cleaving is mediated by nitric oxide, that has been generated by the upregulation of inducible nitric oxide synthase (iNOS). The 130 kDa CPSA is processed to 15 kDa units. After being processed, the endosomes fuse with lysosomes and exocytic vesicles to form MIIC vesicle carrying HLA-DR and the accessory molecule HLA-DM. HLA-DM catalyzes the binding of MHCII to CPSA fragments, which is then presented to the CD4+ T cell receptor (Figure 1.3). This leads to the proliferation of the CD4+ T cell population, that produces IL-10, which is responsible for providing protection against the formation of intra-abdominal abscesses[35].

CPSA can restore the immune system from a variety of autoimmune disorders, making it a promising candidate for a therapeutic drug. Colonization of nude mice with wild type *B. fragilis*, that produces the zwitterionic capsular polysaccharide A, protected animals from antibiotic induced experimental autoimmune encephalomyelitis (EAE), while animals infected with mutant *B. fragilis* deficient in the production of the polysaccharide were not protected[31,39]. In germ free animal models of Inflammatory Bowel Disease (IBD), it was found that CPSA alone without the bacterial carrier was enough to stimulate normal immune system function and prevent intestinal inflammatory disease[39,40]. CPSA has been given therapeutically to decrease pro-inflammatory cytokine production in an experimental model of colonic irritation[13].

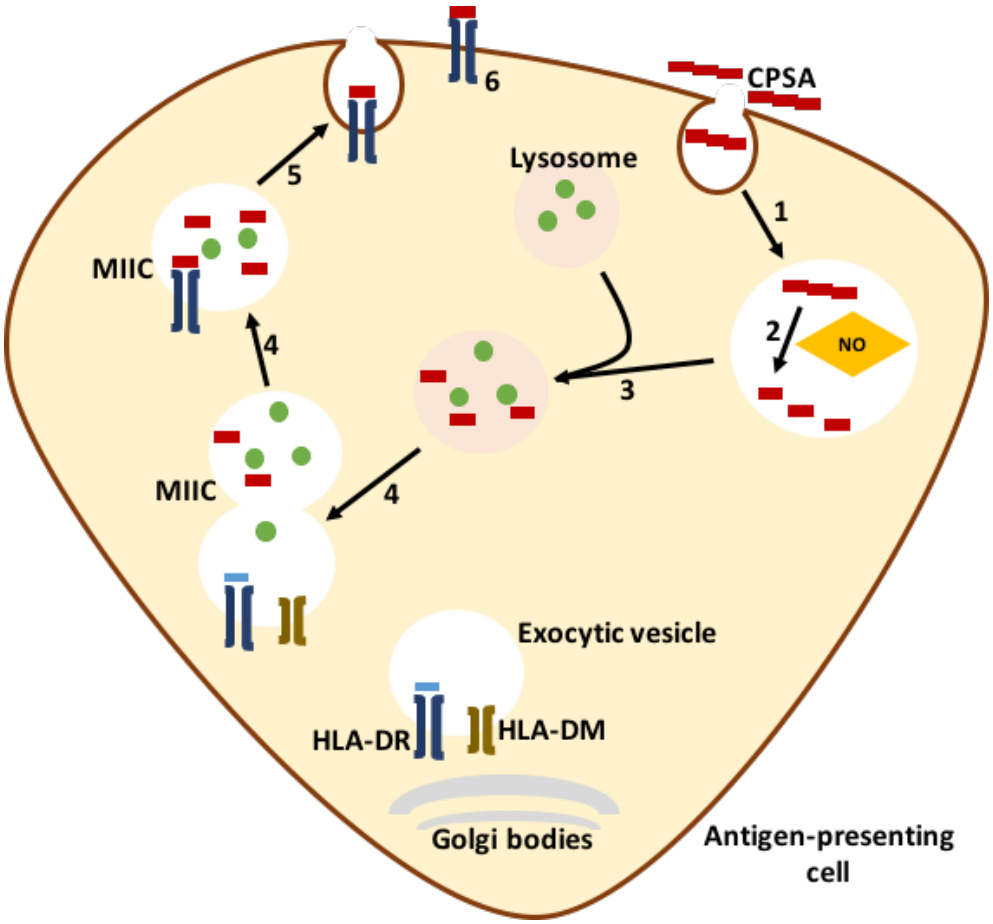


Figure 1.3 Depolymerization of CPSA in antigen presenting cell. 1. Internalization of CPSA in an endosome. 2. iNOS upregulation produces NO, which cleaves 130 kDa CPSA to ~15 kDa units. 3. Endosome fuses with the lysosome. 4. Endo-lysosome fuses with exocytic vesicle to form MIIC vesicle which has HLA-DR, HLA-DM and processed polysaccharide. In here, processed polysaccharide is loaded on HLA-DR with the help of HLA-DM. 5, 6. The loaded HLA-DM is presented on the surface of the antigen presenting cell to be recognized by alpha beta TCR present on CD4+ T-cell[35] (Figure adapted from the review *How bacterial carbohydrates influence the adaptive immune system. Avci, FY and Kasper, DL, Ann. Rev. Immunol. 2010*).

1.3 Production of CPSA

The therapeutic properties of CPSA previously mentioned makes it a very interesting candidate for drug development. Unfortunately, *B. fragilis* is an obligate anaerobe, making it impractical to grow and isolate large quantities from bacteria for

pharmaceutical use. Only two groups have reported CPSA isolation and both utilize the same procedure developed by Kasper et al at Harvard Medical School[25,41,42]. In this method one is entirely dependent on the bacteria to produce the molecule in its pure form and maintain the structural consistency from one batch to the next. According to the most recent publication describing production of CPSA, from a 10 L *B. fragilis* culture, only about ~500 mg was expected to be obtained. It was not clear that the CPSA product was separated from CPSB. In addition, the actual final yield of CPSA was not reported. Moreover, preparation of CPSA by this method leaves little space to develop a methodology to determine the individual role of genes responsible for CPSA biosynthesis.

Alternate approaches, like organic synthesis, for the production of CPSA appear limited due to the numerous reaction steps required for complex tetrasaccharide synthesis. Attempts to achieve the synthesis of the first sugar 2-N-acetylamido-4-amino-6-deoxy-galactopyranose itself required as much as ten synthetic manipulations with an overall yield of 22% to prepare a protected precursor[43]. Recently the Seeberger group published the complete synthesis of the CPSA repeat unit. The synthesis required twenty individual steps with an overall yield of 1.8% [44]. The Andreana group have also demonstrated synthesis of the complete tetrasaccharide repeat in nineteen steps, with an overall yield of 5%[45]. Both of the groups did not polymerize the repeat unit, and studies have suggested a minimum of 10-15 units is required for therapeutic benefit[27,36]. These results suggest that a synthetic sequence to CPSA will not likely be an acceptable route for the production of polysaccharide molecules. The need for a robust method for the generation of CPSA is clear and *in vitro* enzymatic biosynthesis appears promising.

1.4 CPSA gene operon and biosynthesis

The main problem associated with extraction of the capsular polysaccharide from the native bacteria is the inconsistencies in the production of pure polysaccharide as there may be lipopolysaccharides or other surface molecules co-purified with the capsule, and the purity can vary from batch to batch. Also previously stated, the complex nature of the chemical synthesis of polysaccharides makes it unfeasible to go through that route. The aim of this project was to biosynthetically produce polysaccharide capsules *in vitro* by utilizing the native bacterial enzymes that catalyze the formation of these sugar polymers. There are many benefits in making complex polysaccharides with enzymes. It includes the ability to have single pot syntheses, which takes advantage of a natural system that already makes the desired product. Different types of enzymes can be expressed in large quantities to yield the sugar polymer. With the ability to express enzymes, comes the control we can exercise in mix-and-matching similar enzymes from different pathways to form a specific desired sugar product. With the biosynthetic method, the twenty or more synthetic steps to produce one of the sugar substrates can also be avoided.

With all the advantages, the biosynthetic route does have disadvantages as well. It includes the lack of access to the appropriate enzymes needed for a specific CPS, poor knowledge on the function of the enzymes, and the difficulties of scaling up product for characterization (such as by NMR). These challenges can be overcome with homology studies to have an idea of the function of the enzyme, testing with multiple substrates to determine if any new products are formed, designing a marker that can be detected and is associated with the sugar products, and improved analytical isolation methods to separate similar products.

CPSA is a polymer of a tetramer repeated approximately 160 times. Its size is estimated to be 110 kDa[35]. The CPSA tetrameric repeat unit consists of an acetamido-4-amino-6-deoxygalactopyranose (AADGal), 4,6-pyruvate galactose (4,6-pyr-Gal), N-acetylgalactosamine (GalNAc), and a galactofuranose (Gal*f*) sugar[46] (Figure 1.2). The structure of CPSA has previously been well investigated using total correlated spectroscopy and NOESY NMR[34]. Three-dimensional structure of a highly related PSA2 molecule shows a right handed helix with two repeating units per turn, and a pitch of 20 Å. The zwitterionic motif is formed with alternating anionic carboxylate lying in repeated grooves and the cationic-free amines are exposed on the outer surface of the carbohydrate[31,47].

Although the chemical composition of CPSA is known, the biochemical pathway involved in its production is poorly documented. The location of the proposed CPSA operon was knocked out, making a mutant *B. fragilis* which did not express CPSA on its surface, thereby confirming the location of the biosynthetic operon[33] (Figure 1.4). Within the CPSA operon, there are eleven genes, of which nine express proteins similar to other proteins involved in various other polysaccharide biosynthesis pathways (Table 1.1).

Figure 1.4 CPSA operon in the *B. fragilis* genome[33].

Table 1.1 Proposed functions of the gene products in the CPSA biosynthesis operon[33].

ORF	Size (aa)	Size (kDa)	Predicted Role	Accession no.
wzx	482	56	flippase	AAK68914.1
wcfM	364	43	galactopyranose mutase	AAK68915.1
wcfN	291	34	glycosyltransferase	AAK68916.1
wzy	434	43	polymerase	AAK68917.1
wcfO	357	40	pyruvyltransferase	AAK68918.1
wcfP	378	44	glycosyltransferase	AAK68919.1
wcfQ	268	32	glycosyltransferase	AAK68920.1
wcfR	407	45	aminotransferase	AAK68921.1
wcfS	195	23	glycosyltransferase	AAK68922.1

1.5 Previous studies on CPSA biosynthetic pathway

The function for the nine genes found in the CPSA operon, has been proposed based on homology studies and also previously characterized proteins[48]. Based on these similarities, a biosynthetic pathway to construct CPSA has been proposed by the Troutman lab (Figure 1.5). The identity of the genes present in the CPSA gene operon suggests that the most likely route for assembling the complex bacterial polysaccharide is a Wzy-dependent pathway in which the repeat unit oligosaccharides are assembled one sugar at a time on the cytosolic face of the bacterial inner membrane[3]. Assembly of the

oligosaccharide takes place on a C55 isoprenoid bactoprenol[5]. Bactoprenol is a hydrophobic anchor which holds the growing polymer in the cell membrane. Bactoprenyl diphosphate is produced by the condensation of farnesyl diphosphate (FPP) to eight units of isopentenyl diphosphate (IPP), done by the enzyme undecaprenyl diphosphate synthase (UPPS). The bactoprenyl diphosphate is then dephosphorylated by undecaprenyl pyrophosphate phosphatase (UP3) to yield bactoprenyl phosphate (Figure 1.6), which is then used by the bacteria to assemble the oligosaccharide[49].

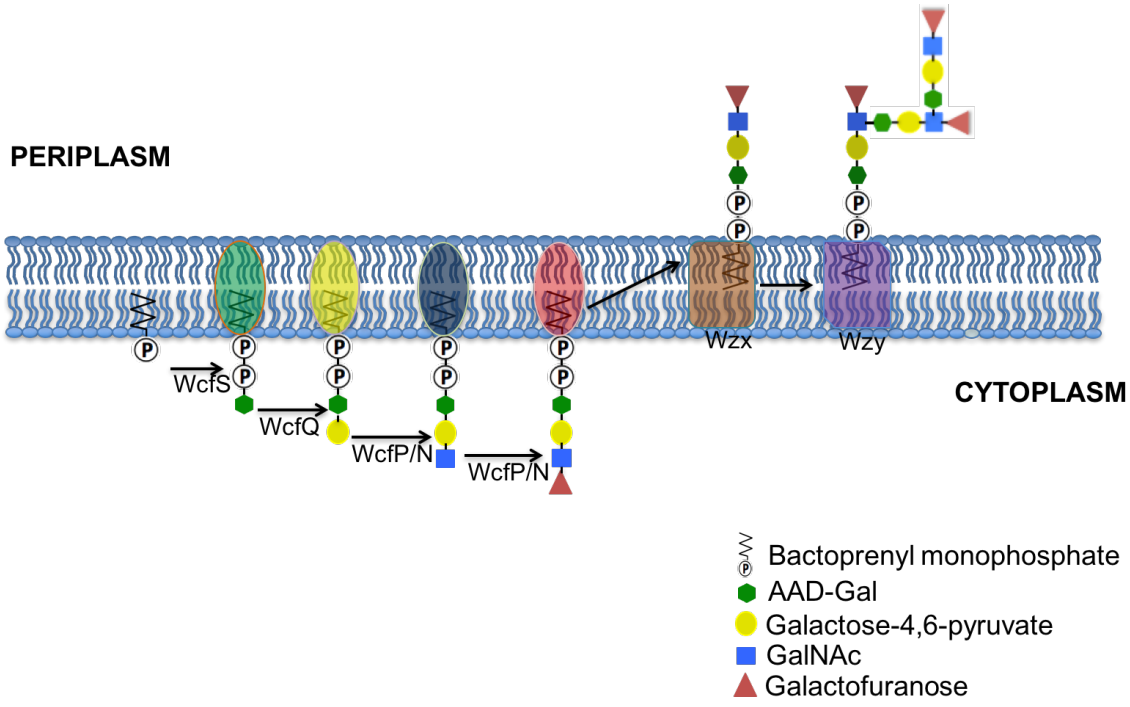


Figure 1.5 Proposed biosynthetic route for CPSA. Enzymes responsible for synthesizing AADGal, have not been shown, as that part has been well characterized.

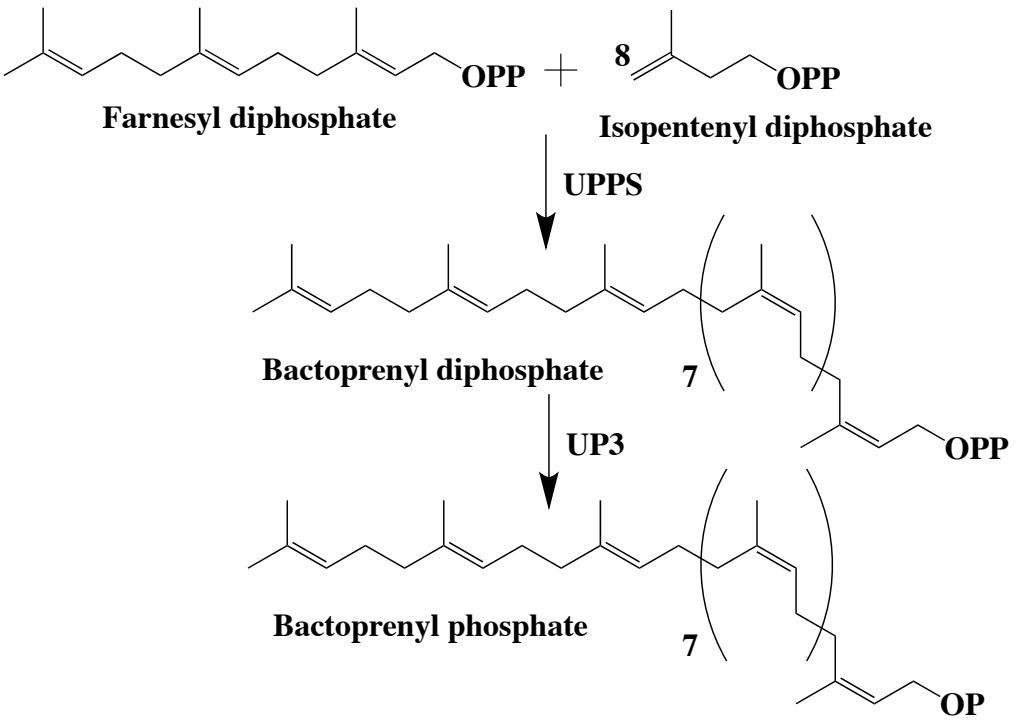


Figure 1.6 Synthesis of bactoprenyl phosphate in bacteria.

The enzymes responsible for the synthesis of the first sugar, AADGal, in the tetrameric repeat, and the enzyme that catalyzes the transfer of this sugar to the bactoprenol anchor have been well characterized by the Troutman lab previously[50]. AADGal is synthesized by the sequential action of a dehydratase and an aminotransferase, which is then transferred to the bactoprenyl monophosphate anchor by a hexose 1-phosphate initiating transferase. Within the CPSA biosynthesis operon, there is a predicted aminotransferase gene, *wcfR*, and a hexose phosphate initiating transferase, *wcfS*, but no predicted dehydratase was found. However, a gene encoding a potential dehydratase, *ungD2*, has been identified elsewhere in the *B. fragilis* genome. When this gene was knocked out by Coyne *et al*, they found that, synthesis of the seven out of the eight capsular polysaccharides was stopped[51]. Initial studies with UngD2 and WcfR did not show any

promise in the synthesis of AADGal. Hence a previously well characterized dehydratase, PglF, from *Campylobacter jejuni* was used to provide the substrate needed for WcfR function[52]. The coupling of these enzymes together led to the production of UDP-AADGal (Figure 1.7). This also points to the notion that depending on homology alone for functional assignment of genes, is not always right, and wet lab results are needed to confirm the function of the gene product.

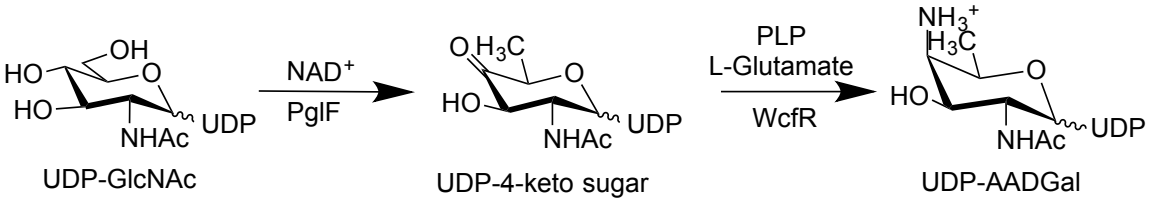


Figure 1.7. Biosynthesis of UDP-AADGal[50].

The synthesized UDP-AADGal was further used as a potential substrate for WcfS, identified as the initiating hexose phosphate transferase. Studies done by Mostafavi *et al* demonstrated that WcfS was indeed the initiating hexose phosphate transferase, which led to the formation of the bactoprenyl linked monosaccharide [50](Figure 1.8).

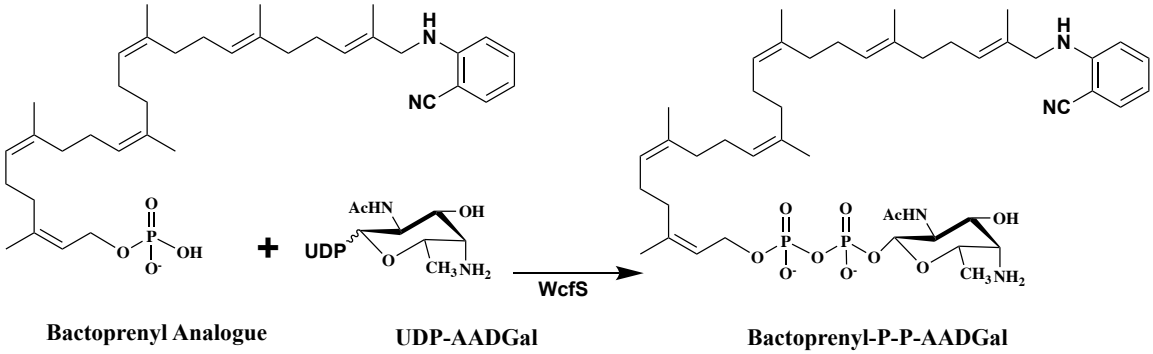


Figure 1.8. Biosynthesis of bactoprenyl linked monosaccharide[50].

As mentioned previously, assembly of the polysaccharides in bacterial cells is done on a C55 bactoprenyl anchor. A major drawback of using this compound in *in vitro* assays is that, it does not have easily distinguishable chromophores associated with it, hence very few rapid assays are available to detect and quantify the activity of enzymes associated with polysaccharide synthesis. The isoprene units, which have double bonds, can be detected at 210 nm, but it is difficult to distinguish it from any other compound with a double bond by absorbance-based methods. In the past, characterization of glycosyltransferases related to biosynthetic pathways has been done using radioisotopic labeling of the sugars attached to the bactoprenyl phosphate, and then tracking the substrate. This method is time consuming and expensive depending on availability of substrates. Another way is to use a plant isoprenoid, which provides a similar functionality but is not an ideal replacement. To circumvent this problem, the Troutman lab has developed fluorescent analogues of the native bactoprenyl phosphate (Figure 1.9), which are easily traceable[53]. Assays done using these analogues take a short time to reveal valuable information about the enzymes when compared to aforementioned traditional assays, which follow the more tedious route of using radioactive labelled substrates. These analogues will be used to functionally characterize the enzymes of the CPSA biosynthetic pathway. They can be easily tracked using HPLC and significantly reduces the time required to monitor the reaction than the traditional assays. Mostafavi *et al* used a p-nitroaniline bactoprenyl phosphate analogue (Figure 1.9) to elucidate the function of WcfS[50] (Figure 1.8).

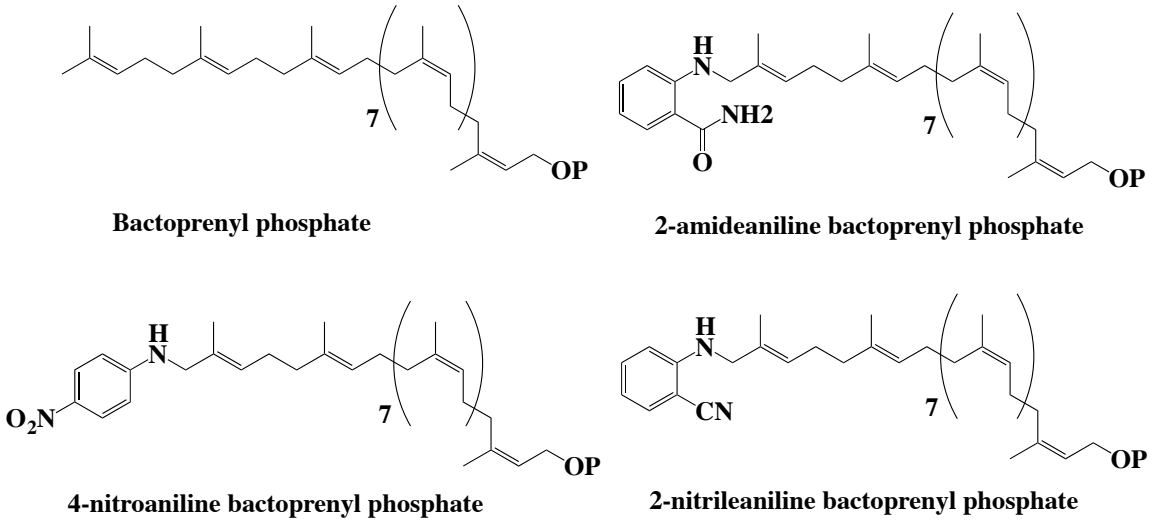


Figure 1.9. Fluorescent analogues of the native bactoprenyl phosphate developed by the Troutman lab.

1.6 Research Objectives

Previous studies by various other groups have demonstrated that CPSA can be developed in a potential therapeutic drug for autoimmune disorders. Studying its biosynthesis would help immensely in future production of this important biomolecule. So far in the biosynthetic pathway, synthesis till the first monosaccharide has been achieved. Once the bactoprenyl linked monosaccharide is formed, the CPSA biosynthetic pathway becomes considerably less clear. According to the NMR structure of CPSA, the next sugar linked to AADGal is a 4,6-pyruvate-galactose (4,6-pyr-Gal) in which the 4- and 6-hydroxyls of galactose are linked to the C2 carbon of pyruvate[34]. One of the objectives of the study will focus on elucidating the sequence of glycosyltransferases that will yield the tetrasaccharide repeat unit. The second goal of the study is to determine whether or not WcfO is responsible for the formation of the pyruvylated sugar. Moreover, our current

understanding of how CPSA is formed in *B. fragilis* does not distinguish between the roles played by the enzymes, and genetic studies performed so far cannot provide the unambiguous characterization of the biochemical pathway that leads to CPSA formation.

The glycosyltransferases involved in CPSA biosynthesis are proposed to be encoded by the *wcfP*, *-N*, and *-Q* genes based on homology studies. Each of these was expected to catalyze only a single sugar transfer from a uridine diphosphate-linked sugar to a bactoprenyl-PP-mono-, di-, or tri- saccharide. Figure 1.5 shows the hypothetical pathway where each glycosyltransferase is expected to transfer a sugar to the repeat unit being assembled on the bactoprenyl anchor. The model (Figure 1.5) for the biosynthesis of the CPSA tetrasaccharide was altered depending on which sugars were accepted as substrates for these glycosyltransferases, for which specificity tests were performed. In order to analyze the glycosyltransferase reactions sequentially, synthesis of the intermediate UDP-linked sugar substrates and the bactoprenyl linked saccharides was also required.

