Harnessing Microbial Biosynthetic Pathways for the Production of Complex Molecules

by

Mohamed Hassan

Thesis submitted
in partial fulfillment of the requirements for the
Doctorate in Philosophy degree in Chemistry

Department of Chemistry and Biomolecular Sciences, Faculty of Science
University of Ottawa

Abstract

Heterologous biosynthetic pathway expression is an essential tool for natural products biochemists. It has provided a powerful methodology for elucidating and characterizing bacterial biosynthetic pathways. In this thesis I will discuss methods to harness biosynthetic pathways for the heterologous production of a monosaccharide natural product, Legionaminic acid (Leg5,7Ac_2). This carbohydrate belongs to a family of sugars called nonulosonic acids (nine carbon α-keto acids) and is a 5,7-diamino derivative of sialic acid (Neu5Ac). It is found in cell surface glycoconjugates of bacteria including pathogens such as *Helicobacter pylori*, *Campylobacter jejuni*, *Acinetobacter baumanii* and *Legionella pneumophila*. Their presence on bacteria has been correlated with virulence in humans by mechanisms that likely involve subversion of the host's immune system or interactions with host cell surfaces due to its similarity to sialic acid. Further investigation into their role in bacterial physiology and pathogenicity is limited as there are no effective methods to produce sufficient quantities of these carbohydrates.

Herein, I harness microbial biosynthetic pathways via metabolic and genetic engineering to produce these complex nonulosonic acids. Leg5,7Ac_2 is produced from *N*-acetylglucosamine using the *Escherichia coli* strain BRL04, which results in substantial over-production (> 100 mg L^{-1} of culture). Pure Leg5,7Ac_2 could be readily isolated and converted into CMP-activated Leg5,7Ac_2 for biochemical applications as well as the phenyl thioglycoside for chemical synthesis applications. A similar strategy was employed to access the related nonulosonic acid pseudaminic acid (Pse5,7Ac_2). A biosynthetic pathway for production of Pse5,7Ac_2 was constructed from *H. pylori* and *C. jejuni* and expressed in *E. coli* BRL04. Unlike Leg5,7Ac_2, Pse5,7Ac_2 was produced in low yields (< 20 mg L^{-1}). A number of modifications were made to the biosynthetic constructs in an effort to enhance production levels yet improved titers were not obtained.

Additionally, this thesis will look at the development of a new strategy for the heterologous expression of biosynthetic pathways in a number of diverse hosts. I will highlight a flexible *in vivo* heterologous expression system that was inspired by viral protein packaging, processing and cleavage to produce violacein, a bright purple pigment with anti-tumor properties. A *de novo* polyprotein design possessing the violacein biosynthetic pathway was shown to work effectively in prokaryotic hosts such as *E. coli* and *S. typhimurium*. Expression of the polyprotein design in eukaryotic hosts like mammalian cells and *S. cerevisiae* were less successful. The ultimate goal of the work presented herein is to highlight the flexibility and powerful nature of synthetic biology for the *in vivo* production of natural products in addition to contributing to the vast arsenal of techniques and strategies that are currently available to researchers in this field.

Acknowledgements

First, I would like to thank my family for their love, care and support over the past several years. They have made my time during graduate school ever more enjoyable.

I would also like to thank Christopher Boddy for his expertise, mentorship and for being excellent PI to work for throughout my degree.

Finally, I would like to thank past and present members of the Boddy lab for being a great group of people to work with.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
Amp^R	Ampicillin resistant
ATP	Adenosine triphosphate
CDP	Cytidine diphosphate
CMP	Cytidine monophosphate
Cm^R	Chloramphenicol resistant
CTP	Cytidine triphosphate
DATDH	2,4-diacetamido-2,4,6-trideoxy hexose
DMB	4,5-methylenedioxy-1,2-phenylenediamine dihydrochloride
DNA	Deoxyribonucleic acid
EFI-EST	Enzyme Function Initiative-Enzyme Similarity Tool
Fa	Forssman Antigen
Fru-6-P	Fructose-6-phosphate
GDP	Guanidine diphosphate
GDP-diNAcBac	GDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucopyranose
GlcN-6-P	Glucosamine-6-phosphate
GalNAc	N-acetyl-D-Galactosamine
GlcNAc	N-acetyl-D-Glucosamine
GPI	Glycosylphosphatidylinositol
His	Histidine
HPLC	High performance liquid chromatography
IPTG	Isopropyl-β-D-thiogalactoside
Km^R	Kanamycin resistant
kDa	Kilo Dalton
Kdn	2-keto-3-3deoxy-D-glycero-D-galacto-nonulosonic acid
LB	Lysogeny Broth
LCMS	Liquid chromatography mass spectrometry
LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
ManNAc	N-acetyl-D-Mannosamine
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NagA	N-acetylmannosamine-6-phosphate deacetylase
NagK	N-acetyl-D-glucosamine kinase
NanA	N-acetylneuraminate lyase
NCBI	National Center for Biotechnology Information
Neu5Ac	Sialic acid, Neuraminic acid
Neu5Gc	Glycolylneuraminic acid

NeuB	N-acetylneuraminic acid synthetase
NeuC	UDP-GlcNAc 2-epimerase
NIS	N-iodosuccinimide
NulO	Nonulosonic Acids
PCR	Polymerase Chain Reaction
PEP	Phosphoenolpyruvate
Pse5,7Ac$_2$	Pseudaminic acid
PTS	phosphotransferase system
UDP	Uridine diphosphate
UDP-GlcNAc	Uridine diphosphate N-acetylglucosamine

Chapter One: Introduction

The complexity and diversity of carbohydrates as biopolymers is evident by their involvement in a number of structural, chemical and biological functions. From their function in cellular structure and integrity to the role carbohydrates play in cell signaling and immune response, a vast amount of research has elucidated the importance and biological impact that carbohydrates have. Carbohydrates found in nature are typically derived from monosaccharide subunits. Monosaccharides are nonhydrolyzable structures, typically possessing a structural core of 5 or 6 carbons, with specific activatable functional groups that allow for the generation of larger units known as oligosaccharides. What makes carbohydrates particularly interesting is the diversity and stereochemical variability that is possible between individual subunits and their linkages. Figure 1.1 highlights a number of common naturally occurring monosaccharides that are involved in a wide range of biological functions including glucose, which plays an essential role in glycolysis and energy production[1], N-acetylglucosamine (GlcNAc) which has a critical signalling role on the cell surface[2] and sialic acid, a key complex carbohydrate with involvement in host-pathogen interactions[3] and immune response.[4]

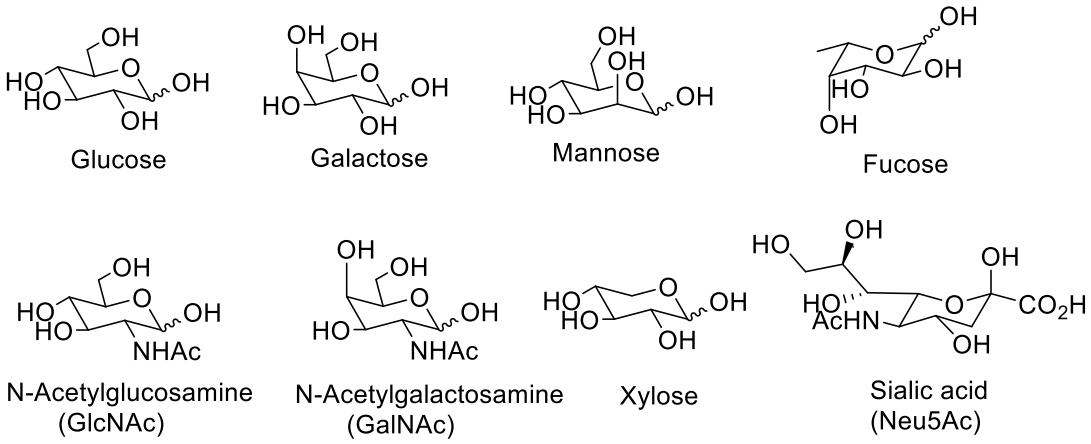


Figure 1.1: Common monosaccharides used as synthetic or biochemical building blocks to generate complex carbohydrates.

Although individual monosaccharides can play key roles in biological function, it is more typical to find biological function associated with glycoconjugates and glycoproteins. The synthesis of these important biomolecules is primarily dependent on the ligation of multiple monosaccharides and/or protein subunits, giving rise to diverse and structurally unique moieties.[5] While glycoconjugates and glycoproteins are shown to be involved in a number of functional and biochemical processes, detailed characterization of their function is frequently hampered by the inability to readily synthesize these compounds in pure form. This review will provide an overview of the current landscape of complex carbohydrates, with particular focus on a family of complex, nine carbon sugars known as nonulosonic acids (NulO's). Current limitations of complex monosaccharide synthesis will be examined. A brief summary of the various methods that have been employed to access these sugars will also be explored, in addition to analyzing areas of glycobiology that have been uncovered due to their availability.

1.1 Accessing complex carbohydrates via chemical synthesis

The structural complexity afforded by glycoconjugates, glycoproteins and glycolipids make them an incredibly interesting group of biopolymers. A significant number of carbohydrates that are of interest to glycobiologists and carbohydrate researchers exist as polysaccharides or glycoconjugates that are linked to one another via O-glycosidic bonds.[6] An inherent challenge in studying oligosaccharides is their presence as complex, heterogeneous mixtures in biology. This results in significant difficulty in their isolation to obtain a pure form for the study of their molecular functions. Thus, a key area of research involves accessing these carbohydrates via synthetic chemistry and generating biologically relevant glycoconjugates. The fundamental mechanisms that drive carbohydrate chemistry are well understood[7] and over the last few decades

an impressive array of complex glycoconjugates have been successfully synthesized.[8] Calicheamicin γ1, an extremely potent antitumor compound with a 9 membered ene-diyne ring is an example of a complex glycoconjugate produced via a demanding chemical synthesis to access a biologically relevant carbohydrate (Figure 1.2).[9] Synthetic strategies to obtain the carbohydrate moiety of calicheamicin in pure form enabled studying of its DNA binding interactions by the ene-diyne warhead. Not only is total synthesis of glycans desired, but generating these compounds in sufficient quantities and with high purity is essential to probe their biological function.

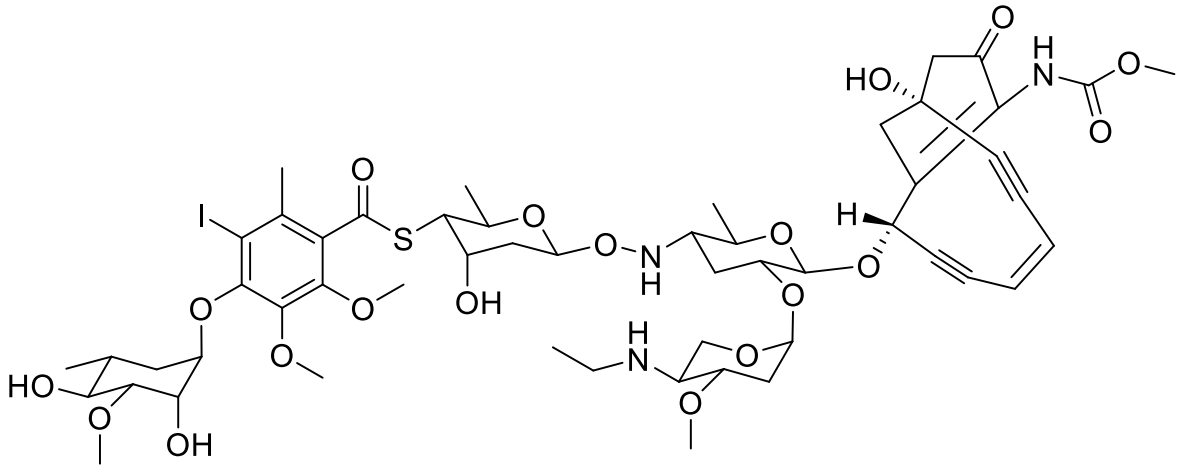


Figure 1.2: Structure of the potent antitumor compound Calicheamicin γ1.

Central to the ability to generate these complex glycoconjugates is a synthetic strategy that shows stereo- and regioselective control in glycosidic bond formation. For example, protecting group strategies can be employed to prevent other alcohols on a carbohydrate from reacting to form undesired glycosidic linkages. Advancements in protecting group strategies for carbohydrate synthesis are thus important tools for the regioselective control of glycosylation reactions and are well reviewed in the literature.[10,11] Guo *et al* provide an overview of protecting group strategies used in glycosylation reactions from a regioselective standpoint.[12] This review highlights limitations in terms of obtaining a desired regiochemical outcome using currently available

protecting group strategies. Additionally, Kulkarni *et al* summarize one-pot protection and deprotection strategies of carbohydrate building blocks, highlighting the flexibility and complexity of oligosaccharide synthesis.[13] Kulkarni *et al* also highlight a number of automated methods (both solid phase and solution phase strategies) that are important methods for rapidly constructing lengthy oligosaccharides, many of which are synthesized via repetitive chemical additions of monosaccharides. An automated solid-phase oligosaccharide assembly system called Glyconeer 2.1 was shown to rapidly assemble oligosaccharides of interest from monosaccharide subunits by Seeberger *et al*.[14] The major drawback to such automated systems is the large excess of monomer, in some cases in excess of 20 molar equivalents that are required to generate an oligosaccharide of interest.

In addition to the regiochemical control associated with protecting group strategies, another key aspect of synthetic carbohydrate chemistry for the generation of multi-subunit oligosaccharides are glycosylation reactions, which involve the ligation of an anomeric carbon atom of one sugar residue acting as an electrophile, to an alcohol group present on a different sugar acting as a nucleophile. This area of carbohydrate synthesis can be attributed to the early syntheses by A. Michael[15] and Emil Fischer[16], but has since expanded to include glycosylation methodologies that enable the generation of incredibly complex carbohydrates. Synthesis of carbohydrates is heavily reliant on a glycosylation method that can generate a desired stereochemical outcome.[17] Synthetic strategies for the formation of a glycosidic bond with a desired α or β stereoselectivity enable carbohydrate chemists to generate specific oligosaccharides of interest. In particular, the C-2 substituent can play a role in controlling the stereochemistry via neighbouring-group participation (Figure 1.3).[18] Equatorial C-2 substituents can use anchimeric assistance to control the stereochemistry of glycosylation, which is taken advantage of by carbohydrate chemists to

generate oligosaccharides possessing β glycosides. Another method of accessing a desired stereochemical outcome at the anomeric position is *in situ* anomerization.[19] This strategy, commonly referred to as halide catalysis, generates α-products in high yields as shown by the synthesis of 2-(α-L-Fucopyranosyl)- 3-(α-D-galactopyranosyl)-D-galactose by Lemieux *et al*.[19] Other methods include the use of solvents such as nitriles in glycosylation reactions to obtain >95% β-product[20], and intramolecular aglycone delivery to generate 1,2-cis-glycosides with β-configuration, as shown by Stork *et al* using a silicon-tethered strategy to produce β-mannopyranosides.[21] Alternative methods to generating glycosidic linkages include harnessing enzymes for use in chemoenzymatic approaches, a topic that will be further elaborated upon in this review.

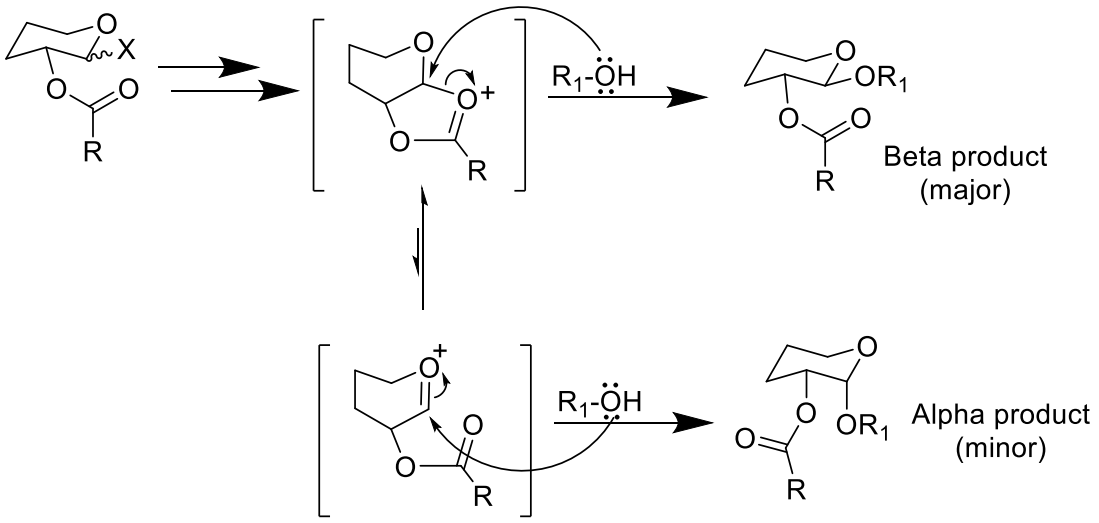


Figure 1.3: Stereochemical outcomes of glycosidic bond formation generated by the anchimeric effect. R designates nonparticipating group; X designates leaving group.

An example of controlling the stereochemistry of the anomeric position for the synthesis of a complex carbohydrate is that of the Glycosylphosphatidylinositol (GPI) membrane anchor that is

present on the cell surface of the malaria pathogen *Plasmodium falciparum* by Fraser-Reid *et al* (Figure 1.4).[22] This tetrasaccharide is comprised of four monosaccharides linked as α glycosides, with an inositol moiety possessing three bulky functional groups.

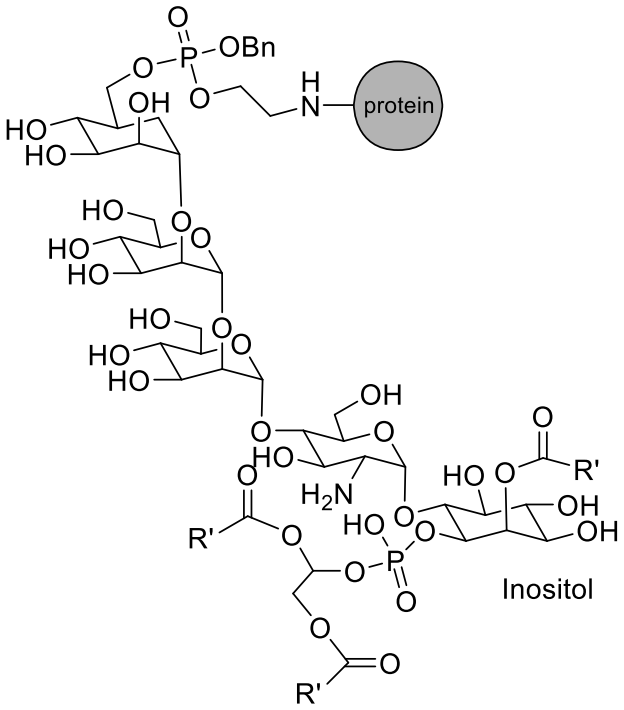


Figure 1.4: Structure of glycosylphosphatidylinositol membrane present on the cell surface of the malaria pathogen *Plasmodium falciparum.*

For a reaction of two carbohydrate moieties to occur, the anomeric carbon must be activated for nucleophilic displacement by the neighbouring group or incoming nucleophile. This is typically done through one of three strategies. The first method involves using a proton catalyst to enhance the leaving group ability of the substituent at the anomeric position, a mechanism that is used in many glycosidases.[23] Hans Peter Wessel prepared allyl-D-glucopyranosides using triflic acid as a catalyst to obtain a 3:1 α/β mixture.[24] Wessel also used concentrated sulfuric acid as a catalyst and produced a 3:1 α/β mixture.[24] The next method involves conversion of the anomeric hydroxyl

group to a halide-leaving group, which can be further activated by metal salts. Glycosyl halides were introduced by Koenigs and Knorr in 1901 and since then, a number of modifications and applications have evolved from their initial methodology, including the use of thioglycosides as donors.[25,26] Thioglycosides are attractive due to their stability under a variety of conditions, making them a robust activation method when considering various deprotection strategies. Base activation strategies that retain the anomeric oxygen atom are attractive alternatives to acid catalyzed methods. A well-studied base-activated strategy is the trichloroacetimidate method, which utilizes the electron-deficient trichloroacetonitrile to generate an O-linked sugar intermediate with a highly reactive trichloroacetimidate group.[27] This group is a bulky and strong electron withdrawing group which results in the acid-catalyzed release of the trichloroacetimidate as a good leaving group.

1.2 Using glycosyltransferases to access diverse, complex carbohydrates

The ability of enzymes to perform enantio-, stereo-, and regioselective modifications is a powerful tool in the development of complex molecules. While chemical reactions have been shown to have the capacity to perform reactions with similar selectivity, the exquisite regio- and chemoselectivity of enzymes, high yielding nature, aqueous solutions, ambient temperature and a lack of protecting groups makes enzymes incredibly useful. An excellent example of enzymes for the synthesis of complex molecules is the chemoenzymatic synthesis of nigelladine A by Frances Arnold and Brian Stoltz.[28] A regioselective C-H oxidation by a P450 enzyme was used in the latter stage of the nigelladine A route.

Chemoenzymatic approaches for the synthesis of carbohydrates provide additional tools for generating some of the complexity required to synthesize certain carbohydrates.[29] Central to chemoenzymatic synthesis is identifying an enzyme and elucidating its mechanism. Advances in

protein isolation technologies such as gene synthesis and recombinant protein expression and purification have made isolating an enzyme for *in vitro* studies relatively straight forward. This review will take a closer look at glycosyltransferases used for accessing carbohydrate complexity, but chemoenzymatic applications of carbohydrates also include phosphorylases[30], sulfotransferases[31], hydrolases[32] and many others.

Glycosyltransferases are a class of enzymes that catalyze the addition of saccharides by the formation of a glycosidic bond between nucleoside sugar donors and glycosyl acceptors.[33] Nucleoside sugar donors are usually activated nucleoside diphosphates (such as UDP-GlcNAc), but are also present as nucleoside monophosphate sugars. A number of activated sugar nucleosides, including CMP, CDP, CTP, GDP, UDP- activated sugars are used in a wide array of biochemical processes. Glycosyltransferases that use nucleotide sugars are referred to as Leloir enzymes, named after Luis Leloir, the 1970 Nobel Prize in Chemistry recipient for his work in sugar metabolism and biosynthesis.[34] In addition to carbohydrates as acceptors, glycosyltransferases can form linkages between sugar donors and proteins or lipids, giving rise to glycoproteins and glycolipids respectively. There are two stereochemical outcomes of glycosidic bond formation by a glycosyltransferase, which is how this class of enzymes are characterized (Figure 1.5). The stereochemical outcome of the anomeric position of the donor group either retains stereochemistry (retaining glycosyltransferases) or an inversion occurs (inverting glycosyltransferases). Additionally, the high stereoselectivity of glycosyltransferases makes them extremely useful for complex carbohydrate synthesis.

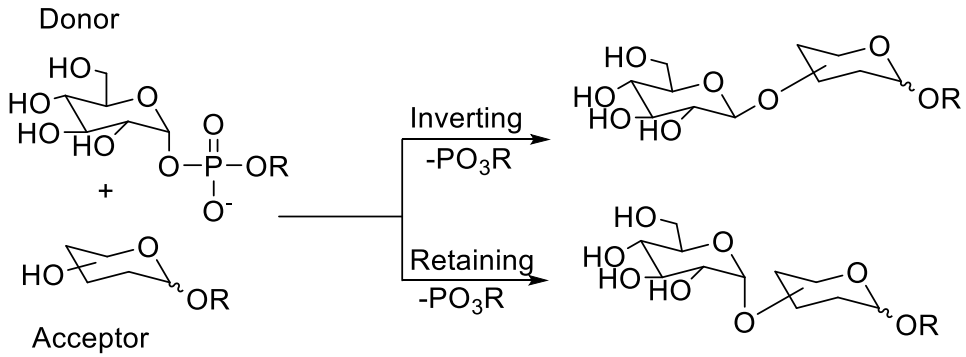


Figure 1.5: Mechanistic outcomes of Leloir glycosyltransferases. Upon formation of the glycosidic bond, the stereochemistry at the anomeric position of the activated donor sugar is either inverted or retained.

Glycosyltransferases are also found in both prokaryotic and eukaryotic glycobiology. These enzymes play a key role in the glycosylation of structural moieties in the outer membrane of gram-negative lipopolysaccharides (LPS).[35] Prokaryotic glycosylation can be found in two forms, depending on the nature of the residues that are attached to the donor carbohydrates. Gram-negative bacteria such as *Campylobacter jejuni* possess heavily glycosylated outer membranes.[36] Some regions are glycosylated with proteins that have carbohydrates attached to asparagine residues, referred to as N-linked glycosylation.[37] These areas are typically conserved with minimal modifications to the residues. Alternatively, carbohydrates can also be linked to serine or threonine residues to form O-linked glycoproteins, as observed in the flagellum of *C. jejuni*, containing variable O-linked residues.[38] The presence of these glycans in the outer membrane of bacteria are of particular interest to glycobiologists. Bacteria have been shown to utilize glycans to avoid host immune detection. *C. jejuni* synthesizes mimics of human GM1 ganglioside as part of its LPS for immune evasion.[39] *Escherichia coli* has also been shown to synthesize polysialic acid for immune evasion, similar to those found in mammalian cells.[40]

The highly specific nature of glycosyltransferases allows for their application for *in vitro* chemoenzymatic synthesis of oligosaccharides. The functionality of these carbohydrates has been of interest to researchers in academia and industry but acquiring sufficient amounts to elucidate their exact role in glycobiology is difficult. Purification of these glycans from the outer membrane of their respective bacterial producers is limited by the amount of sugar that is extracted and the difficulties in obtaining a pure glycan from a heterogenous mixture. Chemoenzymatic synthesis is an attractive alternative to obtain pure oligosaccharides, with the wealth of information that is available regarding the genes that encode the glycosyltransferases used by bacteria for the *in vivo* biosynthesis of these glycans. For example, some blood group antigens, such as P^k and P blood group antigens are precursors to the Forssman antigen (Fa). Fa is synthesized by *C. jejuni* and presumably uses this carbohydrate as a mimic of host cell glycans. Gilbert *et al* identified the three glycosyltransferases, CgtD, CgtE and Pm1138 that produce Fa (Figure 1.6), isolated soluble recombinant proteins and chemoenzymatically synthesized this pentasaccharide starting from *p*-ntirophenyl lactose.[41]

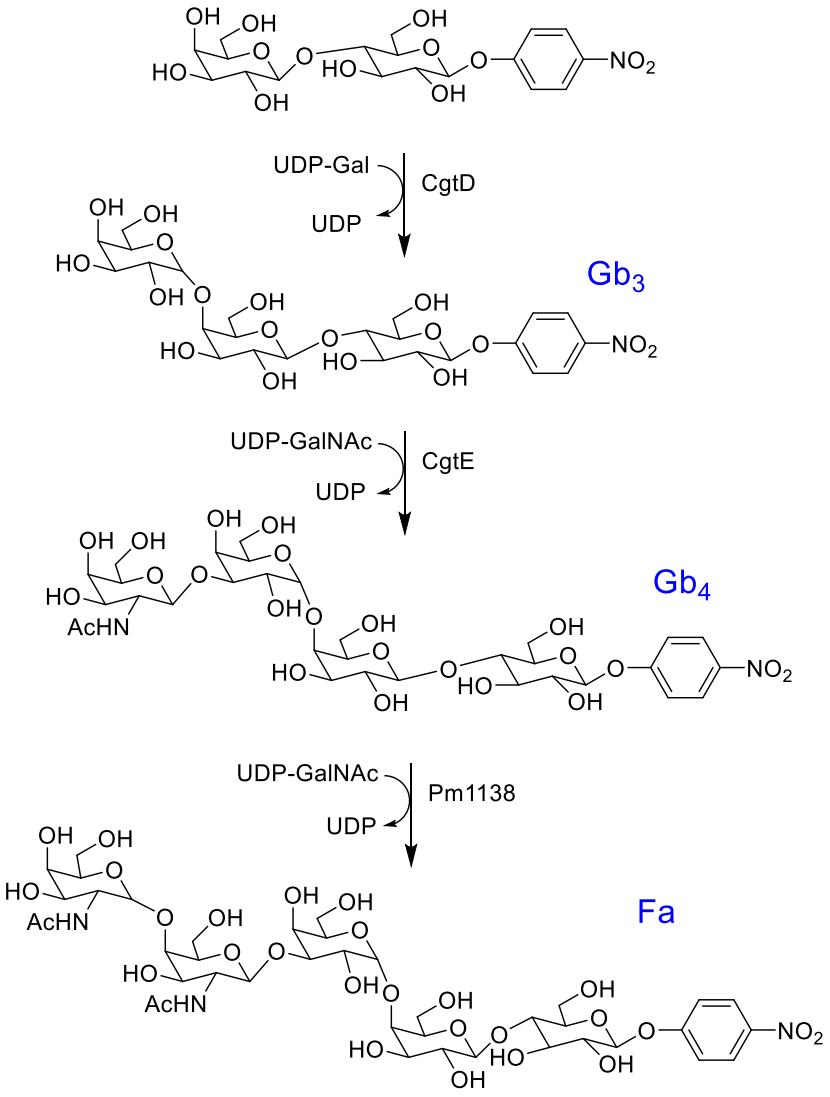


Figure 1.6: Chemoenzymatic synthesis of Fa using three consecutive glycosyltransferases.

An excellent example of coupling traditional carbohydrate synthesis with chemoenzymatic synthesis to produce complex carbohydrates is by Boons *et al*, who synthesized branching glycans by using glycosyltransferases for late-stage modifications.[42] The difficulty in synthesizing an oligosaccharide with over 15 subunits is exacerbated by factors such as protecting group compatibility. A more feasible strategy would incorporate synthetic carbohydrate chemistry to design a smaller core component and tailor the desired glycan with purified enzymes *in vitro*.

Initially, Boons *et al*. chemically synthesized a core pentasaccharide (Figure 1.7) that can be deprotected and selectively extended by glycosyltransferases to produce a library of 15 subunit oligosaccharides, which were then screened for their ability to bind lectins and influenza-virus hemagglutinins.

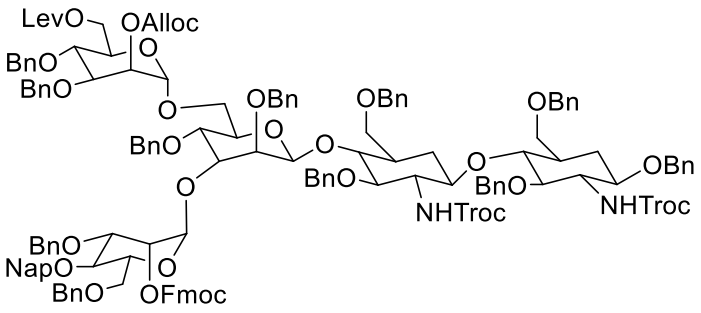


Figure 1.7: Core pentasaccharide that was used by Boons *et al.* as a starting point for chemoenzymatically synthesized oligosaccharides.

Complex glycoconjugates are needed to understand mammalian and bacterial glycobiology. Unlike other biopolymers such as proteins and DNA, complex carbohydrates cannot be readily isolated from organisms in pure form due to their heterogeneous nature, making it difficult to establish a specific role for a given carbohydrate in an impure glycoconjugate mixture. Chemical synthesis and chemoenzymatic approaches play an essential role in generating complex, highly pure glycans with excellent regio- and stereo-selectivity. Chemoenzymatic and synthetic methods require significant amounts of monomer to construct oligosaccharides of interest. The next section of this review will focus on a subset of these highly stereospecific and difficult to produce monomers.

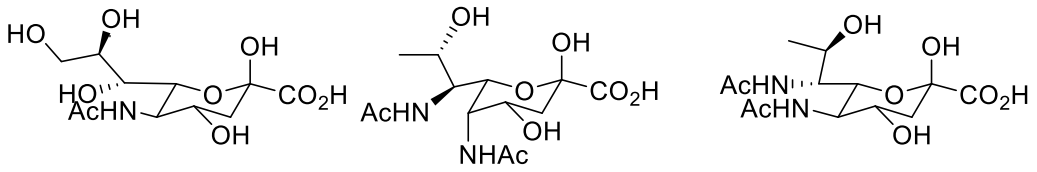
1.3 Nonulosonic acids: A multifaceted approach for the study and synthesis of a class of complex carbohydrates

The complexity of carbohydrates and a desire to develop a better understanding of their structural, signaling and immunological roles in both prokaryotes and eukaryotes has resulted in the chemoenzymatic and total synthesis of a vast amount of biologically relevant glycans. With advancements in molecular biology, sequencing technologies and bioinformatics, the ability to analyze and manipulate large amounts of genetic information enables producing complex carbohydrates via *in vivo* biosynthesis. This section will focus on the identification and the role of a group of complex, nine carbon carbohydrates known as nonulosonic acids (NulO). Various methods of producing these NulO's including chemical, chemoenzymatic and *in vivo* synthesis will be explored.

1.3.1 Sialic acids: A unique complex carbohydrate found in eukaryotes and prokaryotes

Nonulosonic acids (NulO's) are a class of nine carbon α-keto acids that are found present in prokaryotes and eukaryotes (Figure 1.8). The most well-known member of this family of sugars is sialic acid, which is primarily found on the surfaces of eukaryotic cells, but glycosylates bacteria as well.[4] The most common analogue of sialic acid is N-acetylneuraminic acid, or Neu5Ac, which can be derivatized to generate a number of analogues including N-glycolylneuraminic acid (Neu5Gc) and ketodeoxynonulosonic acid (Kdn).[43] Sialic acids are involved in a wide array of biological functions including cell signaling, cell adhesion and immune responses.[4] They are also heavily implicated in a number of human diseases including cancer.[44] The outer surface of malignant cells is heavily sialylated, promoting the invasiveness of cancers and metastatic tumours.[45] Additionally sialic acid metabolism has been found to be upregulated in metastatic

breast tumors.[46] Knocking out of certain biosynthetic genes of sialic acid decreased *in vivo* formation of lung metastases. Sialic acids are also used as neuraminidase inhibitors for the influenza virus.[47]



Sialic Acid (Neu5Ac) | Pseudaminic Acid (Pse5,7Ac$_2$) | Legionaminic Acid (Leg5,7Ac$_2$)

Figure 1.8: Most common derivatives of nonulosonic acids

Unlike sialic acid, little is known about the exclusively bacterial NulO analogues, pseudaminic acid (Pse5,7Ac$_2$) and legionaminic acid (Leg5,7Ac$_2$). This is primarily due to insufficient means of obtaining these complex NulO's with good yields and high purity to decipher their role and function in prokaryotic glycobiology. Sialic acid on the other hand, is readily available through a number of production strategies, ranging from chemical synthesis, to chemoenzymatic approaches, to heterologous expression and isolation from genetically engineered *Escherichia coli*. The first isolation of sialic acid was from sheep saliva by Gunnar Blix in 1936.[48] The term sialic is derived from the Greek word for saliva. Some conventional production strategies are costly or inefficient, such as traditional sialic acid isolation from egg yolk and milk whey sources. These require lengthy, extraneous purification processes that are typically low yielding (<10 %).[49,50] As isolation from natural sources is challenging, synthesis of sialic acid has played an important role in providing material for research efforts.

Chemical synthesis is very challenging due to sialic acid's rare nine carbon backbone and numerous stereocenters that must be accounted for. Formal synthesis of a sialic acid analog, Kdn, was described in 2001.[51] Kdn was synthesized with a 45% overall yield using a novel reaction involving ring closing metathesis (RCM). A follow-up publication in 2002 used the same RCM strategy to synthesize sialic acid.[52] In an elegant synthetic strategy, sialic acid and analogues (Neu5Gc) were synthesized in three steps from L-arabinose.[53] A reaction that was integral to this synthetic approach was an unprecedented base-catalyzed ring opening reaction that converts isoxazolidine to an α-keto acid moiety. Another sialic acid synthetic strategy involved a rhodium-catalyzed addition of a nitrene to a glycal providing a high degree of stereocontrol, which was used in the synthesis of sialic acid.[54] While the chemical synthesis of sialic acid is feasible, it has been shown to be challenging and difficult to scale.

An alternate approach for the production of sialic acids is by *in vitro* chemoenzymatic synthesis. This strategy is particularly viable due to the relatively straightforward biosynthetic pathway of sialic acid in bacteria (Figure 1.9).[55] Intracellular UDP-GlcNAc, a substrate tightly regulated by homeostasis due to its role in cell wall biosynthesis, is converted to ManNAc by a UDP-GlcNAc 2-epimerase. ManNAc is then condensed with phosphoenolpyruvate (PEP) with a sialic acid synthase to produce sialic acid. Large scale chemoenzymatic production of sialic acid was achieved by Maru *et al* in 1998 using the described biosynthetic pathway, and from GlcNAc 29 kg of sialic acid with a conversion rate of 77% was produced.[56] The most common widely used large-scale industrial production of sialic acid is via chemoenzymatic synthesis.[57] This method involves the chemical epimerization of GlcNAc to ManNac followed by condensation with pyruvate by the reverse catalytic reaction of the sialic acid aldolase enzyme, NanA.[57,58]

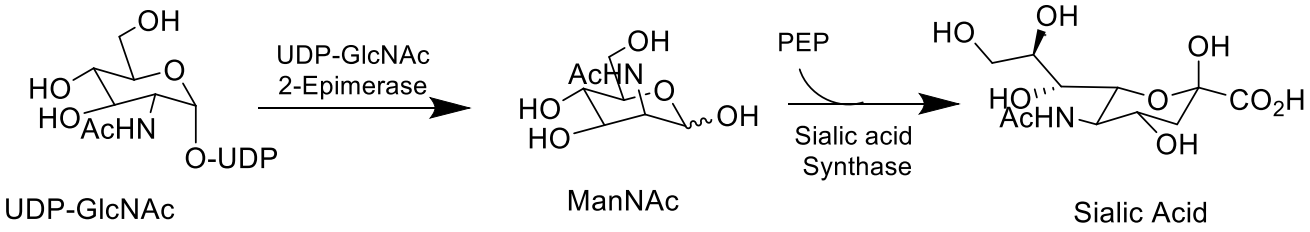


Figure 1.9: 2-step enzymatic conversion of UDP-GlcNAc to sialic acid.

Employing chemoenzymatic approaches does have its drawbacks. First, the large-scale purification of enzymes is labour intensive and expensive. Additionally, expensive cofactors and reagents are necessary for the desired enzymatic activity. Scalability also has its own difficulties; thus, alternative approaches can be desirable. In 2007, Boddy *et al* described a method for the *in vivo* heterologous expression and production of sialic acid using a genetically engineered strain of *E. coli*.[59] *NeuB* and *neuC* from the sialic acid biosynthetic pathway of *Neisseria meningitidis* were cloned and expressed in an *E. coli* strain lacking the sialic acid transporter, *nanT* and the sialic acid aldolase *nanA*, that cleaves the nine carbon sugar into the six carbon ManNAc and three carbon pyruvate precursors. This *in vivo* production strategy resulted in gram scale production of sialic acid with high purity and with a relatively low cost. A number of other *in vivo* sialic acid production strategies have also been reported[60,61], including the use of a bioreactor to increase sialic acid production to > 20 g/L.[62]

1.3.2: Metabolic labelling of glycans: Using click chemistry to understand complex carbohydrates.

Tools to study these important carbohydrate monomers have been limited. One such tool was derived from the ability to selectively label or target a carbohydrate of interest that resides on the outer surface of a bacterial or mammalian cell. This was a highly desired methodology that

would allow for researchers to gain a better understanding into host-pathogen interactions and immune effects. A pioneer in this field is Carolyn Bertozzi, who in 2000, utilized a high specific chemical reaction from the early 1900's to selectively incorporate a modified cell-surface sugar moiety (Figure 1.8).[63,64] Bertozzi *et al* started from N-azidoacetylmannosamine (ManNAz), a C-2 azido derivative of ManNAc, the biosynthetic precursor of sialic acid. Sialic acid is found to heavily glycosylate the outer surface of eukaryotic organisms.[43] Interestingly, when Jurkat cells were supplemented with ManNAz, the biosynthetic machinery converted this exogenous substrate to a sialic acid moiety with an azido group at C-5. This then resulted in the cell surface incorporation of the modified sialic acid to the outer surface of the Jurkat cells (Figure 1.10). The viability of azido-sialic acid production is dependent on two important points. First, the endogenous biosynthetic enzymes of sialic acid should be able to convert a non-native substrate, such as ManNAz, to the sialic acid derivative. This is reliant on the two enzymes involved in sialic acid biosynthesis tolerating a C-2 modification of the acetyl group to an azidoacetyl moiety with sufficient efficiency allowing for a detectable amount of azido-sialic acid biosynthesis. Secondly, side reactions would not arise and be a hindrance in terms of toxicity to the cell. The cell surface modified sialic acids can be tagged with a biotinylated triarylphosphine substrate via a Staudinger reaction, developed by Hermann Staudinger in 1919.[65]

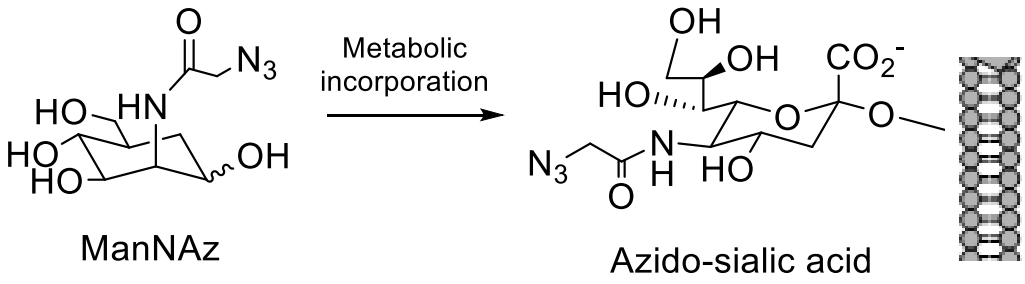


Figure 1.10: Incorporation of azido-sialic acid on the outer surface of Jurkat cells by Bertozzi *et al* starting with ManNAz

This work is an incredibly powerful methodology for metabolic labelling of glycans. Bertozzi *et al* then expanded this work by showing that in *vivo* chemical labelling is not limited to cells, but can be performed in living mice.[66] A phosphine label with a FLAG tag is used for click chemistry, and *in vivo* work in live mice can be performed due to the mild conditions required for the Staudinger ligation. Splenocytes were analyzed 90 mins after the click reaction took place and clear metabolic incorporation of the azido derivative of sialic acid from ManNAz is evident. Over the last two decades the use of this methodology has exploded and click reactions for *in vivo* and *in vitro* labelling have been a staple for work in chemical biology and proteomics. This work has also been used for the metabolic labelling of complex carbohydrates, with a particular focus on sialic acids.

1.3.3 Bacterial nonulosonic acids: Legionaminic acid and pseudaminic acid

A number of gram-negative pathogens such as *Campylobacter jejuni, Helicobacter pylori* and *Legionella pneumophila* are heavily glycosylated by nonulosonic acids. *C. jejuni* is the principal bacterial cause of human gastroenteritis world-wide.[67] *H. pylori* is primarily responsible for intestinal ulcers[68] and *L. pneumophila* is the causative agent of Legionnaires disease, which can lead to deadly pneumonia-like symptoms.[69] Accessing these complex carbohydrates would provide further clarification about their specific role in pathogenic bacteria, which is not well understood. It is well documented that glycosylation is important for the motility and pathogenicity of *C. jejuni*.[70] Knockout mutants of certain NulO's in the outer surface of pathogenic bacteria, in particular the flagellin of *C. jejuni,* have been shown to drastically reduce motility and virulence, two key factors of bacterial pathogenesis.[70] This O-linked glycosylated flagellum provides unique

structural complexity that is not clearly understood and further elucidation may propose potential targets for novel antibiotics.

The glycosylation of gram-negative pathogens contains a number of regions in the outer surface of these bacteria that are variable in terms of carbohydrate composition (Figure 1.11).[71] The O-linked flagellum, the capsular polysaccharides and the lipooligosaccharides of these prokaryotes possess a unique and complex set of carbohydrates that are bacterial analogues of sialic acid. Although there are dozens of bacterial analogues that have been identified, the two most prevalent bacterial NulO's are legionaminic acid (Leg5,7Ac_2, 5,7-diacetamido-3,5,7,9-tetradeoxy-D-glycero-D-galacto-nonulosonic acid) and pseudaminic acid (Pse5,7Ac_2). While the importance of these bacterial sialic acid analogues in terms of their direct involvement in pathogenicity has been shown, the underlying mechanism and the role of NulO's in bacterial physiology, along with their impact in host-pathogen interactions is largely unknown. Leg5,7Ac_2 possesses similar stereochemistry to sialic acid at the 5, 7 and 8-C positions, giving rise to the hypothesis that bacterial pathogens utilize Leg5,7Ac_2 as a molecular mimic of sialic acid. Nevertheless, research has been severely hampered due to the unavailability of these complex carbohydrates, limiting the ability for further research and understanding of these unique and complex nine carbon sialic acid analogues. We will compare the current state of production strategies by both synthetic chemistry and biochemical methods of these bacterial NulO's to highlight the remarkable difficulty that must be overcome to access these carbohydrates.

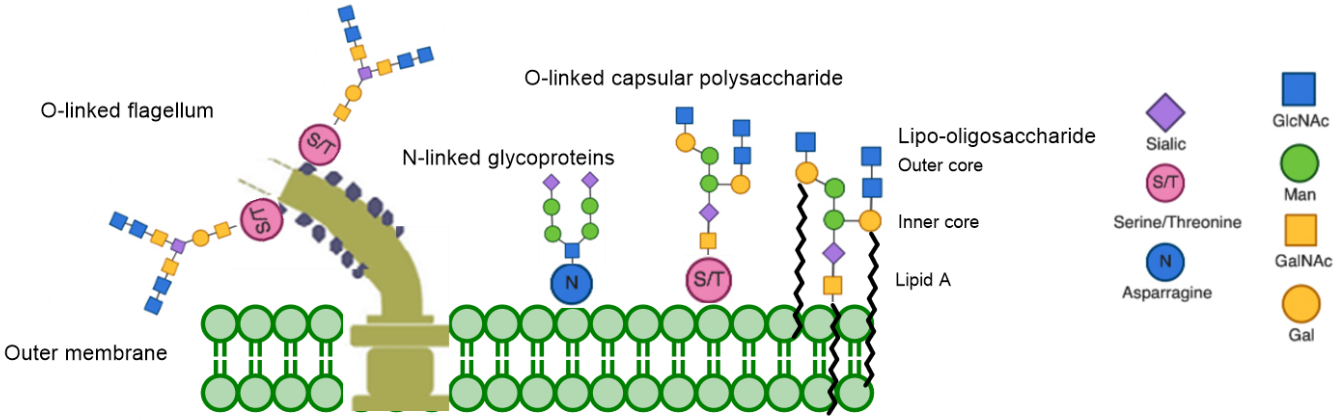


Figure 1.11: Glycosylation of bacterial pathogens. Figure adapted with permission from Van den Steen *et al.*[71]. See appendix for details regarding copyright permissions.

Leg5,7Ac$_2$ was first identified in the O-polysaccharide of LPS in *L. pneumophila* in 1994.[72] Leg5,7Ac$_2$ has also been identified in a number of other pathogenic bacteria such as *Campylobacter jejuni*[73], *Campylobacter coli*[74], *Cronobacter tureicensis*[75], *Acinetobacter baumannii*[76,77] and *Escherichia coli*[78]. The 3-deoxy-D-glycero-D-galacto2-nonulosonic acid skeleton that Leg5,7Ac$_2$ shares with sialic acid along with its exclusive utilization in prokaryotes makes it highly interesting in terms of understanding its role in pathogen colonization and host-pathogen interactions. Although our understanding of the specific role that Leg5,7Ac$_2$ plays is limited, its importance has been highlighted by work on *Vibrio fischeri*, a bacteria that exists in a symbiotic relationship with the Hawaiian bobtail squid, whereby a knockout of a ligase *waaL*, responsible for the synthesis and assembly of the O-antigen LPS limited bacterial motility and thus negatively impacted colonization.[79] The *V. fischeri* mutants were also unable to compete with the wild-type strain in co-colonization studies.

With Leg5,7Ac$_2$ glycosylating a number of bacterial strains, further work is required to elucidate the roles and functions in which this sugar is involved in. This has been a difficult task

considering further work has been limited due to the lack of availability of this complex sugar. A number of synthetic routes have been developed for Leg5,7Ac$_2$, the first of which sought to establish the configuration of this sugar, although this route was low yielding.[80] In 2015, Seeberger *et al* (Figure 1.12) reported a *de novo* total synthesis of Leg5,7Ac$_2$ utilizing an orthogonally protected building block as a glycosylating agent which was used for serological studies.[81] The low-yielding nature of this route limits the ability to utilize it to understand the role of Leg5,7Ac$_2$ in bacterial pathogenesis.

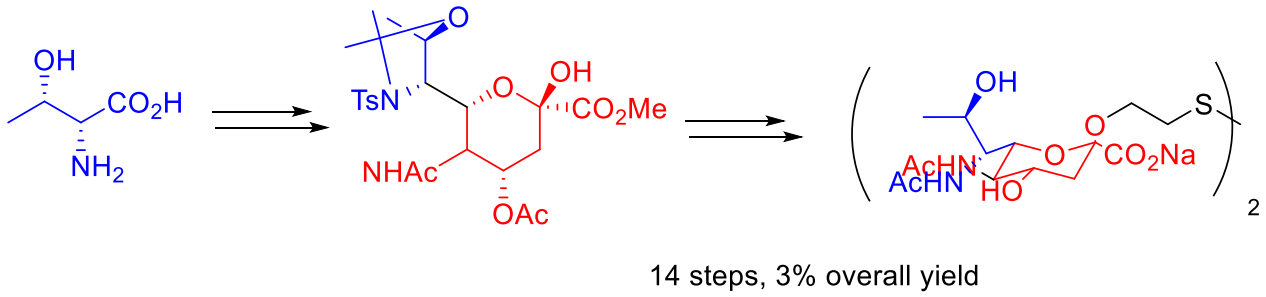


Figure 1.12: Seeberger *et al* route for the synthesis of a Leg5,7Ac$_2$ donor for serological studies.

The complexity involved in the synthesis of Leg5,7Ac$_2$ is highlighted by a route to produce this carbohydrate from the nine-carbon analogue, sialic acid, in 15 steps with an overall yield of 17% (Figure 1.13).[82] A donor Leg5,7Ac$_2$ with an adamantanyl thioglycoside is activated with N-iodosuccinimide (NIS) and trifluoromethanesulfonic acid in the presence of a variety of different alcohols to generate desired glycosides, which is demonstrate to possess good selectivity at the equatorial position.

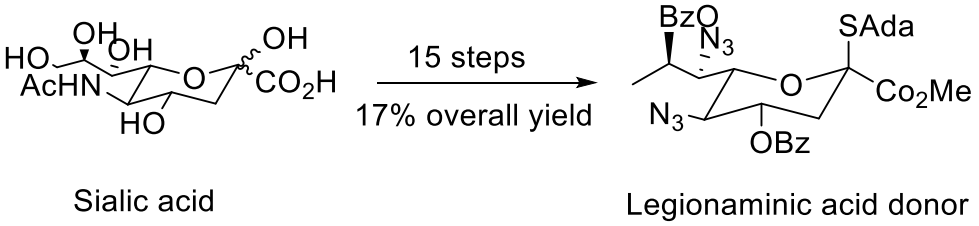


Figure 1.13: Crich *et al's* route for the synthesis of a Leg5,7Ac$_2$ donor from sialic acid. This synthetic route involved 15 steps with an overall yield of 17% to generate a Leg5,7Ac$_2$ donor**.**

Following along with a similar synthetic approach, a route for the production of Leg5,7Ac$_2$ as well as its C-7 analogues from sialic acid has also been described.[83] This strategy involves functionalizing the C-7 nitrogen in the latter stages of the synthesis and providing a suitable starting point for further synthesis of analogues containing C-7 amide derivatives of Leg5,7Ac$_2$. More recently, Leg5,7Ac$_2$ and 4-epi- Leg5,7Ac$_2$ were synthesized starting from a simple D-serine starting material.[84] A protected D-serine is converted to a derivative with a desired C-6 D-rhamno configuration, which is then further elongated to introduce α-ketoacid functionality. Orthogonally protected amines afford a means for further analogues to be synthesized.

In addition to total synthesis strategies, the utilization of a chemoenzymatic approach has been moderately successful for the production of Leg5,7Ac$_2$ and glycoconjugates containing this sugar. Chen *et al* employed such a strategy by initially starting from an inexpensive D-fucose moiety to generate a 6-deoxyMan-2,4diN$_3$ intermediate in 8 steps at 60% yield (Figure 1.14).[85] This was followed by two chemoenzymatic modifications, first to complete the Leg5,7Ac$_2$ backbone by using an aldolase to condense pyruvate with 6-deoxyMan-2,4diN$_3$, and secondly to transfer this 9-C intermediate onto a variety of glycoconjugates, affording flexibility and a means of efficiently modifying and generating a variety of glycoconjugates of interest. Once the

glycoconjugate has been synthesized, the diazido groups were converted to acetamide to generate free Leg5,7Ac$_2$-containing glycoconjugates.

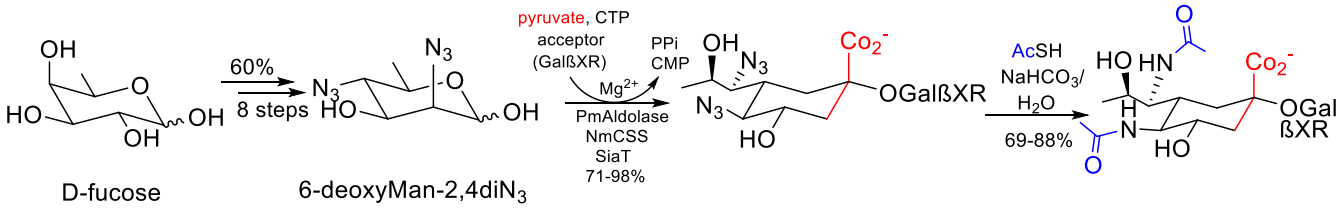


Figure 1.14: Chen *et al's* combined synthetic and chemoenzymatic approach for the synthesis of Leg5,7Ac$_2$ containing glycoconjugates. Starting from D-fucose, 8 steps were required to generate a 2,4-diazidomannose precursor and was followed by 2 chemoenzymatic transformations to produce a glycoconjugate containing Leg5,7Ac$_2$.

An alternative chemoenzymatic approach involving a set of three enzymes responsible for the biosynthesis of CMP-N,N'-diacetyllegionaminic acid (currently known as CMP-Leg5,7Ac$_2$) were identified by Glaze *et al* in 2008 from *L. pneumophila*.[86] First, a sialic acid *NeuC* homolog, was shown to be a hydrolysing 2-epimerase that converts UDP-N,N'-diacetylbacillosamine into 2,4-diacetamido-2,4,6-trideoxymannose. While there are similarities with NeuC from *E. coli*, UDP-GlcNAc was not tolerated by the *L. pneumophila* enzyme, indicating that this enzyme is exclusively used in the biosynthesis of Leg5,7Ac$_2$. Additionally, it was shown that a sialic acid NeuB homolog, which condensed 2,4-diacetamido-2,4,6-trideoxymannose with PEP, similar to synthases shown in the biosynthesis of pseudaminic acid and sialic acid was also present in the biosynthetic pathway. The final enzyme is a NeuA homolog, which readily converted N,N'-diacetyllegionaminic acid in the presence of CTP into CMP-N,N'-diacetyllegionaminic acid, which is the predicted product of a CMP-Leg5,7Ac$_2$ synthetase. These three enzymes, coupled with the three enzymes that convert UDP-GlcNAc into UDP-N,N'-diacetylbacillosamine, established a pathway that produces CMP-Leg5,7Ac$_2$ in *L. pneumophila*.

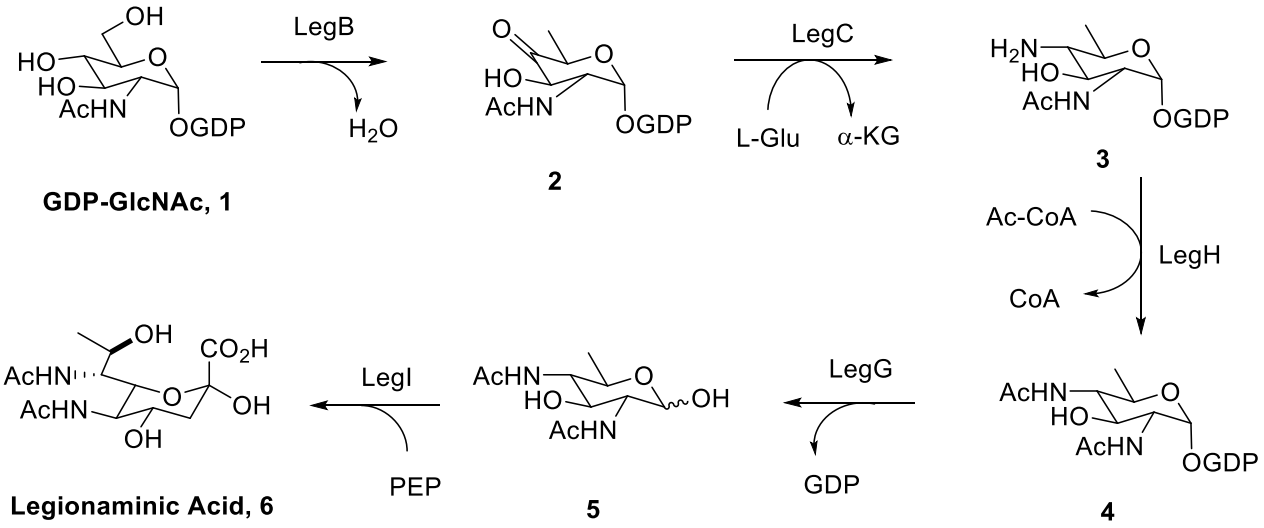


Figure 1.15: Biosynthesis of Leg5,7Ac$_2$ from GDP-GlcNAc was elucidated in *C. jejuni.*

In 2009, the biosynthetic pathway for Leg5,7Ac$_2$ in *Campylobacter jejuni* was elucidated by Schoenhofen *et al*.[73] Unexpectedly this pathway was found to use GDP-GlcNAc (**1,** Figure 1.15) as the key building block, unlike related nonulosonic acid biosynthetic pathways which use UDP-GlcNAc. The native Leg5,7Ac$_2$ pathway starts with the NAD^+-dependent dehydratase LegB, producing the 4-keto intermediate, GDP-2-acetamido-2,6-dideoxy-α-D-*xylo*-hexos-4-ulose, **2**. LegC, a PLP-dependent aminotransferase, catalyzes the transfer of the amino group from L-glutamate to the 4-keto intermediate producing the amino sugar GDP-4-amino-4,6-dideoxy-α-D-GlcNAc, **3.** The acetyltransferase LegH then acylates the C-2 amine to produce GDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucopyranose (**4** GDP-diNAcBac), which is converted into 2,4-diacetamido-2,4,6-trideoxy-D-mannopyranose, **5**, by a hydrolyzing 2-epimerase, LegH. The final step towards the biosynthesis of Leg5,7Ac$_2$ is the condensation of pyruvate with the 6-C backbone containing intermediate by the synthase LegI to generate **6**. Utilizing these five enzymes identified

as the Leg5,7Ac$_2$ pathway in *C. jejuni*, Schoenhofen *et al* employed a one-pot chemoenzymatic strategy supplemented with appropriate cofactors, substrates and reagents to produce Leg5,7Ac$_2$ that was purified with CE-MS. Although this strategy yields highly pure product, costly reagents for enzyme purification prevents it from becoming a feasible strategy for producing mass quantities of this complex sugar. Building off this work, the Boddy lab generated Leg5,7Ac$_2$ from a *de novo* biosynthetic pathway using enzymes from *C. jejuni* and *L. pneumophila* that utilized a UDP-GlcNAc starting substrate, as described in Chapter 2 of this thesis.[87]

Pseudaminic acid (Pse5,7Ac$_2$ 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-α-L-*manno*-nonulosonic acid) is an exclusively prokaryotic nonulosonic acid stereoisomer of Leg5,7Ac$_2$ that was discovered by Knirel *et al* in 1984 from the LPS of *Pseudomonas aeruginosa*.[88] Since then, pseudaminic acid has been found in a number of pathogenic bacteria including *Helicobacter pylori*[89] and *C. jejuni*.[90,91] More recently, the periodontal pathogen *Tanerella forsythia* was shown to be heavily glycosylated with Pse5,7Ac$_2$.[92] Additionally, biosynthetic analysis revealed a Pse5,7Ac$_2$ gene cluster and a knockout mutant of a candidate Pse5,7Ac$_2$ glycosyltransferase from *T. forsythia* resulted in the O-glycans not being capped by this nine-carbon sugar.[93,94] There are a number of structural features that differentiate Pse5,7Ac$_2$ from sialic acids, including the presence of a methyl group at C-9. Stereochemical inversions at carbons five, seven and eight distinguish this sugar from Leg5,7Ac$_2$. As is the case for the vast majority of NulO's (including all bacterial ones) the questions involving their biochemistry and glycobiology remain unanswered due to a lack of a suitable production method.