Pyruvylated sugars are found in various polysaccharides, but how this modification is achieved has not been studied in detail in the past. Since only few studies have been done in the past on pyruvate modification, a handful of literature is available on the family of enzymes that catalyze this reaction. This modification, wherein a pyruvate moiety is attached to a sugar, bestows CPSA with the negative charge. In this study, one of the aims was dedicated towards studying this potential pyruvation enzyme, and finding out its target substrate. Through homology studies *wcfO* gene has been identified to express the pyruvyl transferase protein. So far, either one or two studies have been done on enzymes like this, but these studies have been inconclusive in determining the native target substrate of this

family of enzymes[54,55]. Through this study I will be able to shed light on the native substrate of a pyruvyltransferase with respect to biosynthetic pathway of complex polysaccharides, and make conclusive remarks on the functional characterization of this enzyme, and add a new characterized member to this family.

The last sugar in the CPSA repeat unit, is a five membered rare sugar, which is made from a UDP-linked galactopyranose by a galactopyranose mutase. This conversion from pyranose to furanose is thermodynamically unfavorable due to the conversion from a six-membered ring to a five-membered ring[56]. The CPSA biosynthetic operon is proposed to have a galactopyranose mutase gene *wcfM*, which should catalyze the conversion of the pyranose form to the furanose form, for the last glycosyltransferase to attach it to the trisaccharide, or the proposed mutase may differ to act on the pyranose, once the tetrasaccharide is formed. Galactopyranose mutase reactions have been well characterized previously in *E. coli, Klebsiella pnuemoniae, Mycobacterium tuberculosis* and *C. jejuni*, using both capillary electrophoresis (CE) and high performance liquid chromatography (HPLC), in the forward direction from a six membered ring to a five membered ring[57,58].

CHAPTER 2: GLYCOSYLTRANSFERASES

2.1 Introduction

Cell surface polysaccharides are nothing but complex carbohydrates. They play important roles in a number of biological processes such as cell growth, cell to cell interactions, immune response, and inflammation. The polysaccharides are synthesized by a class of enzymes known as glycosyltransferases[7]. Glycosyltransferases are an enzyme superfamily responsible for the attachment of carbohydrate moieties to a wide array of acceptors that include nucleic acids, polysaccharides, proteins, lipids, and carbohydrates. The majority of glycosyltransferases are sugar nucleotide-dependent enzymes, and utilize nucleoside diphosphate sugars (NDP-sugars) as donors for the glycosidic bond formation. In other cases, the sugar donors can also be lipid phosphates and unsubstituted phosphate[6].

The glycosyltransferases have been classified by sequence homology into 96 families in the Carbohydrate Active enZyme database (CAZY), each of which catalyze the reaction as shown in Figure 2.1[59]. Chain elongation of the oligosaccharide units in complex carbohydrates is achieved by the addition of monosaccharide units through the action of different glycosyltransferases in a specific sequence. The CAZY database provides a highly powerful predictive tool, as the structural fold and mechanism of action remain invariant in most of the families[60]. Therefore, where the structure and mechanism of a glycosyltransferase member for a given family has been reported, some assumptions about

other members of the family can be made. Substrate specificity, however, is more difficult to predict, and requires experimental characterization of the individual glycosyltransferases.

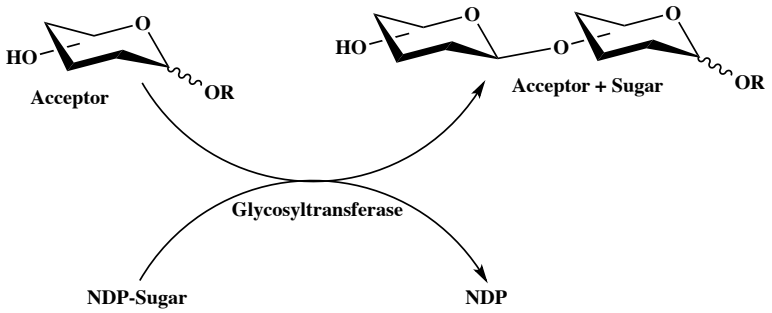


Figure 2.1 General reaction scheme for a Glycosyltransferase (GTs)

Determining both the sugar donor and acceptor for a glycosyltransferase of unknown function can be challenging, and it is one of the reasons there are significantly fewer well characterized isoprenoid linked sugar glycosyltransferases when compared to the glycosyltransferases responsible for synthesizing disaccharides or the oligosaccharides[59]. The less reports on isoprenoid linked sugar transferases can be attributed to the fact that, a high throughput method has not yet been developed which will enable for faster characterization. Another challenge in characterizing the glycosyltransferases is the availability of rare sugars, as most of the bacterial polysaccharides contain rare sugars. Rare sugars, such as rhamnose or fucose, may provide the bacterial polysaccharides with additional biological properties compared to those composed of more common sugar monomers[8,9]. Rare sugars are monosaccharides that are not commonly found in nature and have low abundance, in comparison to common sugars

like D-glucose, D-galactose, D-fructose, D-xylose, D-ribose, and L-arabinose which are more abundant[8]. Moreover, the traditional methods like radioisotopic labelling, thin-layer chromatography (TLC) used to characterize the glycosyltransferase, often tends to be tedious and challenging in tracking the product formation. The Troutman lab has developed fluorescent bactoprenyl analogues, which is much faster and easier to monitor in determining the function of the glycosyltransferases[53].

Glycosyltransferases catalyze glycosidic bond formation with either overall retention or inversion of anomeric configuration when compared to the stereochemistry in the sugar donor. Inverting glycosyltransferases are thought to employ a direct displacement S_N2-like mechanism that involves an enzymatic base catalyst. The active site in one of the side chain serves as a base catalyst which deprotonates the incoming nucleophile from the anomeric carbon of the acceptor, hence facilitating a direct SN2-like displacement of the activated phosphate leaving group[60]. Structural data have shown that several inverting glycosyltransferases, contain no obvious candidate catalytic base indicating these enzymes may use an alternative mechanism[6,7].

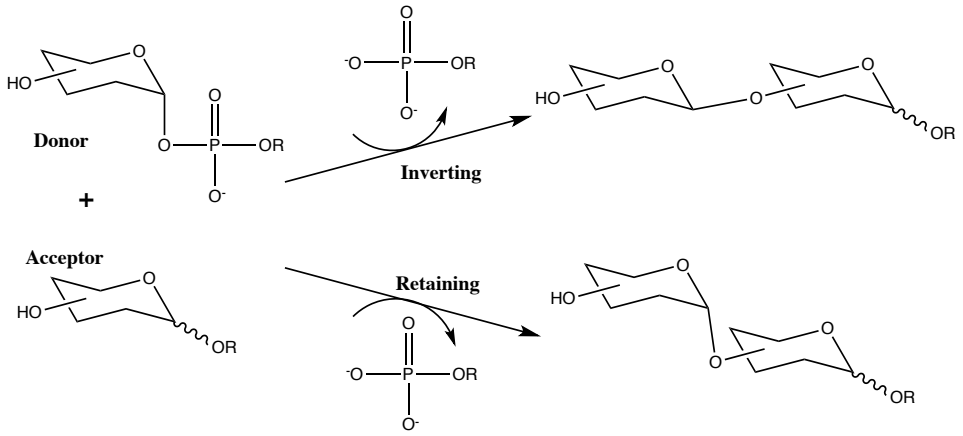


Figure 2.2 Glycosyltransferases catalyze glycosyl group transfer with either inversion or retention of the anomeric stereochemistry with respect to the donor sugar. (Figure adapted from *Glycosyltransferases:Structures, Functions,and Mechanisms. Lairson L.L et al 2008).*

The reaction coordinate employed by retaining glycosyltransferases has been much debated, and it could be possible that the mechanism is not conserved for all retaining enzymes. The most accepted theory currently accepted is a double displacement mechanism via a covalent mechanism, which is analogous to that used by the glycoside hydrolases[60]. A report by Soya *et al*. provided mass spectrometry evidence for the formation of a covalent intermediate between the donor substrate and a cysteine, which had been substituted for the candidate catalytic nucleophile, on two retaining glycosyltransferases[61]. The more favored mechanism in the field is an S_N1 or S_N1-like mechanism, which involves interaction between the leaving group and the attacking nucleophile on the same face. This mechanism has also been supported by kinetic isotope effect studies to analyze the structure of the transition state and by computational modelling[6,7].

2.2 Glycosyltransferases of the CPSA operon

The CPSA gene operon has three genes, *wcfQ, wcfP* and *wcfN*, that putatively encode for glycosyltransferases[33]. Each of these glycosyltransferases is expected to transfer a sugar moiety to the bactoprenyl linked monosaccharide, the disaccharide and the trisaccharide. Based on the CAZy database, and homology studies, WcfQ and WcfN are hypothesized to belong to the glycosyltransferase superfamily A, which follows the inverting mechanism in the sugar transfer. Whereas WcfP is proposed to belong to the glycosyltransferase superfamily B, which follows the retaining mechanism[59].

Table 2.1 Proposed glycosyltransferases in the CPSA biosynthetic operon.

Glycosyltransferase	Superfamily	Proposed Mechanism
WcfN	GT_2	Inverting
WcfP	GT_4	Retaining
WcfQ	GT_2	Inverting

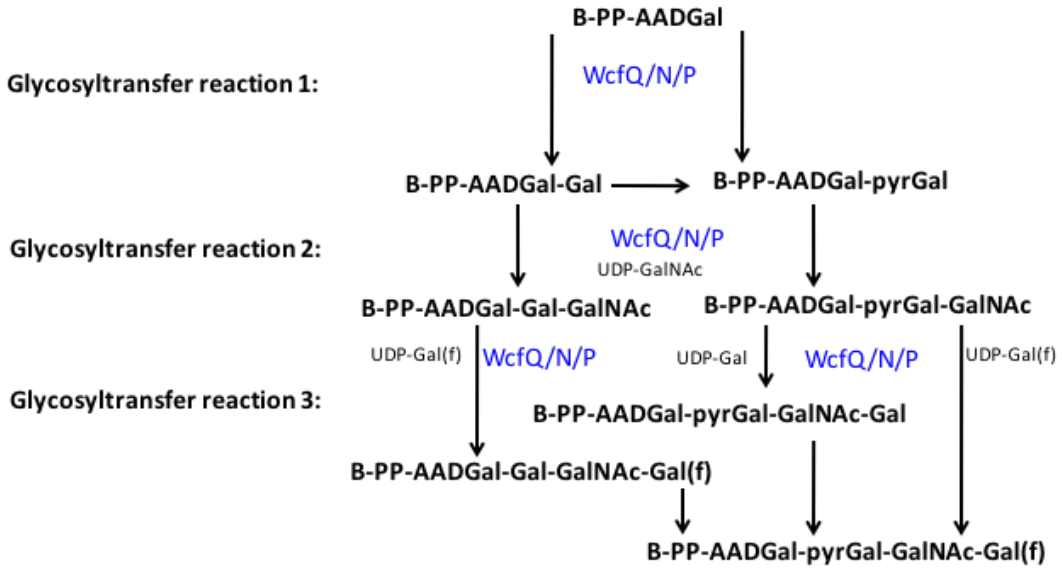
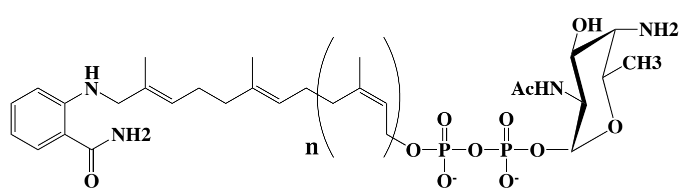


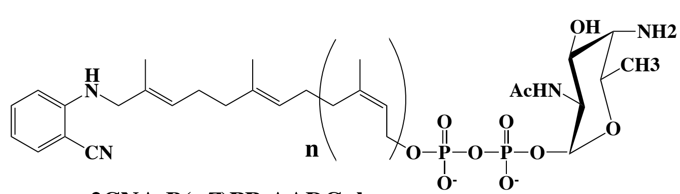
Figure 2.3 The potential routes through which glycosyltransferases can achieve the synthesis of the tetrasaccharide. The glycosyltransferases are represented in blue.

It is not possible to know a priori which glycosyltransferase enzyme will transfer which sugar to the bactoprenyl diphosphate-linked monosaccharide, BPP-AADGal. In addition, there are multiple alternative pathways that could be utilized to biosynthesize the tetrasaccharide repeat unit as shown in Figure 2.2. Glycosyltransferases are highly specific to the donor substrate, hence only one of the enzymes should be capable of transferring the

second sugar (galactose) to the BPP-AADGal to give BPP-AADGal-Gal. Each of the glycosyltransferases, WcfP, Q and N, will be tested with BPP-AADGal and UDP-Gal. The reaction will be analyzed using reverse phase high performance chromatography (RP-HPLC) as the bactoprenyl analogues 2CNA/AA-B(n*Z*)PP-AADGal, where n=4, 6 and 7 (Figure 2.3) are fluorescent and can be easily tracked. In a similar way assays will be performed with the bactoprenyl diphosphate linked disaccharide, trisaccharide and the donor substrates to identify the second and the last glycosyltransferase. All the products will be isolated and characterized through electrospray ionization mass spectrometry (ESI-MS). The structure of CPSA, has a galactofuranose as one of the sugar residues. Assays will be performed to clarify if UDP-Galactose is first converted to the furanose form and then transferred, or when once attached to the growing oligosaccharide, it is converted to the furanose. The CPSA gene operon has been proposed to have a mutase enzyme also, which can catalyze this reaction. Characterization of the mutase and production of the furanose sugar is covered in the fourth chapter



2AA-B(nZ)PP-AADGal



2CNA-B(nZ)PP-AADGal

Figure 2.4 Fluorescent analogues used to determine and monitor the activity of the glycosyltransferase., where n is either 4, 6 and 7.

2.3 Experimental methods

2.3.1 Cloning and overexpression of glycosyltransferases

Polymerase chain reaction amplifications on each gene were performed using *B. fragilis* genomic DNA (ATCC 25285) and the oligonucleotide primers. The *wcfQ* and *wcfP* were digested with BamHI and XhoI, *wcfN* was digested with NheI and XhoI. All three genes were subsequently ligated into a pET-24a vector digested with the appropriate restriction enzyme for each gene. Chemically competent *E. coli* DH5α cells were transformed with each ligated vector and Kanamycin resistant clones were selected. Plasmids were isolated from the Kanamycin resistant clones and sequenced to confirm introduction of each gene. Chemically competent *E. coli* BL-21 RIL cells were then transformed and again positive clones were selected by Kanamycin resistance. The cloning of the genes was done by my advisor Dr. Jerry M. Troutman, prior to me joining the lab.

Cells containing WcfN, WcfP, and WcfQ encoding plasmids were cultured in 5 mL of Luria Broth (LB) overnight at 37 ºC then were used to inoculate 0.5 L of LB. Cells were grown with shaking to an O.D.600 of 0.8 and the temperature was decreased to 25 ºC for 30 min. Isopropylthiogalactoside (IPTG) was added at a final concentration of 1 mM. Cells were allowed to incubate for an additional four hours, then were harvested by centrifugation at 4 ºC for 20 min at 5000 x g, the supernatant was removed and cells were stored at -80 ºC for later use.

2.3.2 Isolation of glycosyltransferases as membrane fraction

Expression cells were thawed in 20 mL of 50 mM Tris–HCl (pH = 8.0) containing 200 mM NaCl and 20 mM imidazole then sonicated on ice for 6 minutes (total) with a pulse of 1 second on and 1 second off. Unbroken cells were then removed by centrifugation

at 5000 x g for 15 min at 4 ºC. The supernatant containing membrane and cytosolic components was then spun at 150,000 x g for 1 hour at 4 ºC. The pelleted membrane components were homogenized into 1 mL of 50 mM Tris-HCl (pH = 8.0) 200 mM NaCl then 10–50 µL aliquots were stored at -80 ºC. Total protein concentration in the membrane fractions was measured using a Bradford assay with BSA as a standard[62]. The presence of the overexpressed proteins was confirmed by SDS–PAGE and Western Blot analysis with an anti-His/anti-rabbit alkaline phosphatase.

2.3.3 General method for glycosyltransferase assays

Screening for the first glycosyltransferase assays were prepared with 1 µL of each membrane fraction (total protein membrane fraction concentrations provided in results section) in a 100 µL solution containing 5 µM 2CNA/AA-(n*Z*)BPP-AADGal, where n=4,6, 1% Triton X-100, 10 mM $MnCl_2$, 25 mM Bicine (pH = 8.3), 500 µM UDP-Galactose, and 10% DMSO. After 30 minutes each mixture was analyzed on HPLC. Once the specific function of an enzyme had been established, the product specific reaction was optimized for full turnover. The screening and optimization for the first glycosyltransferase with various uridine diphosphate linked sugars were performed by my advisor, Dr. Jerry M. Troutman in my absence.

The reaction conditions were kept similar (25 mM Bicine, 10 mM $MnCl_2$, 1% Triton X-100, 10% DMSO) while screening for the second and the last glycosyltransferase, except that the proteins were tested with UDP-N-acetyl galactosamine and UDP-galactofuranose, as these two sugar molecules are the third and last residues in the tetrasaccharide monomer. Furthermore, the enzymes were also tested with other sugar substrates for promiscuity.

A sequential single pot reaction for the synthesis of the tetrasaccharide was also attempted. An HPLC sample was prepared as described below and was allowed to incubate for one hour before HPLC injection of 5 μL 0.5-1 hour after each enzyme addition. The reaction mixture started in 200 μL of 25 mM Bicine (pH=8.3), 1 % Triton-X-100, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 5 μM of 2CN-B(4*Z*)P, 500 μM UDP-AADGal, 0.48 mg/mL total protein WcfS membrane fraction. Next 3 μL of WcfQ membrane fraction (stock concentration 6.9 mg/mL) and 2 μL of UDP-Gal (50 mM) were added, then HPLC analyzed. Next, 1 μL of PEP (1M) and 5 μL of purified WcfO (stock concentration 1.31 mg/mL) were added. After HPLC analysis, 2.6 μL of UDP-GalNAc (115 mM) and 2 μL of WcfP (stock concentration 9.3 mg/mL) were added. Finally, after HPLC analysis of the tetrasaccharide product 2.8 μL of UDP-Gal*f* (105 mM) was added along with 2 μL of WcfN (stock concentration 7.2 mg/mL). This assay was analyzed on the HPLC using 35% 1-propanol with 100 mM ammonium bicarbonate as the mobile phase.

All the reactions were analyzed on HPLC, an Agilent 1100 HPLC system, equipped with diode array and fluorescence detectors. WcfQ assays were analyzed on the HPLC using 42% 1-propanol with 100 mM ammonium bicarbonate, unless stated otherwise. All analysis of WcfP and WcfN assays used isocratic conditions of 32% 1-propanol, 65% 100 mM ammonium bicarbonate for *cis-4* isoprenoids unless mentioned otherwise. All chromatography involving the WcfQ assays was performed on a reverse phase C18 Agilent Eclipse XDB-C18, 5μm, 4.6 x 150 mm column. The WcfP and WcfN assays were performed on a reverse phase C8 Agilent Eclipse XDB-C18, 5μm, 4.6 x 150 mm column.

2.3.4 Characterization of products using MS

To characterize the products produced by WcfQ, WcfP and WcfN, a variation in the reaction mix was done. Instead of using Triton X-100, 3.5 mM sodium cholate was used as a surfactant. This was done as Triton X-100 is a confirmed MS contaminant, hence to get a clear spectrum, a substitute was used. MS was performed on the fractions collected from the HPLC on a Thermo VELOS Pro Dual-Pressure Linear Ion Trap using electrospray ionization (ESI) introduced into the capillary with a 0.200 mL/min flow rate of acetonitrile. For the bactoprenyl-linked sugar products negative mode was used with a capillary temperature of 300 °C and a spray voltage of 4.00 kV. Mass spectrometry on the samples were performed by Katelyn Erickson.

2.4 Results

2.4.1 Overexpressed glycosyltransferases localize to membrane fraction

Few family of glycosyltransferases have been classified to have transmembrane domain either at the C-terminal or the N-terminal[63]. The presence of the transmembrane domains in proteins makes it very hard to overexpress them in the solubilized form. Hence few proteins of this superfamily have been structurally and functionally characterized. It is usually a good approach to analyze the protein sequence before overexpressing it. Analyzing the sequence gives an idea on how many transmembrane domains might be associated with the protein, and it can be predicted if the protein will be associated with the membrane or be soluble when expressed[64]. Using the TMHMM prediction software at least one transmembrane domain was predicted each in WcfP and WcfN. WcfQ (Figure 2.4) was not predicted to have a transmembrane domain, and could be a soluble protein. All the glycosyltransferases were isolated as membrane-bound proteins, and their presence

was confirmed through Western Blot as shown in Figure 2.5. It was surprising that despite not having any predicted transmembrane domains, WcfQ was associated with the membrane, and failed to express as a soluble protein. Total protein concentrations were 6.9, 9.3, and 7.2 mg/mL for WcfQ, WcfP, and WcfN membrane preparations, respectively.

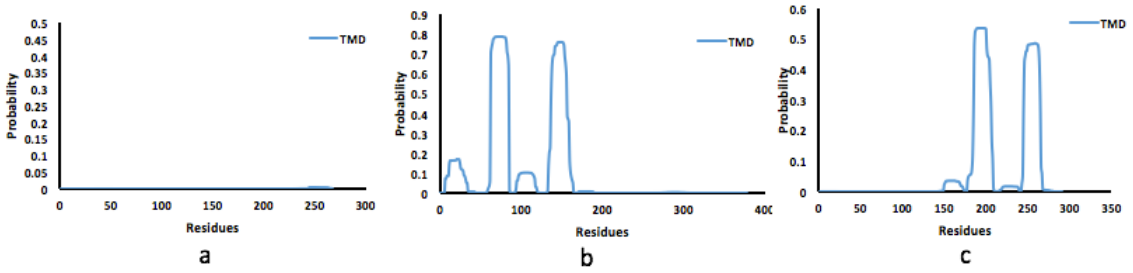


Figure 2.5 Prediction of transmembrane domain using TMHMM[64]. The vertical axis is probability based on hydrophobicity, and horizontal axis is total number of amino acids in the protein. The relative position of transmembrane domain in the protein is shown in red. (A) WcfQ, no transmembrane domain was predicted, whereas (B) WcfP and (C) WcfN were predicted have at least two transmembrane domains in each.

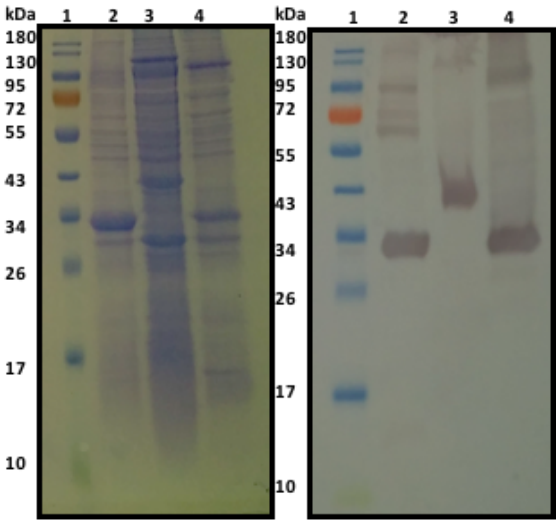


Figure 2.6 Coomassie Stain of SDS-PAGE and Anti-His Tag Western Blot of membrane fractions of glycosyltransferases. Lane 1: Ladder, Lane 2: WcfQ (32 kDa), Lane 2: WcfP (44 kDa) and Lane 2: WcfN (34 kDa). Non-specific streaking is usually observed with membrane fractions.

2.4.2 WcfQ identified as the first glycosyltransferase

In order to test the activity of WcfQ, WcfP, and WcfN, the isolated membrane fractions were mixed with excess uridine diphosphate galactose (UDP-Gal) and the fluorescent 2-cyanoaniline bactoprenyl diphosphate-linked AADGal (2CNA-B(4*Z*)PP-AADGal. Each reaction was then analyzed by HPLC under isocratic conditions with 42% 1-propanol and 100 mM ammonium bicarbonate and fluorescence detection to identify any changes in the retention of the fluorescent substrate. Membrane fractions from the WcfQ and WcfP, but not WcfN expressions, altered the HPLC retention of the 2CNA-B(4*Z*)PP-AADGal from 9.48 min to 9.05 min in the presence of UDP-Gal (Figure 2.6). Since there was no activity from the WcfN containing membrane preparations, it was deduced that the activity observed with WcfP and WcfQ must have been from the membrane fraction of the recombinant overexpressed WcfP and WcfQ[65].

In specificity studies, surprisingly it was found that, WcfQ was also capable of transferring glucose to the monosaccharide. WcfP or WcfN were not able to utilize UDP-glucose (UDP-Glc), UDP-glucuronic acid (UDP-GlcA), UDP-N-acetylgalactosamine (UDP-GalNAc) nor UDP-N-acetylglucosamine (UDP-GlcNAc) as a donor sugar to transfer it to the monosaccharide, suggesting there were no other alternative substrates for either WcfP or WcfN.