There are a number of synthetic strategies that have been employed to produce Pse5,7Ac$_2$. The first reported synthesis was by Knirel *et al,* whereby a 3,4-dibenzoyl rhamnose sugar moiety was used to synthesize the 6-C AltdiNAc intermediate, which was followed by an aldol reaction

with oxaloacetic acid to generate Pse5,7Ac$_2$ (Figure 1.16).[80] In addition to Pse5,7Ac$_2$, Knirel *et al* generated 8 additional 9 carbon analogues including Leg5,7Ac$_2$.

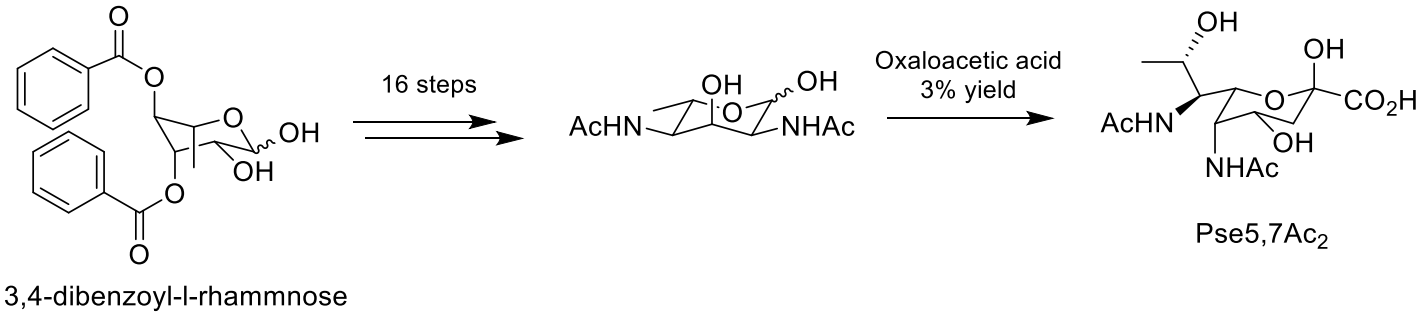


Figure 1.16: Knirel *et al's* synthetic strategy to generate Pse5,7Ac$_2$ from 3,4-dibenzoyl-l-rhammnose.

A decade later, Ito *et al* reported a synthetic strategy starting from GlcNAc to produce 6-deoxy AltdiNAc, a 6-C intermediate in the biosynthetic pathway of Pse5,7Ac$_2$ in 14 steps with 15% yield.[95] This intermediate was then used in an elongation reaction by In-mediated allylation with a bromomethacrylate ester. The final step involved ozonolysis and hydrolysis to produce the Pse5,7Ac$_2$ with a final overall yield of 4%.

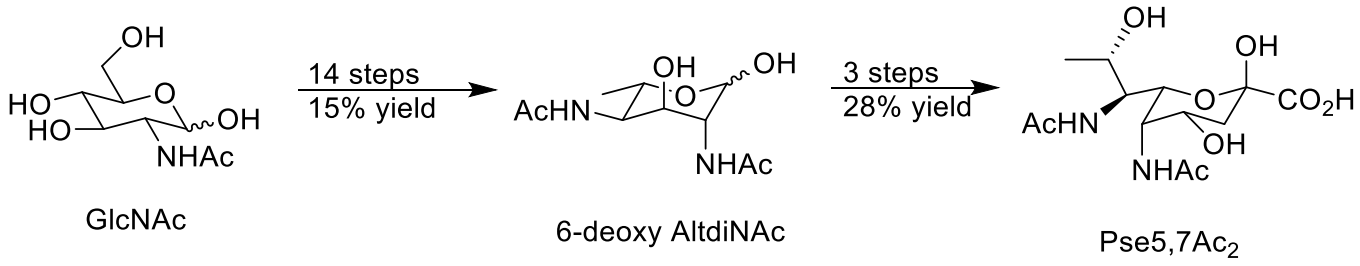


Figure 1.17: Ito *et al's* synthesis of Pse5,7Ac$_2$ by accessing a 6-C intermediate of the Pse5,7Ac$_2$ biosynthetic pathway, 6-deoxy AltdiNAc.

A few years later, Kiefel *et al* synthesized 8-epi Pse5,7Ac$_2$ from sialic acid in 8 steps and an overall yield of 35% (Figure 1.18).[96] This synthesis is particularly interesting due to the functionalization of the nitrogen groups at C-5 and C-7 via bis-azides that were used in the synthetic route. This can lead to the selective synthesis of desired analogues, as shown by the author's work selectively generating a 7-acetamido-5-azido derivative using a Staudinger reaction. The azide groups were then reduced and acetylated to produce the desired acetamido groups. The 9-hydroxy group was also removed by iodination and reduction to generate the free C-9 methyl, resulting in the formation of 8-epi Pse5,7Ac$_2$. Shortly thereafter, Kiefel and Payne unsuccessfully attempted to synthesize Pse5,7Ac$_2$ by inverting the stereochemistry of C-8 from 8-epi Pse5,7Ac$_2$.[97] Instead, they proposed an alternate, 17 step route from commercially available sialic acid.

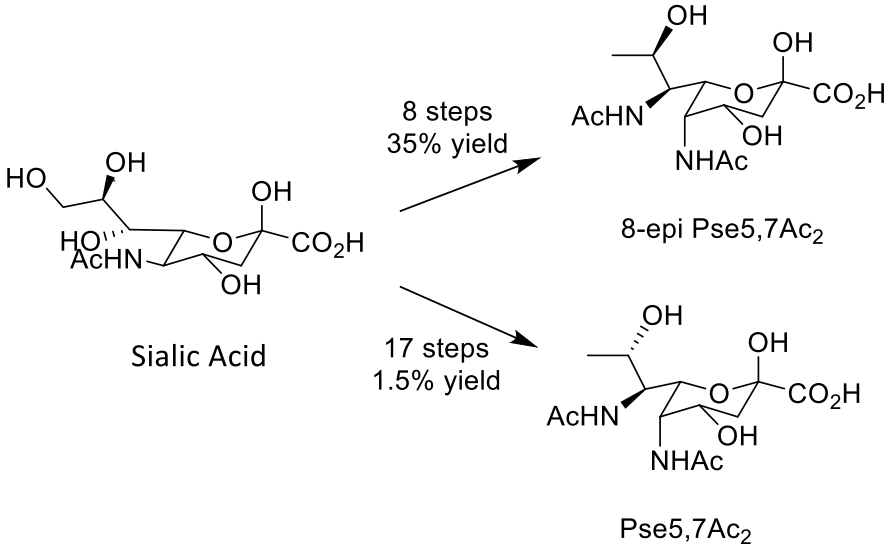


Figure 1.18: Kiefel, Payne *et al's* synthesis of Pse5,7Ac$_2$ and 8-epi Pse5,7Ac$_2$ from sialic acid.

A third synthetic strategy using sialic acid was reported in 2018 by Crich *et al* (Figure 1.19).[98] This synthesis was achieved in 20 relatively straightforward steps, resulting in a 5% yield at sub-gram scale of a Pse5,7Ac$_2$ donor. Glycosylation studies showed that this donor had excellent

equatorial selectivity and suitable conditions that afforded cleavage of 5- and 7-azido groups to release the amines were tested and successfully determined.

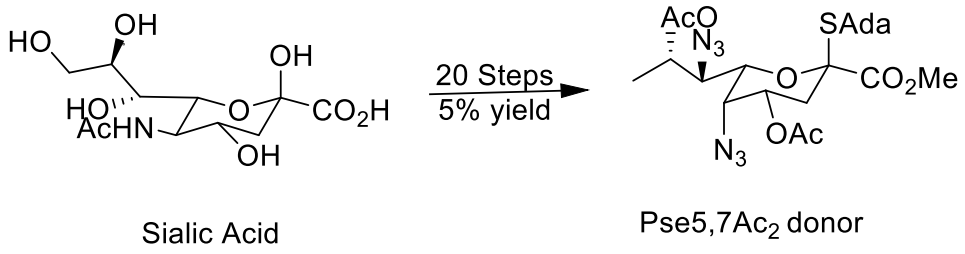


Figure 1.19: Synthesis of a Pse5,7Ac$_2$ donor from sialic acid by Crich *et al* in 20 synthetic steps with a 5% overall yield.

Li *et al* synthesized Pse5,7Ac$_2$ derivatives along with the *P. aeruginosa* 1244 pilin glycan starting from readily available L-threonine (Figure 1.20).[99] The first step is to generate Cbz-L-allo-threonine methyl ester by protecting the hydroxyl and amino groups, reducing the carboxylic acid to an aldehyde and inverting the stereochemistry at C-3. Then, this *de novo* method generates a 1,3-anti-diamino skeleton, followed by a Fukuyuma reduction and an indium-mediated Barbier-type allylation to produce the desired product in 25 steps and 4% overall yield. Additionally, the glycosylation of Pse5,7Ac$_2$ glycosyl donors was examined, leading to the synthesis of the trisaccharide that is present on the pilin *of P. aeruginosa*, Pse5,7Ac$_2$-(2→4)-β-Xyl-(1→3)-FucNAc.

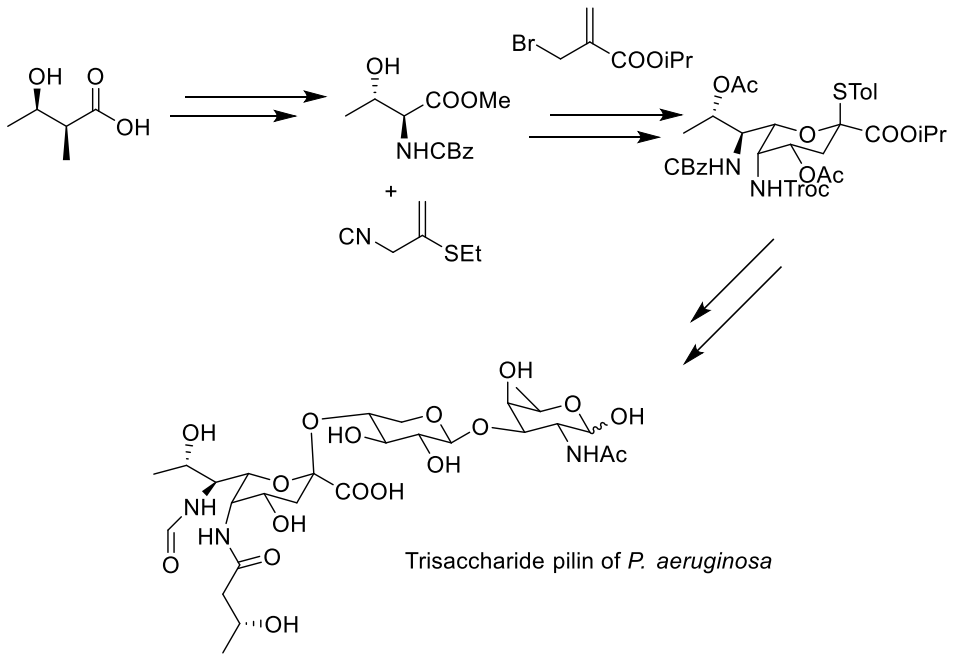


Figure 1.20: Li *et al's* synthesis and utilization of Pse5,7Ac$_2$ as a glycosyl donor to generate atrisaccharide pilin of *P. aeruginosa*.

The complexity of producing NulO's via total synthesis is highlighted by synthetic schemes requiring multi-steps with complex reactions, the vast majority resulting in poor yields. An alternative strategy to access these sugars is to identify and harness the biosynthetic pathways from producing organisms and employ an *in vitro* chemoenzymatic approach. In 2006, the biosynthesis of CMP-Pse5,7Ac$_2$ was elucidated in *H. pylori* by Shoenhofen *et al*.[89] Unlike Leg5,7Ac$_2$ which utilizes GDP-linked precursors, Pse5,7Ac$_2$ is biosynthesized from UDP-GlcNAc (**7**) in a six-step enzymatic transformation (Figure 1.21). The biosynthetic pathways of the two stereoisomers are quite similar, but there are two key distinctions that give rise to each sugar's unique stereochemistry. A NAD(P)+ dependent C-4,6 dehydratase, PseB, carries out an additional C-5 epimerization to produce UDP-2-acetamido-2,6-dideoxy-β-L-arabino-hexos-4-ulose (**8**). PseB works in conjunction with an aminotransferase PseC to convert **8** into UDP-linked UDP-4-

amino-4,6-dideoxy-b-L-AltNAc (**9**).[100] An acetyltransferase PseH results in the N-4 acetylation of **9**, producing UDP-2,4-diacetamido2,4,6-trideoxy-β-L-altropyranose (**10**). A hydrolase PseG removes UDP from C-1 of 4 to produce an unusual 2,4-diacetamido-2,4,6- trideoxy-β-L-altropyranose (DATDH, **11**). Unlike the hydrolase present in the Leg5,7Ac$_2$, PseG is not a hydrolyzing epimerase. The Pse5,7Ac$_2$ synthase PseI, performs the PEP-dependent condensation with DATDH to generate Pse5,7Ac$_2$ (**12**). CMP-activation is performed by the ATP-dependent synthetase, PseF, generating CMP-Ps5,7Ac$_2$ (**13**) which enables the use of this complex carbohydrate as a glycosyl donor. With the elucidated biosynthetic pathway, Schoenhofen *et al* were able to enzymatically synthesize CMP-Pse5,7Ac$_2$ starting from UDP-GlcNAc in a 1-pot manner.

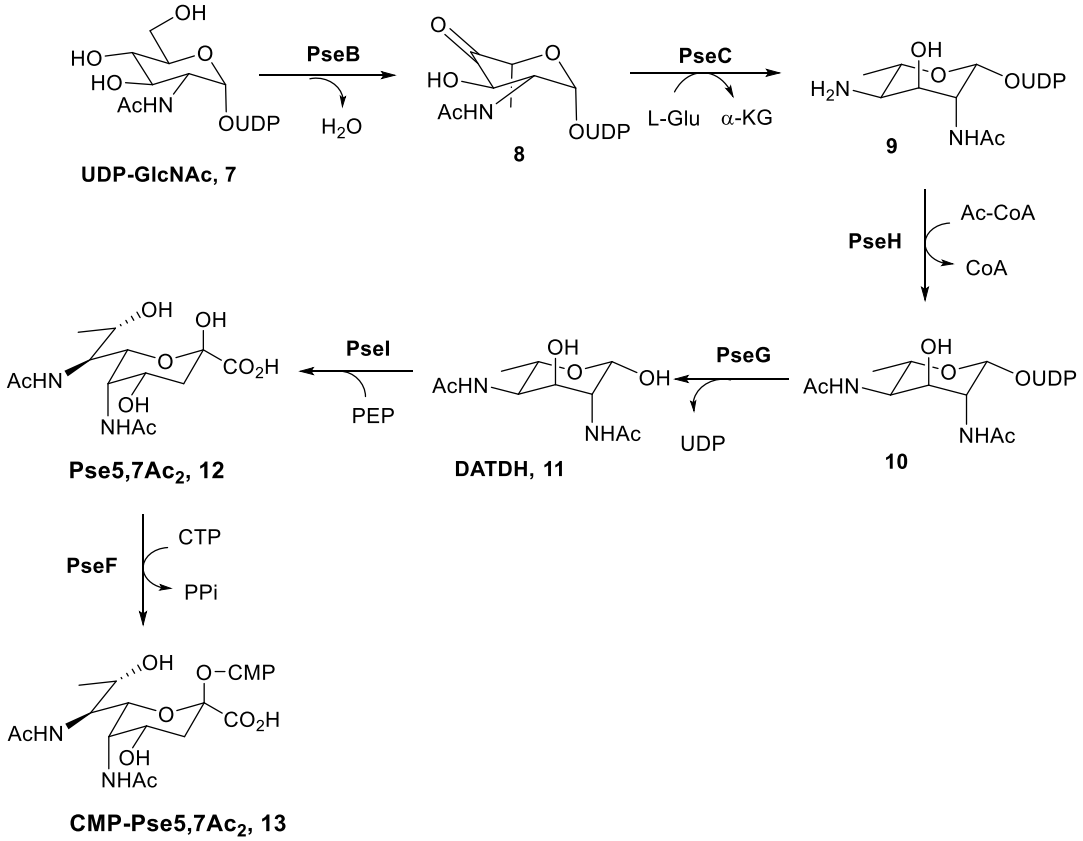


Figure 1.21: Biosynthesis of CMP-Pse5,7Ac$_2$ in *H. pylori* from UDP-GlcNAc.

1.3.4 Unusual bacterial nonulosonic acids: discovery, biosynthesis and synthesis.

In prokaryotes, biologically active natural products use common sugar precursors and their analogues as substituent groups to generate structural diversity. Thibodeaux *et al* summarize the utilization of these sugars in unusual, diverse biosynthetic pathways to generate natural products, with a primary focus on glycosyltransferases to access these sugars.[101] In the context of bacterial NulO's, Leg5,7Ac$_2$ and Pse5,7Ac$_2$ are the most prominent prokaryotic variants of these nine carbon sugars, thus resulting in the majority of the synthetic strategies geared towards accessing them. Nevertheless, there are several other rare bacterial NulO's that have been identified, including Acinetaminic acid (**21**), a 7,8-epi analogue of Leg5,7Ac$_2$ that was found in the capsular polysaccharides of *Acinetobacter baumannii* global clone 1 (GC1) by Hall *et al*.[102] A module containing 10 genes were found, 6 of which were identified as genes for the biosynthesis of Leg5,7Ac$_2$. The other 4 novel genes were predicted to be involved in the conversion of Leg5,7Ac$_2$ into CMP-Acinetaminic acid (Figure 1.22, **18**) and were named *aciABCD.*

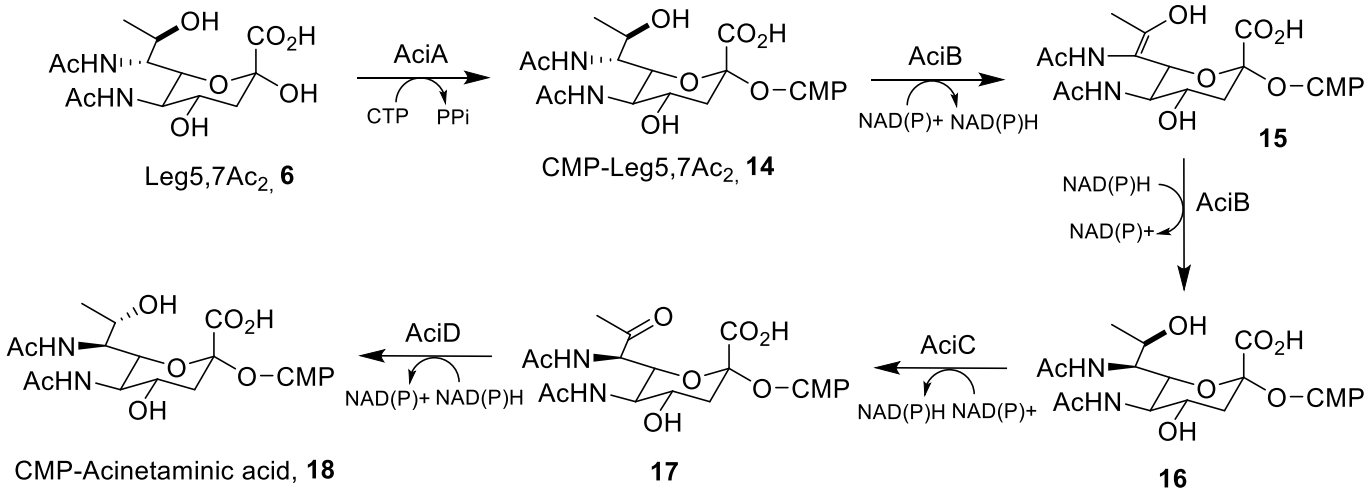
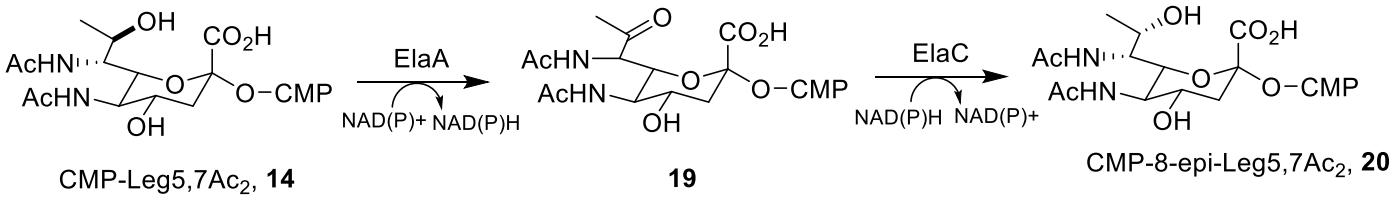


Figure 1.22: Proposed biosynthetic pathway of CMP-Acinetaminic acid from *A. baumannii.*

Hall *et al* also found an additional pair of genes and after sequence analysis they proposed the biosynthesis of another bacterial NulO analogue, CMP-8-epi-Leg5,7Ac$_2$ (Figure 1.23, **14**). From CMP-Leg5,7Ac$_2$, a NADPH dependent dehydrogenase ElaA generates an 8-keto derivative of Leg5,7Ac$_2$ (**19)**. This is followed by a NADPH dependant reductase, ElaC, producing CMP-8-epi-Leg5,7Ac$_2$ (**20**). In 2014, CMP-8-epi-Leg5,7Ac$_2$ was also identified in *A. baumannii* strain LAC-4.[103] Mild acid hydrolysis of the LPS resulted in partial cleavage of LPS complex carbohydrates, and NMR analysis confirmed the presence of CMP-8-epi-Leg5,7Ac$_2$.



CMP-Leg5,7Ac$_2$, **14** **19** CMP-8-epi-Leg5,7Ac$_2$, **20**

Figure 1.23: Proposed biosynthetic pathway of CMP-8-epi-Leg5,7Ac$_2$ from *A. baumannii.*

Other NulO analogues that were found in bacteria include 4-epiLeg5,7Ac$_2$ (**22**), which was isolated from the LPS of *L. pneumophila* serogroup 1 by mild hydrolysis (Figure 1.24).[104] This is the same strain that Leg5,7Ac$_2$ was initially isolated from in 1994. 4-epiLeg5,7Ac$_2$ has also been isolated from several different *L. pneumophila* strains[105] and from the O-specific polysaccharide of *Shewanella japonica*.[106] Additionally, 8-epi-Leg5,7Ac$_2$ (**23**) was identified and characterized from the O-specific polysaccharide of *E. coli* O108,[107,108] *Shewanella putrifaciens*[109], and the O-antigen of *Providencia stuartii.*[110] A 5-acetamidino derivative of 8-epi-Leg5,7Ac$_2$ (**24)** was also identified from the O-specific polysaccharides of *Morganella morganii*.[111] Synthetic strategies have also been utilized to access rare NulO derivatives. Schmid *et al* synthesized 4-epi-Leg5,7Ac$_2$

from D-serine, employing a similar strategy that was used to produce Leg5,7Ac$_2$, as described earlier in this review.[84] The synthesis of 8-epi-Pse5,7Ac$_2$ by Kiefel *et al* has also been highlighted.[96]

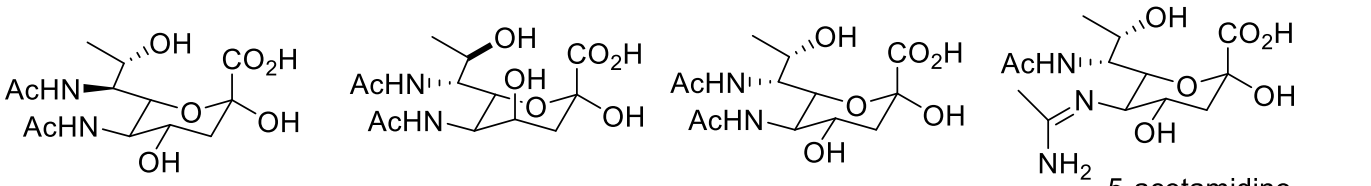


Figure 1.24: Structures of unusual NulO's found to glycosylate a number of different bacteria.

Nonulosonic acids encompass a family of complex monosaccharides found in prokaryotes and eukaryotes. Elucidating their specific roles by synthesizing biologically relevant oligosaccharides is rendered difficult by an inability to access these monosaccharides. Herein I have highlighted current strategies for accessing these important monosaccharides by chemical, chemoenzymatic and biosynthetic methods.

1.4 Scope of thesis

The complexity of NulO's gives rise to technically strenuous, time consuming, expensive and unscalable strategies to access this group of sugars, as shown by the various synthetic routes that were highlighted above. Despite efforts to provide readily available bacterial NulO's, there has not been a synthetic or chemoenzymatic strategy capable of gram-scale production to provide the necessary compounds to further elucidate the role that these complex sugars play in bacterial pathogenicity.

An alternative method of producing bacterial NulO's and their analogues would involve genetically engineering a microbial host for the heterologous expression of a variety of complex

carbohydrate biosynthetic pathways. This is a strategy with strong precedence in the literature for accessing significant quantities of compound, with one of the best examples being the production of an antimalarial drug precursor, artemisinic acid, in *Saccharomyces cerevisiae* by Keasling *et al* in 2006.[112] Malaria is responsible for approximately 1 million deaths per year globally. An effective treatment option against malaria is artemisinin, a sesquiterpene lactone endoperoxide. Synthetic strategies for artemisinin are costly and difficult. Keasling *et al* engineered *S. cerevisiae* to produce an artemisinin precursor, artemisinic acid that can be isolated and converted to artemisinin via chemical synthesis.[113] Central to this engineering strategy is to improve the flux of farnesyl pyrophosphate (FPP), introduce amorphadiene synthase gene ADS from *Artemisia annua*, and clone cytochrome P450 enzymes that convert amorphadiene to artemisinic acid. This strategy successfully produced 100 mg L^{-1} of artemisinic acid in *S. cerevisiae*. Further modifications ultimately resulted in the production of 25 g L^{-1} of artemisinic acid.[114]

Building off this idea of optimizing metabolite production for a certain precursor and coupling this with introducing exogenous genes from prokaryotic or eukaryotic organisms to perform a desired biochemical modification, there have been numerous examples of engineering bacterial or eukaryotic organisms to produce biomolecules. These range from biologically straightforward compounds such as ethanol in *Corynebacterium glutamicum*[115] to more complex molecules such as benzylisoquinalone alkaloids produced in *S. cerevisae*.[116] A significant focus of this thesis is to harness heterologous microbial biosynthetic pathways of Leg5,7Ac_2 and Pse5,7Ac_2 from a variety of gram-negative bacterial pathogens and heterologously express them in engineered *E. coli* strains that are optimized for the production of complex sugars. Multi-gene pathways will be constructed using synthetic biology tools including traditional cloning strategies, Golden gate assembly, and Gibson assembly. Additionally, we will seek to develop a flexible *in*

vivo heterologous expression system based on viral protein packaging, processing and cleavage. This method will enable the rapid movement of natural product production between diverse hosts, with an overarching goal of ultimately enabling synthetic biologists to better produce complex molecules *in vivo.*

1.5 References

(1) Aronoff, S. L.; Berkowitz, K.; Shreiner, B.; Want, L. *Diabetes Spectr.* **2004**, *17* (3), 183–190.

(2) Konopka, J. B. *Scientifica (Cairo).* **2012**, *2012*, 1–15.

(3) Usselman, C. W. N. S. S. J. R. B. *Physiol. Behav.* **2017**, *176* (3), 139–148.

(4) Varki, A.; Gagneux, P. *Ann. N. Y. Acad. Sci.* **2012**, *1253* (1), 16–36.

(5) Crucho, C. I. C.; Correia-Da-Silva, P.; Petrova, K. T.; Barros, M. T. *Carbohydr. Res.* **2015**, *402* (October), 124–132.

(6) Das, R.; Mukhopadhyay, B. *Chem. Rev.* **2016**, *5*, 401–433.

(7) Capon, B. *Chem. Rev.* **1969**, *69* (4), 407–498.

(8) Danishefsky, S. J.; Bilodeau, M. T. *Angew. Chemie (International Ed. English)* **1996**, *35* (13–14), 1381–1419.

(9) Nicolaou, K. C., Hummel, C. W., Pitsinos, E. N., Nakada, M., Smith, A. L., Shibayama, K., & Saimoto, H. *J. Am. Chem. Soc.* **1992**, *114* (25), 10082–10084.

(10) Wang, T.; Demchenko, A. V. *Org. Biomol. Chem.* **2019**, *17* (20), 4934–4950.

(11) Werz, D. B.; Seeberger, P. H. *Chem. - A Eur. J.* **2005**, *11* (11), 3194–3206.

(12) Guo, J.; Ye, X. S. *Molecules* **2010**, *15* (10), 7235–7265.

(13) Kulkarni, S. S.; Wang, C. C.; Sabbavarapu, N. M.; Podilapu, A. R.; Liao, P. H.; Hung, S. C. *Chem. Rev.* **2018**, *118* (17), 8025–8104.

(14) Hahm, H. S.; Schlegel, M. K.; Hurevich, M.; Eller, S.; Schuhmacher, F.; Hofmann, J.; Pagel, K.; Seeberger, P. H. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, 1–5.

(15) Michael, A.; Norton, L. M. *J. Am. Chem. Soc.* **1879**, *1* (11), 484–485.

(16) Fischer, E. *Ueber die Glucoside der Alkohole. Ber. Dtsch. Chem. Ges.* **1893**, *26*, 2400–2412.

(17) Zhu, X.; Schmidt, R. R. *Angew. Chemie - Int. Ed.* **2009**, *48* (11), 1900–1934.

(18) Miljkovz, M.; Gligin, D.; Gligorijevil, M. *J. Org. Chem.* **1975**, *40* (8), 54–57.

(19) Lemieux, R. U.; Driguez, H. *J. Am. Chem. Soc.* **1975**, *97* (14), 4069–4075.

(20) Schmidt, Richard R., Behrendt, Michael; and Toepfer, A. *Synth. Lett.* **1990**, 694.

(21) Stork, G.; La Clair, J. J. *J. Am. Chem. Soc.* **1996**, *118* (1), 247–248.

(22) Lu, J.; Jayaprakash, K. N.; Schlueter, U.; Fraser-reid, B. *J. Am. Chem. Soc.* **2004**, No. 10, 7540–7547.

(23) Bojarová, P.; Křen, V. *Trends Biotechnol.* **2009**, *27* (4), 199–209.

(24) Taylor, P.; Wessel, H. P.; Wessel, H. P. *J. Carbohydr. Chem.* **1988**, *7* (1), 263–269.

(25) Codée, J. D. C.; Litjens, R. E. J. N.; Van Den Bos, L. J.; Overkleeft, H. S.; Van Der Marel, G. A. *Chem. Soc. Rev.* **2005**, *34* (9), 769–782.

(26) Lian, G.; Zhang, X.; Yu, B. *Carbohydr. Res.* **2015**, *403*, 13–22.

(27) Schmidt, R. R.; Zhu, X. *Glycosyl Trichloroacetimidates*; 2008.

(28) Loskot, S. A.; Romney, D. K.; Arnold, F. H.; Stoltz, B. M. *J. Am. Chem. Soc.* **2017**, *139* (30), 10196–10199.

(29) Li, W.; McArthur, J. B.; Chen, X. *Carbohydr. Res.* **2019**, *472* (October 2018), 86–97.

(30) O'Neill, E. C.; Field, R. A. *Carbohydr. Res.* **2015**, *403*, 23–37.

(31) Bowman, K. G.; Bertozzi, C. R. *Chem. Biol.* **1999**, *6* (1), 9–22.

(32) Davies, G.; Henrissat, B. *Structure* **1995**, *3* (9), 853–859.

(33) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. *Annu. Rev. Biochem.* **2008**, *77* (1), 521–555.

(34) Leloir, L. F. *J. Biol. Chem.* **2005**, *280* (19), 158–161.

(35) Cote, J. M.; Taylor, E. A. *Int. J. Mol. Sci.* **2017**, *18* (2256), doi:10.3390/ijms18112256.

(36) Ewing, C. P.; Andreishcheva, E.; Guerry, P. *J. Bacteriol.* **2009**, *191* (22), 7086–7093.

(37) Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-lazar, M.; Haslam, S. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M. *Science (80-.).* **2002**, *298* (November), 1790–1794.

(38) Signe Berg Baldvinsson, Martine C. Holst Sørensen, Christina S. Vegge, Martha R. J. Clokie, L. B. *Appl. Environ. Microbiol.* **2014**, *80* (22), 7096–7106.

(39) N. Yuki, T. Taki, F. Inagaki, T. Kasama, M. Takahashi, K. Saito, S. Handa, T. M. *J. Exp. Med.* **1993**, *178* (November), 2–6.

(40) E. Vimr, S. S. and M. C. *J. Ind. Microbiol.* **1995**, *15*, 352–360.

(41) Houliston, R. S.; Mandrell, R. E.; Jarrell, H. C.; Wakarchuk, W. W.; Gilbert, M. *Glycobiology* **2009**, *19* (2), 153–159.

(42) Wang, Z.; Chinoy, Z. S.; Ambre, S. G.; Peng, W.; Mcbride, R.; Vries, R. P. De; Glushka, J.; Paulson, J. C.; Boons, G. *Science (80-.).* **2013**, *34* (6144), 379–384.

(43) Varki, A. *Trends Mol Med* **2009**, *14* (8), 351–360.

(44) Pearce, O. M. T.; Läubli, H. *Glycobiology* **2015**, *26* (2), 1–65.

(45) Collard, J. G.; Schijven, J. F.; Bikker, A.; Riviere, G. La; Bolscher, J. G. M.; Roos, E.

Cancer Res. **1986**, *46*, 3521–3527.

(46) Teoh, S. T.; Ogrodzinski, M. P.; Ross, C.; Hunter, K. W.; Lunt, S. Y. *Front. Oncol.* **2018**, *8* (May), 1–15.

(47) Glanz, V. Y.; Myasoedova, V. A.; Grechko, A. V; Orekhov, A. N. *Drug, Des. Dev. Ther.* **2018**, *12*, 3431–3437.

(48) Blix, G. *Z. Physiol Chem* **1936**, *240*, 43–54.

(49) Kim, M.; Koketsu, M.; Itoh, T. *Carbohydr. Res.* **1991**, *214* (1), 179–186.

(50) Maru, I., Ohnishi, J., Ohta, Y., & Tsukada, Y. *J. Biosci. Bioeng.* **2002**, *93* (3), 258–265.

(51) Burke, S. D.; Voight, E. A. *Org. Lett.* **2001**, *3* (2), 10–13.

(52) Voight, E. A.; Rein, C.; Burke, S. D. *J. Org. Chem.* **2002**, *67* (24), 8489–8499.

(53) Hong, Z.; Liu, L.; Hsu, C.; Wong, H. *Angew. Chemie (International Ed. English)* **2006**, *45* (44), 7417–7421.

(54) Lorpitthaya, R.; Suryawanshi, S. B.; Wang, S.; Pasunooti, K. K. *Angew. Chemie (International Ed. English)* **2011**, *50* (50), 12054–12057.

(55) Warren, L.; Felsenfeld, H. *J. Biol. Chem.* **1962**, *237* (5), 1421–1432.

(56) Maru, I.; Ohnishi, J.; Ohta, Y.; Tsukada, Y. *Carbohydr. Res.* **1998**, *306*, 575–578.

(57) Blayer, S.; Woodley, J. M.; Dawson, M. J.; Lilly, M. D. *Biotechnol. Bioeng.* **1999**, *66* (2), 131–136.

(58) Xu, P.; Qiu, J. H.; Zhang, Y. N.; Chen, J.; Wang, P. G.; Yan, B.; Song, J.; Xi, R. M.; Deng, Z. X.; Ma, C. Q. *Adv. Synth. Catal.* **2007**, *349* (10), 1614–1618.

(59) Lundgren, B. R.; Boddy, C. N. *Org. Biomol. Chem.* **2007**, *5* (12), 1903.

(60) Tao, F.; Zhang, Y.; Ma, C.; Xu, P. *Sci. Rep.* **2011**, *1*, 1–7.

(61) Chen, F.; Tao, Y.; Jin, C.; Xu, Y.; Lin, B. X. *Appl. Microbiol. Biotechnol.* **2015**, *99* (6), 2603–2611.

(62) Horsman, M. E.; Lundgren, B. R.; Boddy, C. N. *Chem. Eng. Commun.* **2016**, *203* (10), 1326–1335.

(63) Saxon, E.; Bertozzi, C. R. *Science (80-.).* **2000**, *287* (5460), 2007–2010.

(64) Laughlin, S. T., Agard, N. J., Baskin, J. M., Carrico, I. S., Chang, P. V., Ganguli, A. S., … Bertozzi, C. R. *Methods Enzymol.* **2006**, *415* (06), 230–250.

(65) H. Staudinger, J. M. *Helv. Chim. Acta* **1919**, *2* (1), 635–646.

(66) Prescher, J. A.; Dube, D. H.; Bertozzi, C. R. *Nature* **2004**, *430* (7002), 873–877.

(67) Taradon Luangtongkum, Byeonghwa Jeon, Jing Han, Paul Plummer, Catherine M Logue,

Q. Z. *Future Microbiol.* **2009**, *4* (2), 189–200.

(68) Savoldi, A.; Carrara, E.; Graham, D. Y.; Conti, M.; Tacconelli, E. *Gastroenterology* **2018**, *155* (5), 1372-1382.e17.

(69) Sharaby, Y.; Nitzan, O.; Brettar, I.; Höfle, M. G.; Peretz, A.; Halpern, M. *Sci. Rep.* **2019**, *9* (1), 1–10.

(70) Guerry, P.; Ewing, C. P.; Schirm, M.; Lorenzo, M.; Kelly, J.; Pattarini, D.; Majam, G.; Thibault, P.; Logan, S. *Mol. Microbiol.* **2006**, *60* (2), 299–311.

(71) Van Den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G. *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33* (3), 151–208.

(72) KNIREL, Y. A.; RIETSCHEL, E. T.; MARRE, R.; ZÄHRINGER, U. *Eur. J. Biochem.* **1994**, *221* (1), 239–245.

(73) Schoenhofen, I. C.; Vinogradov, E.; Whitfield, D. M.; Brisson, J. R.; Logan, S. M. *Glycobiology* **2009**, *19* (7), 715–725.

(74) McNally, D. J.; Aubry, A. J.; Hui, J. P. M.; Khieu, N. H.; Whitfield, D.; Ewing, C. P.; Guerry, P.; Brisson, J. R.; Logan, S. M.; Soo, E. C. *J. Biol. Chem.* **2007**, *282* (19), 14463–14475.

(75) MacLean, L. L.; Vinogradov, E.; Pagotto, F.; Perry, M. B. *Carbohydr. Res.* **2011**, *346* (16), 2589–2594.

(76) Haseley, S. R.; Wilkinson, S. G. *Eur. J. Biochem.* **1997**, *250* (2), 617–623.

(77) Kenyon, J. J.; Notaro, A.; Hsu, L. Y.; Castro, C. De; Hall, R. M. *Sci. Rep.* **2017**, *7* (11357), 6–11.

(78) Li, X.; Perepelov, A. V.; Wang, Q.; Senchenkova, S. N.; Liu, B.; Shevelev, S. D.; Guo, X.; Shashkov, A. S.; Chen, W.; Wang, L.; Knirel, Y. A. *Carbohydr. Res.* **2010**, *345* (11), 1581–1587.

(79) Post, D. M. B.; Yu, L.; Krasity, B. C.; Choudhury, B.; Mandel, M. J.; Brennan, C. A.; Ruby, E. G.; McFall-Ngai, M. J.; Gibson, B. W.; Apicella, M. A. *J. Biol. Chem.* **2012**, *287* (11), 8515–8530.

(80) Tsvetkov, Y. E.; Shashkov, A. S.; Knirel, Y. A.; Zähringer, U. *Carbohydr. Res.* **2001**, *335* (4), 221–243.

(81) Matthies, S.; Stallforth, P.; Seeberger, P. H. *J. Am. Chem. Soc.* **2015**, *137* (8), 2848–2851.

(82) Popik, O.; Dhakal, B.; Crich, D. *J. Org. Chem.* **2017**, *82* (12), 6142–6152.

(83) Carter, J. R.; Kiefel, M. J. *RSC Adv.* **2018**, *8* (62), 35768–35775.

(84) Gintner, M.; Yoneda, Y.; Schmölzer, C.; Denner, C.; Kählig, H.; Schmid, W. *Carbohydr. Res.* **2019**, *474* (January), 34–42.

(85) Glycosides, S. D.-N. A.; Santra, A.; Xiao, A.; Yu, H.; Li, W.; Li, Y.; Ngo, L.; Mcarthur, J.

B.; Chen, X. *Angew. Chemie (International Ed. English)* **2018**, *2* (57), 2929–2933.

(86) Glaze, P. A.; Watson, D. C.; Young, N. M.; Tanner, M. E. *Biochemistry* **2008**, *47* (10), 3272–3282.

(87) Hassan, M. I.; Lundgren, B. R.; Chaumun, M.; Whitfield, D. M.; Clark, B.; Schoenhofen, I. C.; Boddy, C. N. *Angew. Chemie Int. Ed.* **2016**.

(88) Knirel, Y. A.; Vinogradov, E. V.; L'vov, V. L.; Kocharova, N. A.; Shashkov, A. S.; Dmitriev, B. A.; Kochetkov, N. K. *Carbohydr. Res.* **1984**, *133* (2), 7–10.

(89) Schoenhofen, I. C.; McNally, D. J.; Brisson, J. R.; Logan, S. M. *Glycobiology* **2006**, *16* (9), 8–14.

(90) Stephenson, H. N.; Mills, D. C.; Jones, H.; Milioris, E.; Copland, A.; Dorrell, N.; Wren, B. W.; Crocker, P. R.; Escors, D.; Bajaj-Elliott, M. *J. Infect. Dis.* **2014**, *210* (9), 1487–1498.

(91) Goon, S.; Kelly, J. F.; Logan, S. M.; Ewing, C. P.; Guerry, P. *Mol. Microbiol.* **2003**, *50*, 659–671.

(92) Friedrich, V.; Janesch, B.; Windwarder, M.; Maresch, D.; Braun, M. L.; Megson, Z. A.; Vinogradov, E.; Goneau, M. F.; Sharma, A.; Altmann, F.; Messner, P.; Schoenhofen, I. C.; Schäffer, C. *Glycobiology* **2017**, *27* (4), 342–357.

(93) Posch, G.; Pabst, M.; Brecker, L.; Altmann, F.; Messner, P.; Schäffer, C. *J. Biol. Chem.* **2011**, *286* (44), 38714–38724.

(94) Tomek, M. B.; Janesch, B.; Maresch, D.; Windwarder, M.; Altmann, F.; Messner, P.; Schäffer, C. *Glycobiology* **2017**, *27* (6), 555–567.

(95) Lee, Y. J.; Kubota, A.; Ishiwata, A.; Ito, Y. *Tetrahedron Lett.* **2011**, *52* (3), 418–421.

(96) Zunk, M.; Williams, J.; Carter, J.; Kiefel, M. J. *Org. Biomol. Chem.* **2014**, *12* (18), 2918–2925.

(97) Williams, J. T.; Corcilius, L.; Kiefel, M. J.; Payne, R. J. *J. Org. Chem.* **2016**, *81* (6), 2607–2611.

(98) Dhakal, B.; Crich, D. *J. Am. Chem. Soc.* **2018**, *140* (44), 15008–15015.

(99) Liu, H.; Zhang, Y.; Wei, R.; Andolina, G.; Li, X. *J. Am. Chem. Soc.* **2017**, *139* (38), 13420–13428.

(100) Schoenhofen, I. C.; McNally, D. J.; Vinogradov, E.; Whitfield, D.; Young, N. M.; Dick, S.; Wakarchuk, W. W.; Brisson, J. R.; Logan, S. M. *J. Biol. Chem.* **2006**, *281*, 723–732.

(101) Thibodeaux, C. J.; Melançon, C. E.; Liu, H. *Nature* **2007**, *446* (7139), 1008–1016.

(102) Kenyon, J. J.; Marzaioli, A. M.; De Castro, C.; Hall, R. M. *Glycobiology* **2015**, *25* (6), 644–654.

(103) Vinogradov, E.; Maclean, L.; Xu, H. H.; Chen, W. *Carbohydr. Res.* **2014**, *390*, 42–45.

(104) Knirel, Y. A.; Moll, H.; Helbig, J. H. *Carbohydr. Res.* **1997**, *304*, 77–79.

(105) Knirel, Y. A.; Senchenkova, S. N.; Kocharova, N. A. *Biochem.* **2001**, *66* (9), 1035–1041.

(106) Nazarenko, E. L.; Perepelov, A. V; Shevchenko, L. S.; Daeva, E. D.; Ivanova, E. P.; Shashkov, A. S.; Widmalm, G. *Biochem.* **2011**, *76* (7), 791–796.

(107) Y.A. Knirel, S.N. Senchenkova, A.S. Shashkov, S.D. Shevelev, A.V. Perepelov, L. Bin, L. Feng, L. W. *Adv. Sci. Lett.* **2009**, *2384–387*.

(108) Perepelov, A. V; Liu, B.; Senchenkova, S. N.; Shashkov, A. S.; Shevelev, S. D.; Feng, L.; Wang, L.; Knirel, Y. A. *Biochem.* **2010**, *75* (1).

(109) Shashkov, A. S.; Torgov, V. I.; Nazarenko, E. L.; Zubkov, V. A.; Gorshkova, N. M.; Gorshkova, R. P. *Carbohydr. Res.* **2002**, *337*, 1119–1127.

(110) D, L.; Shashkov, A. S.; Kocharova, N. A.; Zatonsky, G. V; Błaszczyk, A.; Knirel, A.; Rozalski, A. *Carbohydr. Res.* **2007**, *342*, 653–658.

(111) Kilcoyne, M.; Shashkov, A. S.; Senchenkova, A.; Knirel, Y. A.; Vinogradov, E. V; Radziejewska-lebrecht, J.; Galimska-stypa, R.; Savage, V. *Carbohydr. Res.* **2002**, *337*, 1697–1702.

(112) Ro, D.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C. Y.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D. *Nature* **2006**, *440* (April), 3–6.

(113) Roth, R. J. *J. Nat. Prod.* **1989**, *52* (5), 1183–1185.

(114) Paddon, C. J.; Westfall, P. J.; Pitera, D. J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M. D.; Tai, A.; Main, A.; Eng, D.; Polichuk, D. R.; Teoh, K. H.; Reed, D. W.; Treynor, T.; Lenihan, J.; Jiang, H.; Fleck, M.; Bajad, S.; Dang, G.; Dengrove, D.; Diola, D.; Dorin, G.; Ellens, K. W.; Fickes, S.; Galazzo, J.; Gaucher, S. P.; Geistlinger, T.; Henry, R.; Hepp, M.; Horning, T.; Iqbal, T.; Kizer, L.; Lieu, B.; Melis, D.; Moss, N.; Regentin, R.; Secrest, S.; Tsuruta, H.; Vazquez, R.; Westblade, L. F.; Xu, L.; Yu, M.; Zhang, Y.; Zhao, L.; Lievense, J.; Covello, P. S.; Keasling, J. D.; Reiling, K. K.; Renninger, N. S.; Newman, J. D. *Nature* **2013**, *496* (7446), 528–532.

(115) Jojima, T.; Noburyu, R.; Sasaki, M.; Tajima, T. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 1165–1172.

(116) Deloache, W. C.; Russ, Z. N.; Narcross, L.; Gonzales, A. M.; Martin, V. J. J.; Dueber, J. E. *Nat. Chem. Biol.* **2015**, *11* (July), 465–472.

Chapter Two: Total biosynthesis of legionaminic acid, a bacterial sialic acid analogue

The following chapter is adapted with permission from: **Total Biosynthesis of Legionaminic Acid, a Bacterial Sialic Acid Analogue.** Mohamed I. Hassan, Benjamin R. Lundgren, Michael Chaumun, Dennis M. Whitfield, Brady Clark, Ian C. Schoenhofen & Christopher N. Boddy. *Angew. Chem. Int. Ed*., 2016, 55, 12018 –12021 doi:10.1002/anie.201606006

Author Contributions

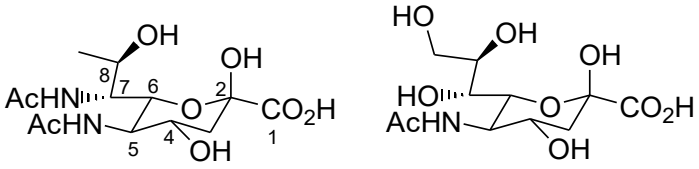
MIH and BRL contributed equally. MIH cloned the biosynthetic pathway of Leg5,7Ac$_2$. BRL engineered the *E. coli* strain BRL04. MIH performed production and purification of Leg5,7Ac$_2$. MIH purified NanA and characterized Leg5,7Ac$_2$ degradation. ICS performed chemoenzymatic activation of purified Leg5,7Ac$_2$. DMW performed the chemical synthesis of the phenylthioglycoside derivative of Leg5,7Ac$_2$. MIH, DMW, ICS and CNB wrote the manuscript with input from all authors.

2.1 Introduction

5-acetamido neuraminic acid, a nine-carbon alpha keto acid sugar (**2**, Neu5Ac, Figure 1), is a member of the nonulosonic acid family of sugars, and plays a major role in prokaryotic and eukaryotic glycobiology. Neu5Ac, along with its approximately 50 derivatives, is commonly referred to as sialic acid. In eukaryotes, Neu5Ac is abundant on cell surface glycoconjugates and is involved in a wide array of processes such as cell-cell interactions and immune response.[1] The amounts and types of sialic acids on eukaryotic cell surfaces are regulated spatially and temporally. These changes are controlled by biosynthetic and catabolic enzymes that are responsive to metabolic controls linked to intracellular *N*-acetylglucosamine levels and to numerous extracellular sialic acid binding proteins such as the SigLec family of proteins.[2] Many human diseases are associated with alterations in sialic acid homeostasis especially diseases associated with immune dysfunction like cancer.[3a] In addition, many bacteria and viruses rely on binding host cell-surface sialic acids as part of their infectious life cycle. Some viruses, such as influenza A, require neuraminidase or sialidase to cleave these host cell-surface sialic acids and release the virus. Selective inhibition of influenza neuraminidase by zanamivir or oseltamivir provides one of the few pharmacological treatment options for combatting an influenza pandemic.[3b] In addition to sialic acid, many bacteria have pathways to produce and use unique nonulosonic acid analogues, notably legionaminic acid (**1**, Leg5,7Ac$_2$[4], Figure 1) and pseudaminic acid.[5]

Legionaminic acid was first identified in 1994 from the lipopolysaccharide of *Legionella pneumophila*, the bacterial agent responsible for Legionnaires' disease.[6] While structurally homologous to sialic acid, the role of legionaminic acid in bacterial physiology

and its impact on host-pathogen interaction is still largely unknown.[7] Studies examining the biology of Leg5,7Ac$_2$ have been critically hampered by the lack of availability of this sugar.



1, Legionaminic Acid (Leg5,7Ac$_2$) **2**, 5-acetamido neuraminic acid (Neu5Ac)

Figure 2.1: Comparison of sialic acid (Neu5Ac) and legionaminic acid (Leg5,7Ac$_2$) structures.

While Leg5,7Ac$_2$ has been produced by total synthesis,[8,9] the synthetic routes are both highly demanding and low yielding. A recent total synthesis of Leg5,7Ac$_2$ enabled the immobilization of the 2-thioethyl glycoside onto maleimide coated glass slides. Probing these slides with human blood sera demonstrated the presence of human antibodies that recognize this bacterial specific sugar presumably due to previous bacterial infections.[9] Despite this advance, the low yield and technical challenge of this route limit its utility for studying the glycobiology of Leg5,7Ac$_2$.

In 2009, the biosynthetic pathway for Leg5,7Ac$_2$ in *Campylobacter jejuni*[10] was elucidated. Unexpectedly this pathway was found to use GDP-GlcNAc (**6**, Figure 2.2) as the key building block, unlike related nonulosonic acid biosynthetic pathways which use UDP-GlcNAc (**7**). The native Leg5,7Ac$_2$ pathway starts with the NAD^+-dependent dehydratase LegB that eliminates water from C4,6 of GDP-GlcNAc, **6**, generating the 4-keto intermediate, GDP-2-acetamido-2,6-dideoxy-α-D-*xylo*-hexos-4-ulose, **8**. LegC, a PLP-dependent aminotransferase, catalyzes the transfer of the amino group from L-glutamate to the 4-keto intermediate producing the amino sugar GDP-4-amino-4,6-dideoxy-α-D-GlcNAc,

10.[11] Acetylation by the acetyltransferase LegH produces GDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucopyranose (**12**, GDP-diNAcBac),[12] which is converted into 2,4-diacetamido-2,4,6-trideoxy-D-mannopyranose, **14**, by LegG, a hydrolyzing 2-epimerase. Finally, the synthase LegI condenses this sugar with pyruvate to produce Leg5,7Ac$_2$, **1**. While this native pathway provides a route to produce Leg5,7Ac$_2$, the rare GDP-linked sugar intermediates suggested that it may be challenging to over-produce significant quantities of the desired compound.

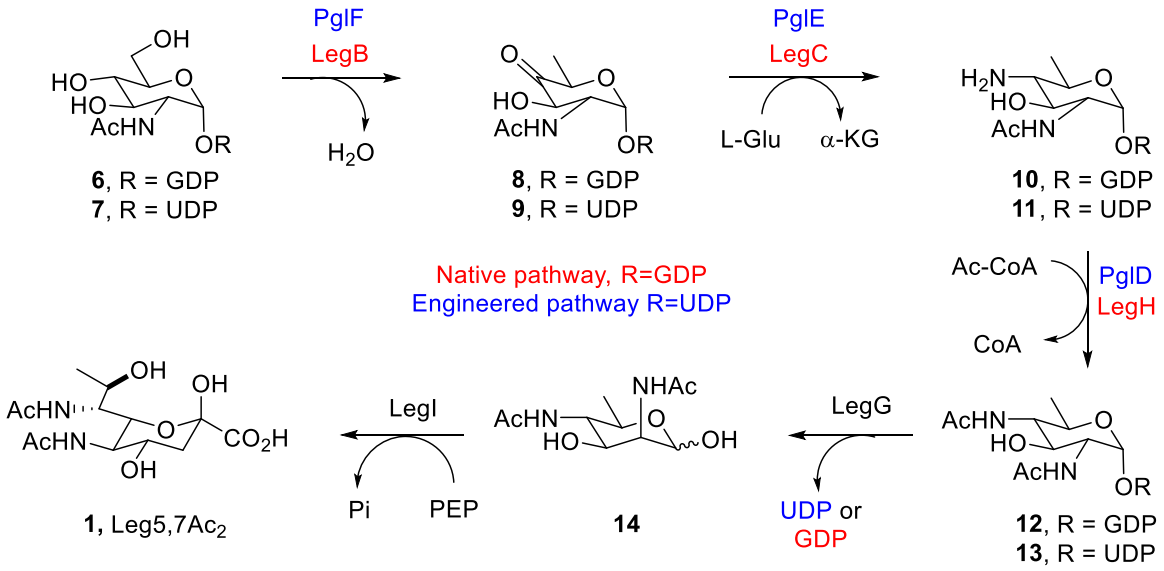


Figure 2.2: Native biosynthetic pathway to produce Leg5,7Ac$_2$ in *C. jejuni* (enzymes in red) and a *de novo* biosynthetic pathway utilizing PglFED from the protein glycosylation pathway from *C. jejuni* (enzymes in blue).