In optimizing the turnover of WcfQ it was found out that WcfQ performed much better when manganese was used as the divalent cation instead of magnesium[65].

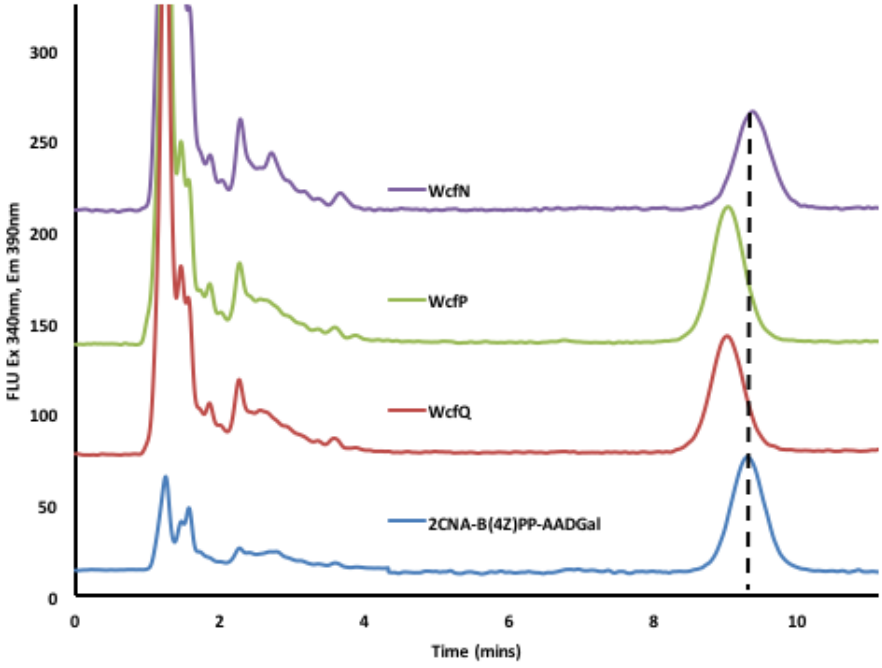


Figure 2.7 Membrane fractions of all three glycosyltransferases were tested with excess UDP-Galactose. Both WcfQ and WcfP were capable of transferring UDP-Galactose to the bactoprenyl linked monosaccharide.

Although WcfQ could catalyze the transfer of UDP-Glucose, more enzyme content and donor sugar were required, hence it was deduced that WcfQ transferred UDP-Galactose much more effectively than UDP-Glucose. Once again, when comparing the transfer of galactose between WcfQ and WcfP, it was observed that, WcfQ was more effective than WcfP. WcfP was required in much more total protein content for the reaction to proceed to completion. Hence this established that WcfQ is the first glycosyltransferase responsible for forming the disaccharide. The WcfQ product, 2AA-B(7)PP-AADGal-Gal, was purified using HPLC and characterized using positive ion mode mass spectrophotometry, in which the theoretical expected mass 1273.72 amu matched with the obtained mass[65].

2.4.3 WcfP is specific for pyruvylated galactose disaccharide to form the trisaccharide

The second sugar in the tetrasaccharide repeat unit has a pyruvate modification, which made identification of the second glycosyltransferase have multistep assays. It remained to be determined whether pyruvylation of the second sugar was necessary for the second glycosyltransferase to initiate the attachment of the third sugar, or the trisaccharide could be produced without having the second sugar undergo the pyruvate modification. Assays of both the expressed WcfP and WcfN with both the galactose and glucose containing disaccharide, 2CN-B(4*Z*)PP-AADGal-Gal and 2CN-B(4*Z*)PP-AADGal-Glc, did not lead to a change in retention time of the substrate peak as shown in figures 2.9 (a,b) and 2.10 (a,b). This suggested that perhaps pyruvylation of the second sugar was an essential step towards the synthesis of the trisaccharide.

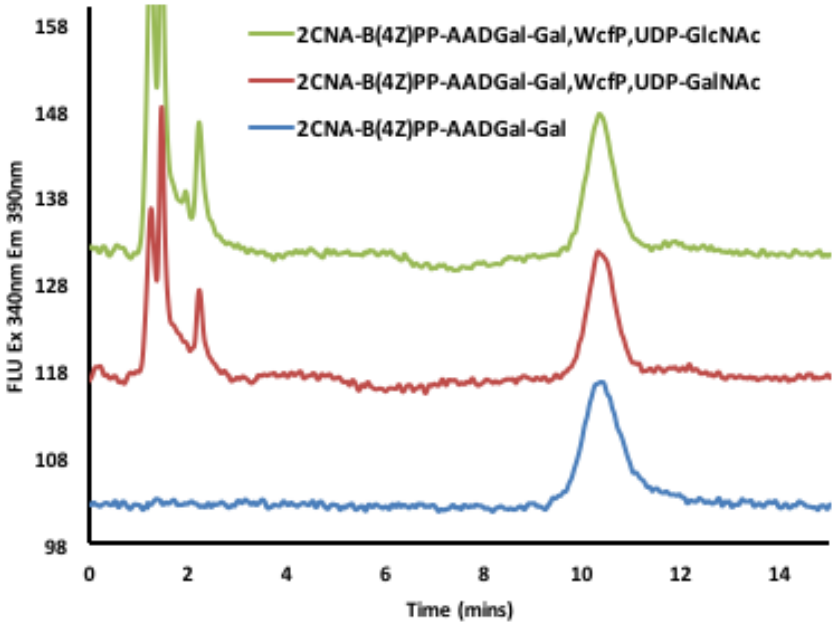


Figure 2.8 (a) WcfP does not accept unpyruvylated disaccharide, 2CN-B(4*Z*)PP-AADGal-Gal as a acceptor substrate. HPLC conditions used were 42% 1-propanol with 100 mM ammonium bicarbonate as the mobile phase.

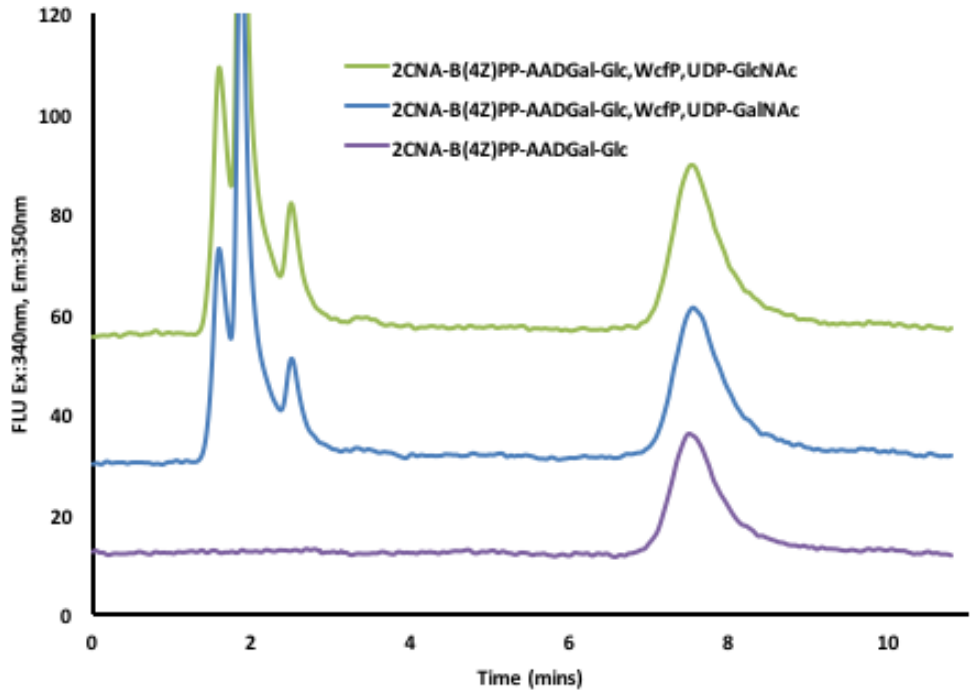


Figure 2.8 (b) WcfP does not accept unpyruvylated disaccharide, 2CN-B(4*Z*)PP-AADGal-Glc as a acceptor substrate. HPLC conditions used were 45% 1-propanol with 100 mM ammonium bicarbonate as the mobile phase.

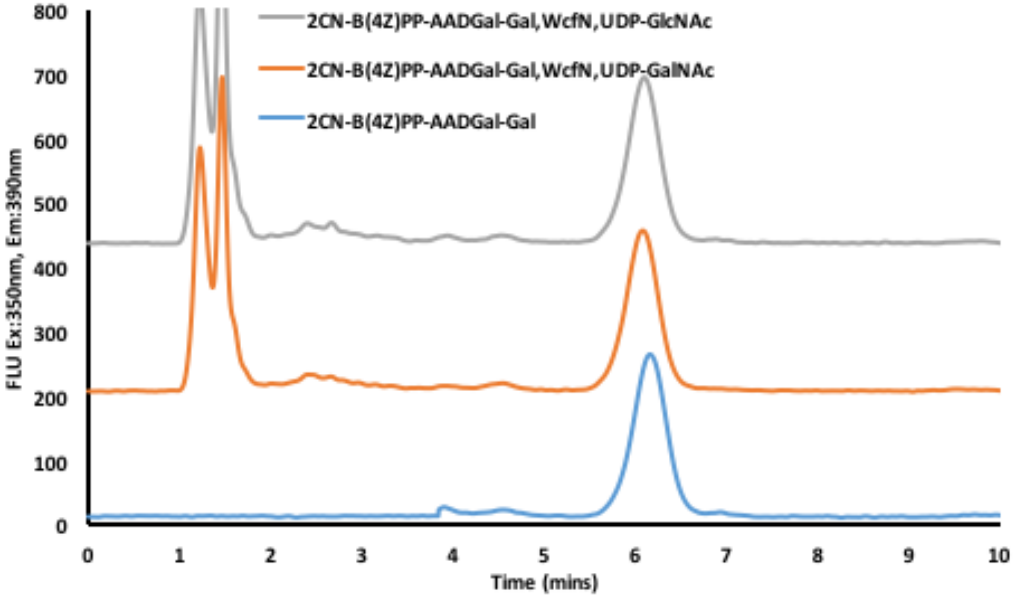


Figure 2.9 (a) WcfN does not accept unpyruvylated disaccharide, 2CN-B(4*Z*)PP-AADGal-Gal as a acceptor substrate. HPLC conditions used were 49% 1-propanol with 100 mM ammonium bicarbonate as the mobile phase.

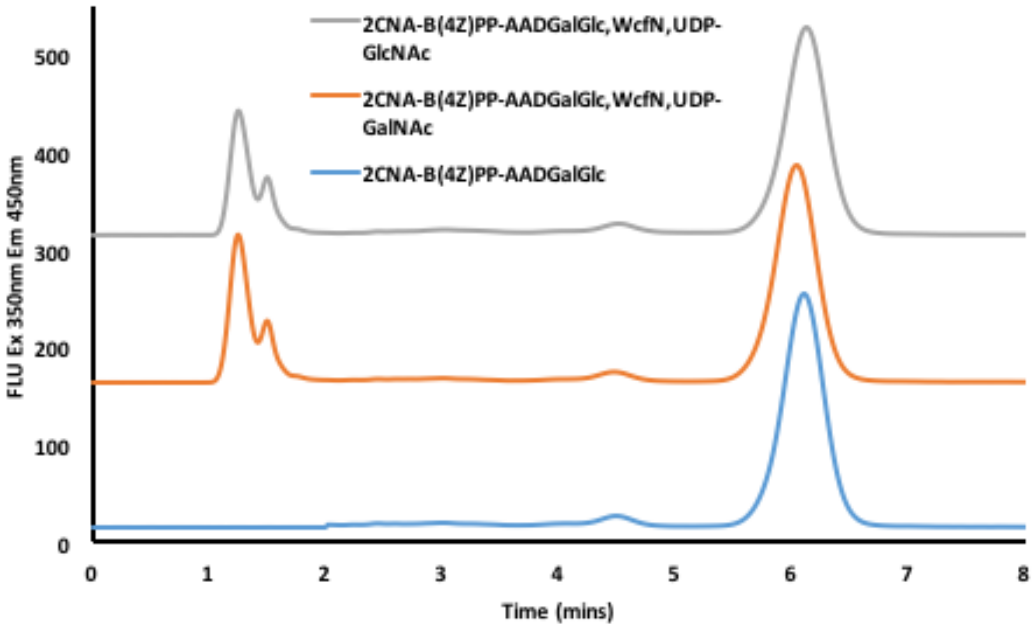


Figure 2.9 (b) WcfN does not accept unpyruvylated disaccharide, 2CN-B(4*Z*)PP-AADGal-Glc as a acceptor substrate. HPLC conditions used were 49% 1-propanol with 100 mM ammonium bicarbonate as mobile phase.

Moving forward, the previous assays suggested that, the unpyruvylated disaccharides were not being accepted as a substrate by either WcfP or WcfN. In parallel it was established that WcfO was indeed pyruvylating the disaccharides BPP-AADGal-Gal and BPP-AADGal-Glc (Chapter 3). Since WcfO was capable of producing two substrates, it remained to be deduced, whether the next glycosyltransferase would also be able to accept both the pyruvylated disaccharides as a substrate, as the published structure of CPSA suggested a 4,6-pyuvylated galactose. Since there did not appear to be any discrimination in the pyruvylation of the two disaccharides by WcfO it was quite possible that the next step in the pathway would only use one or the other as a substrate. When WcfN was incubated in the presence of UDP-N-acetylgalactosamine (GalNAc), it was once again ineffective in altering the retention time of either of the pyruvylated products

(galactose or glucose linked) obtained from WcfO, 2CNA-B(4*Z*)PP-AADGal-pyrGal or 2CNA-B(4*Z*)PP-AADGal-pyrGlc (Figure 2.11 a,b). However, incubation of WcfP with the isoprenoid linked substrates and the donor sugar did, we found that it catalyzed the transfer of UDP-GalNAc to 2CNA-B(4*Z*)PP-AADGal-pyrGal, but did not the transfer any sugar to the glucose containing pyruvylated disaccharide (Figure 2.12 a,b).

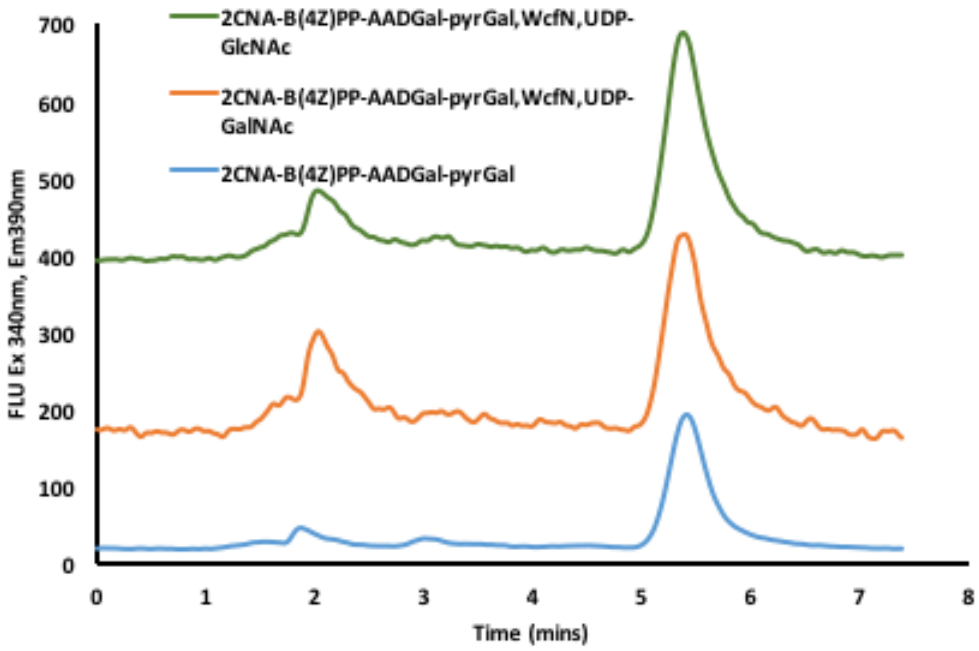


Figure 2.10 (a) WcfN does not accept pyruvylated disaccharide, 2CN-B(4*Z*)PP-AADGal-pyrGal as a acceptor substrate. HPLC conditions used were 32% 1-propanol with 100 mM ammonium bicarbonate as mobile phase.

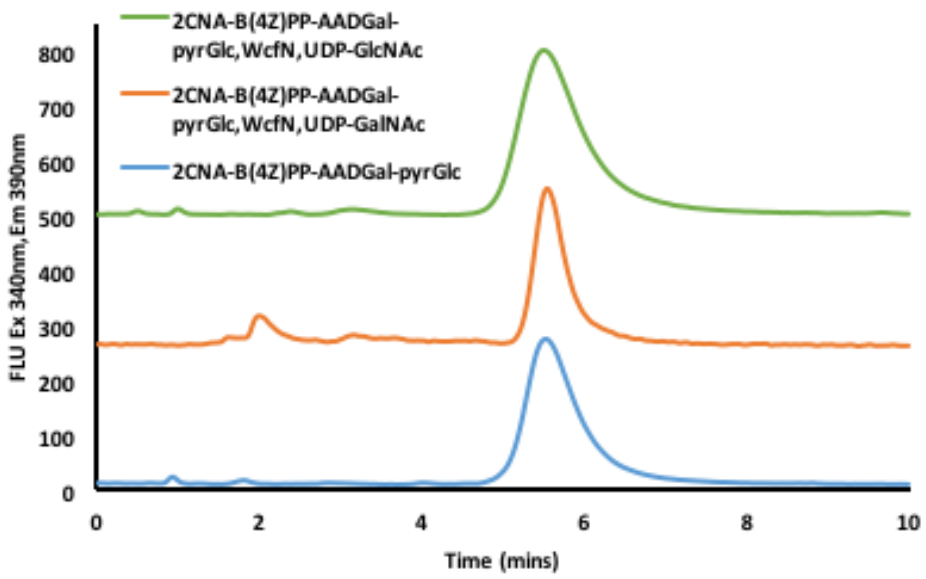


Figure 2.10 (b) WcfN does not accept pyruvylated disaccharide, 2CN-B(4*Z*)PP-AADGal-pyrGlc as a acceptor substrate. HPLC conditions used were 32% 1-propanol with 100 mM ammonium bicarbonate as mobile phase.

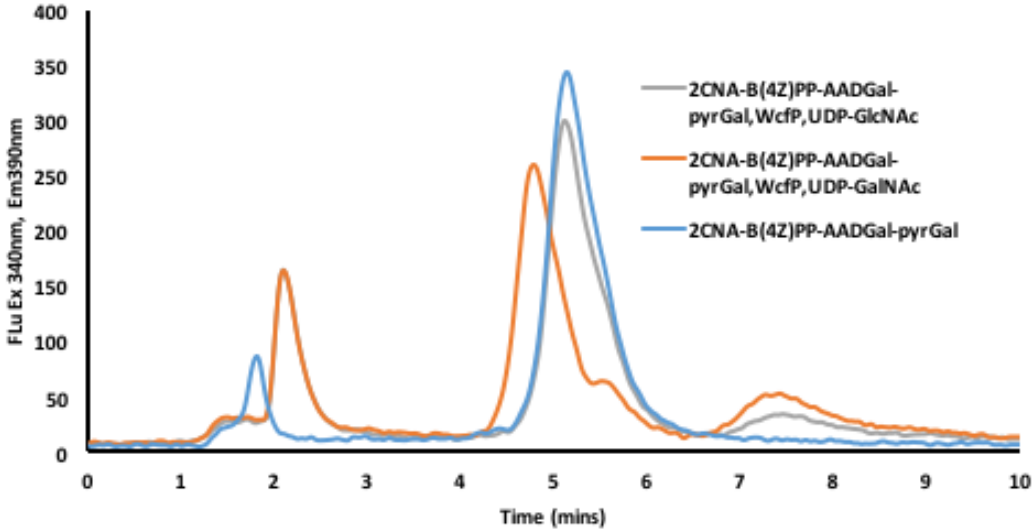


Figure 2.11 (a) WcfP accepts pyruvylated disaccharide, 2CN-B(4*Z*)PP-AADGal-pyrGal as an acceptor substrate. It does not use UDP-GlcNAc as sugar donor. HPLC conditions used were 32% 1-propanol with 100 mM ammonium bicarbonate as mobile phase.

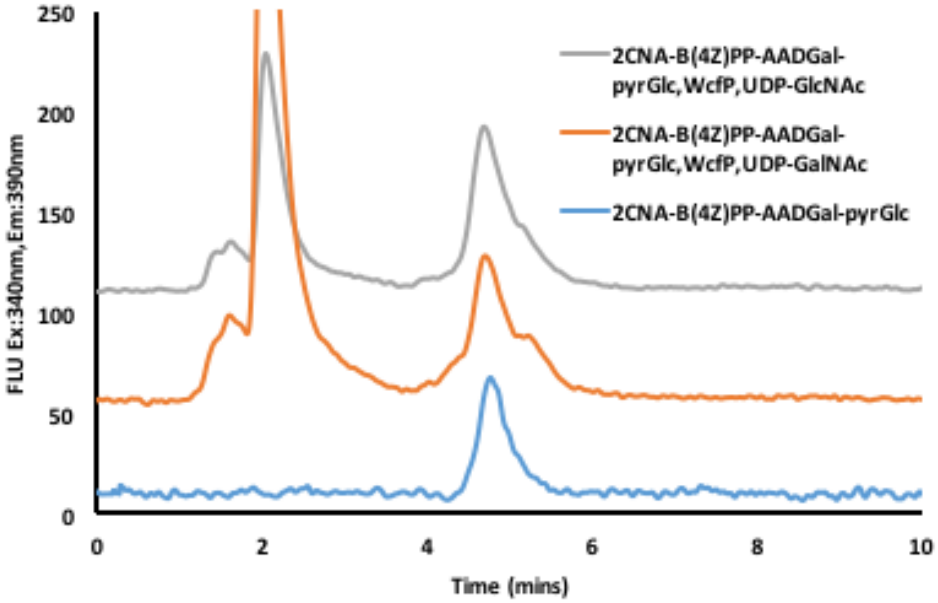


Figure 2.11 (b) WcfP did not accept pyruvylated disaccharide, 2CN-B(4*Z*)PP-AADGal-pyrGlc as an acceptor substrate with either of the sugar substrates. HPLC conditions used were 32% 1-propanol with 100 mM ammonium bicarbonate as mobile phase.

Product from the WcfP reaction was isolated and then characterized by negative ion mode ESI-MS and the mass observed was consistent with the expected mass for 2CNA-B(4*Z*)PP-AADGal-pyrGal-GalNAc (Figure 2.13). UDP-N-acetylglucosamine (GlcNAc) was also tested with WcfP and WcfN, but no product formation was observed, suggesting that WcfP was selective for the pyruvylated galactose acceptor and UDP-GalNAc donor.

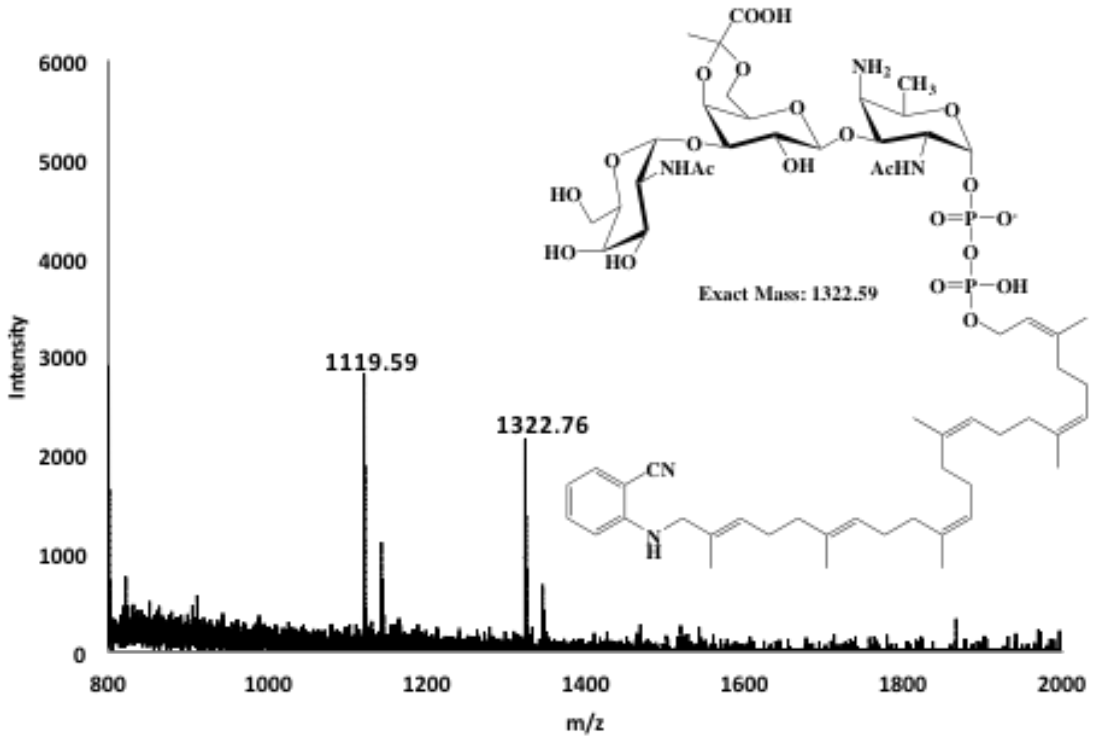


Figure 2.12 MS of 2CNA-B(4*Z*)PP-AADGal-pyrGal-GalNAc. 2CNA-B(4*Z*)PP-AADGal-pyrGal-GalNAc was purified using RP-HPLC. The purified product was characterized for mass on Thermo VELOS Pro Dual-Pressure Linear Ion Trap in negative mode. The exact mass of the product is 1322.59, which matched with the 1322.76, mass obtained from the VELOS. The WcfO product can also be seen in the spectrum.