Previously the Boddy lab had over-produced multiple grams per litre of the biosynthetically less complex sugar Neu5Ac from UDP-GlcNAc in *Escherichia coli*.[13] We thus hypothesized it would be possible to produce significant quantities of Leg5,7Ac$_2$ in *E. coli* by creating a *de novo* pathway that relied on UDP-GlcNAc rather than the rare GDP-GlcNAc. This required accessing the key intermediate UDP-2,4-diacetamido-2,4,6-trideoxy-α-D-glucopyranose (UDP-diNAcBac, **13**), which is a known metabolite from the *N*-linked protein glycosylation pathway (Pgl) in *C.*

jejuni.[14] This pathway converts UDP-GlcNAc, **7**, to UDP-diNAcBac, **13**, where the latter is converted to a membrane-bound undecaprenyl diphosphate-diNAcBac intermediate and serves as the core for subsequent glycan assembly and transfer to Asn residues on proteins.[15] The design of our *de novo* biosynthetic pathway thus relied on the dehydratase PglF, the aminotransferase PglE and the acetyltransferase PglD from the *C. jejuni* Pgl pathway to convert UDP-GlcNAc, **7**, into UDP-diNAcBac, **13** (Figure 2.2). At this point **13** must intercept the Leg5,7Ac$_2$ pathway and be converted by LegG and LegI into the final product. While *in vitro* data with purified LegI and LegG from *L. pneumophila* suggested this was possible, the reported efficiencies were very low.[10,16] Herein we show that our *de novo* pathway is highly effective, producing Leg5,7Ac$_2$ with a titer of ~120 mg L^{-1} of *E. coli* culture broth. This is the first *in vivo* production system for accessing this key carbohydrate and demonstrates the power of *de novo* biosynthetic pathway design.

To construct this system, the individual genes *pglFED* and *legGI* were codon optimized for expression in *E. coli* and cloned into pKH22, a low copy expression vector.[13] This installed a C-terminal or N-terminal hexa-His tag into each gene product and enabled all five genes to be combined into a single operon under the control of an inducible T7 promoter with ribosome binding sites directly upstream of each start codon. This polycistronic plasmid, pMIH37 was introduced into the robust protein expression strain *E. coli* BL21 (λDE3). Western blot analysis of the cell lysate 24 h post induction with an anti-hexa-Histidine antibody showed all five proteins were produced (Figure 2.3).

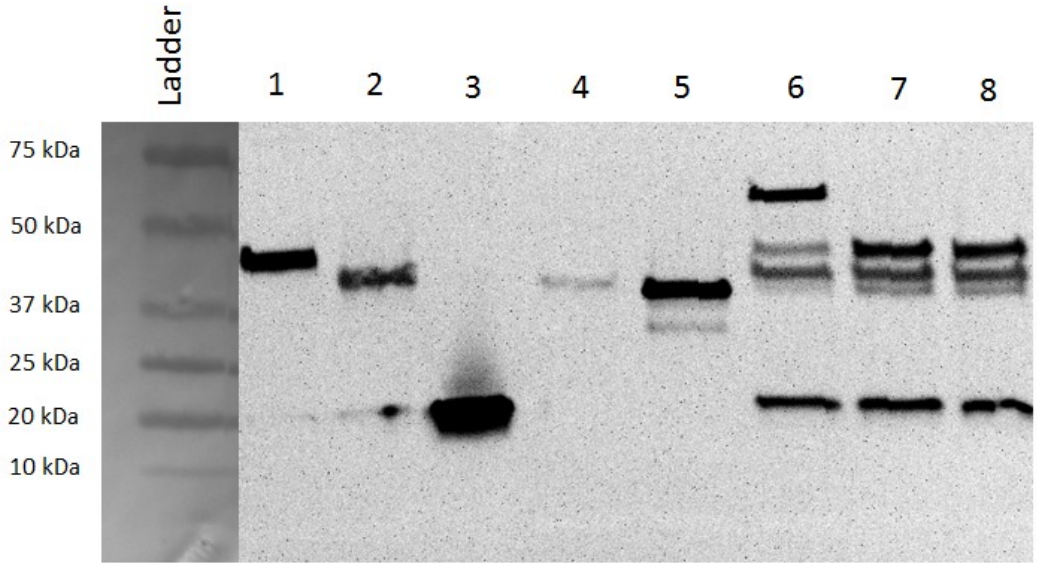


Figure 2.3: Western blot analysis of proteins involved in the de novo biosynthetic pathway for legionaminic acid production from Campylobacter jejuni and Legionella pneumophila using a HRP-conjugated primary hexa-his tag antibody. Lysates from BRL04, BRL02 and BL21 expressing pMIH37 are also shown. Lane 1, His_6PglF; lane 2, His_6PglE; lane 3, His_6PglD (10-fold dilution of lysate); lane 4, His_6LegG; lane 5, His_6LegI; lane 6, His_6pMIH37 in BRL04 (protein at 60 kDa is Agm1 from pBRL178); lane 7, His_6pMIH37 in BRL02; lane 8, His_6pMIH37 in BL21. On the left side, Precision Plus dual color standards (Bio-Rad) labeled next to the image of the protein standard that transferred onto the membrane (not chemiluminescent).

A reliable and efficient detection and quantification method of Leg5,7Ac_2 was necessary to evaluate expression levels of this sugar. Direct analysis by LCMS is not feasible due to poor retention of Leg5,7Ac_2 on reverse phase columns. Additionally, this sugar has a poor chromophore making it difficult to visualize by HPLC. A well-established method for sialic acid detection involves derivatization with a fluorogenic reagent, DMB (4,5-Methylenedioxy-1,2-phenylenediamine dihydrochloride), which has been shown to improve retention on C18 columns and enable UV-vis detection with the addition of the aromatic moiety.[32] Because Leg5,7Ac_2 possesses the α–keto acid functionality that is required to react with DMB, we hypothesized that a similar detection method can be utilized for this sugar (Figure 2.4). Our approach involved

centrifuging and removing the cells from the culture broth followed by direct derivatization of the clarified culture broths. Using a pure Leg5,7Ac$_2$ standard that was generously gifted to us by Ian Schoenhofen at the NRC, we were able to derivatize Leg5,7Ac$_2$ and generate a standard curve that was used to quantify *in vivo* production in *E. coli*.

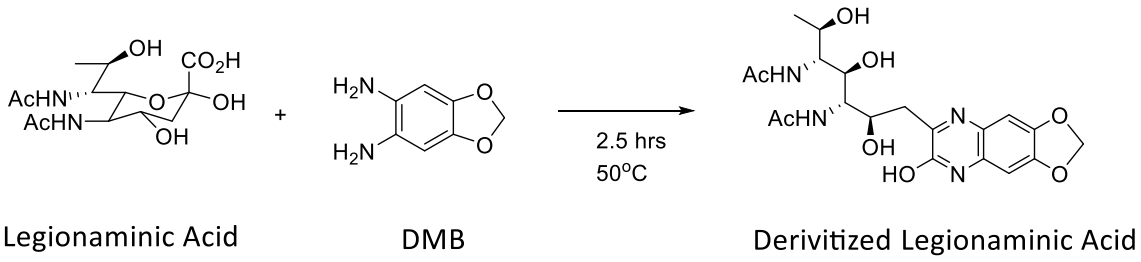


Figure 2.4: DMB derivatization of Leg5,7Ac$_2$ to generate a derivatized product that improves retention on C-18 columns for improved analysis by HPLC.

The polycistronic plasmid pMIH37 was introduced into *E. coli* BL21 (λDE3) and production of Leg5,7Ac$_2$ was probed. Surprisingly, no Leg5,7Ac$_2$ was detected in the culture broth (Figure 2.8). We hypothesized that this result was due to degradation of Leg5,7Ac$_2$ as it was being produced. In the case of Neu5Ac production in *E. coli*, genes encoding the inner membrane sialic acid transporter NanT and the degrading sialic acid aldolase NanA needed to be inactivated to prevent Neu5Ac catabolism.[13] While there is no known Leg5,7Ac$_2$ specific transporter and degradation pathway in *E. coli,* we hypothesized that this Neu5Ac catabolism pathway could also act on Leg5,7Ac$_2$ due to the structural homology between the two molecules.[17] To evaluate this hypothesis, we investigated if NanA was capable of degrading Leg5,7Ac$_2$. Treatment of Leg5,7Ac$_2$ with recombinant purified NanA (Figure 2.5C) showed consumption of Leg5,7Ac$_2$ by HPLC. 5 μM of purified NanA showed complete degradation of 1 mM sialic acid, the native substrate for this enzyme, after 18 h (Figure 2.5A). 25 μM of NanA was required for an appreciable amount of degradation of Leg5,7Ac$_2$ to be detected. It is evident that while Leg5,7Ac$_2$ is a substrate

for this enzyme, its degradation is nowhere near as efficient as it is for its native substrate sialic acid (Figure 2.5B).

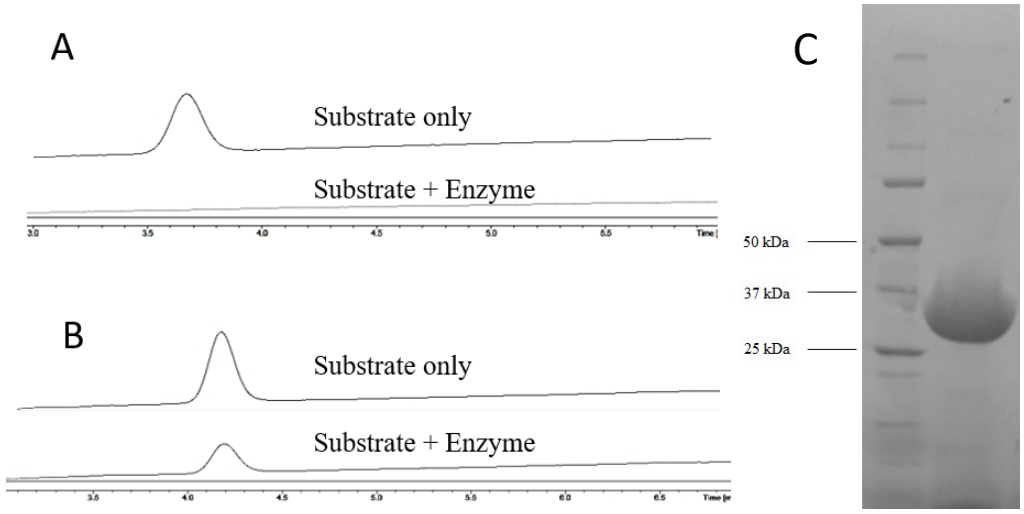


Figure 2.5: (**A**) HPLC analysis (λ= 350 nm) for the degradation of sialic acid with NanA. Complete consumption of sialic acid is detected after 18 hrs. 5 µM NanA, 1 mM Sialic acid, 50 mM Tris was used for this assay. Sialic acid elutes at 3.7 min when derivatized with DMB. (**B**) HPLC analysis (λ= 350 nm) for the degradation of Leg5,7Ac$_2$ with NanA. Approx. 50% Leg5,7Ac$_2$ degradation is obtained after 18 h. 1 mM Leg5,7Ac$_2$, 25 µM NanA and 50 mM Tris was used for this assay. (**C**) Sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis of NanA from *Escherichia coli* after nickel-nitrilotriacetic acid purification. Precision Plus unstained standards (Bio-Rad) was used as the molecular mass standards shown on the left. The expected size of NanA is 32.5 kDa.

Quantification of the pyruvate produced from the aldolase activity enabled determination of the kinetic parameters of NanA-mediate degradation of Leg5,7Ac$_2$ (k_{cat} = 0.52 ± 0.1 min^{-1}, K_M = 1.80 ± 0.5 mM, k_{cat}/K_M = 4.8 ± 2.2 M^{-1}s^{-1}, Appendix Figure S2.1). To test if deletion of the Neu5Ac catabolism genes impacted Leg5,7Ac$_2$ production, *E. coli* BRL02, which is NanT$^-$ and NanA$^-$ was transformed with pMIH37 encoding the synthetic *pglFED-LegGI* operon and cultivated in a shake flask. [13] Consistent with our hypothesis, Leg5,7Ac$_2$ was readily detected at 52 ± 11 mg L^{-1} when this strain was fed 0.3 % GlcNAc and 0.3 % glycerol (Figure 2.8).

To further improve Leg5,7Ac$_2$ production, we focused on maximizing intracellular UDP-GlcNAc pools (Figure 2.6). The key feedstock GlcNAc is transported by the phosphotransferase system NagE and concomitantly phosphorylated to generate intracellular GlcNAc-6-phosphate (GlcNAc-6-P, **4**).[18] GlcNAc-6-P binds the transcriptional repressor, NagC, relieving repression of the *nag* operon and leads to conversion of GlcNAc-6-P into fructose-6-phosphate.[19] The first step in this catabolic pathway is NagA catalyzed deacetylation of the GlcNAc-6-P 2-acetamido group forming GlcN-6-P. Deletion of *nagA* was thus expected to eliminate catabolism of GlcNAc-6-P.[13d] To ensure robust conversion of intracellular GlcNAc-6-P into UDP-GlcNAc, **7**, the *Saccharomyces cerevisiae* enzymes Agm1, a GlcNAc-6-P to GlcNAc-1-P mutase, and Uap1, a GlcNAc-1-P uridyltransferase, were added to our *de novo* pathway.[20] These enzymes were particularly advantageous since they circumvented the highly regulated metabolic node focused on GlcN-6-P that connects central metabolism to cell wall biosynthesis in *E. coli* and ultimately limits UDP-GlcNAc pools.[13d,21]

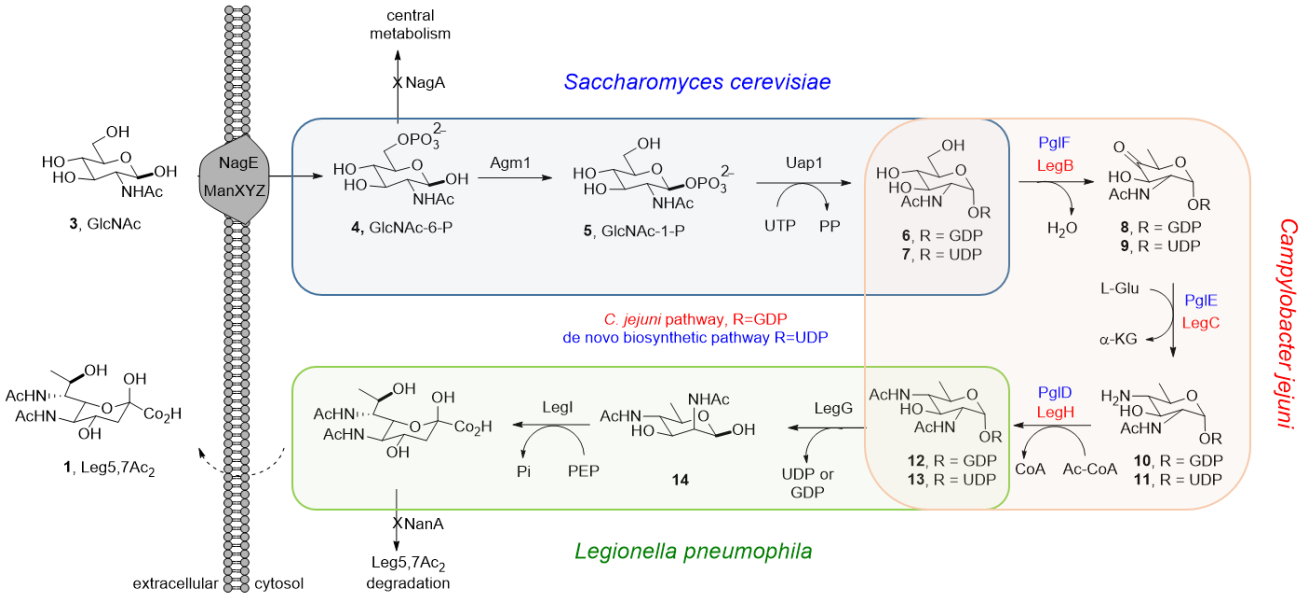


Figure 2.6: *De novo* biosynthetic pathway of Leg5,7Ac$_2$ production in *E. coli*. Enzymes listed in blue are from the engineered UDP-linked pathway and those in red from the native *C. jejuni* GDP-linked biosynthetic pathway.

Both *agm1* and *uap1* were cloned onto an inducible T7-controlled expression vector (pBRL178). Co-transformation of the *nagA* deletion strain (BRL04)[13c] with pBRL178 and pMIH37 generated strains capable of producing high titers of Leg5,7Ac$_2$ (121 ± 19 mg L^{-1}, Figure 2.8). ^{1}HNMR and 1D TOCSY analysis directly from culture broth confirmed the production of Leg5,7Ac$_2$ (Appendix I). Production experiments to isolate Leg5,7Ac$_2$ were performed on a 500 mL scale (Figure 2.7). Strains were grown in minimal media and expression was induced with IPTG. The cultivation was supplemented with antibiotics, 0.3 % glycerol (*w*/*v*) and 0.3% GlcNAc (*w*/*v*) at 0 h, 18 h, 36 h and 54 h. Production continued to 168 h post induction after which the culture was lyophilized, triturated with methanol and processed through a DOWEX 1x8 ion exchange column in the formate form. Upon ion exchange purification, roughly 50-70% purity is obtained, but to carry out further biochemical and synthetic activation of this sugar, a higher level of purity must be achieved. A suitable solvent system for silica gel column chromatography was found, which then afforded nearly 50 mg of pure Leg5,7Ac$_2$ (as judged by ^{1}H NMR analysis,[22] see Appendix S2.3) from 500 mL of production broth.

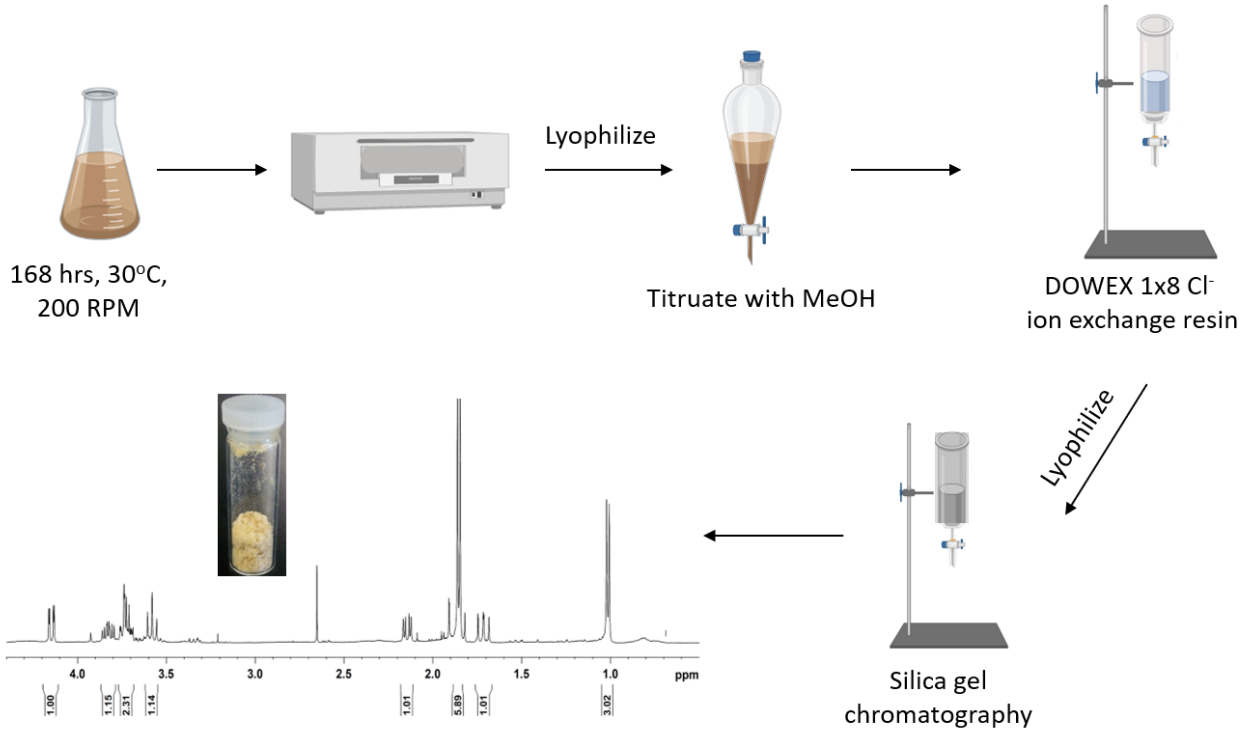


Figure 2.7: Isolation strategy to obtain highly pure Leg5,7Ac$_2$ from engineered *E. coli*.

To increase the yields obtained from shake flask production, we collaborated with Jason Zhang's lab in the Department of Engineering at the University of Ottawa to perform the production of Leg5,7Ac$_2$ in a bioreactor. After screening temperature and feeding rates, optimal bioreactor titers after 60 h production runs yielded 5.6 g L^{-1} of Leg5,7Ac$_2$ from *E. coli*. This greatly simplified the isolation procedure for Leg5,7Ac$_2$ by providing culture broth with an initial higher purity, as highlighted in Figure 2.7.

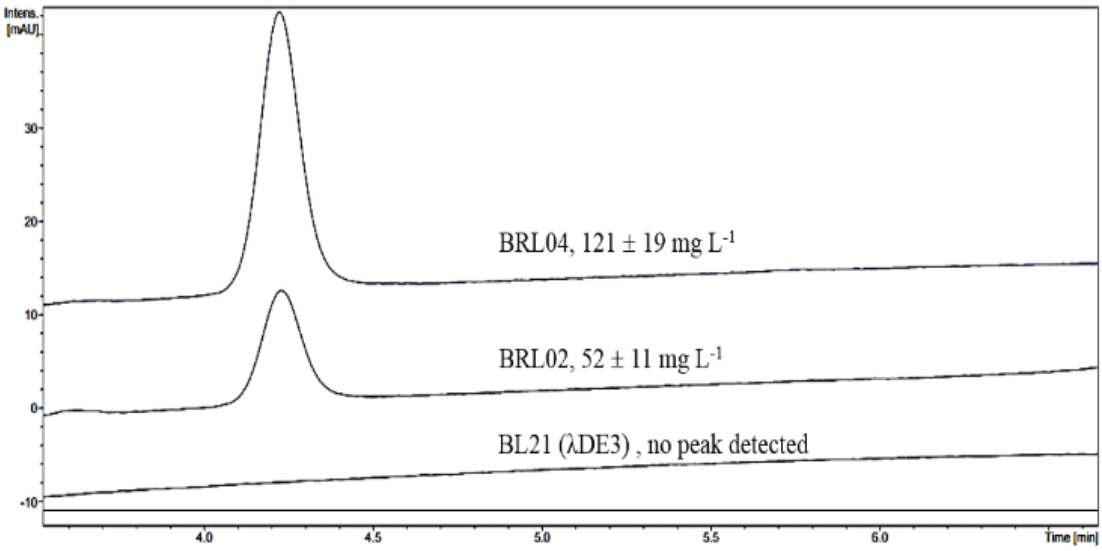


Figure 2.8: Leg5,7Ac$_2$ production in *E. coli* strains. The supernatant from culture broths of BRL04, BRL02 and BL21(λDE3) were reacted with 1,2-diamino-4,5-dimethoxybenzene to label 1,2-diketo compounds and analyzed via HPLC (λ= 350 nm). The peak for derivatized Leg5,7Ac$_2$ was observed eluting at 4.20 minutes.

Many complex cell surface glycoconjugates containing Neu5Ac have been synthesized chemoenzymatically or chemically.[23] These compounds have been essential for understanding sialobiology. The analogous synthesis of Leg5,7Ac$_2$ containing complex glycoconjugates is thus a necessity for further understanding its role in glycobiology.[24] To evaluate whether CMP-activated Leg5,7Ac$_2$, which can be used with glycosyltransferases (GT) to generate complex glycoconjugates, could be produced from our purified Leg5,7Ac$_2$, we treated our product with a recombinant purified CMP-legionaminic acid synthetase, LegF, and CTP (Figure 9).[10] The expected CMP-Leg5,7Ac$_2$ was readily produced, isolated, and characterized by 1D and 2D NMR (see Appendix S2.3 for NMR). Near complete enzymatic conversion is also evident by CE-MS (Appendix Figure S2.2).

As some known sialyltransferases[25] tolerate CMP-Leg5,7Ac$_2$ as a donor, this nucleotide activated sugar can be readily used for glycoconjugate synthesis.[26] In addition it will prove indispensable in efforts to identify the first legionaminic specific GTs, such as the candidate GTs from the lipopolysaccharide (LPS) pathway in *Acinetobacter baumannii* and flagellar glycosylation in *C. jejuni*.[27] Furthermore CMP-Leg5,7Ac$_2$ has proven to be a promising antibiotic for the treatment of *Neisseria gonorrhoeae* infections in a mouse model.[28] Presumably Leg5,7Ac$_2$ replaces some of the native Neu5Ac during *N. gonorrhoeae* LPS biosynthesis, leading to the bacteria becoming sensitive to serum factors and being eliminated from the host.

To obtain activated Leg5,7Ac$_2$ for chemical glycosylation, we generated the thioglycoside, **16**. Leg5,7Ac$_2$, **1**, was converted into its methyl ester **15** under standard conditions and purified by silica gel chromatography (Figure 2.9). Like Neu5Ac, **15** is predominantly its β-anomer in solution.[29] The protected, activated α-phenylthioglycoside **16** was generated via the β-chloride. Thus, our system enables the synthesis of a key intermediate, which will be essential for the development of efficient chemical glycosylation protocols.[30]

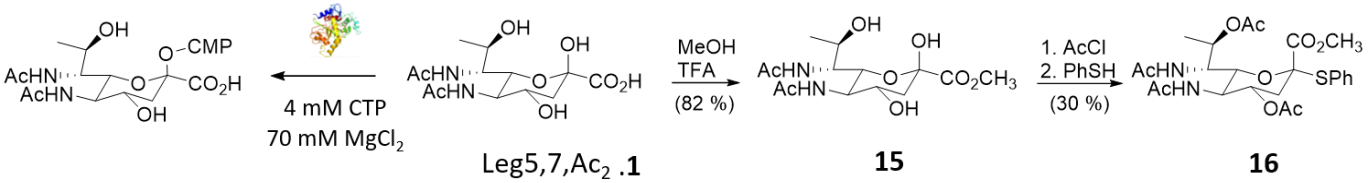


Figure 2.9: Synthesis of Leg5,7Ac$_2$ derivatives. CMP-activated Leg5,7Ac$_2$ was obtained with near 100% efficiency using a CTP synthetase from *N. gonorrhoeae*. A thioglycoside derivative that can be used for chemical synthesis was also generated.

In summary, we have developed the first cell-based production system for Leg5,7Ac$_2$. Our *E. coli* system relies on a *de novo* biosynthetic pathway created from three different metabolic

modules obtained from three different organisms. We show that manipulations of *E. coli* catabolic pathways capable of degrading Leg5,7Ac$_2$ and its biosynthetic precursor GlcNAc-6-P are essential for increasing titers of Leg5,7Ac$_2$. Our system now provides for the first-time access to sufficient purified Leg5,7Ac$_2$ to begin elucidating it biological role. We have shown that Leg5,7Ac$_2$ can be used to enzymatically synthesize CMP-Leg5,7Ac$_2$, which is being tested as a novel antibiotic against *N. gonorrhoeae* and affords a route to numerous biologically relevant glycoconjugates. Leg5,7Ac$_2$ can also be chemically activated as the thioglycoside for further synthetic work. The presence of antibodies to Leg5,7Ac$_2$ in human serum[9] underscores its biological relevance in human health and disease and our study provides the necessary tools to begin understanding the function of this important carbohydrate.

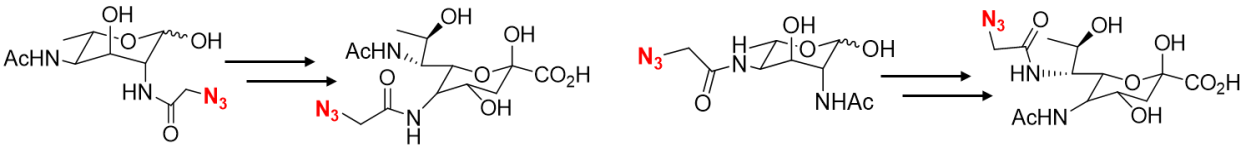


Figure 2.10: Azidoacetyl-Leg5,7Ac$_2$ derivatives to be produced *in vivo* by using azidoacetyl-AltNAc as a starting substrate.

Future studies will focus on using the framework established in this work to generate both biologically relevant and novel Leg5,7Ac$_2$ analogues. The biosynthetic pathway of 8-epi-Leg5,7Ac$_2$ has been elucidated and will be cloned and expressed in the system designed in this chapter to generate sufficient quantities of this sugar, enabling understanding of its biological roles and functions. Additionally, we will attempt to produce 5-azidoacetyl-Leg5,7Ac$_2$ and 7-azidoacetyl-Leg5,7Ac$_2$ (Figure 2.10). The synthesis of the starting substrates to be used for this work (2-azidoacetyl-AltNac or 4-azidoacetyl-AltNAc) have been established.[31] This would

generate Leg5,7Ac_2 analogues containing residues that would enable *in vivo* click chemistry, thus provide a method to elucidate further biological roles of this complex sugar, including identifying novel glycosyltransferases that are involved in glycosylating the outer surface of bacterial pathogens with Leg5,7Ac_2.

2.2 Materials and Methods

A. Molecular Biological Methods.

Construction of Leg5,7Ac$_2$ biosynthetic plasmid pMIH37.