Since WcfP did not accept B(4*Z*)PP-AADGal-pyrGlc as a potential substrate, it will be an interesting study to find out whether the native *B. fragilis* produces an oligosaccharide which has a pyruvate glucose modification also, or the modification we have seen in this analysis is just an *in vitro* artifact.

2.4.4 WcfN completes the tetrasaccharide and is a galactofuranosyltransferase

The final step in the assembly of the CPSA repeat unit was expected to be the addition of a galactofuranose (Gal*f*) residue to the pyruvylated trisaccharide. Only one glycosyltransferase, WcfN, remained uncharacterized in the CPSA biosynthesis gene cluster. As mentioned previously, the last sugar residue is a galactofuranose, a rare sugar.

Hence, a chemically synthesized UDP-Gal*f*, generously given by the Todd Lowary lab was used with WcfN and the bactoprenyl linked trisaccharide, 2CNA-B(4*Z*)PP-AADGal-pyrGal-GalNAc, to verify the role of WcfN in the biosynthesis pathway (Figure 2.13).

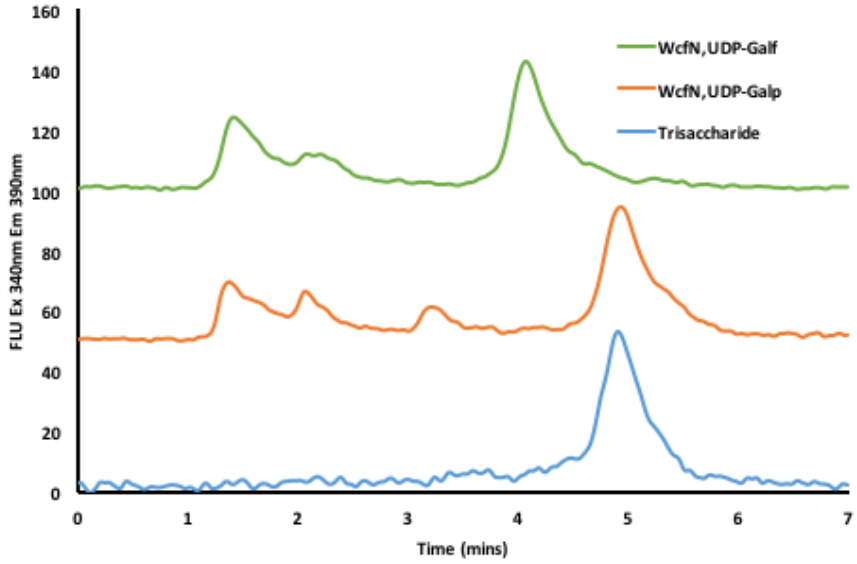


Figure 2.13 WcfN utilizes UDP-Gal*f* as the donor sugar to complete the synthesis of the tetrasaccharide. WcfN does not use UDP-Gal*p*. HPLC conditions used were, 32% 1-propanol with 100 mM ammonium bicarbonate as the mobile phase.

From the above mentioned assay, it was demonstrated that WcfN catalyzed the addition of the Gal*f* residue to form the complete repeat unit of the CPSA tetrasaccharide. WcfN was also tested to see if it could use Gal*p,* and the results exhibited that it did not use Gal*p* as a donor sugar, confirming that it transferred only the furanose form of galactose. Furthermore, when WcfN was coupled with WcfM, putative galactopyranose mutase, in the presence of UDP-galactopyranose (Figure 2.14) the expected product was observed only in the presence of WcfM. This assay also strengthened the hypothesis that

WcfM is the galactopyranose mutase encoded by the CPSA gene operon, and its product is required by WcfN. WcfN product was purified using HPLC, and the product identity was confirmed by negative ion mode ESI-MS (Figure 2.15). In the MS analysis one can also see peaks corresponding to the trisaccharide at 1322.67, the pyruvylated disaccharide at 1119.67. The peak corresponding to 1200.59 most likely is the unpyruvylated trisaccharide, where the pyruvyl group was knocked off during ionization. Similarly, the peak at 1033.17 corresponds to just the bactoprenyl associated disaccharide.

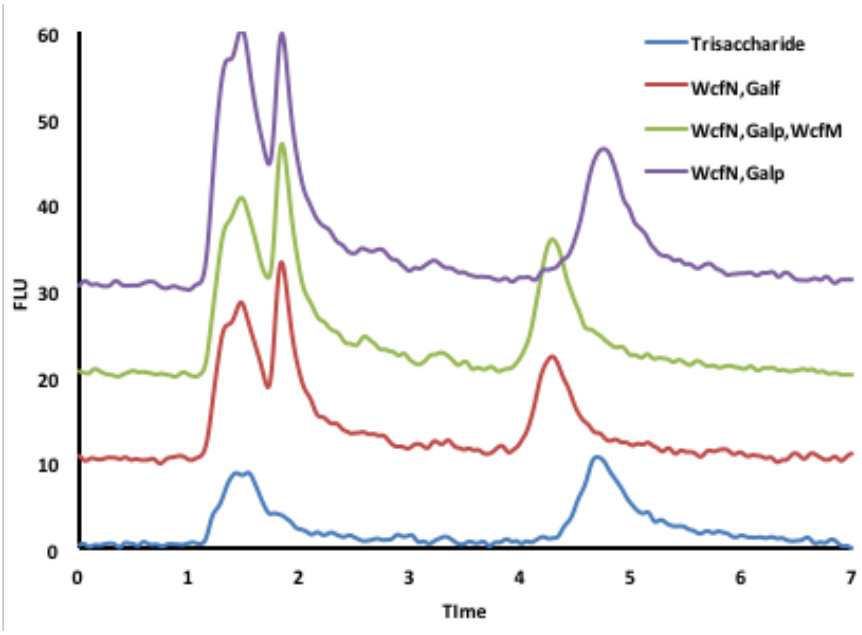


Figure 2.14 WcfN utilizes UDP-Gal*f* as the donor sugar to complete the synthesis of the tetrasaccharide. WcfN does not use UDP-Gal*p*. HPLC conditions used were 32% 1-propanol with 100 mM ammonium bicarbonate as mobile phase.

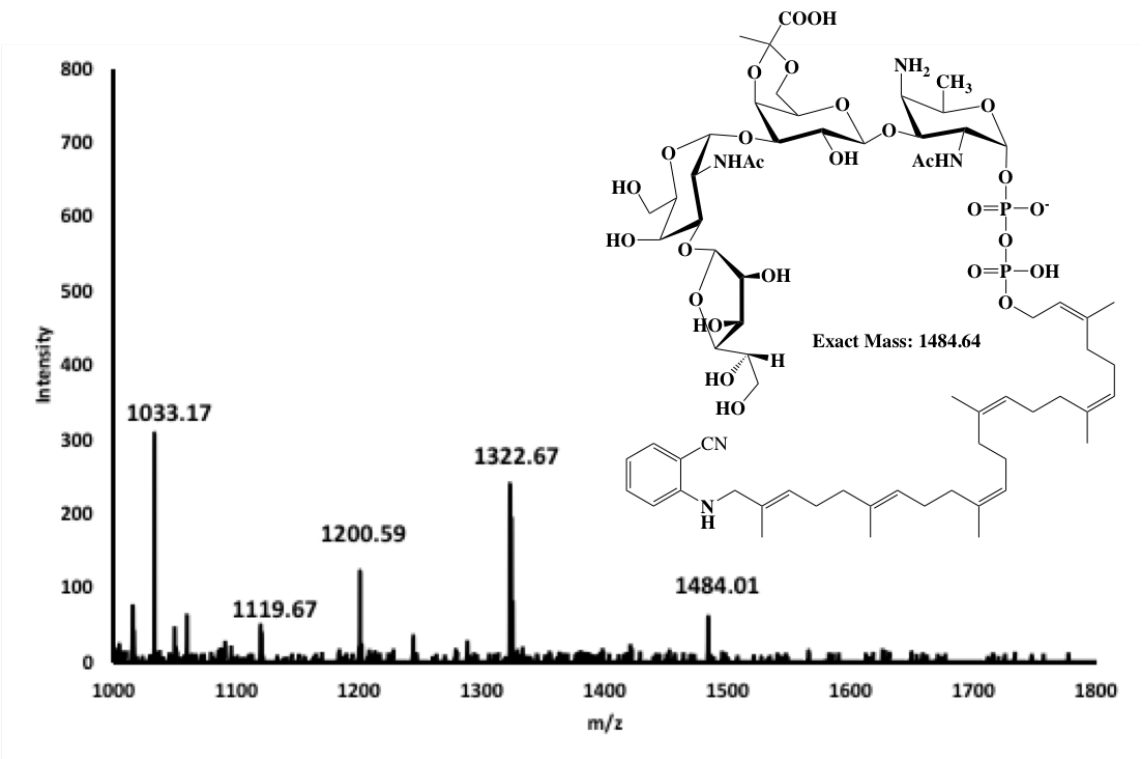


Figure 2.15 MS of 2CNA-B(4*Z*)PP-AADGal-pyrGal-GalNAc-Gal(*f*). 2CNA-B(4*Z*)PP-AADGal-pyrGal-GalNAc-Gal(*f*) was purified using RP-HPLC. The purified product was characterized for mass on Thermo VELOS Pro Dual-Pressure Linear Ion Trap in negative mode. The exact mass of the product is 1484.65, which matched with the 1484.01, mass obtained from the VELOS.

2.4.5 Single pot synthesis of the isoprene-linked tetrasaccharide.

A key advantage of the enzymatic synthesis of polysaccharides is the possibility that one can take advantage of the specificity of these proteins to do multiple synthetic transformations in a single pot reaction, without purification of the intermediates. Previously a one pot synthesis of the bactoprenyl diphosphate linked monosaccharide (4NABPP-AADGal) was done[50]. On the same line, in this study we were successful in producing the full tetrasaccharide sequentially in a single pot. Sequential addition of the enzymes was done at an interval of every 30 minutes, starting with WcfS, 2CNA-BP, and

UDP-AADGal. Then WcfQ, WcfO, WcfP and WcfN along with the appropriate donor sugars were added stepwise to give the full tetrasaccharide (Figure 2.16).

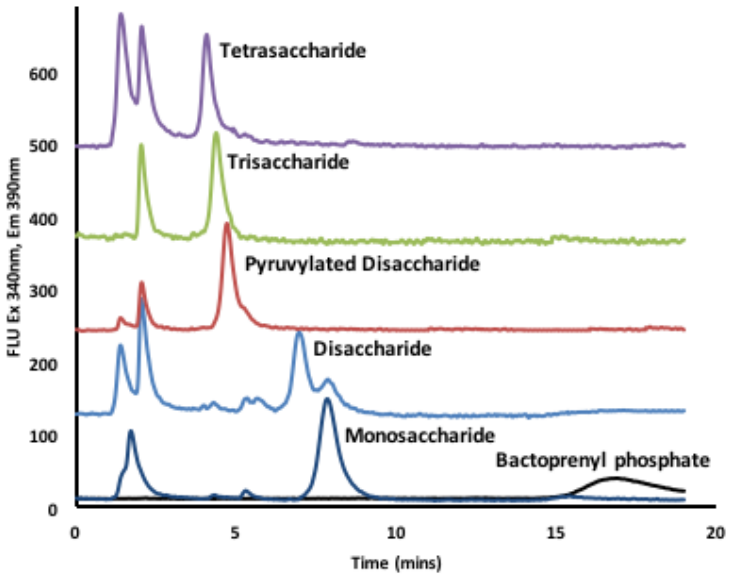


Figure 2.16 Sequential addition of enzymes in CPSA biosynthetic pathway leading to the formation of the tetrasaccharide. HPLC conditions used were 35% 1-propanol with 100 mM ammonium bicarbonate as mobile phase.

2.5 Summary

The project had proposed a different route towards the synthesis of CPSA. The proposal was to build CPSA biosynthetically in an *in vitro* system with starting materials that are commercially available. The study was an attempt to gain an insight into the biosynthetic pathway of CPSA and also allow the characterization of previously uncharacterized enzymes with functions that are currently poorly understood. A hypothetical biosynthetic pathway leading to CPSA was constructed, starting with the assembly of the tetrasaccharide unit on a bactoprenyl phosphate anchor, and then polymerization of the tetrasaccharide. During the course of the project, specific role for

each gene involved in the biosynthesis of CPSA was identified, leading to the assembly of the repeat unit *in vitro*.

CPSA is assembled on a bactoprenyl anchor inside a bacterial cell. In the *in vitro* synthesis, fluorescent bactoprenyl analogues were used to construct CPSA tetrasaccharide unit. The initiating hexose phosphate transferase was already identified and characterized previously. The assembly of the remaining three sugars in the CPSA unit was performed in this study. WcfQ was identified as the first glycosyltransferase, which transferred galactose to the isoprenoid linked monosaccharide. Although WcfP was also capable of transferring galactose, WcfQ was more effective in it, as both less total protein content and less donor sugar was required in the reaction assay to form the disaccharide[65]. Hence it was identified as the galactosyltransferase in the CPSA biosynthetic pathway. WcfQ was also hypothesized to be an inverting glycosyltransferase. Since the published structure of CPSA states the linkage between the first and second sugar is in a beta configuration with respect to the second sugar, WcfQ must be inverting the configuration of the donor substrate. This further strengthened WcfQ position as the galactosyltransferase.

As stated previously in the introduction, WcfP is related to the GT_4 family of proteins, suggesting that it is a retaining glycosyltransferase. It was therefore more likely that WcfP catalyzed GalNAc transfer to the galactose, but it was not known if it transferred GalNAc to the unpyruvylated disaccharide or the pyruvylated disacsharide. Analyzing both WcfN and WcfP, it was demonstrated that WcfP transferred only UDP-GalNAc to the pyruvylated disaccharide. WcfP was also found to be specific in transferring the sugar only to the pyruvylated galactose disaccharide and not the other substrate.

In homology studies, WcfN was predicted to be a member of the GT_2 family. In

this family, few members have already been identified to transfer the furanose. It was also found to be homologous to the well characterized protein GlfT2. GlfT2 is the galactofuranosyltransferase found in *Mycobacterium tuberculosis*, and is involved in the synthesis of arabinogalactan cell wall[66]. WcfN was also hypothesized to be an inverting transferase, which inverts the stereochemistry of the anomeric carbon. WcfN was found to transfer the galactofuranose to the trisaccharide, hence completing the *in vitro* synthesis of the tetrasaccharide.

The biosynthesis of the tetrasaccharide was achieved in a single pot sequential reaction where the enzymes were incubated in a sequence, starting with the bactoprenyl phosphate and the sugar substrates. This marks a very significant step towards polysaccharide biosynthesis and the ease with which glycosyltransferases can be characterized using fluorescent analogues instead of studying their function using radiolabeled substrates and TLC, which consumes more time.

CHAPTER 3: PYRUVYLTRANSFERASE

3.1 Introduction

Polysaccharides of various prokaryotes are covalently linked with variable combinations of sulfates and pyruvates, for example, *Rhizobium leguminosarum*: 4,6-pyrGalactose and 4,6- pyrGlucose, *Bacillus anthracis*: 4,6-pyrManNAc, and *Xanthomonas campestris*: 4,6-pyrMannose, more modifications provided in Table 3.1[17]. These modifications provide a highly negative charge of these polysaccharides, which is often essential for function[67]. For example, when the pyruvyltransferase PssM, responsible for the pyruvate modification in the *R. leguminosarum* exopolysaccharide was deleted, the bacterium was found to be ineffective in infecting pea plants to initiate the formation of root nodules. This led to formation of aberrant root nodules, which were unable to fix nitrogen[15]. Moreover, some studies have linked the pyruvic acetals in oligo- and polysaccharides to their immunological properties[17,20].

Pyruvylation of sugars is fairly common yet an extensive search of the literature reveals little on successful isolations of an enzyme responsible for this sugar modification. However, very recently a family of genes has been identified that appear to be involved in pyruvate transfer reactions in prokaryotes. A publication in 2013 showed successful purification of pyruvyltransferase Pvg1p from the eukaryote *Schizosaccharomyces pombe*. This group demonstrated the activity of Pvg1p on beta-nitrophenyl galactose, a substrate analogue of galactose[54]. Apart from this eukaryotic pyruvyltransferase Pvg1p and the

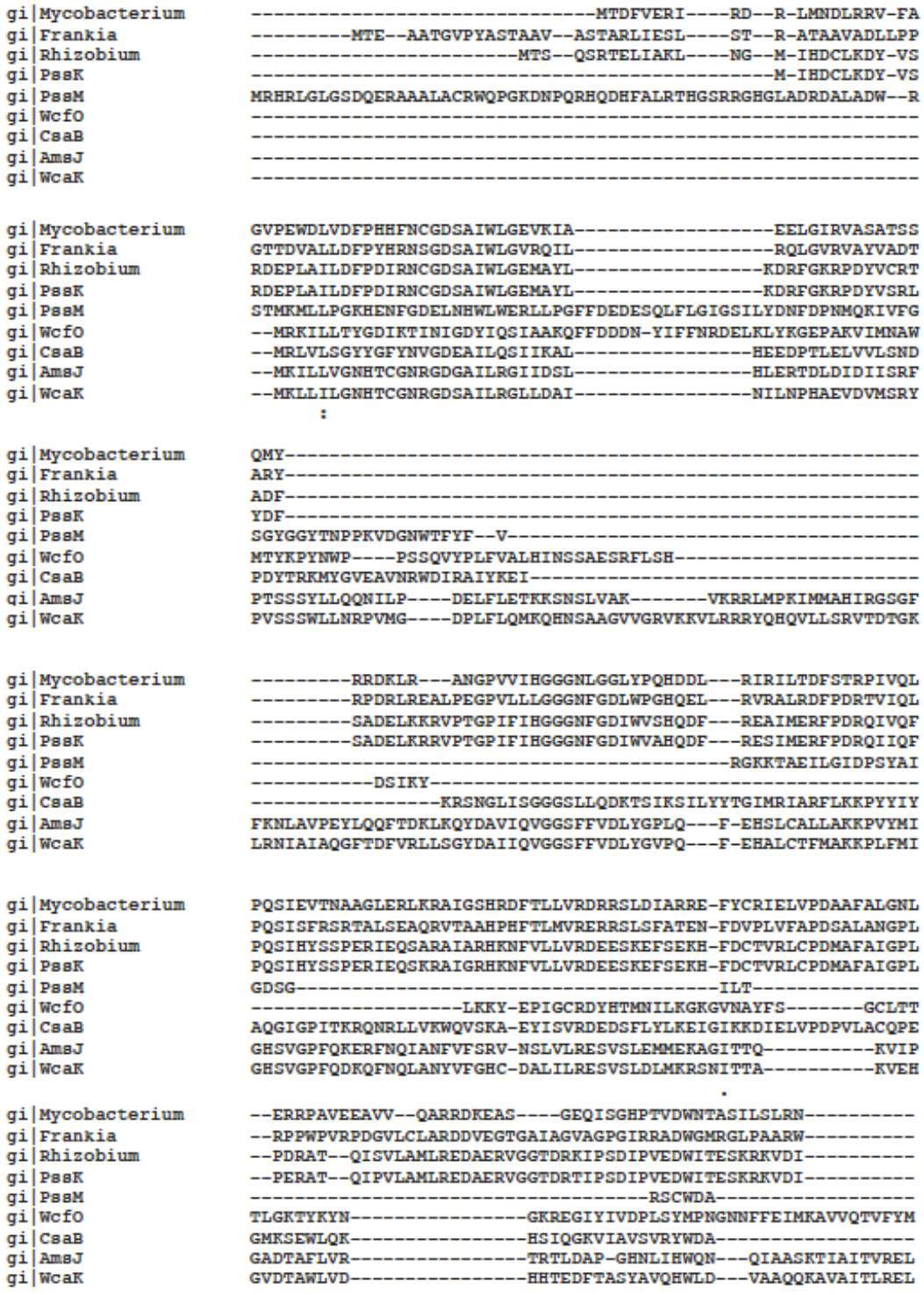
prokaryotic pyruvyltransferase PssM from *R. leguminosarum*, no other pyruvyltransferases have been characterized[55]. In both of these studies it was not elucidated clearly what is the natural substrate of the pyruvyltransferase, and where does a pyruvyltransferase function in context of a biosynthesis pathway. Furthermore, multiple sequence alignment studies showed very little similarity in the sequences of various hypothesized pyruvyltransferases (Figure 3.1). Although recently the Schaeffer group has reported the functional characterization of the pyruvyltransferase CsaB from *Paenibacillus alvei*. It modifies the single cell wall polymer, which enables the binding of S-layer proteins through SLH domain interaction.

3.2 Putative pyruvyltransferase in CPSA gene operon

Among the eleven proteins encoded in the CPSA gene operon, one of the genes transcribes a hypothesized pyruvyltransferase based on homology studies performed using pBLAST[48]. There is little sequence similarity to other known proteins with the *wcfO* gene product. WcfO has very minimal sequence identity to the two characterized pyruvyltransferases Pvg1p from *S. pombe* and PssM from *R. leguminosarum*. The activity of CPSA is dependent on its zwitterionic character in which the –AADGal amino group is positively charged while the pyruvate is negatively charged[36]. Due to the fact that all other sugar modifying enzymes and glycosyltransferases required for CPSA biosynthesis have been located in the CPSA biosynthesis operon, it is proposed that the *wcfO* gene product is likely responsible for the pyruvylation modification required for the formation of the second sugar in the CPSA tetrasaccharide repeat unit. Hence through this study, it is an endeavor to study a potential member of this new family of proteins.

Table 3.1 The pyruvate modifications present in polysaccharides reported in literature. For most of them, the protein responsible for the modification has not been identified yet. (pyr-pyruvyl, GlcA - Glucouronic Acid, Gal - Galactose, Glc - Glucose, GlcNAc - N acetyl Glucosamine, ManNAc- N acetyl Mannose, CPS – capsular polysaccharide, EPS - exopolysaccharide)

Organism	Structure Type	Modification	Gene
PLA Klebsiella pneumoniae	CPS	2,3-pyrGlcA	*orf8*
Rhizobium phaseoli	EPS	3,4-pyrGal	-
Klebsiella serotype 13	CPS	3,4-pyrGal	-
Escherichia coli K47	CPS	3,4-pyrGal	-
Rhizobium trifolii	EPS	4,6-pyrGlc	*pssk*
Rhizobium trifolii	EPS	4,6-pyrGal	*pssm*
Mycobacterium smegmatis	Glycopeptidolipid	4,6-pyrGlc	-
Mycobacterium avium	Glycopeptidolipid	4,6-pyrGlc	-
Schizosaccharomyces pombe	Oligosaccharide	4,6-pyrGal	*pvg1p*
Escherichia coli	CPS	4,6-pyrGal	*wcak*
Gelidium	EPS	4,6-pyrGal	-
Streptococcus pneumoniae	CPS	4,6-pyrGlcNAc	-
Bacillus anthracis	Cell wall	4,6-pyrManNAc	*csaB*
Xanthomonas campestris	EPS	4,6-pyrMannose	-


gi|Mycobacterium
gi|Frankia
gi|Rhizobium
gi|PssK
gi|PssM
gi|WcfO
gi|CsaB
gi|AmsJ
gi|WcaK

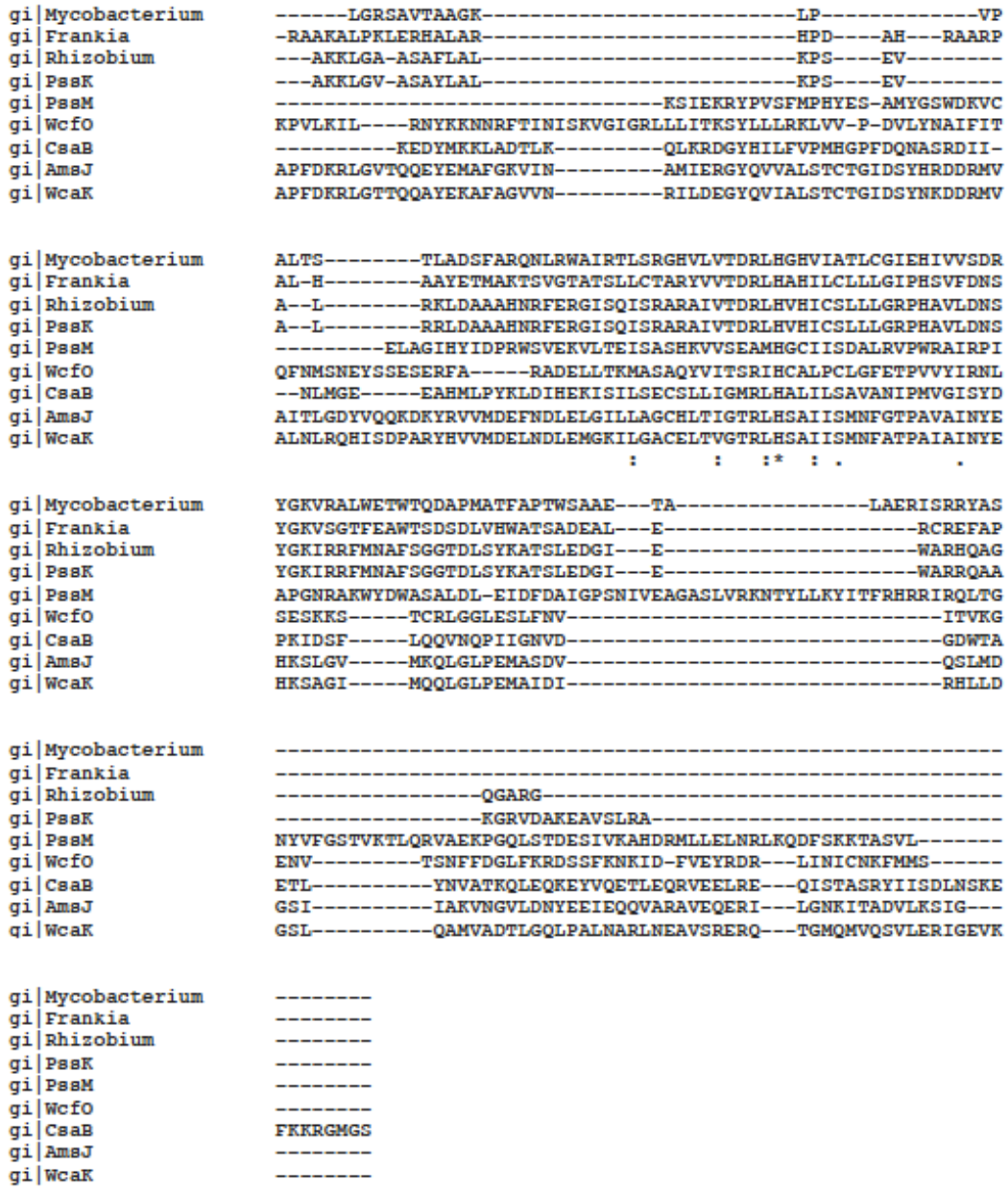


Fig 3.1 Sequence analysis of WcfO.