E. coli TOP10 and *E. coli* XL1Blue were used for cloning, following standard recombinant DNA techniques. DNA restriction enzymes were used as recommended by the manufacturer (New England Biolabs, NEB). PCR was performed using Herculase® High-Fidelity DNA Polymerase (New England Biolabs, NEB). PCR products were confirmed by DNA sequencing (Genome Quebec, Montreal QC). Codon optimized synthetic GeneArt™ Strings™ DNA fragments containing N- or C-terminal hexa-histidine tags were purchased from Life Technologies (Sequences available in Appendix I). The gene fragments *PglF*, *PglE* and *PglD* are from *Campylobacter jejuni* while LegG and LegI are from *Legionella pneumophila*. To construct the polycistronic plasmid, the gene fragments *PglF, PglE, PglD, LegG* and *LegI* were amplified by PCR (Primers listed in Appendix) and individually cloned into PCR-Blunt vector using the Zero Blunt PCR Cloning Kit (Life Technologies, Carlsbad CA). All 5 gene fragments were then enzyme digested from PCR-Blunt using NdeI and EcoRI restriction enzymes and individually cloned into an NdeI/EcoRI digested pKH22 (Amp^R), a low copy expression vector under the control of a T7 promoter. To generate the polycistronic arrangement, the plasmid containing *PglE* in a pKH22 backbone (denoted as *pglE*-pKH22) was doubly digested with AvrII and XbaI. This digested *PglE* fragment was cloned into *pglF*-pKH22 plasmid at its AvrII site to generate a *PglF-PglE*-pKH22 plasmid. PglD was cloned into the polycistronic vector in the same manner as *PglE*, thus generating *PglF-PglE-PglD*-pKH22. In parallel, *LegI* (digested with AvrII/XbaI) was cloned into the AvrII site of *LegG*-pKH22 to generate *LegG-LegI*-pKH22. Finally, *LegG-LegI*-pKH22 was digested with AvrII/XbaI and the gene fragments *LegG-LegI* were cloned into the AvrII site of the

PglF-PglE-PglD-pKH22 plasmid to generate *PglF-PglE-PglD-LegG-LegI*-pKH22 (pMIH37). The correct orientation of the inserted vector(s) was confirmed by both EcoRI digestion and sequencing (Genome Quebec, Montreal QC).

Leg5,7Ac$_2$ production in shake flasks.

Starter cultures were grown in LB media, supplemented with the necessary antibiotics, at 37 °C, 200 rpm, for 18 h. F2 minimal media consisting of per litre, 12.24 g of K_2HPO_4, 6.0 g of KH_2PO_4, 4.0 g $(NH_4)_2SO_4$ and 175.5 mg of $MgSO_4$, supplemented with the necessary antibiotics (100 μg mL^{-1} ampicillin and 40 μg mL^{-1} chloramphenicol), 0.25% casitone and 0.5% (*v*/*v*) glycerol was inoculated with 0.5% starter culture. Production cultures were grown at 37°C, 200 rpm, until an OD_{600} of 0.5 was reached. At this point, legionaminic acid production was induced with 0.2 mM isopropyl-1-thio-β-D-galactopyranoside and the incubation temperature was lowered to 30°C. Production cultures were grown for 168 h and supplemented with 0.3% (*w*/*v*) glycerol, 0.3% (*w*/*v*) GlcNAc and antibiotics at 0, 18, and 36 h post induction. At 54 h post induction, an additional feeding of 0.3% GlcNAc, and antibiotics were supplemented to the culture. At 72 and 120 h post induction, only antibiotics were added. The production experiment was stopped at 168 h and the cells were removed from the sample by centrifugation at 4000 RPM for 30 min. The resulting supernatant was collected and lyophilized.

Isolation and purification of Leg5,7Ac$_2$.

500 mL of supernatant from the production experiment was lyophilized, followed by methanol trituration overnight. 2 mL of methanol is added per 20 mL of lyophilized culture broth. The mixture is filtered and the resulting extract is evaporated and loaded onto a DOWEX 1x8 Formate form 100-200 mesh, 100 g resin (Sigma Aldrich). The resin (chloride form) is first incubated in distilled deionized water. The fines were decanted from the resin and the procedure repeated 4×

more with water. The resin was then loaded onto a column and washed with 3 M NaOH (250 mL), water (2 L), 3 M formic acid (250 mL) and water (3 L). The following solutions were made: water (1 L), 5% (0.05 M HCOOH, 250 mL), 10% (0.1 M HCOOH, 250 mL), 25% (0.25 M HCOOH, 500 mL), 50% (0.5 M HCOOH, 250 mL), 100%(1 M HCOOH, 250 mL) by diluting 1.5 L of 1 M HCOOH. This was made from 90% reagent grade formic acid (62.9 mL for 1.5 L). The residue dissolved in water (5 mL) and centrifuged at 3000 rpm for 5 min and the supernatant was loaded onto the column. The column was eluted using the above solutions in this order. Two 25% fractions are collected in 250 mL round bottom flasks, concentrated to 25 mL, frozen at -80^{o}C, then lyophilized. The 10% and 50% (250 mL each) were kept in the refrigerator in case of column failure. All other solutions were discarded. The column was washed with at least 4 L of water after use. Typically a ^{1}H NMR spectrum of the crude product is obtained in D_2O. The crude product (156.2 mg) is dissolved in methanol and applied to a silica column equilibrated in ethyl acetate:methanol:water:acetic acid 6:2.4:1.4:0.2 by volume. The sample was eluted with this solvent mixture, azeotroped with toluene and lyophilized to give pure legionaminic acid. 46.5 mg of pure legionaminic acid was isolated (NMR shown in Section C). ^{1}H NMR D_2O: β-anomer 4.29 dd (1H, $J_{6,7}$ = 1.9, H-6), 3.98 ddd (1H, $J_{4,5}$ = 10.2, H-4), 3.88 ddd (1H, $J_{7,8}$ = 10.6, H-7), 3.86 dq (1H, $J_{8,9}$ = 6.0, H-8), 3.73 brt (1H, $J_{5,6}$ = 10.5, H-5), 2.30 dd (1H, $J_{3eq,4}$ = 4.9, H-3$_{eq}$), 2.01 s (3H, NC(=O)C**H$_3$**), 2.00 s (3H, NC(=O)C**H$_3$**), 1.87 dd (1H, $J_{3ax,4}$ = 11.6, $J_{3eq,3ax}$ = 13.1, H-3$_{ax}$), 1.16 d (3H, H$_3$-9); α-anomer 2.74 dd (1H, $J_{3eq,4}$ = 4.6, H-3$_{eq}$), 2.06 s (3H, NC(=O)C**H$_3$**), 1.97 s (3H, NC(=O)C**H$_3$**), 1.68 dd (1H, $J_{3ax,4}$ = 11.5, $J_{3eq,3ax}$ = 12.6, H-3$_{ax}$);^{13}C NMR D_2O: β-anomer 175.4 (C-1), 175.2 (**C**(=O)N), 175.0 (**C**(=O)N), 96.9 (C-2) 70.9 (C-6), 68.5 (C-4), 67.5 (C-8), 54.5 (C-7), 53.9 (C-5), 40.5 (C-3), 23.4 (NC(=O)**C**H$_3$), 23.0 (NC(=O)**C**H$_3$), 20.4 (C-9);

HPLC detection and quantification of Leg5,7Ac$_2$.

DMB (4,5-Methylenedioxy-1,2-phenylenediamine dihydrochloride) derivatization is a common method for quantifying sialic acids and was used in this study for legionaminic acid quantification.[32] Leg5,7Ac_2 is mixed with a DMB solution (8 mM DMB (Sigma-Aldrich), 1.5 M acetic acid, 14 mM sodium hydrosulfite, 0.8 M 2-mercaptoethanol) in a 1:1 volume mixture. The sample is placed in a heat block at 50ºC for 2.5 h and shielded from light due to the sensitivity of DMB. The sample is then placed in a 1260 Infinity High Performance Liquid Chromatography system (Agilent Technologies). Conditions were: Prontosil C18, 5 µm, 125 mm x 4 mm, flow rate: 1.00 mL min^{-1}, λ= 350 nm. Mobile phase: 84% water: 9% acetonitrile: 7% methanol constant for 12 minutes, followed by a linear gradient of acetonitrile from 9% to 100% over 1 minute until completion. A Leg5,7Ac_2 standard curve was generated using a known concentration of an authentic standard (R^2 =0.993). Production experiments in BRL04, BRL02 and BL21 were performed in triplicate and upon completion the resulting supernatant was quantified using the area under the curve of the peak corresponding to derivatized Leg5,7Ac_2.

Western blotting to detect for his-tagged proteins.

Plasmids encoding *PglF, PglE, PglD, LegG and LegI* were each individually transformed into chemically competent BL21 *E.coli*. Protein expression was performed using the fermentation conditions previously described in A.2. The only modification was that proteins were grown at 30ºC for 24 h post induction without additional feedings. Production broths were centrifuged at 3000 × *g* for 20 minutes. Pellets were resuspended with 0.1 mL of Lysis buffer (100 mM sodium phosphate, 300 mM NaCl, 10% (*v*/*v*) glycerol, 1 mg/mL lysozyme, 1 µg/mL pepstatin A, 1 µg/mL leupeptin, pH 8.0) per 1 mL of culture. Cell lysis was done with sonication on ice (3 pulses, 30 seconds each). The cellular suspension was centrifuged for 1 h at 9000 × *g*. The crude lysates were subject to western blot analysis. Samples were run on Precise Tris-HEPES pre-cast gels

(ThermoFisher Scientific, 10 cm x 8.5 cm). Proteins were then transferred onto a PVDF membrane (Immobilon, 0.45 µm, 10 cm x 7 cm). HRP- conjugated Anti-His monoclonal antibody (Genescript) reacting specifically with N-terminal and C-terminal 6-his tags of proteins was used. Blocking, transfer and antibody dilution buffers were prepared as recommended by the antibody kit. Detection was performed with an Immobilon^{TM} Western Chemiluminescent Horse Radish Peroxidase substrate (ThermoFisher Scientific) following the manufacturer's instructions. The resulting membrane was visualized using an imager without a filter.

Kinetic analysis of NanA degradation of Leg5,7Ac$_2$.

The expression vector containing NanA (pBRL008, possesses a pET28b backbone) was transformed into chemically competent *E. coli* BL21 (DE3) for protein expression. 400 mL of LB media was supplemented with 50 µg/mL of kanamycin, and inoculated with 0.5% *v*/*v* of an overnight seed culture consisting of *E. coli* BL21/pBRL008. The culture was grown at 37ºC at 200 RPM until an OD_{600} of 0.5 was reached. Protein expression was induced with 0.1 mM of isopropyl-1-thio-β-D-galactopyranoside. The culture was grown at 30 ºC at 200 RPM for 12 h. The culture was spun down at 4000 RPM for 30 minutes and the collected cell pellets were resuspended with 0.1 mL of Lysis buffer (100 mM sodium phosphate, 300 mM NaCl, 10% (*v*/*v*) glycerol, 1 mg/mL lysozyme, 1 µg/mL pepstatin A, 1 µg/mL leupeptin, pH 8.0) per 1 mL of culture. Cell lysis was performed by sonication on ice (3 pulses, 30 seconds each). The cellular suspension was centrifuged for 1 h at 9000 × *g*. The cleared lysate was incubated in a 15 mL falcon tube with 400 µL of nickel-nitrilotriacetic acid (Ni-NTA) resin (QIAGEN, Valenica, CA) at 4 ºC. The lysate was then loaded onto a column. The resin was first washed with wash buffer (100 mM Tris, 300 mM NaCl, pH 8.0) and then sequentially eluted with wash buffer supplemented with 20 mM, 100 mM

and 250 mM imidazole. The wash buffers containing 100 mM and 250 mM were performed twice. The 250 mM imidazole washes were combined and the purified protein was concentrated by centrifugation and buffer exchanged using dialysis buffer (100 mM Tris, 300 mM NaCl, pH 7.4). The concentrated protein was flash frozen and stored at -80 °C. A Bradford assay (Bio-Rad) was used to quantify protein concentration. Approximately 9.7 mg/mL of NanA was purified. Leg5,7Ac_2 degradation by NanA was analyzed by quantifying the pyruvate produced by the aldolase activity using a lactate dehydrogenase (LDH) assay. Continuous enzymatic assays were carried out in 100 mM Tris buffer (pH 8.0), 10 μM NanA, 0.05 mM - 1 mM legionaminic acid, 0.15 mM NADH and 25 μg/mL LDH. Absorbances (at 340 nm) were measured in an Evolution 300 UV-VIS spectrophotometer (ThermoFisher Scientific). The amount of Leg5,7Ac_2 degradation was determined by comparison to a pyruvate standard curve (R^2 =0.99) that was generated using the LDH assay. Non-linear regression analysis of NanA rate functions was performed in GraphPad Prism 6).

CMP-activation of Leg5,7Ac_2.

CMP-activation of Leg5,7Ac_2 was performed enzymatically using a CMP-legionaminic acid synthetase, LegF, from *C. jejuni*.[10] Reactions contained 50 mM Tris pH 9, 70 mM $MgCl_2$, 4 mM CTP, 4 mM Leg5,7Ac_2, 1 unit pyrophosphatase per 12 μmoles of CTP, and sufficient quantities of LegF enzyme to obtain optimal conversion at 4–6 h. CMP-Leg5,7Ac_2 enzymatic reaction mixtures were then passed through an Amicon Ultra-15 (10,000 Da molecular weight cut-off) filter membrane before purification. Next, filtered CMP- Leg5,7Ac_2 samples were diluted (~1/20) in cold 1 mM NaCl and applied to 100 ml Q Sepharose Fast Flow (GE Healthcare Life Sciences) column at 4 °C. After washing in 1 CV of 1 mM NaCl, CMP-Leg5,7Ac_2 was eluted in a 1.5 CV

100 mM NaCl step. To achieve additional purity and for desalting, this step elution was lyophilized and subjected to a Superdex Peptide 10/300 GL (GE Healthcare) column using 10 mM ammonium bicarbonate. Quantification of CMP- Leg5,7Ac_2 was determined using the molar extinction coefficient of CMP (ε260 = 7,400). Prior to final lyophilization of purified material, NaCl was added to CMP-Leg5,7Ac_2 preparations in a molar ratio of 2:1 (salt : NulO).For structural characterization of CMP-Leg5,7Ac_2**,** purified material was dissolved in >99% D_2O. Structural analysis was performed using a Varian Inova 500 MHz (^{1}H) spectrometer with a Varian Z-gradient 3 mm probe, or a Varian 600 MHz (^{1}H) spectrometer with a Varian 5 mm Z-gradient probe. Spectra were referenced to an internal acetone standard (δ_H 2.225 ppm and δ_C 31.07 ppm). Results are shown in Section C verifying its production.[10]

B. Synthetic Methods

General synthetic methods.

The ^{1}H NMR spectra were obtained on a Bruker-400 (400 MHz) with tetramethylsilane or the residual signal of the solvent as the internal standard. The ^{13}C NMR spectra were recorded at Bruker-400 (100 MHz). ^{1}H and ^{13}C NMR were obtained in $CDCl_3$ solution (referenced to residual $CHCl_3$ at 7.26 ppm ^{1}H and 77.0 ppm central resonance ^{13}C) or were obtained in CD_3OD (referenced to residual CHD_2OD at 3.31 ppm ^{1}H and 49.15 ppm central resonance ^{13}C). Chemical shifts are in ppm and coupling constants in Hz. ^{13}C resonances are reported to 1 decimal place except to indicate the separation of closely separated resonances where 2 decimal points are given. Thin-layer chromatography was performed on precoated plates of silica gel (60-F_{254}, E. Merck, Darmstadt) and visualized with H_2SO_4-H_2O (1:20 *v*/*v*) followed by heating. Unless otherwise stated, flash column chromatography was performed on silica gel 60 (230-400 mesh, Merck). All

solvents and reagents were purified and dried according to standard procedures. For example, methanol was dried over activated 3Å molecular sieve pellets.

Synthesis of Leg5,7Ac$_2$ analogues.

Methyl-5-acetamido-7-acetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulopyranosiduronate, **15**. Partially purified legionaminic acid (70% absolute purity by NMR with an internal standard, 50 mg = 35 mg; 0.11 mmol) was dissolved in dry methanol (4.2 mL) with stirring under an atmosphere of nitrogen at RT. To this mixture was added by syringe trifluoroacetic acetic acid (0.8 mL). The flask was stoppered and the stirring was continued for 16 h. At this time TLC in ethyl acetate:methanol:water:acetic acid 6:3:1:0.1 by volume indicated that the starting material had disappeared. To the mixture was added toluene (about 20 mL) and the mixture evaporated on a rotary evaporator. A further portion of toluene was added and the evaporation repeated. The residue was taken up in a small volume of ethyl acetate:methanol 2:1 and applied to a short silica column equilibrated in ethyl acetate:methanol:water 207:90:3 by volume. The sample was eluted with this solvent mixture and subsequently with ethyl acetate:methanol:water 198:90:12 by volume to yield the ester.[32] (30 mg, 82%) ^{1}H NMR CD_3OD:4.352 dd (1H, $J_{6,7}$ = 2.3, H-6), 3.971 ddd (1H, $J_{7,8}$ = 9.1, H-7), 3.903 ddd (1H, $J_{4,5}$ = 10.1, H-4), 3.816 s (3H, OC**H$_3$**), 3.711 ddd (1H, $J_{5,6}$ = 10.5, H-5), 3.700 dq (1H, $J_{8,9}$ = 6.4, H-8), 2.251 dd (1H, $J_{3eq,4}$ = 4.8, H-3$_{eq}$), 1.958 s (3H, NC(=O)C**H$_3$**), 1.937 s (3H, NC(=O)C**H$_3$**), 1.815 dd (1H, $J_{3ax,4}$ = 11.3, $J_{3eq,3ax}$ = 12.8, H-3$_{ax}$), 1.159 d (3H, H_3-9); ^{13}C NMR CD_3OD:173.8 (**C**(=O)N), 173.6 (**C**(=O)N), 172.7 (C-1), 97.0 (C-2), 71.1 (C-6), 69.0 (C-4), 67.9 (C-8), 54.9 (C-7), 54.5 (C-5), 53.6 (O**C**H$_3$), 41.3 (C-3), 23.2 (NC(=O)**C**H$_3$), 22.8 (NC(=O)**C**H$_3$), 20.7 (C-9); HR-MS 371.1420 ESI$^+$ calcd. for $C_{14}H_{24}N_2O_8Na$ 371.1430.

Phenyl methyl (5,7-diacetamido 4,8-di-*O*-acetyl-3,5,7,9-tetradeoxy-D-*glycero*-α-D-*galacto*-2-thio-nonulopyranosyl)onate, **16** Methyl ester (30 mg; 0.09 mmol) was dissolved with vigorous magnetic stirring in acetyl chloride (2 mL) under an atmosphere of nitrogen at RT. The flask was stoppered and left to stir for 24 h and toluene (about 20 mL) was added and the mixture evaporated. TLC analysis in ethyl acetate:dichloromethane:methanol 8:1:1 by volume indicated a new spot with R_f about 0.18. The toluene co-evaporation was repeated and the residue dried at high vacuum for 16 h. partial ^{1}H NMR and MS indicated formation of methyl (5-acetamido-7-acetamido 4,8-di-*O*-acetyl-3,5,7,9-tetradeoxy-D-*glycero*-β-D-*galacto*-2-nonulopyranosylchloride)onate $C_{18}H_{27}N_2O_9Cl_1$. Samples left in $CDCl_3$ solution anomerized to the α-isomer partial ^{1}H NMR; ^{1}H NMR $CDCl_3$: 5.804 d (1H, $J_{NH,7}$ = 10.1, NH-7), 5.485 d (1H, $J_{NH,5}$ = 9.5, NH-5), 4.894 ddd (1H, $J_{4,5}$ = 10.0, H-4), 4.874 dq (1H, $J_{8,9}$ = 6.2, H-8), 4.534 dd (1H, $J_{6,7}$ = 1.8, H-6), 4.480 ddd (1H, $J_{5,6}$ = 10.0, H-5), 3.861 m (1H, $J_{7,8}$ = 8.4, H-7), 3.870 s (3H, OC**H$_3$**), 2.826 dd (1H, $J_{3eq,4}$ = 4.9, H-3$_{eq}$), 2.086 s (3H, OC(=O)C**H$_3$**), 2.026 s (3H, OC(=O)C**H$_3$**), 2.026 s (3H, NC(=O)C**H$_3$**), 1.946 s (3H, NC(=O)C**H$_3$**), 1.920 dd (1H, $J_{3ax,4}$ = 10.6, $J_{3eq,3ax}$ = 14.0, H-3$_{ax}$), 1.260 d (3H, H$_3$-9). The residue was dissolved in ethyl acetate (3 mL) to which was added 1 M Na_2CO_{3aq} (3 mL) followed by thiophenol (50 μL) and tetrabutylammonium hydrogen sulfate (20 mg). After stirring, TLC in the above solvent mixture indicated a new UV active spot with R_f about 0.22. The mixture was diluted with ethyl acetate followed by water to allow transfer to a separatory funnel. The layers were separated and the organic layer washed with further water (about 20 mL). After the emulsion separated the organic layer was dried with Na_2SO_4, filtered and evaporated. The residue was dissolved in dichloromethane and applied to a short silica column equilibrated in ethyl acetate:dichloromethane:methanol 8.5:1:0.5 by volume. The sample was eluted with this solvent mixture and subsequently with ethyl acetate:dichloromethane:methanol 8.25:1:0.75 and then ethyl

acetate:dichloromethane:methanol 8:1:1 by volume to yield the thioglycoside.[33] (14 mg, 30%) ^{1}H NMR $CDCl_3$:7.50 m (2H, Ph$_o$), 7.45 - 7.35 m (3H, Ph$_m$, Ph$_p$), 5.648 (1H, $J_{NH,7}$ = 10.2, NH-7), 5.304 d (1H, $J_{NH,5}$ = 9.6, NH-5), 4.981 ddd (1H, $J_{4,5}$ = 10.4, H-4), 4.898 dq (1H, $J_{8,9}$ = 6.4, H-8), 4.732 ddd (1H, $J_{5,6}$ = 10.5, H-5), 4.313 ddd (1H, $J_{7,8}$ = 8.9, H-7), 4.079 dd (1H, $J_{6,7}$ = 1.6, H-6), 3.543 s (3H, OC**H$_3$**), 2.865 dd (1H, $J_{3eq,4}$ = 4.7, H-3$_{eq}$), 2.023 s (3H, OC(=O)C**H$_3$**), 2.019 s (3H, OC(=O)C**H$_3$**), 2.016 s (3H, NC(=O)C**H$_3$**), 1.981 dd (1H, $J_{3ax,4}$ = 11.6, $J_{3eq,3ax}$ = 12.9, H-3$_{ax}$), 1.901 s (3H, NC(=O)C**H$_3$**), 1.246 d (3H, H$_3$-9); ^{13}C NMR $CDCl_3$: 170.7 (**C**(=O)O), 170.3 (**C**(=O)O), 170.16 (**C**(=O)N), 170.08 (**C**(=O)N), 167.7 (C-1), 136.2 (Ph$_o$), 130.0 (Ph$_p$), 128.9 (Ph$_m$), 128.7 (Ph$_{ip}$), 86.8 (C-2), 73.9 (C-6), 69.4 (C-8), 69.1 (C-4), 52.7 (O**C**H$_3$), 51.0 (C-7), 50.6 (C-5), 38.7 (C-3), 23.28 (NC(=O)**C**H$_3$), 23.26 (NC(=O)**C**H$_3$), 21.3 (OC(=O)**C**H$_3$), 20.9 (OC(=O)**C**H$_3$), 17.0 (C-9); MS-ESI$^+$ 525.4 calcd. for $C_{24}H_{32}N_2O_9S$

2.3 References

(1) (a) SialoGlyco Chemistry and Biology I Biosynthesis, structural diversity and sialoglycopathologies Eds. Gerardy-Schahn, R., Delannoy, P., von Itzstein, M. *Top. Curr. Chem*. **2015** *366* b) SialoGlyco Chemistry and Biology II Tools and Techniques to Identify and Capture Sialoglycans Eds. Gerardy-Schahn, R., Delannoy, P., von Itzstein, M. *Top. Curr. Chem*. **2015** *367*.

(2) Varki, A., Schauer. R. *Essentials of Glycobiology*. 2nd edition **2009** Chapter 14 Eds. Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P. Bertozzi, C.R., Hart, G.W., Etzler. M. E. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

(3) (a) O'Reilly, M.K., Paulson, J.C. *Trends Pharmacol Sci*. **2009** *30*, 240-248. (b) von Itzstein, M. *Nature Rev. Drug Dis.* **2007** *6*, 967-974.

(4) We have adopted the nomenclature Leg5,7Ac_2, as opposed to the previously published nomenclature Leg or Leg5Ac7Ac to better mimic the nomenclature that is already established for sialic acids.

5 (a) Michael J. Morrison, M.J; Imperiali, I. *Biochemistry* **2014**, 53, 624-638. (b) Lewis, A.L., Desa, N., Hansen, E.E., Knirel, Y.A, Gordon, J.I., Pascal Gagneux, P., Nizet, V., Varki, A. *Proc. Nat. Acad. Sci*. **2009** *106*, 13552-13557. (c) Zunk, M., Kiefel, M.J. *RSC Adv.* **2014**, *4*, 3413-3421.

(6) Knirel, Y. A., Rietschel, E. T., Marre, R., and Zahringer, U. *Eur. J. Biochem*. **1994** *221,* 239-245.

(7) (a) Khan, M.A., Knox, N., Prashar, A., Alexander, D., Abdel-Nour, M., Duncan, C., Tang, P., Amatullah, H., Dos Santos, C.C., Tijet, N., Low, D.E., Pourcel, C., Van Domselaar, G., Terebiznik, Ensminger, A.W., Guyard, C. *PloS One* **2013** *8*(*6*): e67298. (b) Kodali, S., Vinogradov, E., Lin, F., Khoury, N. Hao, L., Pavliak, V., Hal Jones, C. Laverde, D., Huebner, J., Jansen, K.U., Anderson, A.S., Donald, R.G.K. *J. Biol. Chem*. **2015** *290*, 19512-19526.

(8) (a) Tsvetkov, Y.E., Shashkov, A.S., Knirel, Y.A., Zähringer, U. *Carbohydr. Res*. **2001** *335*, 221-243. (b) Knirel, Y.A., Shashkov, A.S., Tsvetkov, Y.E. Jansson, P.-E., Zähringer, U. *Adv. Carbohydr. Chem. Biochem.* **2003** *58*, 371-417.

(9) Matthies, S.; Stallforth, P.; Seeberger, P.H, *J. Am. Chem. Soc.* **2015,** 137, 2848

10 Schoenhofen, I. C., Vinogradov, E., Whitfield, D. M., Brisson, J.R., Logan, S.M, *Glycobiology*. **2009**, 19, 715

(11) Schoenhofen, I.C., McNally, D.J., Vinogradov, E., Whitfield, D., Young, N.M., Dick, S., Wakarchuk, W.W., Brisson, J.R., Logan, S.M. *J. Biol. Chem*. **2006**, *281*, 723

(12) Demendi, M., Creuznet, C. *Biochem Cell Biol*. **2009**, *87*, 469

(13) (a) Lundgren, B.R., Boddy, C.N. *Org. Biomol. Chem.* **2007**, *5*, 1903. (b)Boddy, C. N.; Lundgren B. L. US 8,722,365 (c) Boddy, C. N.; Lundgren B. L. US 9,243,240 (d) Horsman, M. E.; Lundgren B. R.; Boddy, C. N. *Chem. Eng. Commun.* **2016**, in press. DOI:10.1080/00986445.2016.1188293

(14) Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd S., Jagels, K., Karlyshev, A. V., Moule, S., Pallen, M. J., Penn, C. W. Quail, M. A., Rajandream, M. A., Rutherford, K. M., van Vliet, C. W. Whitehead, S., and Barrell, B. G., *Nature*, **2000** 403, 665

(15) Lizak, C., Gerber, S., Numao, S., Aebi, M., Locher, K.P. *Nature* **2011** *474*, 350-355.

(16) Glaze, P.A.; Watson, D. C.; Young, N. M.; Tanner, M. E. *Biochemistry* **2008**, *47*, 3272-3282.

(17) Almagro-Moreno, S., Boyd , E.F. *Gut Microbes*, **2010** *1:1*, 45-50.

(18) (a) Plimbridge, J. *J. Bacteriol.* **2009**, *191*, 5641-5647. (b) Mukhija, S.; Erni, B. *J. Biol. Chem.* **1996**, *271*, 14819-14824.

(19) Plumbridge, J. *Mol. Microbiol.* **1991**, *5*, 2053-2062.

(20) Wendland, J,; Schaub, Y; Walther, A. *Appl. Env. Microbiology* **2009**, 75, 5840

(21) Plumbrideg, J. *J. Mol. Microbiol. Biotechnol.* **2015**, *25*, 154-67.

(22) (a) Shashkov, A.S., Kenyon, J.J., Senchenkova, S.N., Shneider, M.M., Popova, A.V., Arbatsky, N.P., Miroshnikov, K.A., Volozhantsev, N.V., Hall, R.M., Knirel,Y.A. *Glycobiology*, **2016**, *26*, 501-508.(b) See NMR Tables in Ref. 8b for comparison.

(23) (a) Hemali D. Premathilake, H.D., Gobble, C.P., Pornsuriyasak, P., Hardimon, T., Demchenko, A.V., De Meo, C. *Org. Lett*., **2012**, *14*, 1126-1129. (b) Noel, A., Delpech, B., Crich, D. *Org. Lett*., **2012** *14*, 1342-1345. (c) Hsu, C.-H, Chu, K.-C., Lin, Y.-S, Han, J.-L., Peng, Y.-S., Ren, C.-T., Wu, C.-Y, Wong, C.-H. *Chem. Eur. J*. **2010**, *16*, 1754-1760. (d) Boons, G.-J., Demchenko, A.V., *Chem. Rev*. **2000** *100*, 4539-4566. (e) Ress, D.K., Linhardt, R.J. *Current Organic Synthesis,* **2004**, *1,* 31-46.