There are two potential routes through which WcfO may be able to pyruvate galactose. The first hypothesis is transformation of uridine diphosphate linked galactose (UDP-Gal) to UDP- 4, 6-pyruvate-galactose (Figure 3.2). Phosphoenolpyruvate (PEP) was used as the donor molecule to provide the pyruvate moiety. It would be readily able to

provide the energy upon phosphate cleavage for the masking of the 4- and 6- hydroxyls of the galactose sugar. Hence the hypothesis that WcfO might use PEP as a pyruvate donor. The second hypothesis is that, WcfO does the pyruvate addition once the galactose is attached to the bactoprenyl linked monosaccharide. The modification can also happen on the assembled trisaccharide or the tetrasaccharide (Figure 3.3).

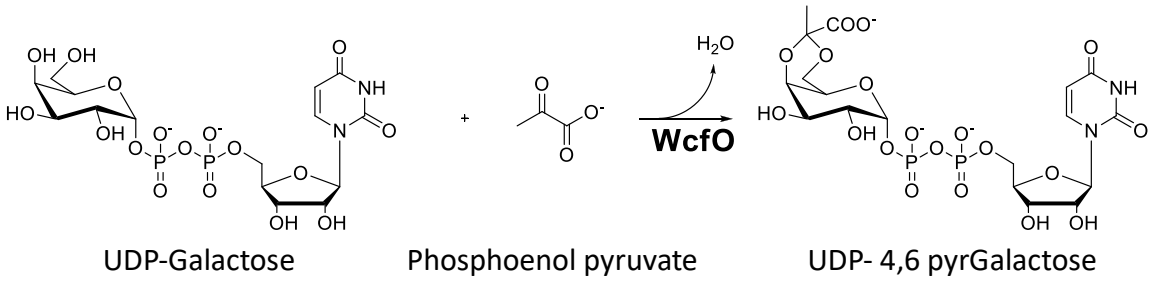


Figure 3.2 Potential hypothesis one for pyruvate modification of UDP-Galactose by WcfO

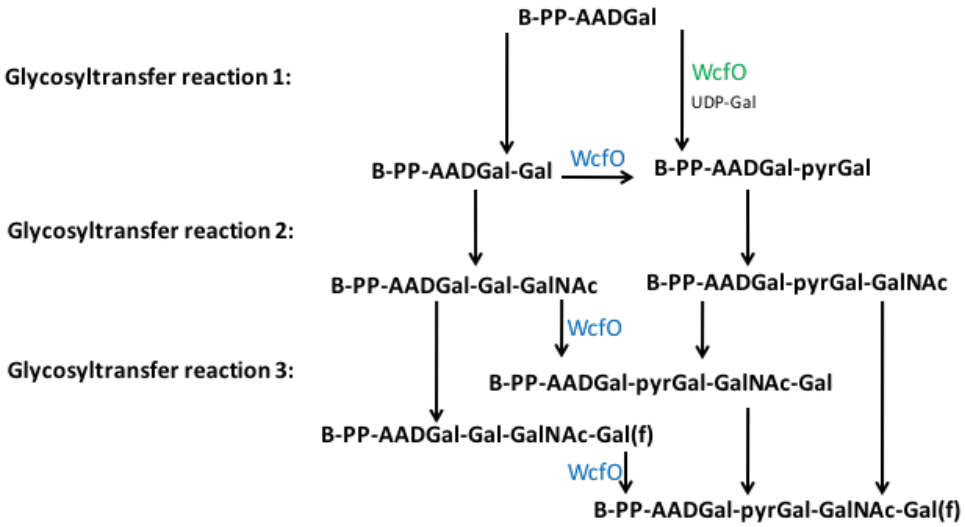


Figure 3.3 Second hypothesis for pyruvate modification of UDP-Galactose by WcfO. The WcfO shown in blue is where it can potentially pyruvate the galactose (Gal) residue. WcfO in green is hypothesis one, where it pyruvylates just UDP-Galactose.

3.3 Experimental methods

3.3.1 Cloning and expression of WcfO

Polymerase chain reaction amplification of *wcfO* was performed using *B. fragilis* genomic DNA (ATCC 25285). The *wcfO* gene was digested with BamHI and XhoI. It was next ligated into a pET-24a vector digested with the appropriate restriction enzyme for each gene. Chemically competent *E. coli* DH5α cells were transformed with the ligated vector and Kanamycin resistant colonies were selected. Plasmids were isolated from the kanamycin resistant colonies and sequenced to confirm introduction of each gene (Eurofins-Operon). Chemically competent E. coli BL-21(DE3) RIL cells (Agilent) were then transformed and again positive colonies were selected by Kanamycin resistance. Cells containing WcfO encoding plasmids were cultured in 5 mL of Luria Broth (LB) overnight at 37° C and later used to inoculate 0.5 L of LB. Cells were grown to an O.D.600 of 0.6 and the temperature was decreased to 26.5° C for 30 min. Isopropylthiogalactoside (IPTG) was added at a final concentration of 1 mM. Cells were allowed to incubate for an additional three hours, then were harvested by centrifugation at 4° C for 20 min at 5000 x G, the supernatant was removed and cells were stored at -80° C for later use. The cloning and initial expression of WcfO was performed by my advisor Dr. Jerry M. Troutman.

3.3.2 WcfO isolation as membrane fraction

Protein expressing cells were thawed in 20 mL of 50 mM Tris-HCl (pH=8.0) containing 200 mM NaCl and 20 mM imidazole then sonicated on ice for 3 minutes (total) with a pulse of one second on and one second off. Unbroken cells were removed by centrifugation at 5000 x G for 15 minutes at 4° C. The supernatant containing membrane and cytosolic components was further spun at 150,000 x G for an hour at 4° C. The pelleted

membrane components were homogenized into 2 mL of 50 mM Tris-HCl (pH=8.0) 200 mM NaCl then 10-50 μL aliquots were stored at -80° C. Total protein concentration in the membrane fraction was measured using a Bradford assay with BSA as a standard[62]. The presence of the overexpressed protein was confirmed by SDS-PAGE and Western Blot analysis with an anti-His antibody.

3.3.3 Isolation of WcfO from membrane fraction

Membrane fraction (1 mL) was homogenized in 20 mL of Tris-HCl supplemented with 1% Triton-X-100 (or 0.1% DDM) and 200 mM NaCl. The homogenized fraction was then spun at 150,000 X G for an hour at 4° C. The pellet after centrifugation was again homogenized as described above. The supernatant from the high speed spin was mixed with 1 mL Ni-NTA Agarose (5 Prime) for one hour at 4° C then was poured through a column. The resin was washed with a solution of 50 mM Tris-HCl (pH=8.0), 50 mM imidazole, 200 mM NaCl 4 x 3 mL. After washing, 0.35 mL of elution buffer containing 50 mM Tris-HCl (pH=8.0), 500 mM imidazole, 200 mM NaCl was passed through the column. Another 1.75 mL of elution buffer was then passed through the column and collected 5 times to obtain purified protein. The presence of purified protein was confirmed by Western Blot analysis with an anti-His antibody.

3.3.4 General method for WcfO assays

Assay for pyruvyltransferase activity was set with 1 μL of membrane fraction of total protein membrane fraction (10.5 mg/ml) in a 100 μL solution containing 5 μM 2CNA-B(4Z)PP-AADGal-Gal, 1% Triton, 10 mM MgCl2, 10% DMSO, 25 mM Bicine (pH=8.3) and 10 mM phosphoenolpyruvate. After 1 hour each mixture was analyzed by HPLC with a mobile phase of 35% 1-propanol and 100 mM ammonium bicarbonate.

A reaction solution was also prepared containing 5 μM 2CNA-B(4Z)PP-AADGal-Gal, 10 mM phosphoenolpyruvate, 25 mM Bicine, 10 mM MgCl2, 1% Triton-X-100, 10% DMSO, and 10 μg/mL purified WcfO then was analyzed by HPLC for product turnover. A similar reaction was also set to test if WcfO catalyzed pyruvate modification of 2CNA-B(4Z)PP-AADGal-Glc.

All the reactions were analyzed on HPLC, an Agilent 1100 HPLC system, equipped with diode array and fluorescence detectors. All analysis used isocratic conditions of 35% 1-propanol with 100 mM ammonium bicarbonate for cis-4 isoprenoids, and 50% 1-propanol with 100 mM ammonium bicarbonate for cis-6 isoprenoids. All chromatography involving WcfO assays was performed on a reverse phase C18 Agilent Eclipse XDB-C18, 5μm, 4.6 x 150 mm column.

3.3.5 WcfO assay with unlinked UDP-Galactose

Assay was set with 10 μg/mL of membrane fraction and purified WcfO in a 100 μL containing 25 mM Bicine, 10 mM MgCl2, 1% Triton-X-100, 300 μM UDP-Gal, 10 mM phosphoenolpyruvate. The reactions were incubated at room temperature for 4 hours and analyzed on CE for product formation. After the incubation 8 μL of the reaction components were injected by pressure injection and separated at 15 kV for 25 minutes through a 48.0 μm inner diameter capillary at a length of 40.2 cm with 25 mM sodium tetraborate (pH 9.26) running buffer. The starting materials as well as any products covalently linked to UDP were detected at 260 nm. NAD+ was added as an internal standard to normalize electrophoretic mobility between analyses allowing us to account for electrophoretic drift from injection to injection.

3.3.6 WcfO assay with α and β nitrophenyl sugars

A solution was prepared containing 1 mM α-p-nitrophenyl galactopyranoside or β-o-nitrophenyl galactopyranoside, 5 mM phosphoenolpyruvate, 25 mM Bicine, 10 mM MgCl2, 1% Triton-X-100, and 10 μg/mL purified WcfO, which was then analyzed by HPLC. HPLC analysis was performed on the same C18 column as above using a gradient method in which over 30 minutes 12 % 1-propanol 88 % 100 mM ammonium bicarbonate was raised to 100 % 1-propanol. Sugar was detected at an absorbance of 405 nm.

3.3.7 WcfO product turnover optimization

A standard reaction was prepared containing 25 mM Bicine, 10 mM MgCl2, 1% Triton-X-100, 10% DMSO, 10 mM phosphoenolpyruvate, 5 μM 2AA-B(6Z)PP-AADGal-Gal. Each reaction component was varied from these concentrations and the reactions contained 5 μg/mL of pure WcfO protein. Various detergents like, 0.5% octylthioglucoside (OTG), 0.5 mM Tween, and 0.1 mM Brij-35 were also used to test the WcfO activity. The concentration of the detergents was used based on their critical micellar concentration.

Divalent metals were screened with WcfO in reactions containing 25 mM Bicine, 10% DMSO, 10 mM MgCl2, MnCl2 or CaCl2, 1% Triton-X-100, 10 mM, 5 μM 2AA-B(6Z)PP-AADGal with 5 μg/mL pure protein.

All alternative buffers HEPES (pH=7.4), MOPS (pH=6.3) and Tris-HCl (pH=8.0) were of 25 mM final concentration and were screened in reactions containing 5 μg/mL pure protein WcfO, 10% DMSO, 1% Triton-X-100, 10 mM MgCl2, 5 μM 2AA-B(6Z)PP-AADGal and 10 mM phosphoenolpyruvate.

3.3.8 Kinetic assay of WcfO

Reaction mixture of 500 µl volume containing 0.45 µM of 2AA-B(6Z)PP-AAD-Gal-Gal or 2AA-B(6Z)PP-AAD-Gal-Glc and varying concentrations of PEP (0.1 mM, 0.25 mM, 0.5 mM, 0.75 mM, 1 mM, 3 mM and 5 mM) were incubated with 0.0095 µM of enzyme, and aliquots were taken at intervals of 15 seconds for one and half minutes. The aliquots were quenched with 1-propanol. The quenched aliquots were then analyzed by HPLC with fluorescence detection at excitation 350 nm and emission 450 nm. The column used was the same C18 described above, and the mobile phase was 50 % 1-propanol in 100 mM ammonium bicarbonate. The product turnover was calculated based on peak integrals and the initial rate was determined. Initial rates were fit to the Michaelis-Menten equation to determine apparent kinetic constants kcat and Km. Fits and error were based on all data from seven PEP concentrations performed three times. The isoprenoid linked substrates used for the assays were prepared as described previously.

3.3.9 Characterization of products using MS

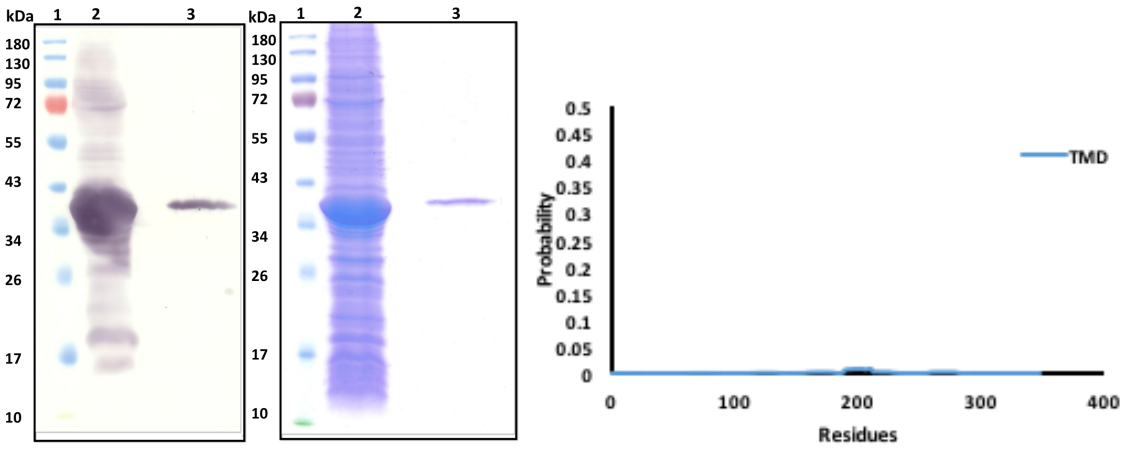
58 nanomoles of 2CNA-B(4Z)PP-AADGal-Gal, 3.5 mM sodium cholate, 10 mM MgCl2, 10% DMSO, 25 mM Bicine (pH=8.3) and 10 mM phosphoenolpyruvate was incubated with 5 µl of WcfO membrane fraction (total protein concentration 10.5 mg/ml). The product was isolated using 35% 1-propanol with 100 mM ammonium bicarbonate. A similar one pot reaction was set up to isolate the 2CNA-B(4Z)PP-AADGal-Glc. MS was performed on the fractions collected containing the bactoprenyl-linked sugar products from the HPLC on a Thermo VELOS Pro Dual-Pressure Linear Ion Trap using electrospray ionization (ESI) introduced into the capillary with a 0.200 mL/min flow rate of acetonitrile.

Negative mode was used with a capillary temperature of 300 °C and a spray voltage of 4.00 kV. Mass spectrometry on the samples were performed by Katelyn Erickson.

3.4 Results

3.4.1 WcfO isolated as a membrane fraction and pure protein

The CPSA biosynthesis gene cluster has one gene that emerges homologous to pyruvyltransferase family members based on a pBLAST search on the encoded protein sequences. This gene, wcfO, encodes a 347 amino acid protein with less than 45% sequence identity to any other protein in the NCBI database, and has no significant similarity to a biochemically characterized protein. However, since there were no other probable candidates in the CPSA biosynthesis gene cluster and there was an apparent homology to the pyruvyltransferase family of proteins, the *wcfO* gene was amplified from genomic DNA, and inserted into a pET-24a vector for expression in *E. coli* BL-21(DE3) cells. As recently observed with other proteins encoded by the CPSA biosynthesis gene cluster, WcfO fractionated upon cell lysis with the bacterial cell membrane even though there were no clear transmembrane domains based on sequence analysis (Figure 3.4 b). Total protein concentration was 10.5 mg/mL in WcfO membrane preparation. Unlike the glycosyltransferases encoded by the cluster, WcfO could be separated from the membrane fraction and purified to homogeneity (Figure 3.4 a). The total purified protein yield from a 500 ml expression culture was 1.31 mg/ml.



a b

Figure 3.4 (a). Lane 1, 2 and 3 represent protein ladder, membrane fraction and solubilized protein respectively. On the left is the Western Blot with Anti-Histidine specific antibody. The right image is the Coomassie stain. The protein from the membrane fraction was solubilized to purity. (b). The TMHMM server did not predict any clear transmembrane domains in the protein sequence

3.4.2 WcfO catalyzes the addition of pyruvate to isoprenoid-linked disaccharide

The hypothesized function of WcfO was to transfer pyruvate moiety to the bactoprenyl diphosphate-linked disaccharide, BPP-AADGal-Gal or to pyruvylate only the unlinked UDP-Galactose. In the study performed by Yoritsune et al in 2013, they had reported using phosphoenolpyruvate as the pyruvate donor, and were able to utilize a nitroaromatic-linked galactose as the acceptor for the transferase Pvg1p. To test the hypothesis that WcfO might also be able to pyruvylate an unlinked UDP-Galactose, α-nitrophenyl galactopyranoside and β-nitrophenyl galactopyranoside were used as potential acceptors for the pyruvate, and they were incubated with WcfO and phosphoenolpyruvate. The assay was easily monitored because of the nitroaromatic sugars (Figure 3.5). The data obtained demonstrated no change in the HPLC retention times of these sugars, which

suggested that either WcfO was not a pyruvyltransferase as hypothesized or, unlike the previously characterized pyruvyltransferase Pvg1p, it was not able to recognize the analogues of the UDP-Galactose and modify it.

In addition, when WcfO was mixed with the potential pyruvate donor, phosphoenolpyruvate, and uridine diphosphate (UDP)-linked galactose, again no product formation was observed by capillary electrophoresis (CE) (Figure 3.6) or by HPLC analysis (Figure 3.7). These two results combined with the results of the nitro-aromatic linked galactose suggested that, there might be a different substrate which is pyruvylated by WcfO, because in the previous chapter we noted that neither WcfP nor WcfN were able to use the bactoprenyl diphosphate linked disaccharide as the acceptor substrate.

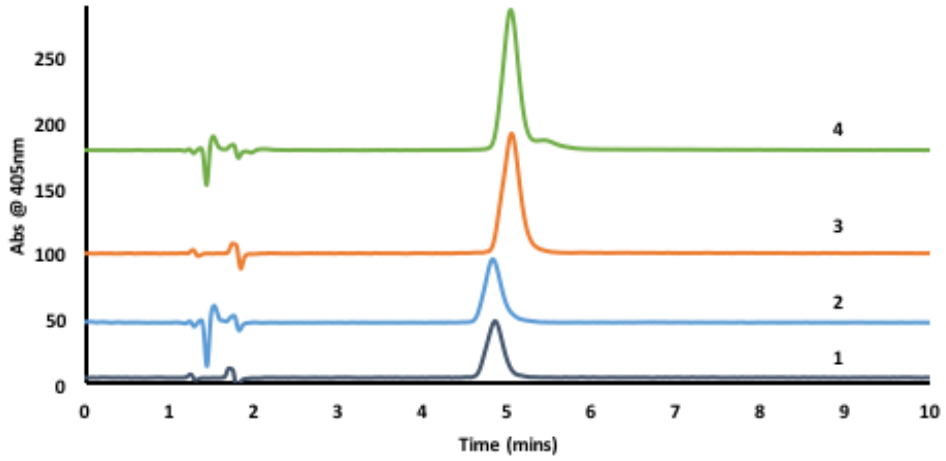


Figure 3.5 WcfO was functionally tested with β-o-NO2-phenyl-Gal (1) and α-p-NO2-phenyl-Gal (3). WcfO failed to show any product formation with these two substrates, β-o-NO2-phenyl–Gal and WcfO (2) and, α-p-NO2-phenyl-Gal and WcfO (4) indicating that it is likely not accepting unlinked sugar.

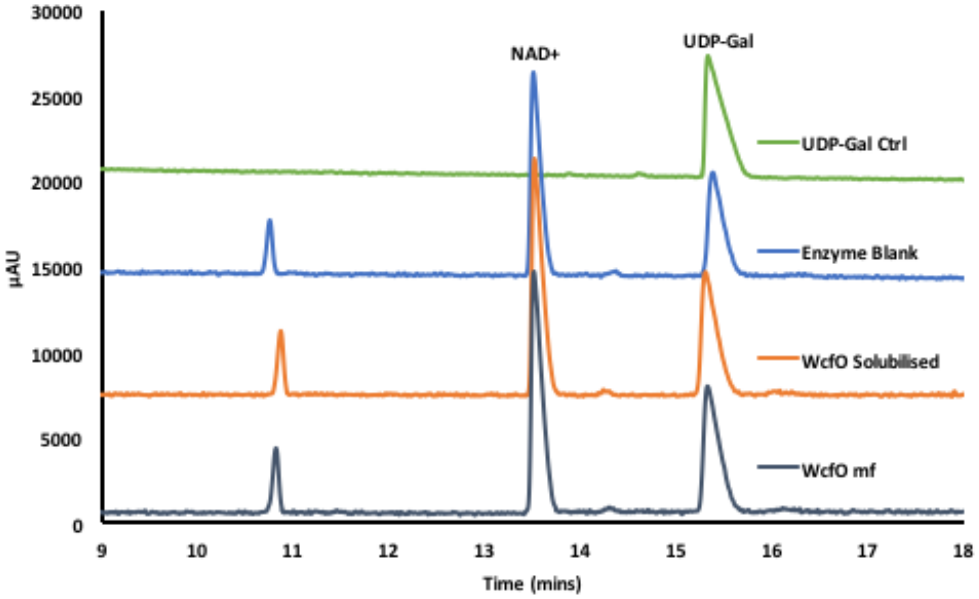


Figure 3.6 WcfO was tested with unlinked UDP-Gal and samples were analyzed using CE. No new peak was observed showing that substrate was not being utilized.

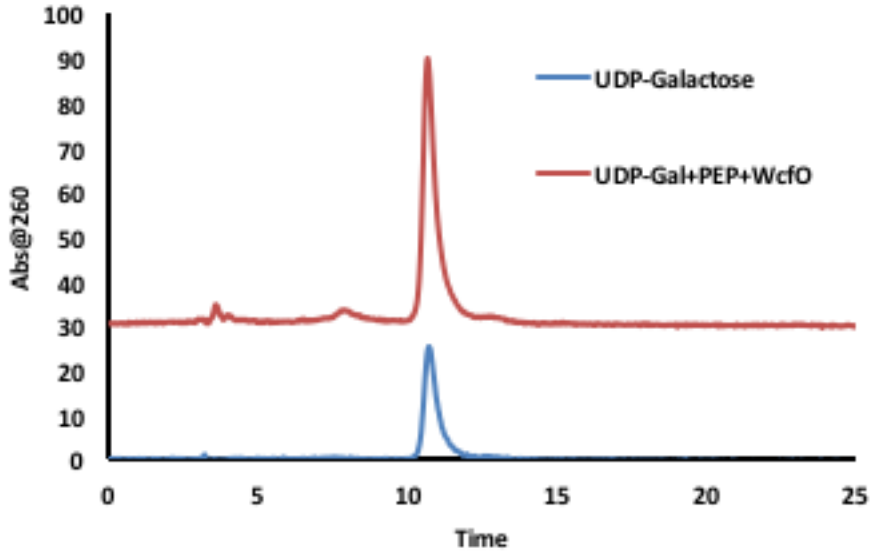


Figure 3.7 WcfO was tested with unlinked UDP-Galactose and samples were analyzed using HPLC. Once again, the reaction did not display any formation of a new peak suggesting that WcfO did not use unlinked galactose, unlike Pvg1p.