(24) (a) Chen, X., Varki, A. *ACS Chem Biol*. **2010** *5*, 163-176. (b) Ito, Y., Paulson, J.C. *J. Am. Chem. Soc*., **1993** *115*, 1603-1605.

(25) Li, Y. Chen, X. *Appl Microbiol Biotechnol*. **2012** *94*, 887-905.

(26) (a) Watson, D.C., Wakarchuk, W.W., Gervais, C., Durocher, Y., Robotham, A., Fernandes, S.M., Schnaar, R.L., Young, N.M., Gilbert, M. *Glycoconj. J.* **2015** *32*, 729-734. (b) Watson, D.C., Leclerc, S., Wakarchuk, W.W., Young, N.M. Glycobiology **2011**, 99-108. (c)Watson, D.C., Wakarchuk, W.W., Leclerc, S., Schur, M.J., Schoenhofen, I.C., Young, N.M., Gilbert, M. *Glycobiology* **2015** *25*, 767-773.

(27) (a) Giguère, D. *Carbohydr. Res.* **2015** *418*, 29-43. (b) Shashkov, A.S., Kenyon, J.J., Senchenkova, S.N., Shneider, M.M., Popova, A.V., Arbatsky, N.P., Miroshnikov, K.A., Volozhantsev, N.V., Hall, R.M., Knirel, Y.A. *Glycobiology* **2016** doi:10.1093/glycob/cwv168

(28) Gulati, S., Schoenhofen, I.C., Whitfield, D, Cox, A.D., Zheng, B., Ohnishi, M., Unemo, M., Lewis, L.A., Taylor, R.E., Landig, C., Diaz, S., Reed, G., Varki, A., Rice, P.A., Ram, S., Li, J., St. Michael, F., Vinogradov, E., Stupak, J. *PLOS Pathog*. **2015** *11*, e1005290 doi:10.1371/journal.ppat.1005290.

(29) Klepach, T., Carmichael, I., Serianni, A.S. *J. Am. Chem. Soc*. **2008**, *130*, 11892-11900.

(30) Kancharla, P. K., Crich, D. *J. Am. Chem. Soc*. **2013**, *135*, 18999-19007.

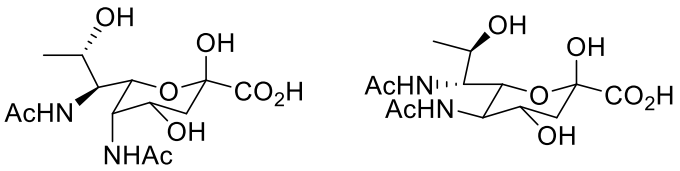
(31) Andolina, G, *et al ACS Chem. Biol.* **2018**, 13, 3030−3037

(32) Martin, M.J., Vazquez, E., Reuda, R. *Anal. Bioanal. Chem*. **2007,** 387, 8, 2943-2949

Chapter Three: Heterologous expression of pseudaminic acid using a genetically engineered strain optimized for complex sugar production

3.1 Introduction

The structural similarities between bacterial nonulosonic acids (NulO's) such as Leg5,7Ac_2 and sialic acid, which is found in both prokaryotes and eukaryotes makes this class of nine-carbon sugars particularly interesting for researchers in the fields of glycobiology and carbohydrate chemistry.[1] A significant obstacle in elucidating the biological role of these unusual bacterial sugars has been a lack of a reliable access to these compounds. The Boddy lab has previously developed a microbial production method to produce sialic acid in *Escherichia coli*,[2] and a similar strategy has been employed for the production of Leg5,7Ac_2 (Chapter 2 of this thesis).[3] Both these NulO's were produced in gram quantities, highlighting the powerful nature of heterologous expression of microbial biosynthetic pathways in an engineered bacterial strain. While Leg5,7Ac_2 and sialic acid differ in their functional groups at C-7 and C-9, they do share relative stereochemistry at all positions (Figure 3.1). This gives rise to the hypothesis that bacteria use sialic acid analogues such as Leg5,7Ac_2 partly to contribute to the evasion of the host's immune response. There are other bacterial NulO's that possess unique stereochemistry at a number of positions, such as pseudaminic acid (Pse5,7Ac_2), which is an isomer of Leg5,7Ac_2 (Figure 2.1). Pse5,7Ac_2 and Leg5,7Ac_2 are stereoisomers at C-5, C-7 and C-8.



Pseudaminic Acid
(Pse5,7Ac$_2$)

Legionaminic Acid
(Leg5,7Ac$_2$)

Figure 3.1: Structure of bacterial NulO's, Pse5,7Ac$_2$ and Leg5,7Ac$_2$

While the presence of these bacterial NulO analogues in glycoconjugates have been verified *in vivo*[4], the impact of these unique sugars on bacterial physiology and pathology is poorly understood. Elucidation of their biological function has been severely hampered by the lack of suitable production strategies to access these sugars. While there has been a number of synthetic methods to produce Pse5,7Ac$_2$ (highlighted in Chapter 1), they are low yielding and involve 12 or more synthetic steps rendering them technically challenging to reproduce.[5,6,7,8,9] Chapter 2 of this thesis detailed a method that is capable of producing Leg5,7Ac$_2$ at gram scale, an appealing approach to accessing Pse5,7Ac$_2$. Thus, in this chapter we will explore utilizing a similar strategy for the heterologous expression of microbial biosynthetic pathways in *E. coli* for the production of Pse5,7Ac$_2$.

In 2006, the chemoenzymatic transformation of UDP-GlcNAc to CMP-Pse5,7Ac$_2$ in six steps was elucidated from a biosynthetic pathway of *H. pylori* by Schoenhofen *et al*.[10] First, a dehydratase/aminotransferase pair, PseB and PseC convert UDP-GlcNAc to UDP-4-amino-4,6-dideoxy-β-L-AltNAc.[11] This intermediate is modified by an acetyltransferase PseH and a hydrolase PseG to generate 2,4-diacetamido-2,4,6- trideoxy-β-L-altropyranose (DATDH), an

unusual 6 carbon intermediate that is unique to the biosynthetic pathway of Pse5,7Ac$_2$. Next, a synthase, PseI condenses PEP with the aforementioned DATDH to generate Pse5,7Ac$_2$. Finally, CMP-Pse5,7Ac$_2$ is synthesized by an ATP-dependent synthetase, PseF, in the presence of CTP to produce CMP-Pse5,7Ac$_2$. Schoenhofen *et al* obtained highly pure CMP-Pse5,7Ac$_2$ via this *in vitro* chemoenzymatic approach. There are limitations to this strategy though, including costly reagents and scalability. Nevertheless, this work provides a biosynthetic pathway that can be harnessed for the *in vivo* production of Pse5,7Ac$_2$, a strategy that has not been successfully employed before for this compound.

To achieve our goal of producing Pse5,7Ac$_2$, we hypothesized that harnessing the first 5 enzymes of the biosynthetic pathways from *H. pylori* and *C. jejuni* without the synthetase PseF, and expressing them in *E. coli* would afford a means of generating this complex sugar (Figure 3.2). The full six-enzyme pathway would generate an activated sugar that would necessitate additional intracellular extraction strategies, which can be circumvented by synthesizing the free sugar by excluding PseF instead.

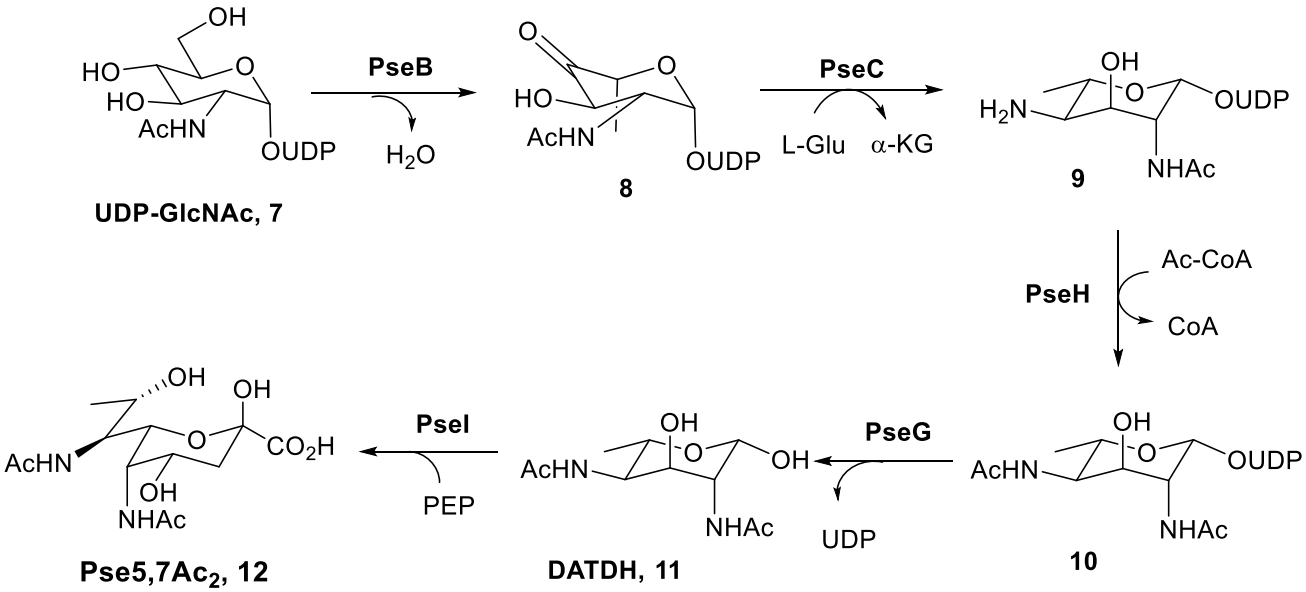
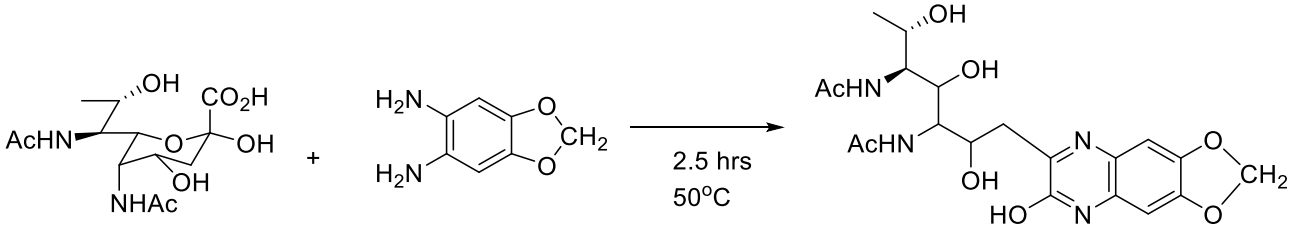


Figure 3.2: Designed biosynthetic pathway to produce Pse5,7Ac$_2$ in *E. coli* to be cloned from *H. pylori* and *C. jejuni*. UDP-GlcNAc is the starting substrate of this pathway.

The plasmid containing the four enzymes of the Pse5,7Ac$_2$ biosynthetic pathway from *H. pylori* (pse*BCHI*) and the hydrolase from *C. jejuni* (*pseG*)[4] was cloned by Dr. Benjamin Lundgren into a low copy, IPTG-inducible Amp^R expression vector under the control of a T7 promoter and designated as pBRL175. To test for the production of Pse5,7Ac$_2$, we initially expressed the biosynthetic pathway in the commercially available *E. coli* heterologous expression strain, BL21 λ(DE3). Minimal media was supplemented with various concentrations of glycerol as a carbon source, initially starting with 1% (w/v). Ultimately glycerol concentration was adjusted to 0.3% after NMR analysis showed a large quantity of glycerol remaining in the spent culture broth at the end of 1% glycerol production runs. GlcNAc was also fed to the cultures as it serves as the starting substrate and is converted to UDP-GlcNAc intracellularly via *E. coli's* primary metabolism,

providing the required precursor for the pathway. Initially, production experiments were run for 72 h, with supplemental feedings of GlcNAc, glycerol and necessary antibiotics. Production was concluded at the 72 h mark and Pse5,7Ac$_2$ titers were analyzed by the DMB derivatization method that was previously developed to accurately detect and quantify Leg5,7Ac$_2$ (Figure 3.3).



Pse5,7Ac$_2$, **12**
334.14 g/mol

DMB
152.06 g/mol

Derivatized Pse5,7Ac$_2$
450.18 g/mol

Figure 3.3: DMB derivatization of Pse5,7Ac$_2$ to generate a derivatized product that improves retention on C-18 columns for improved analysis via LC-ESI-MS/MS.

Validating the DMB derivatization method by LC-ESI-MS/MS and developing a standard curve to quantify production levels was a necessary first step before production experiments could be analyzed. To this end, an authentic Pse5,7Ac$_2$ standard was supplied by our collaborator Ian Schoenhofen at the National Research Council (Ottawa, Canada). DMB derivatization was readily quantifiable (Figure 3.4). The mass spectrum showed peaks corresponding to $[M\text{-}OH]^+$, $[M+H]^+$ and $[M+Na]^+$.

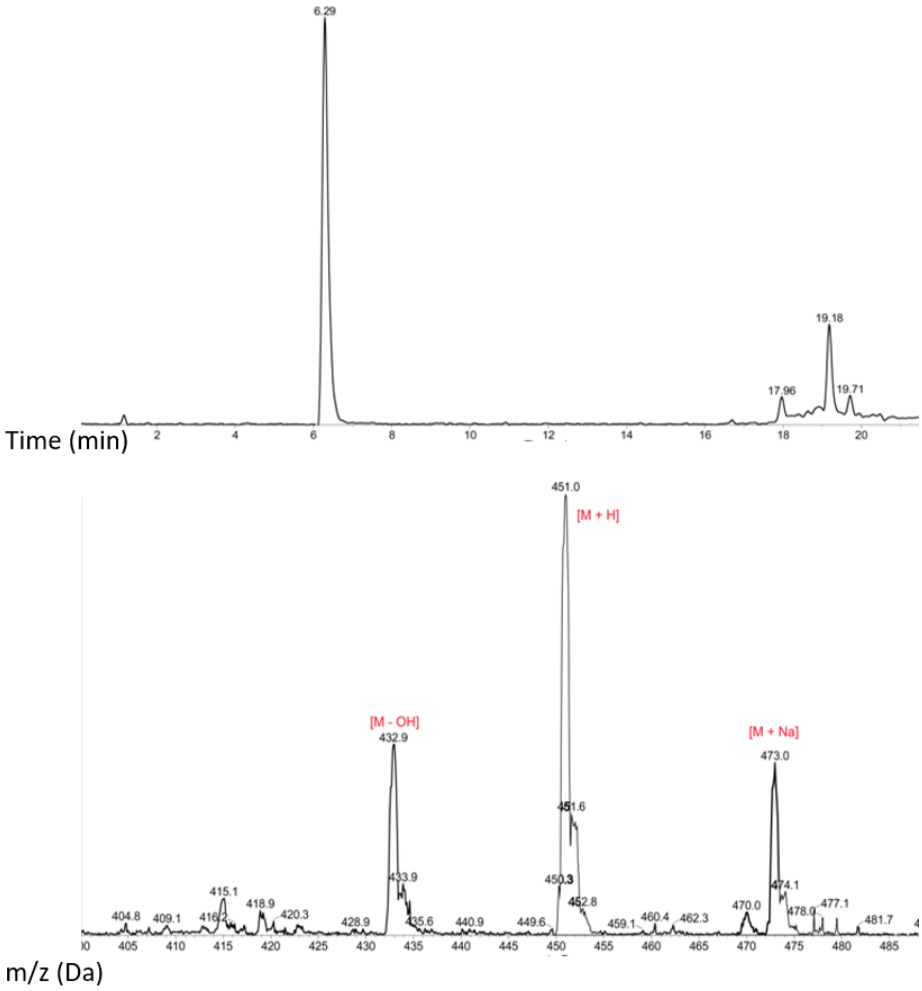


Figure 3.4: LC-ESI-MS/MS analysis of DMB derivatized Pse5,7Ac$_2$. **Top panel:** Extracted ion chromatogram (451 to 452 Da) of a DMB derivatized authentic Pse5,7Ac$_2$ standard of 75 mg L^{-1}. Peak at 6.29 min in the top panel is mass spectrum analyzed in the lower panel. **Bottom panel:** Mass Spectrum of Pse5,7Ac$_2$ standard. Peaks at *m/z* of 432.9 $[M\text{-}OH]^+$, 451.0 $[M+H]^+$ and 473.0 $[M+Na]^+$ were observed.

Upon derivatizing the Pse5,7Ac$_2$ culture broth after expression of pBRL175 in BL21 and analyzing these samples via LC-ESI-MS/MS, no production of this complex carbohydrate was detected (Figure 3.5). A similar phenomenon was observed when the biosynthetic pathway of the

related sugar Leg5,7Ac$_2$ was heterologously expressed in *E. coli* (chapter 2). Previous studies have shown that Pse5,7Ac$_2$ proteins express well in BL21, thus a protein expression problem was unlikely.[10]

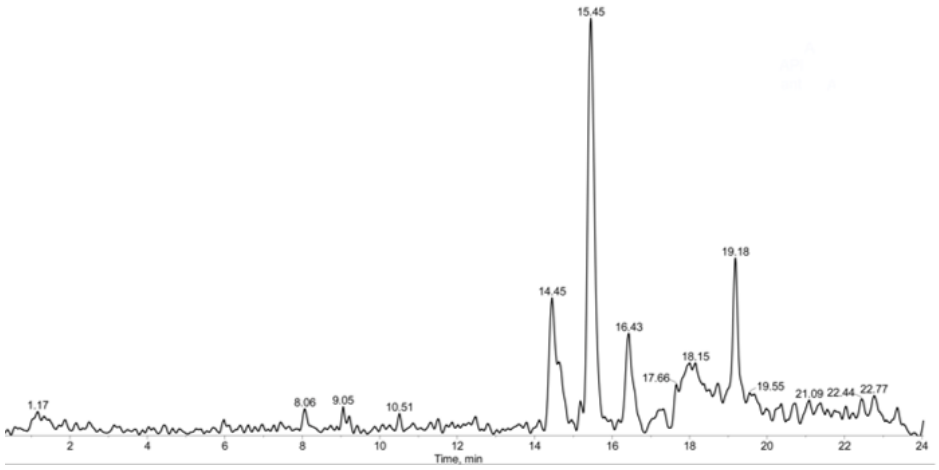


Figure 3.5: LC-ESI-MS/MS extracted ion chromatogram (*m/z* 451-452) trace of Pse5,7Ac$_2$ production broth 72 h post induction. Expression of BL21/pBRL175 failed to produce detectable Pse5,7Ac$_2$, an expected peak around 6.29 mins was not observed. Samples were derivatized with DMB reagent for 2 h prior to analysis.

A lack of carbohydrate production in BL21 can be rationalized by the presence of native carbohydrate reuptake and catabolic genes in BL21. There is no known Pse5,7Ac$_2$-specific degradation pathway in *E. coli* but as previously stated, a sialic acid degradation pathway is present. NanA and NanT work in conjunction to cleave sialic acid into ManNAc and pyruvate, which are then shuttled into the various metabolic pathways of *E. coli* to be used as a carbon source. Due to the structural similarities between sialic acid and Pse5,7Ac$_2$, it's possible that the same degradation pathway is enacting upon both monosaccharides. NanA and NanT can potentially cleave Pse5,7Ac$_2$ into the six-carbon DATDH intermediate and pyruvate, which are

subsequently incorporated into *E coli* metabolism, providing a hypothesis for the lack of Pse5,7Ac$_2$ production in *E. coli* BL21. A similar observation was also made for Leg5,7Ac$_2$ production in BL21.

To test our hypothesis, we introduced the Pse5,7Ac$_2$ biosynthetic plasmid pBRL175 into BRL02 (*nanT* $^-$*, nanA* $^-$) an *E. coli* strain that was engineered to lack these catabolic genes. pBRL175 was subsequently transformed and expressed in BRL02 following the previously outlined 72 h fermentation method. A small Pse5,7Ac$_2$ peak with a retention time of 6.25 minutes was detected via LC-ESI-MS/MS (Figure 3.6). pBRL175/BRL02 produced a quantity of Pse5,7Ac$_2$ that is below the limit of quantification of the mass spectrometer used in this study (< 5 mg L^{-1}). This low titer was somewhat disconcerting considering the related NulO analogue Leg5,7Ac$_2$ produced over 50 mg L^{-1} in this same strain. Further modifications to improve Pse5,7Ac$_2$ yields were required.

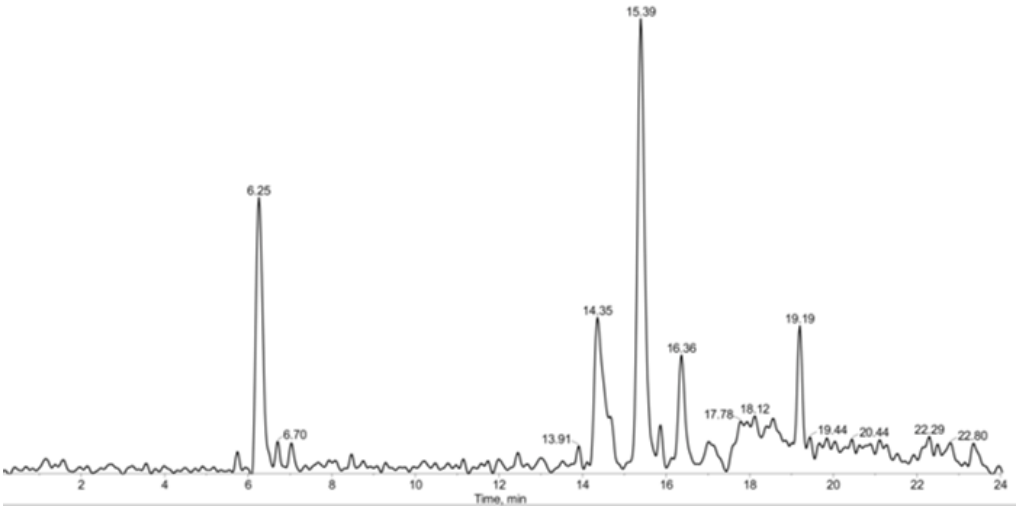


Figure 3.6: Extracted ion chromatogram (*m/z* 451-452) LC-ESI-MS/MS trace of Pse5,7Ac$_2$ production broth derivatized with DMB prior to analysis. Expression of BRL02/pBRL175 produced trace amounts of this complex sugar. Peak observed at 6.25 min corresponds to

Pse5,7Ac$_2$, as determined by MS analysis. *m/z* of 432.9 $[M\text{-}OH]^+$, 451.0 $[M+H]^+$ and 473.0 $[M+Na]^+$ were observed.

To enhance Pse5,7Ac$_2$ productivity in *E. coli* and improve current titers of < 5 mg L^{-1}, we adopted the production strategy that produced optimal titers for Leg5,7Ac$_2$. This necessitated the use of BRL04, an *E. coli* strain that was generated with a *nagA* deletion to BRL02. *nagA* is GlcNAc-6-P deacetylase that is the third gene of the *nag* operon, which is a set of genes required for N-acetylglucosamine utilization in *E. coli*.[12] These genes are responsible for the catabolism of acetamido sugars. NagA acts by deacetylating the C-2 position of GlcNAc-6-P to produce GlcN-6-P.[13] Upon further enzymatic modifications, GlcN-6-P forms Fru-6-P, which then enters *E. coli* central metabolism.[14,15] Thus, nag mediated catabolism can reduce GlcNAc-6-P levels ultimately limiting production of Pse5,7Ac$_2$. A deletion of *nag*A in the chromosome of BRL02 was made by Dr. Ben Lundgren to generate the strain BRL04.[16]

An additional modification included the addition and overexpression of a set of enzymes to convert GlcNAc-6-P into UDP-GlcNAc, the starting substrate of the Pse5,7Ac$_2$ biosynthetic pathway. UDP-GlcNAc is a key diverging point of the Pse5,7Ac$_2$-biosynthetic pathway from *E. coli* metabolism. This intermediate is essential for peptidoglycan and LPS biosynthesis, and its synthesis is tightly regulated by homeostatic mechanisms.[17] The dependency of Pse5,7Ac$_2$ production on the intracellular supply of UDP-GlcNAc can be overcome by the co-expression of a yeast-derived UDP-GlcNAc biosynthetic pathway along with the *pseBCHGI* genes (Figure 3.7). Similar to the production of Leg5,7Ac$_2$ that was described earlier, the two foreign genes from *S. cerevisiae* that code for Agm1 (GlcNAc-6-P mutase)[17] and Uap1 (GlcNAc-1-P

uridyltransferase)[18,19] were cloned into a low copy expression plasmid and designated as pBRL178. Addition of these foreign genes would circumvent the tight regulatory GlcN-6-P metabolic node[20] in *E. coli* by the combined effort of knocking out *nagA* and adding Agm1 and Uap1. This optimized production pathway starts with exogenously fed GlcNAc which is taken up by PTS-transporters such as ManXYZ, then phosphorylated to GlcNAc-6-P. This intermediate is directly converted to UDP-GlcNAc by the *S. cerevisae* enzymes Agm1 and Uap1, thus initiating the biosynthesis of Pse5,7Ac$_2$.

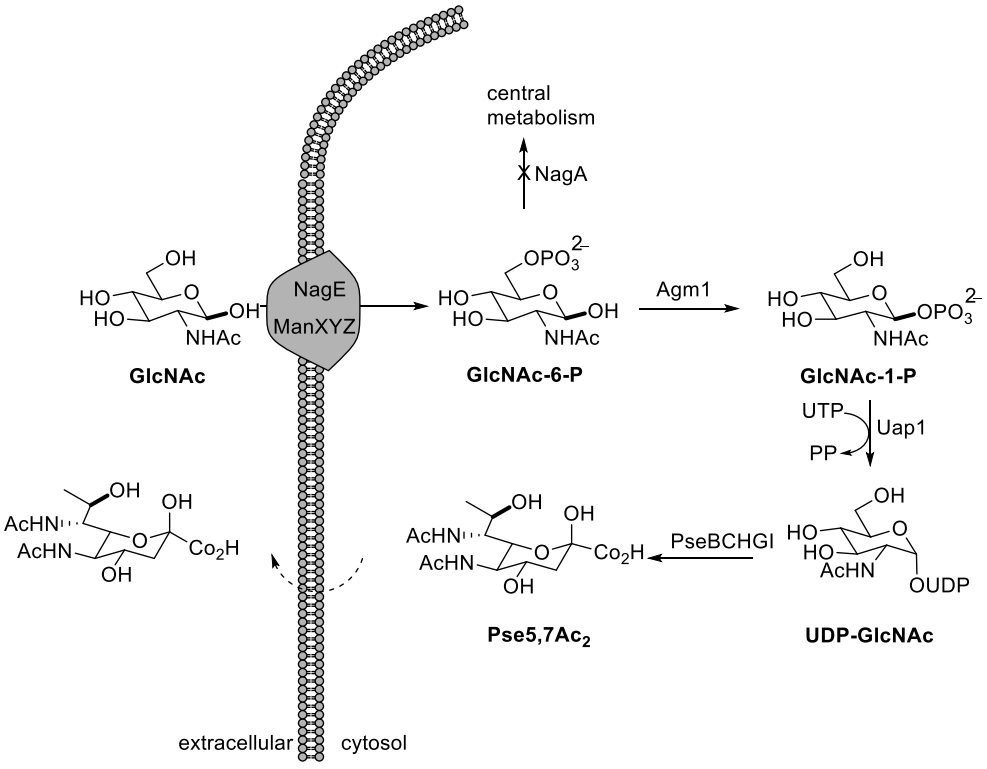


Figure 3.7: Optimized Pse5,7Ac$_2$ production strategy in the engineered *E. coli* strain BRL04. The non-native *S. cerevisae* enzymes Agm1 and Uap1 were used to convert exogenously fed GlcNAc into UDP-GlcNAc. This substrate then enters Pse5,7Ac$_2$ biosynthesis via PseBCHGI.

With a production strategy in place, the *pseBCHGI* pathway (pRBL175, Amp^R) on a single vector was co-transformed with a bi-cistronic operon containing the UDP-GlcNAc overexpression genes *agm1-uap1* (pBRL178, Cm^R) into the BRL04 *E. coli* strain. This optimized production strategy resulted in the production of 15-20 mg L^{-1} of Pse5,7Ac_2 (Figure 3.8), which is a substantial increase from earlier Pse5,7Ac_2 strategies but falls well short of Leg5,7Ac_2, which under the exact same production conditions generated >120 mg L^{-1}.