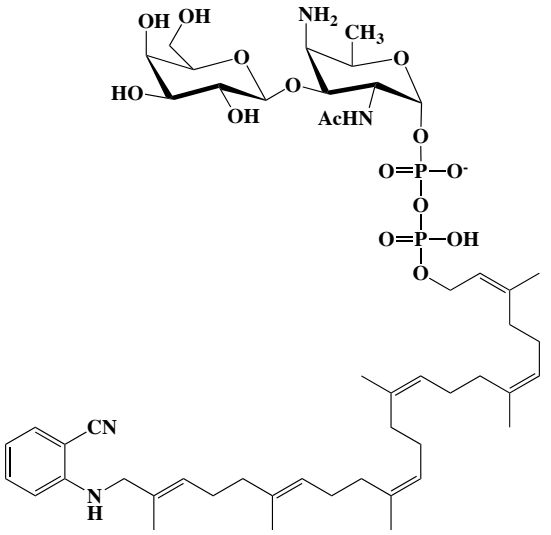


Figure 3.8 2CNA-B(4Z)PP-AADGal-Gal, substrate used to test WcfO.

Hence, when phosphoenolpyruvate was mixed with a fluorescent 2-nitrileanilino-bactoprenyl diphosphate-linked AADGal-Gal (2CNA-B(4Z)PP-AADGal-Gal) (Figure 3.8) a major shift in retention was observed in the presence of WcfO which was consistent with the introduction of the negatively charged pyruvate moiety (Figure 3.9). The product associated with this new HPLC chromatogram peak was isolated, and then characterized by electrospray ionization mass spectrometry (ESI-MS). The mass of the new product was consistent with the addition of a pyruvate acetal to 2CNA-B(4Z)PP-AADGal-Gal to form 2CNA-B(4Z)PP-AADGal-pyrGal (Figure 3.10).

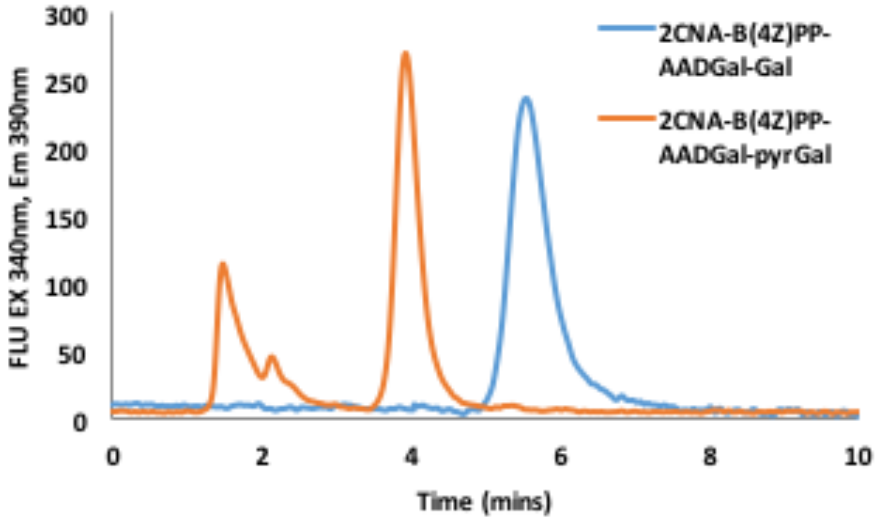


Figure 3.9 When WcfO was incubated with 2CNA-B(4Z)PP-AADGal-Gal and PEP, a new peak with a different retention time was observed, suggesting that the disaccharide might be its potential substrate.

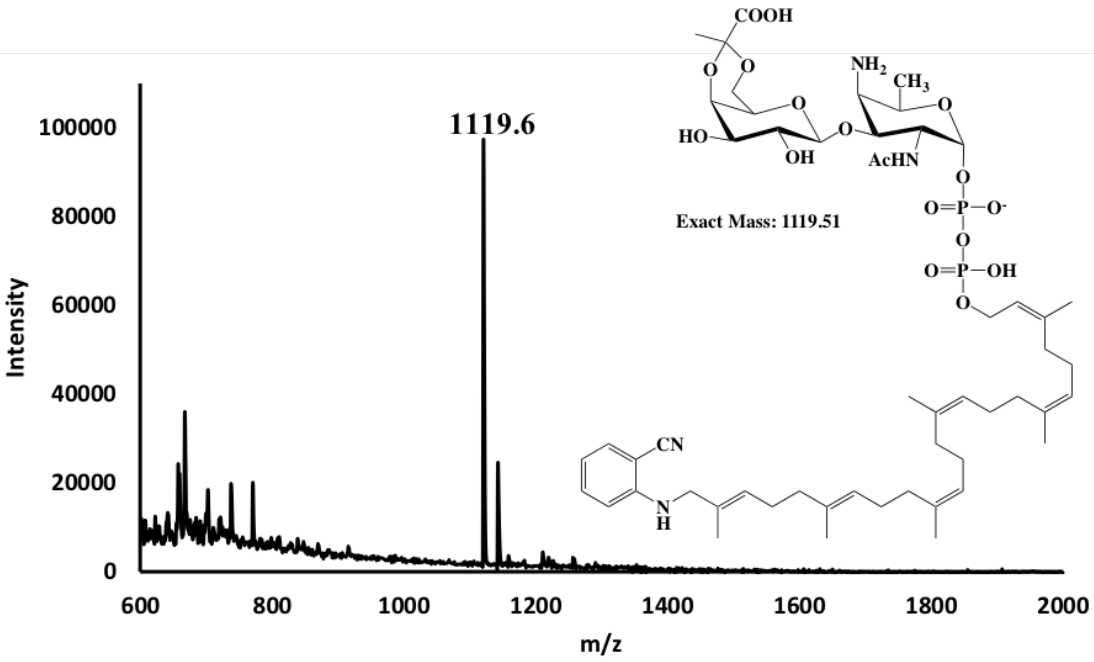


Figure 3.10 MS of 2CNA-B(4Z)PP-AADGal-pyrGal. The purified product 2CNA-B(4Z)PP-AADGal-pyrGal was analyzed for mass on Thermo VELOS Pro Dual-Pressure Linear Ion Trap in negative mode. The exact mass of the product is 1119.51, which matched with the 1119.6, mass obtained from the VELOS.

3.4.3 WcfO is not specific to the sugar attached to the isoprenoid linked AADGal

From the above experiments, it became clear that WcfO was selective for the structure of the pyruvate acceptor as no activity was observed on galactose that was not linked to AADGal. In the previous chapter, it was stated that the glycosyltransferase WcfQ could be forced to transfer glucose to AADGal using high concentrations of donor sugar and protein. WcfO was tested with a 2CNA-B(4Z)PP-AADGal-Glc substrate using conditions identical to those described above with the galactose substrate. Surprisingly it was found that while WcfO was selective for what the galactose was linked to, it was not selective for galactose over glucose (Figure 3.11). The retention time of the WcfO product with galactose was identical to the retention time of the glucose product suggesting similar structural characteristics. In addition, ESI-MS analysis was consistent with the formation of the pyruvate acetal on the glucose (Figure 3.12).

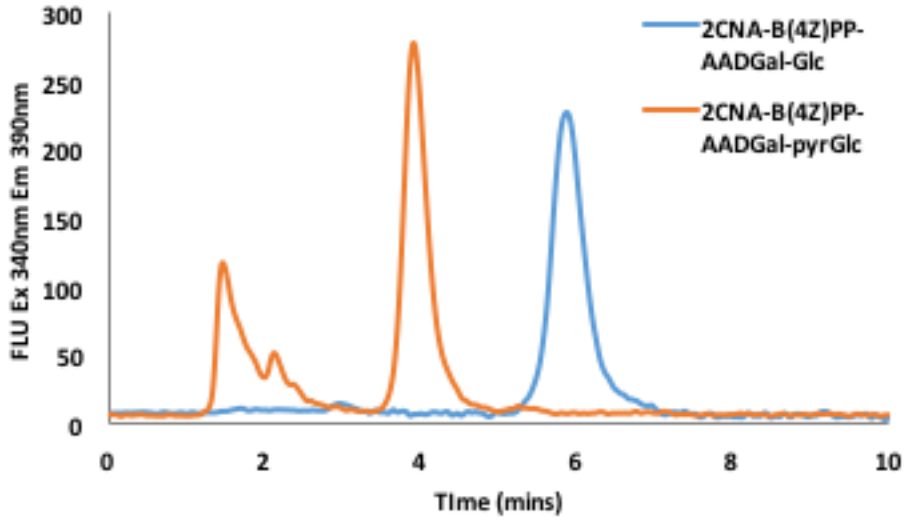


Figure 3.11 WcfO was also capable of pyruvating 2CNA-B(4Z)PP-AADGal-Glc, suggesting it does not discriminate between glucose or galactose.

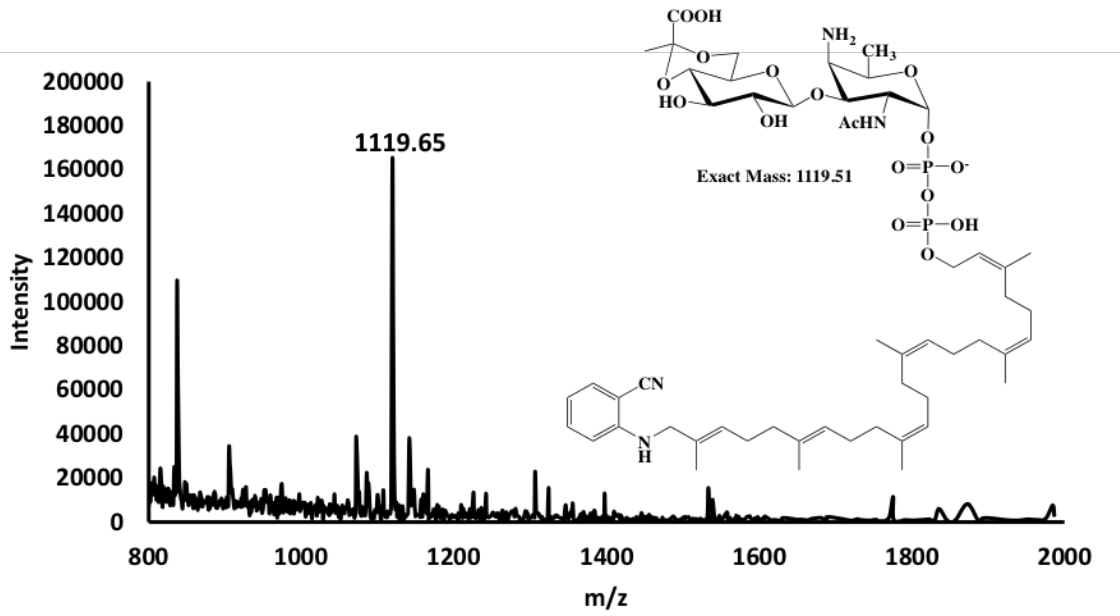


Figure 3.12 MS of 2CNA-B(4Z)PP-AADGal-pyrGlc. The purified product, 2CNA-B(4Z)PP-AADGal-pyrGlc was characterized for mass on Thermo VELOS Pro Dual-Pressure Linear Ion Trap in negative mode. The exact mass of the product is 1119.51, which matched with the 1119.65, mass obtained from the VELOS.

3.4.4 Optimization of WcfO reaction conditions

Reaction conditions were varied for WcfO assay to find out the most optimum conditions for complete product turnover. Of all the buffers tested, WcfO was observed to perform best in Bicine (pH=8.3), and had a preference for magnesium as the divalent cation. When homology studies was done with the putative pyruvyltransferase, one of the classification revealed that, WcfO might belong to a glycosyltransferase family. Therefore, it did not come as surprising that it required a divalent cation for its function. Presence of monovalent cations like sodium or potassium did not affect the turnover capacity of WcfO. The overall turnover of the substrate to the product had very little change when (dimethyl sulfoxide) DMSO was varied. Triton X-100 was the most suitable detergent for achieving full turnover from the substrate to product (Figure 3.13 a, b and 3.14). Various

concentrations of Triton X-100 were also tested to find the ideal concentration, which was found to be at 1%.

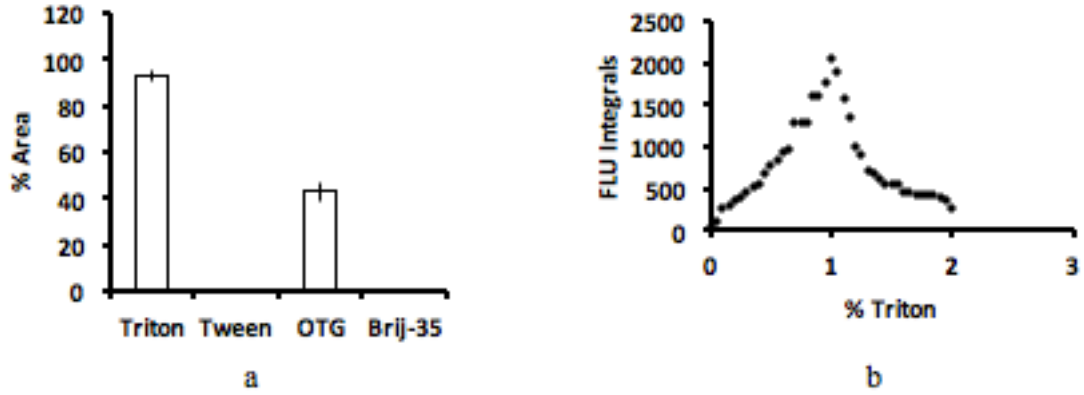


Figure 3.13 (a). Product turnover when different detergents are used. (b). Product turnover when Trition X-100 is varied.

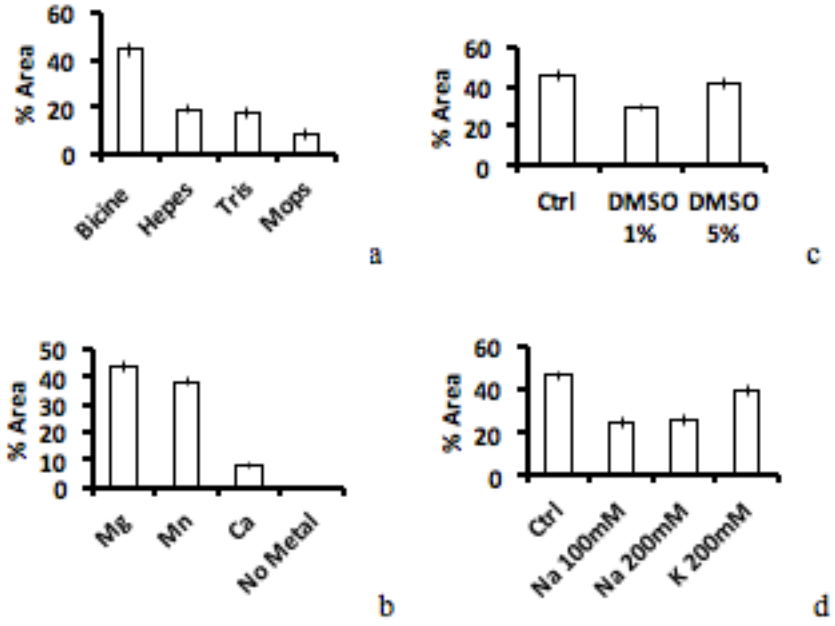


Figure 3.14 Optimizing conditions for WcfO Reaction. (a). Bicine had the most turnover of the four buffers tested. (b). A slight decrease was seen when using Manganese (Mn), while very less product was seen when using Calcium (Ca). (c). Increasing or decreasing DMSO did not have any dramatic effect. (d). With adding monovalent salts, extraordinary results were not seen.

3.4.5 Kinetic studies of WcfO with the two different disaccharides

WcfO demonstrated the capability of utilizing both galactose and glucose disaccharide for pyruvylation. To test whether bactoprenyl diphosphate linked AADGal-Gal was the preferred substrate of WcfO and therefore the more likely substrate in the bacterial cells, a kinetic analysis was performed using the fluorescent disaccharide with UDP-galactose and UDP-glucose as the sugar donors. Using a timed quench assay, initial reaction rates were determined then fit to the Michaelis-Menten equation. No significant difference was observed in the catalytic efficiency when 2AA-B(6Z)PP-AAD-Gal-Gal was used as substrate in comparison to 2AA-B(6Z)PP-AAD-Gal-Glc (Table 3.2).

Table 3.2 Kinetic evaluation of WcfO with respect to using the different disaccharides

	2AA-BPP(6Z)AADGal-Gal	BPP(6)AADGal-Glc
K_m	0.299 mM ± 0.049	0.291 mM ± 0.046
k_{cat}	0.249 /s ± 0.010	0.231 /s ± 0.009
k_{cat}/K_m	833 /M*s ± 141	795 /M*s ± 129

3.5 Summary

Pyruvate modifications have been frequently observed in various bacterial polysaccharides and also in few marine sponges and algae[17,68]. In spite of its ubiquitous presence, and role in various physiological and immunological processes, there is a dearth of literature on characterization of the proteins that are responsible for this modification. Although a couple of studies do exist that have biochemically characterized

pyruvyltransferases, but in these studies the function of pyruvyltransferases in context of a biosynthesis pathway was not explicitly demonstrated. The eukaryotic pyruvyltransferase purified from the yeast *S. pombe*, was shown to pyruvate a beta nitrophenyl galactose, and membrane fraction from *R. leguminosarum* was used to demonstrate pyruvylation but the substrate was inconclusive. In this report the proposed pyruvyltransferase WcfO encoded in the CPSA gene cluster was shown to have the pyruvate transfer activity. Unlike the other pyruvyltransferase, WcfO was not able to pyruvate unlinked uridine diphosphate galactose but under the same conditions it was able to add a pyruvate moiety to the galactose in the isoprenoid linked disaccharide. This is the first study in which a pyruvyltransferase was demonstrated to accept a native like substrate. We have demonstrated that WcfO was capable of attaching pyruvate to BPP-AADGal-Gal.

The study also uncovered a rather promiscuous WcfO when it came to accepting substrates for pyruvylation. WcfO seemed not to discriminate between the glucose or the galactose attached to the isoprene linked monosaccharide, and pyruvylated both the substrates equally. Both of these products were characterized using mass spectrophotometry. Kinetic studies with both the pyruvylated disaccharides did not show that WcfO preferred one substrate over the other. It will be rather interesting to know if the capsule of *B. fragilis* has a polysaccharide with this specific modification, or it is just an artifact of *in vitro* studies.

CHAPTER 4: GALACTOPYRANOSE MUTASE

4.1 Introduction

Polysaccharides composed of furanosyl residues are important constituents of many bacteria, protozoa, fungi, plants and archaebacteria[66,69]. The furanosyl constituents have also been identified in glycopeptides, glycolipids as well as nucleotide sugars. D-Galactose is by far the most widespread hexose in the furanose form in naturally occurring polysaccharides, and the most impressive examples of these glycans are encountered in mycobacteria[56,69]. Galactofuranose, (Gal*f*), which is thermodynamically less stable than galactose, is essential for the viability of several pathogenic species of bacteria and protozoa. It is absent in this form in mammalian cell structure, hence the biochemical pathways by which galactofuranose containing glycans are assembled have been attractive sites for drug action[69,70]. This potential has led to an increased interest in the synthesis of molecules containing galactofuranose residues, and their subsequent use in studies directed towards understanding of the enzymes that process these residues and the identification of potential inhibitors of these pathways[66].

The enzyme UDP-galactopyranose mutase is central to galactofuranose metabolism. Most organisms cannot use exogenous galactofuranose, and UDP-galactofuranose appears to be the biological source of galactofuranose residues in polysaccharides[66]. The major structural component of the *Mycobacterium tuberculosis* cell wall contains a galactan chain of approximately thirty-five galactofuranose units, and the

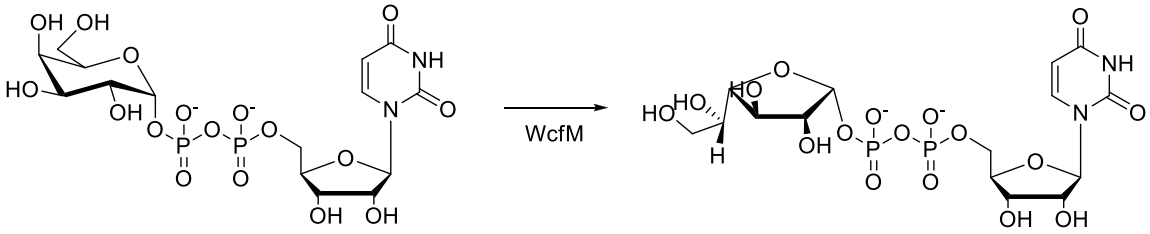
biosynthesis of the galactan is essential for viability[69]. The O-antigens of both *Escherichia coli* and *Klebsiella pneumoniae* contain galactofuranose as a component of lipopolysaccharide[70]. Several galactofuranose containing glycoconjugates have been found in *Trypanosoma cruzi*, the causative agent of Chagas disease, including glycoinositolphospholipids, lipopeptidophosphoglycans and mucin-like proteins. The galactomannan of *Aspergillus fumigatus* also contains galactofuranose, and this polysaccharide is used for clinical detection of fungal infections. Finally, it is also known that stopping galactofuranose biosynthesis in Leishmania major attenuates its virulence[56,66,69,70]. The above-mentioned pathogenic organisms all use the same building block for synthesizing galactofuranose-containing polysaccharides: uridine diphosphogalactofuranose (UDP-galactofuranose). This sugar nucleotide is produced from UDP-Glcp by the enzymes UDP-Glucose 4-epimerase (generating UDP-Galp,) and UDP-galactopyranose mutase (UGM), which catalyzes the transformation of UDP-Galp to UDP-galactofuranose. The gene encoding UGM was first identified in *E. coli* in 1996, followed shortly by its identification in *K. pneumoniae* and *M. tuberculosis*[56,57,71]. More recently, UGM was identified in the eukaryotes *A. fumigatus*, *Cryptococcus neoformans*, *L. major* and *T. cruzi*.

In the past several years' major milestones have been achieved, which include an in-depth understanding of the mechanism of UDP-galactopyranose mutase (UGM), the enzyme which produces UDP-galactofuranose, and is the donor species used by galactofuranosyltransferases. A number of methods for the synthesis of galactofuranosides have also been developed[58]. UDP-galactofuranose has also been prepared by a number of

approaches, and currently it appears that a chemoenzymatic approach is the most viable method for producing multi-milligram amounts of this important rare sugar[58,66].

4.2 UDP-galactopyranose mutase in CPSA biosynthetic operon

The biosynthetic gene operon of CPSA encodes a *wcfM* gene, which was found to be homologous to other galactopyranose mutases. It is homologous to two known UDP-galactopyranose mutases, one from *Streptococcus pneumonia* (Cps33fN: 66% identity and 82% similarity) and the other from *E. coli* (59% identity and 79% similarity). The gene encodes a 43 kDa protein with one potential N-terminal transmembrane domain. Like other galactopyranose mutases, the protein is hypothesized to catalyze the reaction as shown in Figure 4.1. The product of WcfM is required for the final step in the synthesis of the CPSA tetrasaccharide repeat unit. The last glycosyltransferase transfers UDP-galactofuranose to the trisaccharide, as explained in chapter 2.



UDP-Galp UDP-Galactofuranose

Figure 4.1 Reaction catalyzed by UGM

Experimental methods

4.3.1 Cloning and overexpression of WcfM

The expression and isolation of the galactopyranose mutase enzyme from *B. fragilis* has been completed using a pET24(a) vector in a BL21(DE3) RIL *E. coli* cell line. This vector encodes both a His-Tag and a T7-Tag that were utilized in the isolation of the protein using affinity chromatography or in Western blot analysis of expressed product. The cloning of the *wcfM* gene was done by my advisor Dr. Jerry M. Troutman prior of my joining the lab. Two 500ml batches were inoculated with an overnight culture (1 mL each), and the cells were grown at 37 °C with vigorous agitation till OD600 of 0.6 was reached, upon which temperature was reduced to 26 °C, and the culture was induced with 1mM IPTG. The cells were harvested by centrifugation (5000g, 10 min), 4 hours after induction. The harvested cells were washed with 50 mM potassium phosphate buffer (pH 7.4), collected again by centrifugation (5000 x g, 10 min), and stored at -80 °C. Later, the cells were resuspended in lysis buffer (50 mM potassium phosphate buffer, pH 7.4, 20 mM imidazole, 200 mM sodium chloride), lysed by sonication, and pelleted at 2800 x g to remove cell debris. The supernatant was centrifuged again at 18000 x g for an hour to separate the membrane fraction. The supernatant was used to isolate the soluble protein, which was carried out using nickel nitrilotriacetic acid (Ni-NTA), resin that chelates to the His tag on the protein. The protein was eluted off the Ni-NTA using a high concentration of imidazole (500 mM) and then dialysis was performed to remove the imidazole from the isolated protein. The purified mutase, WcfM, was found to be stable in phosphate buffers containing 15% glycerol. An SDS-PAGE demonstrated the purity of the expressed proteins, which were the most prominent with low amounts of other proteins. An anti-His

Western Blot was performed to show that the protein was obtained at the correct molecular weight.

4.3.2 WcfM activity assay

WcfM assay was performed in a 100 mM potassium phosphate buffer (pH 7.4) with 20mM sodium dithionite. 3μM WcfM was incubated with 3.5 mM UDP-Gal(p). After 10 minutes the reaction was analyzed on a reverse phase C18 Agilent Eclipse XDB-C18, 5μm, 4.6 x 150 mm column. The mobile phase used was 1.5% acetonitrile in 50mM triethylammonium acetate buffer (pH = 6.8), with a flow rate of 0.5 ml/min. WcfM was also assayed coupled with WcfN for the production of the tetrasaccharide (Chapter 2). Pure UDP-Galactofuranose was kindly provided by the Todd Lowary Lab.

4.4 Results

4.4.1 Overexpression of WcfM

Overexpression of WcfM was achieved both as a membrane fraction and solubilized protein. This correlates with the prediction of one transmembrane domain in WcfM (Figure 4.2), and also previous studies, where WcfM was isolated as soluble protein. The concentration of solubilized WcfM was determined to be 1.5 mg/mL using a UV-Vis spectrophotometer with an extinction coefficient of 66,825 M-1 cm-1 and molecular weight of 42 kDa at 280 nm. Presence of both the membrane fraction and soluble protein was confirmed by SDS-PAGE and Western Blotting (Figure 4.3).

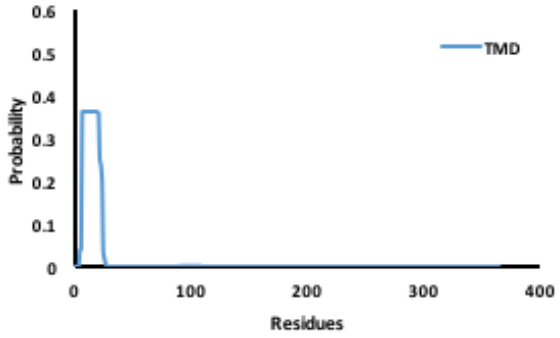


Figure 4.2 WcfM is predicted to have one transmembrane domain, using the TMHMM server.

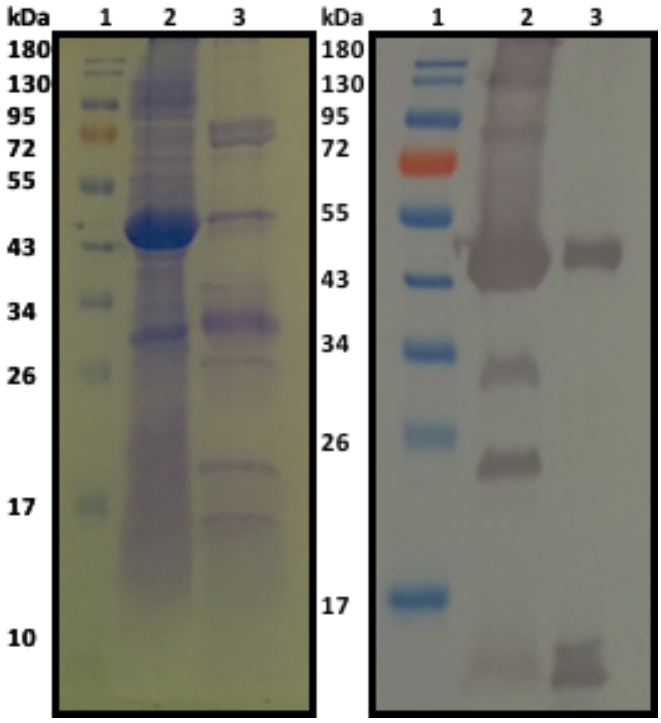


Figure 4.3 Lanes 1, 2 and 3 represent protein ladder, membrane fraction and soluble protein respectively. On the right is the Western Blot with Anti-Histidine specific antibody. The left image is the Coomassie stain. WcfM can also be isolated as soluble protein from membrane fraction

4.4.2 WcfM is the galactopyranose mutase in CPSA biosynthetic pathway

The last sugar in the CPSA tetrasaccharide unit is a galactofuranose. In chapter 2 it was shown that this last sugar was transferred by WcfN to complete the CPSA tetrasaccharide repeat. When WcfM is incubated just by itself in a reaction mix containing UDP-Gal*p*, the conversion equilibrium is in favor of UDP-Galp, with just approximately 7% being converted to UDP-galactofuranose, as shown in figure 4.4. Similar percentage turnover were seen in previous studies also[58], as the six membered ring pyranose form is more thermodynamically favorable and stable than its counterpart furanose form, notably because steric interactions are minimized in six membered rings.

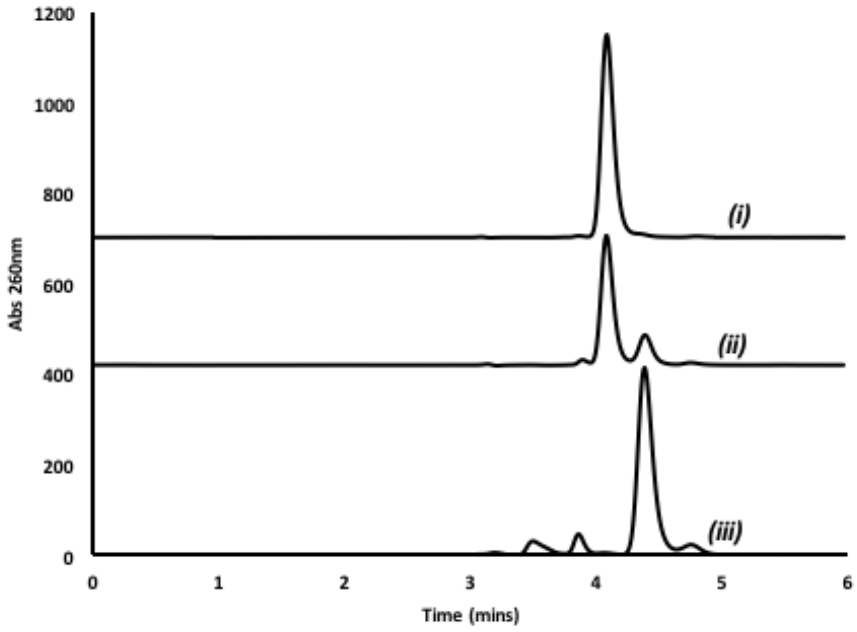


Fig 4.4 WcfM is a galactopyranose mutase. In the chromatograms shown above, (i) and (iii) are UDP-Gal(p) and UDP-Gal(f) controls respectively. Chromatogram (iii) is WcfM reaction with UDP-Gal(p), showing emergence of a new peak in line with the UDP-Gal(f) control.

4.5 Summary

Galactofuranoses are an important component of many polysaccharides found in various bacterial species, protozoa and fungi[66]. Since exogenous galactofuranose cannot be used by these microorganisms for the biosynthesis of these complex molecules, a biosynthetic system must be present in these organisms to produce galactofuranose. In 1996, the enzyme responsible for the synthesis of galactofuranose from galactopyranose (Gal*p*) was isolated and characterized[57]. Since then various other studies have identified galactopyranose mutases in several other organisms.

In this study I have successfully identified and characterized WcfM as the galactopyranose mutase in the CPSA biosynthetic pathway. WcfM is responsible for the conversion of Gal*p* to galactofuranose, which is then subsequently used by the galactofuransoyltransferase WcfN to complete the CPSA tetrasaccharide. Since the more thermodynamically favored form is the pyranose form, approximately 7% of the pyranose is converted to the furanose form. This thermodynamic hump is overcome by coupling the mutase with the glycosyltransferase in the synthesis of the complex polysaccharides, as detailed by Whitfield et al.

Furanoses are widely distributed in nature, but more predominantly in pathogenic organisms[66]. Moreover, their absence in mammalian glycoconjugates suggests that the enzymes involved in the metabolism of such sugars in bacteria, fungi and protozoa would constitute a good target for the design of new drugs[69]. Therefore, they have raised the interest of glycoscientists globally due to their potential biological applications.

CHAPTER 5: DISCUSSION

5.1 Introduction

Every bacterium, pathogenic, symbiotic or benign, have a variety of polysaccharides present on their cell surface. Capsular polysaccharides are one of the many polysaccharides found on the surface of certain bacteria. Cell surface polysaccharides play diverse roles in bacterial ecology and physiology. They serve as a barrier between the cell wall and the environment, and mediate host-pathogen interactions. Some of the polysaccharides on these bacteria also may be beneficial in the development of the immune system in humans, as is the case with the Gram-negative bacterium *Bacteroides fragilis*. The capsular polysaccharide A (CPSA) found on the surface of *B. fragilis* has been identified as a potential therapeutic biomolecule. Growing evidence has supported the notion that CPSA plays a role in normal immune system function of host mammals. Studies in animal models have suggested that it may be useful as a therapeutic for diseases such as multiple sclerosis and inflammatory intestinal disorders. Currently, CPSA isolation involves culture growth, isolation and purification. Isolation of CPSA in this way is tedious and often may be contaminated with other polysaccharides present on the cell surface. This process also may lead to batch-to-batch variation. An alternative synthetic approach was taken by Seeberger group to make the tetrasaccharide unit of CPSA. Their carbohydrate synthesis process is a series of more than 20 chemical transformations with an overall yield of 1.8%. The Andreana group also managed to chemically synthesize CPSA repeat unit

with a 5% total yield. This study was aimed to provide insight into CPSA biosynthesis and isolation of this important polymer. The proposal was to build CPSA biosynthetically in an *in vitro* system with starting materials that are commercially available, giving insights into the biosynthetic pathway of CPSA and also allowing the characterization of previously uncharacterized enzymes with functions that are currently poorly understood. During the course of this study, we were able to successfully synthesize the whole tetrasaccharide repeat unit of CPSA. The enzymes responsible for transfer of each sugar were characterized, and we were also able to specifically demonstrate the pyruvate modification on the second sugar in the repeat unit, along with the characterization of a galactopyranose mutase.

5.2 Importance of fluorescent analogues

Bacteria use the C55 bactoprenyl phosphate anchor to build the polysaccharide sugar by sugar. Each sugar addition is catalyzed by a different glycosyltransferase. The native bactoprenyl phosphate has a major drawback of not having any associated chromophores, hence reaction assays cannot be monitored easily. Compounding the issue is that, there is a limited availability of the bacterial bactoprenyl phosphate. Sometimes a substitute, plant undecaprenol from *Magnolia kobus* is used, though there is a difference between the both. The plant undecaprenol is a 3 trans 7-*cis* isoprene, whereas the bactoprenyl phosphate is a 2 trans 8-*cis* isoprene[72]. Hence there was a need to develop more robust and trackable analogues of the native bactoprenyl phosphate.

In the past five years, the Troutman lab has been able to develop various fluorescent analogues of the bactoprenyl phosphate. These analogues have enormous detection sensitivity, making the analysis more trackable. Low turnover could also be observed by

using the fluorescent analogues, hence dispensing with the need of using excess quantities of reacting substrates. Starting from setting up the assays and monitoring them on the HPLC, the entire process could be completed within an hour compared to exhaustive traditional studies that involved using of radiolabeled substrates and extractions of the lipid product. A major advantage to the fluorescent isoprenoid was that specificity studies of the glycosyltransferases and steady state kinetic studies with the pyruvyltransferase could be conducted without requiring access to expensive radiolabeled nucleotide-linked sugars, and was achieved in a short time when compared to similar studies done previously. In addition, multiple small-scale reactions could be prepared, for example when optimizing turnover conditions for a specific enzyme, and then left for automated HPLC analysis without requiring extractions or thin layer chromatography (TLC)[65]. Another advantage of using the fluorescent analogues was the ease in purifying the enzyme products using HPLC for characterization through mass spectrometry.

5.3 CPSA biosynthesis operon encodes a galactosyltransferase, a N-acetylgalactosyltransferase, a galactofuranosyltransferase and a galactopyranose mutase

CPSA is assembled on a bactoprenyl phosphate anchor inside a bacterial cell. In the *in vitro* synthesis, fluorescent bactoprenyl analogues were used to construct CPSA tetrasaccharide unit. The initiating hexose phosphate transferase, WcfS, had already been identified and characterized by a member of the lab previously. WcfQ was identified as the first glycosyltransferase, which transferred galactose to the isoprenoid linked monosaccharide, even though it was observed that, WcfQ could also transfer glucose to the bactoprenyl linked monosaccharide. This is because WcfQ required glucose in much excess when compared to galactose. It was also found out that even though WcfP had the

capability of transferring galactose, WcfQ was more efficient in it, hence it was identified as the galactosyltransferase in the CPSA biosynthetic pathway. Moreover, based on the Carbohydrate-Active enZYmes (CAZY) database the WcfQ sequence matched the GT_2 family of glycosyltransferases which invert the configuration of the anomeric carbon of the donor, while WcfP was similar to a GT_4 family glycosyltransferase, which retains the anomeric stereo-configuration of the donating sugar. The published structure of the CPSA tetrasaccharide unit suggests that the linkage should be in a beta configuration. This supported the conclusion that WcfQ is the protein responsible for introducing galactose, and that it introduces the sugar in the appropriate beta configuration.

As stated above, WcfP is related to the GT_4 family of proteins suggesting that it is a retaining glycosyltransferase, it was therefore more likely that WcfP catalyzed UDP-GalNAc transfer to the galactose, but it was not known if it transferred UDP-GalNAc to the unpyruvylated disaccharide or the pyruvylated disaccharide. Analyzing both WcfN and WcfP with the pyruvylated and the unpyruvylated disaccharides, it was demonstrated that WcfP transfers only UDP-GalNAc to the pyruvylated disaccharide.

In homology studies, WcfN was predicted to be a member of the GT_2 family, whose members have been identified to transfer furanose residues. WcfN was also hypothesized to be an inverting transferase, which inverts the stereochemistry of the anomeric carbon. Since the linkage between the third and the fourth sugar in the tetrasaccharide repeat unit is in the beta configuration, WcfN fitted the role of being the last glycosyltransferase. WcfN was found to transfer the galactofuranose to the trisaccharide, hence completing the *in vitro* synthesis of the tetrasaccharide.

In this study, three new glycosyltransferases were characterized functionally.

Glycosyltransferases comprise a super family of proteins that have been grouped in to a total of 106 distinct sub-families, (CAZY, Carbohydrate-Active Enzymes database)[59,63,73], based on sequence. They can also be grouped on structural basis, with either having a GT-A or GT-B fold[6]. Two new folds have also been described for CAZY family GT_51[74] and GT_66[75]. Finally, glycosyltransferases may also be termed retaining or inverting based on the configuration of the sugar nucleotide-derived carbohydrate in the final product[6,76]. No correlation has been found between the fold type (GT-A or GT-B) and the stereochemical outcome of the reaction. Instances of both inverting and retaining glycosyltransferases have been found in both families. However, all members of a CAZY family are expected to have the same fold[76].

Leloir glycosyltransferases[6], named after Luis Leloir whose group isolated and characterized sugar nucleotides[60] and also reported the first glycosyltransferases[77], are known for being very selective for their substrates, the nucleotide sugar donor and the acceptor. In fact, this led to the widely accepted concept that each glycosidic linkage is a product of a single enzyme, a hypothesis that was advanced by Saul Roseman, Moffat and Khorana[77,78]. A brilliant example for strict specificity is the human B blood group α1-3 galactosyltransferase. This enzyme catalyzes a glycosylation reaction only when the galactose residue of the acceptor substrate has fucose in α1–2 linkage[78]. There are exceptions though, wherein glycosyltransferases have been shown to use different donor sugars. The enzyme EXTL2, involved in glucosaminoglycan (GAG) synthesis, where it can attach either uridine linked *N*-acetylgalactosamine or *N*-acetylglucosamine in α linkage to glucuronic acid[79]. In our studies, WcfQ was found to attach uridine linked galactose or glucose to the undecaprenyl diphosphate disaccharide. Both EXTL2 and

WcfQ belong to GT-A family, the only difference being that the former is retaining and the latter inverting. Studies through crystal models have suggested that the GT-A fold lacks sugar signatures which can influence sugar specificity, while in fold GT-B, the residue adjacent to the strictly conserved E/D-Q of Loop C5 can influence the C2-C4 specificity of a sugar[59,80]. It has been suggested that interactions with donor sugar are often minimal and mostly dominated by steric obstruction, leading to glycosyltransferases often being sugar indiscriminant (studies on directed evolution mutant glycosyltransferase OleD by Gantt et al)[81,82] and they may even be able to transfer non-carbohydrates like glycerol. TagF is a good example for this[74].

Despite the above-mentioned findings, literature review has revealed that instances for specificity for multiple sugar donors in nature are very rare. Hence, it is all the more intriguing that CPSA operon codes for two promiscuous glycosyltransferases. WcfP was found to be very specific in transferring the sugar only to the pyruvylated galactose disaccharide and not the pyruvylated glucose disaccharide which is interesting, as WcfP has shown more plasticity in accepting different acceptor substrates and also different donor sugar substrates than WcfQ. WcfQ differs only in using two different donor sugar substrates. I used the PHYRE^2 (Protein Homology/analogy Recognition Engine v 2.0) automatic fold recognition server for predicting the structure of WcfP[83]. It was modeled on the template MshA with 97% coverage of the sequence and 16% sequence identity; but many studies have pointed that glycosyltransferases are structurally conserved rather than sequentially except key residues. Both WcfP and MshA have the GT-B fold and belong to GT_4 family. GT-B fold exhibits a global domain movement after binding of the substrate. This leads to a transition from an open to close confirmation, involving a

shift of approximately 10-12°, which brings the acceptor and the donor in close proximity[80]. MshA, which catalyzes the first step of mycothiol biosynthesis in *Cornybacterium glutamicum,* is an exception in this instance[84]. It shows a 97° rotational reorientation of N-terminal domain relative to the C-terminal domain. MshA was also shown of being capable of using UDP-GlcNAc and UDP-fluroGlucose as sugar donors by Vetting et al. It can be argued that the large 97° movement of the N-domain and C-domain when the sugar nucleotide binds to MshA highlights the flexibility of the hinge region of GT-B glycosyltransferases. This also points to the potential dynamic range of flexibility other family members may experience on binding of the substrate. The structural data obtained so far through various studies indicate that the interdomain flexibility is an important feature of sugar nucleotide recognition of the GT-B family. Since WcfP structure was modeled with 97% sequence coverage, it could be plausible that WcfP may have domain orientations similar to MshA and this may be the reason for its promiscuity. It should be noted that threading models serve the purpose of only forming a testable hypothesis. They do not reveal anything with certainty. This behavior of WcfP can be best explored through crystallographic studies.

Glycosyltransferases create a diverse range of saccharides and glycoconjugates in nature. Understanding glycosyltransferases at the molecular level, through structural and kinetic studies, is important for gaining insights into their function. In addition, this understanding can help identify those enzymes which are involved in diseases, or that could be engineered to synthesize biologically or medically relevant molecules. Moreover, in bacteria the majority of glycosyltransferases are involved in the synthesis of glycolipids, peptidoglycans, and lipopolysaccharides, and can be suitable target for drug development

against bacterial pathogens.

Finally, based on the homology data obtained through pBLAST[48], WcfM was identified as a potential galactopyranose mutase. Galactopyranose mutases are a well characterized family of enzymes. According to previous literature on similar galactopyranose mutases, WcfM was hypothesized to convert the pyranose form of UDP-Galactose to the furanose form, which is then utilized by the glycosyltransferase. In this study we were successful in demonstrating that WcfM is indeed a galactopyranose mutase encoded in the CPSA biosynthesis operon. We were also able to show that, in conjunction with WcfM, WcfN transferred galactofuranose to the trisaccharide, leading to the formation of the tetrasaccharide.

This is the first study where the CPSA tetrasaccharide repeat subunit has been synthesized *in vitro* using overexpressed enzymes. There have been other studies also where cell wall components have been reconstituted *in vitro*. The Imperiali group have demonstrated the *in vitro* biosynthesis of the N-linked glycan (heptasaccharide) from *Campylobacter jejuni*. The heptasaccharide lipid-linked donor was synthesized by coupling the action of eight enzymes (PglF, PglE, PglD, PglC, PglA, PglJ, PglH, and PglI) in the Pgl pathway in a single reaction. The product was analyzed using HPLC and MALDI-MS[52]. Gale et al reconstituted the biosynthesis of wall teichoic acid *in vitro* by using a semi synthetic approach. They utilized undecaprenol isolated from *Laurus nobilis* leaves, which was subjected to sequential modification to yield undecaprenyl diphosphate N-glucosamine, which was then used as a substrate by downstream enzymes in the wall teichoic acid biosynthesis. They used a combination of TLC, high resolution mass spectrometry (HR-MS) and LC/MS for analysis of the product[85]. Recently, the Schaeffer

group characterized the biosynthesis of the *Paenibacillus alvei* secondary cell wall polymer. They reconstituted the enzymatic pathway *in vitro* using a synthetic 11-phenoxyundecyl-diphosphoryl-α-GlcNAc acceptor[86]. They were also able to highlight pyruvylation of the secondary cell wall polymer, which I will discuss in the next section. Mass spectrometry, and nuclear magnetic resonance spectroscopy were used for product characterization. In the mentioned above studies, various lipid carriers were used to build the polymers, I utilized fluorescent undecaprenyl analogues developed by the Troutman lab as discussed previously. HPLC and MS have been used to analyze the reaction products, and is common through all the studies.

5.4 WcfO identified as the pyruvyl transferase

Multiple sequence alignment of some of the known and hypothesized pyruvyltransferases showed that there are no concrete conserved domains having residues which potentially might be involved in the catalysis process. The motif "TSRIH" seemed to be present in all the pyruvyltransferases, with R and H being conserved, though T, S and I residues varied a lot. Yoritsune et al proposed that, the arginine residue is most likely to be a catalytic residue[87]. These domains if well identified can serve to identify the pyruvyltransferases, as is the case with the glycosyltransferase superfamily. Instead, there are only a couple of conserved residues throughout the sequence which limits targeting specific residues for mutagenesis studies, and observe the actual importance of the residue in the reaction process and for the enzyme. This limiting aspect can be overcome by structural studies. Having a crystal structure with the substrate, can help in better identification of the residues involved in interaction with the substrate and also understand why WcfO is not able to selectively differentiate between the galactose or the glucose, as

it pyruvates both the disaccharides equally and does not discriminate between them. Till now there is just one published structure of a pyruvyltransferase, that too from an eukaryote and not a prokaryote[87]. This pyruvyltransferase Pvg1b as previously mentioned, pyruvates only the galactose molecule, which is in stark contrast with WcfO. WcfO is only capable of modifying galactose or glucose when they are linked to the isoprenoid lipid carrier. This points to the direction that, there may be sub-families within the pyruvyltransferase family that utilize different substrates. WcfO structure was threaded using PHYRE[2], which used Pvg1b as a template. It revealed that WcfO structure resembles the GT-B superfamily. It has two domains, resembling the Rossmann fold, separated by a cleft. The cleft is supposed to have a high degree of flexibility and believed to serve as the catalytic center.

Pyruvyltrasferases and pyruvylation have been less studied in prokaryotes, despite a burgeoning evidence of its presence in bacteria. Addition of pyruvate moiety gives a negative charge to the polymer and is utilized by the bacteria in various functions[88]. An example of this is the pyruvylation of ManNAc residue by the enzyme CsaB in the secondary cell wall polymer of *Bacillus anthracis* and *Paenibacillus alvei*[89,90]. This pyruvylated residue comes in use in anchoring the S-layer proteins in Gram positive bacteria by binding to the SLH domains of the S-layer proteins[91]. Knocking out the CsaB has led to a lethal phenotype, which suggests that, pyruvylation of the secondary cell wall polymer is essential to the growth and survival of the bacteria[90]. CsaB was recently characterized by the Schaffer group[86]. Including WcfO, a total of three pyruvyltransferases have now been functionally characterized. Pvg1b is from an eukaryote, and whose crystal structure has been solved. These studies hopefully will help make further inroads towards characterization of more pyruvyltransferases present in other biosynthesis pathways.

For further studies, also of great interest is the difference in the stereochemistry of the two pyruvated disaccharides BPP-AADGalpyrGal and BPP-AADGalpyrGlc. In the published structure of CPSA, pyruvated galactose is the real structure, and WcfP infact accepts it as the substrate over the pyruvated glucose. Knowing the stereochemical difference between the both will help us gain further understanding on why BPP-AADGalpyrGlc is rejected by WcfP, which seems very much in contradiction to the way it has been able to accept different acceptor substrates. This insight can be gained through NMR studies, but obtaining enough pure product has been a tough challenge, either for mass spectrometry or NMR. This will be discussed in detail in the next section.

5.5 Difficulty in characterizing the bactoprenyl linked oligosaccharides

The enzymes required for the assembly of CPSA tetrasaccharide unit have been sequentially identified in this study. To confirm that each enzyme was making the product it was hypothesized to produce, the product needed to be isolated in a scale wherein they could be characterized using mass spectrometry. The use of the fluorescent analogues of the bactoprenyl phosphate greatly helped in purifying the compounds produced by the enzymes. A major issue we faced while purifying these compounds was obtaining enough sample. We faced roadblocks in losing a fraction of the sample at each step of the purification process, starting from the isolation using HPLC, drying it down in centrifugal evaporator, and finally re-suspending it in a solution amenable to characterization through mass spectrometry. Initially not much success was achieved in obtaining mass on the samples when the products were completely dried down and re-suspended in a solution. We achieved more success when the samples that were analyzed on the mass spectrometer were the HPLC fractions containing the product isolate was used. We had good success

with the fractions of pyruvylated disaccharides. With the trisaccharide fraction, even though the HPLC fraction contained higher percentage of the trisaccharide, the disaccharide showed up prominently too in the spectra. We had major difficulty in obtaining mass on the tetrasaccharide. The HPLC fraction method did not work, but re-suspending the sample and analyzing that in the mass spectrometer worked. Here too the trisaccharide showed up as the major product as compared to the tetrasaccharide. Observing all the data obtained from the mass spectrometer, there is a need to develop better method to characterize oligosaccharides having more than three sugar residues. Recently, the Troutman lab have acquired a LC-MS system, which has circumvented the need of requiring to isolate the product for analysis using mass spectrometry with direct injection. This method was successfully used in mapping the biosynthesis pathway of the colanic acid repeat unit[92]. Since we had trouble obtaining larger quantities of the product for mass analysis itself, it was difficult to obtain sufficient sample for NMR spectra on the products made by the enzymes. However, as it has been reported in the above study, when used together, LC-MS and fluorescence analysis can be an efficient tool in characterizing pathways.

5.6 Clinical significance of CPSA

In this study I have managed to synthesize *in vitro* the tetrasaccharide repeat unit of CPSA. Literature study has revealed that for CPSA to be effective as a therapeutic molecule, the tetrasaccharide repeat needs to be a polymer of ten repeat units or longer. If shorter than that, it fails to activate the immune system. CPSA operon encodes for a flippase *wzx*, which takes the repeat unit and flips it from the cytoplasmic space to the periplasmic space, where the polymerase *wzy*, utilizes the repeat unit and polymerizes it

till it reaches a length of approximately 130 repeat units. Wzy has been successfully cloned by me, but its expression has proven to be difficult. Therefore, producing oligomers of CPSA that will be clinically significant is still a challenge, as it requires expressing a functional polymerase enzyme and also sufficient availability of the synthesized tetrasaccharide unit.

Recent successes in cancer vaccines and in monoclonal antibody cancer immunotherapy have given the impetus towards development of vaccines targeting cancer-associated carbohydrates. The Andreana group have been developing carbohydrate immunogens to elicit a T-cell dependent immune response. CPSA is known to stimulate a strong T-cell mediated response. They have successfully linked CPSA to the tumor-associated carbohydrate antigen (TACA), Sialyl Thomsen-nouveau (STn) and were able to obtain a robust immune response to the antigen[45,93-96]. They have further reported total synthesis of the CPSA unit in 19 steps with a final yield of 5%[45]. Chemoenzymatic assembly is a faster and scalable approach, that can be used as an alternative or in combination with chemical synthesis. CPSA obtained in this way, can then be linked to the antigen. The chemoenzymatic method has also been used to create capsule polysaccharide based glycoconjugates for *Neisseria meningitidis* serotypes A, C and X[97-99]. In some cases, recombinant glycosyltransferases can be used to assemble non-native carbohydrate antigens in compliant host organisms like *Escherichia coli*. This method has been successfully used by the Brendan W. Wren lab for the *in vivo* assembly of capsular polysaccharide from several serotypes of *Streptococcus pneumoniae*[100]. A similar approach is also being applied in the Troutman lab also, wherein the whole CPSA biosynthesis and assembly will be done inside *E. coli*. This will also allow us to have access to longer

oligomers of CPSA, which can be helpful in studies toward size requirement in eliciting immune response. So far there have been no reports of CPSA unit being polymerized synthetically.

5.7 Summary

The project started with the aim of synthesizing the CPSA tetrasaccharide repeat unit *in vitro* using enzymes from the CPSA gene operon. It was proposed that the polysaccharide unit is assembled in a Wzy dependent pathway. The initiating hexose-1-phosphate transferase, WcfS transferred the first sugar AADGal to the bactoprenyl phosphate, done in previous studies. For the synthesis of the disaccharide it was demonstrated that, WcfQ transferred galactose, which was then pyruvylated by the pyruvyltransferase, WcfO to make the pyruvylated disaccharide. It was also established that the third sugar UDP-GalNAc was added to the disaccharide by WcfP only if the disaccharide was pyruvylated and not vice-versa. WcfP also exhibited substrate selectivity when adding the third sugar. It differentiated between the pyruvylated glucose and pyruvylated galactose residues. Finally, WcfN working in combination with the galactopyranose mutase WcfM, transferred the last sugar galactofuranose to the trisaccharide. The products of the glycosyltransferases and the pyruvyltransferases were purified using HPLC and characterized using mass spectrometry. This study is the first instance wherein a whole polysaccharide repeat unit of CPSA has been assembled *in vitro* using enzymes. Through this study it was also determined, the site of pyruvate modification in polysaccharides, in context of their biosynthesis, and also the characterization of three new glycosyltransferases.

REFERENCES

1 Lipke, P. N. & Ovalle, R. Cell Wall Architecture in Yeast: New Structure and New Challenges. *Journal of Bacteriology* **180** (1998).

2 Silhavy, T. J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb Perspect Biol* **2**, (2010).

3 Whitfield, C. Biosynthesis and Assembly of Capsular Polysaccharides in Escherichia coli. *Annual review of biochemistry* **75**, 39-68 (2006).

4 Guo, H., Yi, W., Song, J. K. & Wang, P. G. Current Understanding on Biosynthesis of Microbial Polysaccharides. *Current Topics in Medicinal Chemistry* **8**, 11 (2008).

5 Troy, F. The chemistry and biosynthesis of selected bacterial capsular polymers. *Annual Review of Microbiology* **33** (1979).

6 Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annual review of biochemistry* **77**, 521-555 (2008).

7 Wagner, G. K. & Pesnot, T. Glycosyltransferases and their assays. *Chembiochem* **11**, 1939-1949, (2010).

8 Roca, C., Alves, V. D., Freitas, F. & Reis, M. A. Exopolysaccharides enriched in rare sugars: bacterial sources, production, and applications. *Front Microbiol* **6**, 288, (2015).

9 Emmadi, M. & Kulkarni, S. S. Recent advances in synthesis of bacterial rare sugar building blocks and their applications. *Nat Prod Rep* **31**, 870-879, (2014).

10 Tabaqchali, S., Fiddian, P. A., El-Hag, K. & Kasper, D. L. Capsular and 'O' serotype determinants of bacteroides fragilis. *Infection* **10**, 333-337 (1982).

11 Erturk-Hasdemir, D. & Kasper, D. L. Finding a needle in a haystack: Bacteroides fragilis polysaccharide A as the archetypical symbiosis factor. *Ann N Y Acad Sci* **1417**, 116-129, (2018).

12 Donlan, R. M. Biofilms: Microbial Life on Surfaces. *Emerging Infectious Diseases* **8** (2002).

13 Mazmanian, S. K. Capsular polysaccharides of symbiotic bacteria modulate immune responses during experimental colitis. *J Pediatr Gastroenterol Nutr* **46 Suppl 1**, E11-12, (2008).

14 Kasper, D. L. & Onderdonk, A. B. Infection with Bacteroides fragilis: pathogenesis and immunoprophylaxis in an animal model. *Scand J Infect Dis Suppl* **31**, 28-33 (1982).

15 Ivashina, T. V. & Ksenzenko, V. N. Exopolysaccharide Biosynthesis in Rhizobium leguminosarum: From Genes to Functions. (2012).

16 Kuo, Y. L., Shieh, S. J., Chiu, H. Y. & Lee, J. W. Necrotizing fasciitis caused by Vibrio vulnificus: epidemiology, clinical findings, treatment and prevention. *Eur J Clin Microbiol Infect Dis* **26**, 785-792 (2007).

17 Ziegler, T. Pyruvated Saccharides - Novel Strategies for Oligosaccharide Synthesis.

18 Onderdonk, A. B., Kasper, D. L., Cisneros, R. L. & Bartlett, J. G. The capsular polysaccharide of Bacteroides fragilis as a virulence factor: comparison of the pathogenic potential of encapsulated and unencapsulated strains. *J Infect Dis* **136**, 82-89 (1977).

19 Lindberg, A. A., Weintraub, A., Kasper, D. L. & Lonngren, J. Virulence factors in infections with bacteroides fragilis: isolation and characterization of capsular polysaccharide and lipopolysaccharide. *Scand J Infect Dis Suppl* **35**, 45-52 (1982).

20 Mazmanian, S. K. & Kasper, D. L. The love-hate relationship between bacterial polysaccharides and the host immune system. *Nat Rev Immunol* **6**, 849-858 (2006).

21 Edelman, S. M. & Kasper, D. L. Symbiotic commensal bacteria direct maturation of the host immune system. *Curr Opin Gastroenterol* **24**, 720-724, (2008).

22 Jimenez, J. & Sanchez-Perez, M. The immune system and microorganisms: a love-hate relationship revisited. *Int Microbiol* **6**, 3-4, (2003).

23 Tzianabos, A., Wang, J. Y. & Kasper, D. L. Biological chemistry of immunomodulation by zwitterionic polysaccharides. *Carbohydr Res* **338**, 2531-2538, (2003).

24 Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313-323 (2009).

25 Kasper, D. L. *et al.* Isolation and identification of encapsulated strains of Bacteroides fragilis. *J Infect Dis* **136**, 75-81 (1977).

26 Tzianabos, A. O., Kasper, D. L. & Onderdonk, A. B. Structure and function of Bacteroides fragilis capsular polysaccharides: relationship to induction and prevention of abscesses. *Clin Infect Dis* **20 Suppl 2**, S132-140 (1995).

27 Kalka-Moll, W. M. *et al.* Immunochemical and biological characterization of three capsular polysaccharides from a single Bacteroides fragilis strain. *Infect Immun* **69**, 2339-2344, (2001).

28 Krinos, C. M. *et al.* Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414**, 555-558, (2001).

29 Pantosti, A., Tzianabos, A. O., Reinap, B. G., Onderdonk, A. B. & Kasper, D. L. Bacteroides fragilis strains express multiple capsular polysaccharides. *J Clin Microbiol* **31**, 1850-1855 (1993).

30 Tzianabos, A., Kasper, D. & Onderdonk, A. B. Structure and Function of Bacteroides fragilis Capsular Polysaccharides: Relationship to Induction and Prevention of Abscesses. *Clinical Infectious Diseases* **20** (1995).

31 Troy, E. B. & Kasper, D. L. Beneficial effects of Bacteroides fragilis polysaccharides on the immune system. *Front Biosci* **15**, 25-34 (2010).

32 Kasper, D. L., Onderdonk, A. B., Crabb, J. & Bartlett, J. G. Protective efficacy of immunization with capsular antigen against experimental infection with Bacteroides fragilis. *J Infect Dis* **140**, 724-731 (1979).

33 Coyne, M. J. *et al.* Polysaccharide biosynthesis locus required for virulence of Bacteroides fragilis. *Infect Immun* **69**, 4342-4350 (2001).

34 Tzianabos, A. O. *et al.* Structural characterization of two surface polysaccharides of Bacteroides fragilis. *Trans Assoc Am Physicians* **104**, 285-295 (1991).

35 Avci, F. Y. & Kasper, D. L. How bacterial carbohydrates influence the adaptive immune system. *Annu Rev Immunol* **28**, 107-130 (2010).

36 Cobb, B. A. & Kasper, D. L. Zwitterionic capsular polysaccharides: the new MHCII-dependent antigens. *Cell Microbiol* **7**, 1398-1403 (2005).

37 Cobb, B. A., Wang, Q., Tzianabos, A. O. & Kasper, D. L. Polysaccharide processing and presentation by the MHCII pathway. *Cell* **117**, 677-687 (2004).

38 Duan, J., Avci, F. Y. & Kasper, D. L. Microbial carbohydrate depolymerization by antigen-presenting cells: deamination prior to presentation by the MHCII pathway. *Proc Natl Acad Sci U S A* **105**, (2008).

39 Tzianabos, A. O., Gibson, F. C., 3rd, Cisneros, R. L. & Kasper, D. L. Protection against experimental intraabdominal sepsis by two polysaccharide immunomodulators. *J Infect Dis* **178**, 200-206 (1998).

40 Onderdonk, A. B., Shapiro, M. E., Finberg, R. W., Zaleznik, D. F. & Kasper, D. L. Use of a model of intraabdominal sepsis for studies of the pathogenicity of Bacteroides fragilis. *Rev Infect Dis* **6 Suppl 1**, S91-95 (1984).

41 Comstock, L. E. *et al.* Analysis of a capsular polysaccharide biosynthesis locus of Bacteroides fragilis. *Infect Immun* **67**, 3525-3532 (1999).

42 Chatzidaki-Livanis, M., Coyne, M. J., Roche-Hakansson, H. & Comstock, L. E. Expression of a uniquely regulated extracellular polysaccharide confers a large-capsule phenotype to Bacteroides fragilis. *J Bacteriol* **190**, 1020-1026, (2008).

43 Pragani, R., Stallforth, P. & Seeberger, P. H. De Novo Synthesis of a 2-Acetamido-4 amino-2,4,6-trideoxy-D-galactose (AAT) Building Block for the Preparation of a Bacteroides fragilis A1 Polysaccharide Fragment. *Org Lett* **12**, 1624-1627 (2010).

44 Pragani, R. & Seeberger, P. H. Total Synthesis of the Bacteroides fragilis Zwitterionic Polysaccharide A1 Repeating Unit. *Journal of American Chemical Society* **133**, 102-107 (2011).

45 Eradi, P., Ghosh, S. & Andreana, P. R. Total Synthesis of Zwitterionic Tetrasaccharide Repeating Unit from Bacteroides fragilis ATCC 25285/NCTC 9343 Capsular Polysaccharide PS A1 with Alternating Charges on Adjacent Monosaccharides. *Org Lett* **20**, 4526-4530, (2018).

46 Tzianabos, A. O. *et al.* The capsular polysaccharide of Bacteroides fragilis comprises two ionically linked polysaccharides. *J Biol Chem* **267**, 18230-18235 (1992).

47 Baumann, H., Tzianabos, A. O., Brisson, J. R., Kasper, D. L. & Jennings, H. J. Structural elucidation of two capsular polysaccharides from one strain of Bacteroides fragilis using high-resolution NMR spectroscopy. *Biochemistry-Us* **31**, 4081-4089 (1992).

48 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool *Journal of Molecular Biology* **215**, 7 (1990).

49 El Ghachi, M., Bouhss, A., Blanot, D. & Mengin-Lecreulx, D. The bacA gene of Escherichia coli encodes an undecaprenyl pyrophosphate phosphatase activity. *J Biol Chem* **279**, 30106-30113, (2004).

50 Mostafavi, A. Z. & Troutman, J. M. Biosynthetic assembly of the Bacteroides fragilis capsular polysaccharide A precursor bactoprenyl diphosphate-linked

acetamido-4-amino-6-deoxygalactopyranose. *Biochemistry-Us* **52**, 1939-1949 (2013).

51 Coyne, M. J., Chatzidaki-Livanis, M., Paoletti, L. C. & Comstock, L. E. Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont Bacteroides fragilis. *Proc Natl Acad Sci U S A* **105**, 13099-13104, (2008).

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

53 Mostafavi, A. Z., Lujan, D. K., Erickson, K. M., Martinez, C. D. & Troutman, J. M. Fluorescent probes for investigation of isoprenoid configuration and size discrimination by bactoprenol-utilizing enzymes. *Bioorg Med Chem* **21**, 5428-5435 (2013).

54 Yoritsune, K., Matsuzawa, T., Ohashi, T. & Takegawa, K. The fission yeast Pvg1p has galactose-specific pyruvyltransferase activity. *FEBS Lett* **587**, 917-921 (2013).

55 Gardiol, A. E. & Dazzo, F. B. Biosynthesis of Rhizobium trifolii Capsular Polysaccharide: Enzymatic Transfer of Pyruvate Substitutions into Lipid-Bound Saccharide Intermediatest. *Journal of Bacteriology* **168**, 1459-1462 (1986).

56 Guan, S., Clarke, A. J. & Whitfield, C. Functional analysis of the galactosyltransferases required for biosynthesis of D-galactan I, a component of the lipopolysaccharide O1 antigen of Klebsiella pneumoniae. *J Bacteriol* **183**, 3318-3327 (2001).

57 Nassau, P. M. *et al.* Galactofuranose Biosynthesis in Escherichia coli K-12: Identification and Cloning of UDP-Galactopyranose Mutase. *Journal of Bacteriology* **178**, 6 (1996).

58 Poulin, M. B. *et al.* Characterization of a bifunctional pyranose-furanose mutase from Campylobacter jejuni 11168. *J Biol Chem* **285**, 493-501 (2010).

59 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495, (2014).

60 Gloster, T. M. Advances in understanding glycosyltransferases from a structural perspective. *Curr Opin Struct Biol* **28**, 131-141 (2014).

61 Soya, N., Fang, Y., Palcic, M. M. & Klassen, J. S. Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases. *Glycobiology* **21**, 547-552 (2011).

62 Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding *Analytical Biochemistry* **72**, 7 (1976).

63 Kapitonov, D. & Yu, R. K. Conserved domains of glycosyltransferase. *Glycobiology* **9** (1999).

64 Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-580, (2001).

65 **Troutman, J. M., Sharma, S., Erickson, K. M. & Martinez, C. D. Functional identification of a galactosyltransferase critical to Bacteroides fragilis Capsular Polysaccharide A biosynthesis. *Carbohydr Res* 395, 19-28 (2014).**

66 Richards, M. R. & Lowary, T. L. Chemistry and biology of galactofuranose-containing polysaccharides. *Chembiochem* **10**, 1920-1938, (2009).

67 Yang, F. L. *et al.* Structure and immunological characterization of the capsular polysaccharide of a pyrogenic liver abscess caused by Klebsiella pneumoniae: activation of macrophages through Toll-like receptor 4. *J Biol Chem* **286**, 21041-21051, (2011).

68 R.Gemmill, T. & B.Trimble, R. All pyruvylated galactose in Schizosaccharomyces pombe N-glycans is present in the terminal disaccharide, 4,6-O-[(R)-(1-carboxyethylidine)]-Galb1,3Gala1-. *Glycobiology* **8**, 1087–1095 (1998).

69 Peltier, P., Euzen, R., Daniellou, R., Nugier-Chauvin, C. & Ferrieres, V. Recent knowledge and innovations related to hexofuranosides: structure, synthesis and applications. *Carbohydr Res* **343**, 1897-1923, (2008).

70 Koplin, R., Brisson, J. R. & Whitfield, C. UDP-galactofuranose Precursor Required for Formation of the Lipopolysaccharide O Antigen of Klebsiella pneumoniae Serotype O1 Is Synthesized by the Product of the rfbDKPO1 Gene. *Journal of Biological Chemistry* **272**, 8 (1997).

71 Lee, R. *et al.* Enzymatic Synthesis of UDP-Galactofuranose and an Assay for UDP-Galactopyranose Mutase Based on High-Performance Liquid Chromatography. *Analytical Biochemistry* **242**, 7 (1996).

72 **Lujan, D. K., Stanziale, J. A., Mostafavi, A. Z., Sharma, S. & Troutman, J. M. Chemoenzymatic synthesis of an isoprenoid phosphate tool for the analysis of complex bacterial oligosaccharide biosynthesis. *Carbohydr Res* 359, 44-53 (2012).**

73 Coutinho, P. M., Deleury, E., Davies, G. J. & Henrissat, B. An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology* **328**, 307-317, (2003).

74 Lovering, A. L., de Castro, L. H., Lim, D. & Strynadka, N. C. J. Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis. *Science* **315**, 1402-1405 (2007).

75 Igura, M. *et al.* Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *The EMBO Journal* **27**, 234–243, (2008).

76 Breton, C., Snajdrova, L., Jeanneau, C., Koca, J. & Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**, 29R-37R, (2006).

77 Roseman, S. Reflections on glycobiology. *Journal of Biological Chemistry* (2001).

78 Varki, A. *et al.* Essentials of Glycobiology. *CSH Press*, 64-65 (2011).

79 Negishi, M., Dong, J., Darden, T. A., Pedersen, L. G. & Pedersen, L. C. Glucosaminylglycan biosynthesis: what we can learn from the X-ray crystal structures of glycosyltransferases GlcAT1 and EXTL2. *Biochem Bioph Res Co* **303**, 393-398, (2003).

80 Chang, A., Singh, S., Phillips, G. N., Jr. & Thorson, J. S. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *Curr Opin Biotechnol* **22**, 800-808, (2011).

81 Gantt, R. W., Gof, f. R. D., Williams, G. J. & Thorson, J. S. Probing the aglycon promiscuity of an engineered glycosyltransferase. *Angew Chem Int Ed Engl* **47**, 8889-8892 (2008).

82 Williams, G. J., Zhang, C. & Thorson, J. S. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nature chemical biology* **3**, 657-662 (2007).

83 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858, (2015).

84 Vetting, M. W., Frantom, P. A. & Blanchard, J. S. Structural and enzymatic analysis of MshA from Corynebacterium glutamicum: substrate-assisted catalysis. *J Biol Chem* **283**, 15834-15844, (2008).

85 Gale, R. T., Sewell, E. W., Garrett, T. A. & Brown, E. D. Reconstituting poly(glycerol phosphate) wall teichoic acid biosynthesis in vitro using authentic substrates. *Chemical Science* **5**, (2014).

86 Hager, F. F. *et al.* Functional Characterization of Enzymatic Steps Involved in Pyruvylation of Bacterial Secondary Cell Wall Polymer Fragments. *Front Microbiol* **9**, 1356, (2018).

87 Higuchi, Y. *et al.* A rationally engineered yeast pyruvyltransferase Pvg1p introduces sialylation-like properties in neo-human-type complex oligosaccharide. *Sci Rep* **6**, 26349, (2016).

88 Cava, F., de Pedro, M. A., Schwarz, H., Henne, A. & Berenguer, J. Binding to pyruvylated compounds as an ancestral mechanism to anchor the outer envelope in primitive bacteria. *Mol Microbiol* **52**, 677-690, (2004).

89 Ahlgren, J. A. Purification and Characterization of a Pyruvated-Mannose-Specific Xanthan Lyase from Heat-Stable, Salt-Tolerant Bacteria. *Appl Environ Microbiol* **57**, 2523-2528 (1991).

90 Forsberg, L. S. *et al.* Localization and structural analysis of a conserved pyruvylated epitope in Bacillus anthracis secondary cell wall polysaccharides and characterization of the galactose-deficient wall polysaccharide from avirulent B. anthracis CDC 684. *Glycobiology* **22**, 1103-1117, (2012).

91 Mesnage, S. *et al.* Bacterial SLH domain proteins are non covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *The EMBO Journal* **19**, 4473-4484 (2000).

92 Scott, P. M., Erickson, K. M. & Troutman, J. M. Identification of the Functional Roles of Six Key Proteins in the Biosynthesis of Enterobacteriaceae Colanic Acid. *Biochemistry-Us* **58**, 1818-1830, (2019).

93 De Silva, R. A., Wang, Q., Chidley, T., Appulage, D. K. & Andreana, P. R. Immunological Response from an Entirely Carbohydrate Antigen: Design of Synthetic Vaccines Based on Tn-PS A1 Conjugates. *Journal of American Chemical Society* **131**, 9622-9623 (2009).

94 Shi, M., Kleski, K. A., Trabbic, K. R., Bourgault, J. P. & Andreana, P. R. Sialyl-Tn Polysaccharide A1 as an Entirely Carbohydrate Immunogen: Synthesis and Immunological Evaluation. *J Am Chem Soc* **138**, 14264-14272, (2016).

95 Nishat, S. & Andreana, P. R. Entirely Carbohydrate-Based Vaccines: An Emerging Field for Specific and Selective Immune Responses. *Vaccines (Basel)* **4**, (2016).

96 Ghosh, S., Nishat, S. & Andreana, P. R. Synthesis of an Aminooxy Derivative of the Tetrasaccharide Repeating Unit of Streptococcus dysgalactiae 2023 Polysaccharide for a PS A1 Conjugate Vaccine. *J Org Chem* **81**, 4475-4484, (2016).

97 Fiebig, T. *et al.* Molecular cloning and functional characterization of components of the capsule biosynthesis complex of Neisseria meningitidis serogroup A: toward in vitro vaccine production. *J Biol Chem* **289**, 19395-19407, (2014).

98 Fiebig, T. *et al.* An efficient cell free enzyme-based total synthesis of a meningococcal vaccine candidate. *NPJ Vaccines* **1**, 16017, (2016).

99 McCarthy, P. C. *et al.* Chemoenzymatic synthesis of immunogenic meningococcal group C polysialic acid-tetanus Hc fragment glycoconjugates. *Glycoconj J* **30**, 857-870, (2013).

100 Kay, E. J., Yates, L. E., Terra, V. S., Cuccui, J. & Wren, B. W. Recombinant expression of Streptococcus pneumoniae capsular polysaccharides in Escherichia coli. *Open Biol* **6**, 150243 (2016).