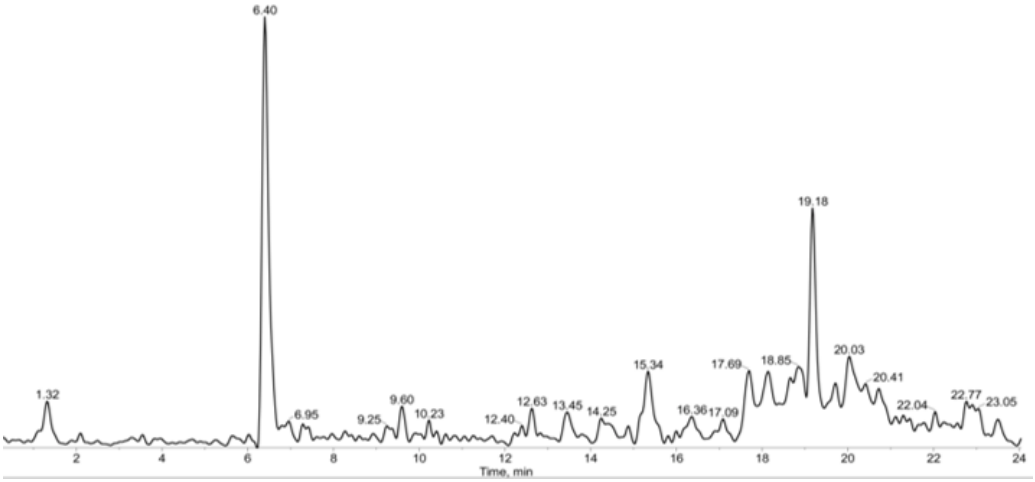


Figure 3.8: LC-ESI-MS/MS extracted ion chromatogram (*m/z* 451-452) trace of Pse5,7Ac_2 production broth after 1:1 derivitization with DMB reagent prior to analysis. Expression of BRL04/pBRL175/pBRL178 produced roughly 20 mg L^{-1} of Pse5,7Ac_2. Peak observed at 6.40 min is derivatized Pse5,7Ac_2.

Evaluating the function of Pse5,7Ac_2 in prokaryotic glycobiology requires sufficient quantities of this sugar for further experimentation. Unfortunately, an *in vivo* titer of 20 mg L^{-1} would necessitate the need to harness enormous quantities of production media for purification of reasonable quantities of the final compound. Additionally, isolation of the desired sugar increases in difficulty when working with low titers due to the decreased initial purity.

Our first step at developing a solution to increase Pse5,7Ac$_2$ titers would be to confirm that suitable expression levels of all five proteins of the biosynthetic pathway of this sugar are achieved. Western blot analysis with an anti-His$_6$ antibody was performed to verify successful protein expression. When designing pBRL175 (*pseBCHGI*), some of the individual genes were first cloned into a pET-based expression vector that does not possess 6x-Histidine tags (formerly pKH22, now designated as pMGX-A). As these constructs lacked a His tag, western blotting with this specific antibody would not be possible. BRL04 and BRL02 are also kanamycin resistant due to the insertion of the Km^R maker into *nanT* therefore the commercially available pET28b expression vector containing both N and C-terminal His tags could not be used. Additionally, stop codons at the C-terminus of the Pse5,7Ac$_2$ genes prevented the use of the Amp^R pET21 expression vector to generate C-terminal His tagged constructs. To circumvent these issues, a hybrid plasmid containing a pET21 backbone with the N-terminal tag of pET28 was designed and cloned. This vector provided an Amp^R vector that contains an N-terminal 6x-Histidine tag suitable for the purposes of this experiment and was designated as pMIH02. *PseH, pseB* and *pseG* were cloned into pMIH02, and all five Pse5,7Ac$_2$ genes were ready for analysis (Table 1)

Table 3.1: List of plasmids used for western blot analysis. pMIH02 is a pET-based Amp^R expression vector with an N-terminal 6x-Histidine tag.

Plasmid	**Inserted Gene(s)**	**Resistance marker**	**Backbone**	**Bacterial source**
pMIH03	*pseH*	Amp^R	pMIH02	*H. pylori*
pNRC37.1	*pseC*	Amp^R	pFO4	*H. pylori*
pMIH04	*pseB*	Amp^R	pMIH02	*H. pylori*
pMIH05	*pseG*	Amp^R	pMIH02	*C. jejuni*
pNRC36.3	*pseI*	Amp^R	pFO4	*H. pylori*

Unexpectedly, two of the five proteins, PseG and PseH, failed to express in BL21. Expression of PseG[21], which is responsible for hydrolyzing the high energy UDP-linked intermediate and produce a free six-carbon precursor, was not detected by western blotting. Sequencing of this gene confirmed a frameshift mutation due to an additional A, which introduced a premature stop codon. The resulting truncated protein was likely misfolded and subjected to cytoplasmic proteolytic degradation in *E. coli*.[22] Additionally, sequencing confirmed that the mutation was present in the *pseBCHGI* expression vector, pBRL175 (Appendix Figure S3.1). Biochemical characterization of PseG has shown the substrate 2,4-diacetamido-2,4,6- trideoxy-β-L-altropyranose can undergo background hydrolysis to generate the expected product, which may account for the production of Pse5,7Ac$_2$ in spite of a mutation rendering this enzyme non-functional.[23]

To remove the stop codon and restore the functionality of PseG, site-directed mutagenesis to delete the additional nucleotide was performed using a splice by overlap extension (SOE) technique (Appendix Figure S3.2). Two internal primers with a 20 bp overlap were created and the mutation of interest was introduced in this overlap region. A series of PCR amplifications successfully removed the mutation and produced the wild type *pseG* sequence. Expression of the wild type gene in BL21 followed by western blotting, showing successful production of PseG (Figure 3.9).

Fully functional PseG was cloned into a vector with all 5 genes in a sequential manner using the plasmid pMGX-A (pMIH11). This plasmid was used in Pse5,7Ac$_2$ production

experiments in the *E. coli* strain BRL04, which had produced the highest Pse5,7Ac$_2$ titers in the previous iterations of the biosynthetic pathway. In the presence of UDP-GlcNAc optimizing enzymes, Agm1 and Uap1, no Pse5,7Ac$_2$ was observed. Inability to detect full length acetyltransferase protein PseH thus may be contributing to a lack of product formation. Solving the expression issues of this enzyme was the next logical step.

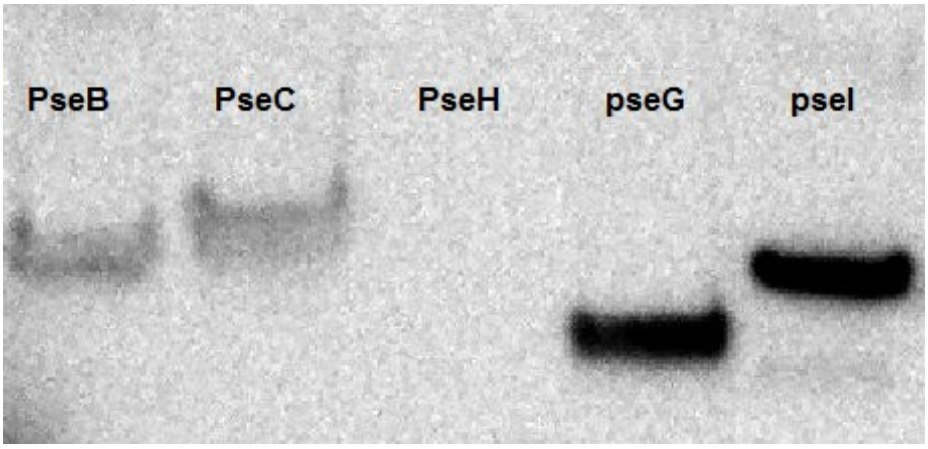


Figure 3.9: Expression of Pse5,7Ac$_2$ proteins detected via western blotting. Proteins were assessed via western blotting using an Anti-6xHis HP-conjugated antibody. Expected size of proteins: pseB: 37 kDa, pseC: 42 kDa, pseH: 21 kDa, pseG: 28 kDa, pseI: 33 kDa

After repeated attempts, no detectable expression of the acetyltransferase PseH from *H. pylori* was observable by western blotting. Additionally, sequencing showed that no mutations were present in the gene. This necessitated the use of alternative acetyltransferases, thus PglD from the protein glycosylation pathway (*pgl*) of *C. jejuni* was used in the next iteration of the Pse5,7Ac$_2$ biosynthetic pathway constructs. PglD can reportedly acetylate the Pse5,7Ac$_2$ precursor (**10**, Figure 3.2) making it suitable to use for our purposes.[24] After confirming the expression of *pglD* in *E. coli* (Appendix Figure S3.3) a *pseBC-pglD-pseGI* construct was sequentially cloned using

the pMGX-A vector and designated as pMIH36. Once again, expression in BRL04 yielded no detectable Pse5,7Ac$_2$ by LC-ESI-MS/MS after 72 h of production. Numerous growth conditions were screened in an attempt to generate the desired product, including increasing the time of production from 72 h to 168 h, testing various media conditions, and increasing/reducing feed rates of the carbon source glycerol and the starting substrate GlcNAc. These efforts were deemed fruitless as Pse5,7Ac$_2$ was not produced with the acetyltransferase from the *pgl* pathway of *C. jejuni*.

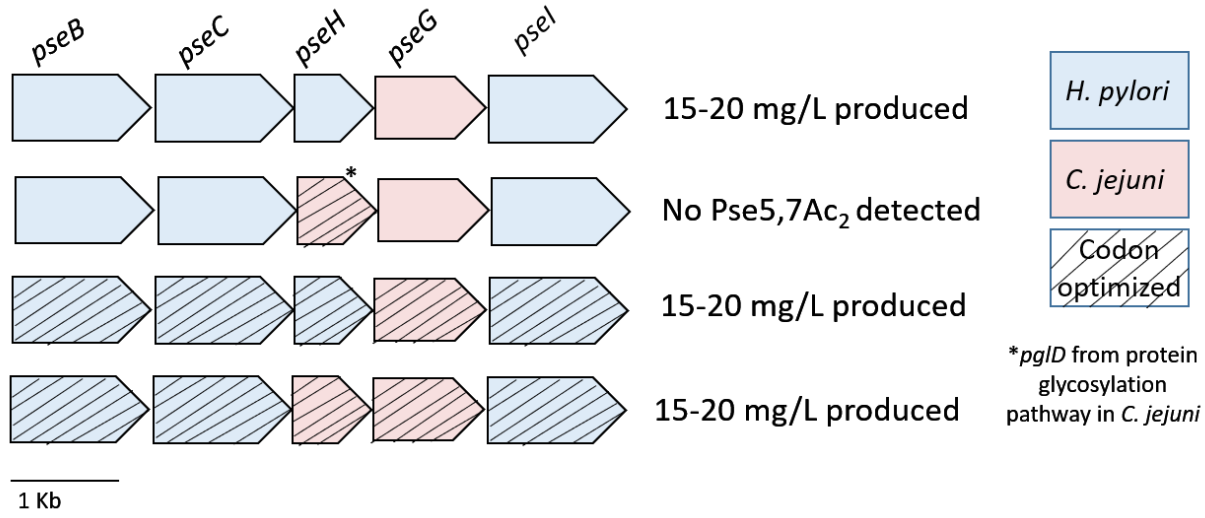


Figure 3.10: Representative diagram of some of the pathways assembled for the production Pse5,7Ac$_2$.

The initial pathways that were constructed were using Pse5,7Ac$_2$ genes that were PCR amplified from the genomic DNA of *H. pylori* and *C. jejuni*. In an effort to improve protein expression levels, all five genes in the Pse5,7Ac$_2$ biosynthetic pathway from *H. pylori* were codon optimized[25] and ordered as synthetic GeneArt strings from Life Technologies (Thermo Fisher Scientific). Initially these five genes were cloned into pMGX-HA, an in-house multi-gene

expression vector with an N-terminal 6x-Histidine tag.[26] The multigene Pse5,7Ac$_2$ pathway was assembled sequentially (pMIH51) and all proteins were expressed successfully. Yet this construct did not produce Pse5,7Ac$_2$ in the engineered *E. coli* strain BRL04. In contrast to the N-terminal tags present in this construct, PseB and pseG were cloned with C-terminal 6x-Histidine tags for *in vitro* studies[10] thus the N-terminal tags present on these two proteins in our construct were removed. Since the ability to detect by western blotting with an anti-His6 antibody was still desired, *pseB* and *pseG* were cloned into pET21b, a commercially available expression vector possessing C-terminal 6x-Histidine tags.

Upon cloning the individual genes into the pET21b vector, full length protein was readily detected by western blotting of the soluble fraction (Appendix Figure S3.4). For the generation of a pentacistronic arrangement of the Pse5,7Ac$_2$ pathway, sequential insertions are time consuming, taking approximately 3-4 weeks to clone all five genes of the desired pathway into a single plasmid. Additionally, replacing one of the first few genes of the Pse5,7Ac$_2$ pathway would necessitate near total re-cloning of the system, adding to the total time spent cloning. If multiple Pse5,7Ac$_2$ pathways were to be cloned and tested, a more efficient pathway cloning strategy would be required.

Gibson assembly is a cloning method that enables the assembly of multiple linear DNA fragments into an intact plasmid.[27] This method employs 20-40 bp overhangs that are homologous to adjacent fragments for each piece of DNA (generated by primer design and PCR amplification). These DNA fragments are then joined together, first by an exonuclease that chews back the 5' ends, a polymerase that fills in any gaps and a ligase that covalently joins adjacent segments with

homologous overlapping regions. This strategy has been used to clone together several hundred kilobases of DNA.[28]

While an appealing strategy, we reasoned that this method would have a high ratio of false positive clones due to each gene in the *de novo* Pse5,7Ac$_2$ biosynthetic pathways possessing either an N-or C-terminal 6x-Histidine tags. This would thus require extensive screening to identify suitable clones. Instead, we used Golden Gate Assembly, a cloning strategy that enables simultaneous assembly of multiple DNA fragments using a Type IIS restriction enzyme (BsaI) and a T4 DNA ligase.[29] Type IIS restriction enzymes produce 4 base pair overhangs on adjacent DNA fragments that can be covalently attached together with a DNA ligase. This method should not be hindered by the multiple 6x-Histidine tags that are present in the Pse5,7Ac$_2$ DNA fragments. Using Golden Gate assembly, the codon optimized Pse5,7Ac$_2$ pathway with correct placement of the N- or C-terminal 6x-Histidine tags was cloned in a pentacistronic arrangement (pMIH63), and protein expression levels of the full pathway in BRL04 was confirmed (Appendix Figure S3.5).

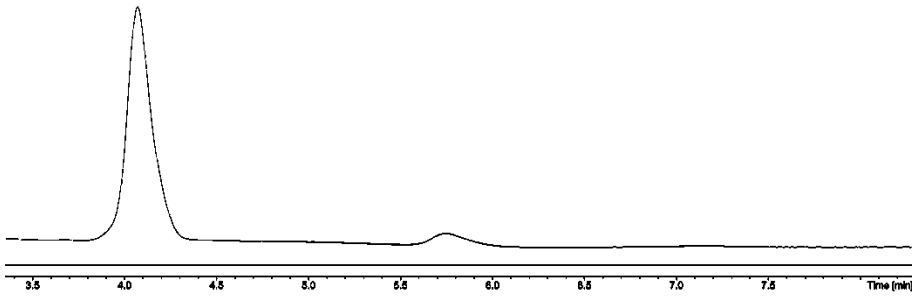


Figure 3.11: HPLC analysis (λ= 350 nm) of Pse5,7Ac$_2$ production broth after 1:1 derivatization with DMB reagent for 2.5 h prior to analysis. Expression of BRL04/pMIH67/pBRL178 produced roughly 20 mg L^{-1} of Pse5,7Ac$_2$. Peak observed at 4.10 min is derivatized Pse5,7Ac$_2$.

With the pathway assembled using Golden Gate assembly, production experiments were performed in the *E. coli* strain BRL04. DMB derivitization followed by quantification using a HPLC were performed and approximately 15-20 mg L^{-1} of Pse5,7Ac_2 was detected (Figure 3.11), which is comparable to the non-codon optimized biosynthetic pathways that were initially tested. The acetyltransferase *pseH* from *H. pylori* was replaced with the acetyltransferase from *C. jejuni* in order to see if *H. pylori* PseH was the source of the low Pse5,7Ac_2 titer. Golden Gate assembly was once again utilized to generate the five gene vector containing the desired biosynthetic pathway and was designated as pMIH67. Expression of all proteins in a polycistronic manner was confirmed by Western blotting in the *E. coli* strain BRL04 (Appendix Figure S3.6). HPLC analysis of the culture broth confirmed that the acetyltransferase from *C. jejuni* also produced 15-20 mg L^{-1} of Pse5,7Ac_2 after shake flask production for 72 h. The inability to surpass a titer of 20 mg L^{-1} despite using a number of different pathways, both native nucleotide sequences and codon optimized variants, was the driving force behind seeking alternative hypotheses for the unexpectedly poor production levels of Pse5,7Ac_2 relative to its analogue, Leg5,7Ac_2.

We initially proposed that a fundamental difference in the biosynthesis of these two carbohydrates may be contributing to the difference in the quantified titers. The first enzyme of Pse5,7Ac_2 biosynthesis, PseB, carries out a 4,6-dehydration reaction followed by a C-5 epimerization to generate an axial methyl substituent at C-5. This is in contrast to Leg5,7Ac_2, where only dehydration takes place and producing an equatorial C-5 methyl substituent. This difference introduces a few possibilities that can account for poor production of Pse5,7Ac_2 relative to Leg5,7Ac_2. First, it is possible that the 4,6-dehydration product is forming more rapidly than

the epimerization reaction, and in doing so the 2nd enzyme of the biosynthetic pathway, the PLP-dependent aminotransferase PseC, utilizes non-epimerized equatorial C-5 products as a substrate thus reducing the total substrate pool for Pse5,7Ac$_2$ biosynthesis. Alternatively, the dehydratase PseB may catalyze dehydration and epimerization at a faster rate than the aminotransferase PseC. This would result in the accumulation of the intermediate for PseC, which could undergo epimerization of the axial C-5 substituent to the more stable equatorial isomer. This would again negatively affect the substrate flux of Pse5,7Ac$_2$ by generating unwanted shunt products and could rationalized the observed low titers. To circumvent this issue, alternative dehydratase/aminotransferase pairs will be tested for enhanced *in vivo* production of Pse5,7Ac$_2$, and initial experiments will be performed with the enzymes found in the biosynthetic pathway of Pse5,7Ac$_2$ in *C. jejuni*.[11] This pair of enzymes have been shown in *in vitro* studies to convert UDP-GlcNAc to UDP-4-amino-4,6-dideoxy-β-L-AltNAc in a stable and consistent manner.[11] Cloning and testing of a Pse5,7Ac$_2$ pathway containing both these genes will be a target for future studies.

When considering pathway design for Pse5,7Ac$_2$, utilizing biosynthetic genes that are clustered together in the genome of producing organisms can reduce the incompatibility between biosynthetic enzymes. The majority of the genes that were cloned and tested in this work were from *H. pylori*. Clustering analysis for the genes associated with Pse5,7Ac$_2$ biosynthesis was performed by the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST).[30] This is a tool that can generate a sequence similarity network for a cluster or closest neighbours of a protein sequence from the UniProt database. Initially, the synthase *pseI* from *C. jejuni* was used to generate a sequence similarity network and to generate a database containing the genes that are 10 Kb

upstream and 10 Kb downstream from each node in the sequence similarity network. As expected, analysis of the sequence similarity network shows that synthases from the same genus cluster together (Figure 3.12A), consistent with low levels of recent horizontal transmission of *pseI*. Similar patterns were observed for the dehydratase *pseB* and the synthetase *pseF* (Appendix Figures S3.7 and S3.8) although the genus of the predominant clustering bacteria was different.

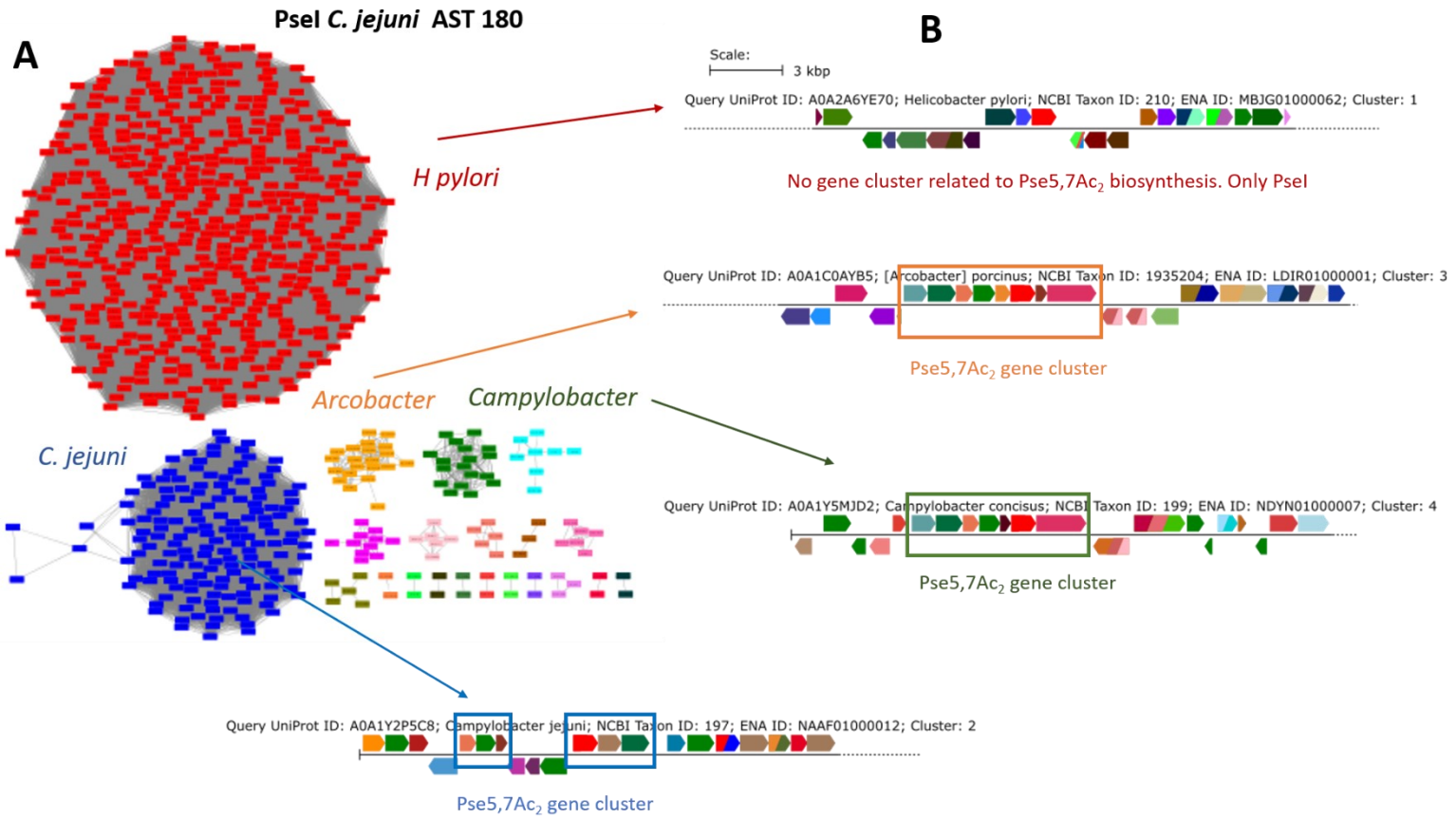

PseI C. jejuni AST 180
A
B
H pylori
Arcobacter
Campylobacter
C. jejuni
Scale:
3 kbp
Query UniProt ID: A0A2A6YE70; Helicobacter pylori; NCBI Taxon ID: 210; ENA ID: MBJG01000062; Cluster: 1
No gene cluster related to Pse5,7Ac2 biosynthesis. Only PseI
Query UniProt ID: A0A1C0AYB5; [Arcobacter] porcinus; NCBI Taxon ID: 1935204; ENA ID: LDIR01000001; Cluster: 3
Pse5,7Ac2 gene cluster
Query UniProt ID: A0A1Y5MJD2; Campylobacter concisus; NCBI Taxon ID: 199; ENA ID: NDYN01000007; Cluster: 4
Pse5,7Ac2 gene cluster
Query UniProt ID: A0A1Y2P5C8; Campylobacter jejuni; NCBI Taxon ID: 197; ENA ID: NAAF01000012; Cluster: 2
Pse5,7Ac2 gene cluster

Figure 3.12 A: A sequence similarity network analysis of Pse5,7Ac2 synthase *pseI* generated using EFI-Enzyme similarity tool with *C. jejuni pseI* as the seed sequence, shows four main clusters, one each for *H. pylori*, *C. jejuni, Arcobacter porcinus* and *C. concisus* derived *PseI* sequences. Each square in each cluster (node) corresponds to a single *pseI* sequence. Each line connecting a node (edge) corresponds to sequences with an E value of $\leq 1 \times 10^{-180}$. **B:** EFI-EST "Closest neighbour" analysis for the Pse5,7Ac2 synthase *pseI*. Gene clusters directly upstream and downstream of pseI are shown for *H. pylori, C. jejuni, Arcobacter porcinus* and *C. concisus*. No Pse5,7Ac2 biosynthetic gene cluster is present aroud *H. pylori pseI.*

Interestingly, when the genes that are 10 Kb directly upstream and downstream of the synthase *pseI* were queried, the entire Pse5,7Ac$_2$ biosynthetic pathway was clustered together (Figure 3.12B) in *Campylobacter jejuni* genomes. The same phenomenon was observed with *Arcobacter porcinus* and *Campylobacter concisus*. The opposite was observed with *H. pylori,* where the *pseI* gene from the Pse5,7Ac$_2$ biosynthetic pathways was always at least 10 kb away from the additional Pse5,7Ac$_2$ biosynthetic genes. Examination of the dehydratase *pseB*, aminotransferase *pseG*, acetyltransferase *pseH* and CMP-synthetase *pseF* gene networks and closest neighbours further supported a lack of clustering for *H. pylori* Pse5,7Ac$_2$ biosynthesis. Manual analysis of the sequenced genome of *H. pylori* in the NCBI database confirms that the synthase *pseI* occurs in one locus, the hydrolase *pseG* and acetyltransferase *pseH* occurs in another with the aminotransferase *pseC* approximately 30 Kb away, and the dehydratase/epimerase *pseB* occurs in a third distinct locus.

A number of Pse5,7Ac$_2$ genes that were cloned and expressed in this work were from distinct locuses in the genome of *H. pylori*. While the biosynthetic pathway from *H. pylori* produced Pse5,7Ac$_2$ *in vitro,*[10] there is a likelihood that utilizing a clustered biosynthetic pathway

such as the one present in *C. jejuni* may yield higher production levels *in vivo*. This is further supported by the stable dehydratase/aminotransferase pair from *C. jejuni* that moving forward we would like to construct. Taking it one step further, a future solution to improving Pse5,7Ac$_2$ titers may be to clone a clustered biosynthetic pathway and because the enzymes from *C. jejuni* were experimentally tested and verified, it would be a suitable bacterial strain to utilize for cloning purposes.

In conclusion, we have developed an *in vivo* microbial production method for the unusual nine carbon sugar, Pse5,7Ac$_2$. The biosynthetic genes (*pseBCHGI*) from *H. pylori* and *C. jejuni* were cloned into a low copy expression vector under the control of a T7 polymerase. Initally, expression in *E. coli* BL21 did not produce detectable amounts of this sugar. Next, production was tested in *E. coli* BRL02 (*nanAT*-), where less than 5 mg L^{-1} was detected by LC-ESI-MS/MS. The highest titer levels observed in this study was in the region of 15-20 mg L^{-1} and were obtained using an *E. coli* strain lacking sugar catabolism genes, BRL04 (*nanAT-, nagA*-) coupled with the overexpression of UDP-GlcNAc biosynthetic genes, *agm1* and *uap1*, resulted in a 4-fold increase in production of Pse5,7Ac$_2$. Despite multiple attempts at optimizing production by either testing different biosynthetic pathways that were cloned with Golden Gate assembly or modifying fermentation conditions, a titer greater than 20 mg L^{-1} was not obtained, which is in stark contrast to the related sugar Leg5,7Ac$_2$, where a titer of 120 mg L^{-1} was achieved in identical fermentation conditions. As mentioned earlier, a contributing factor to this observed phenomenon may be due to the primary biosynthetic genes that were used in this study were derived from non-clustering Pse5,7Ac$_2$ pathway in *H. pylori*.

Future work will involve cloning pathways from *C. jejuni*, which is shown to have a more stable dehydratase and aminotransferase pair (*pseBC*) coupled with the fact that the biosynthetic pathway clusters together and may be more suitable working together *in vivo*. Additionally, a modified screening strategy incorporating cell-free protein synthesis (CFPS) will be examined.[31] Instead of cloning the entire pathways for *in vivo* analysis, the individual genes to be tested will be cloned and used with a commercial reagent to detect for protein expression and Pse5,7Ac_2 production. This will potentially reduce workload and improve throughput by quickly mixing and matching desired biosynthetic genes without investing several days cloning, expressing and detecting of products. Initial tests to determine the feasibility of such a system are underway using Leg5,7Ac_2 as a proof of principle (Appendix Figure S3.9). *In vitro* protein expression is observed but further optimizations are required to generate detectable quantities of these complex carbohydrates.

3.2 Materials and Methods

General Methods

Reagents were purchased from Thermo Fisher Scientific and Sigma Aldrich and were used without additional purification steps. Synthetic DNA fragments were codon optimized and purchased from Thermo Fisher Scientific. Oligonucleotides used for Golden Gate Assembly, mutagenesis and PCR were purchased from Integrated DNA Technologies. Plasmid purification and Gel extraction kits were purchased from Omega Bio-Tek. PCR reactions for cloning were performed using Herculase II Fusion DNA polymerase and Phusion High-Fidelity DNA Polymerase from Agilent following the manufacturer's recommended protocol. The Zero Blunt PCR Cloning Kit from Invitrogen was used for subcloning amplified DNA fragments. Restriction endonuclease cloning was performed using standard methods. Plasmids were maintained in and replicated using *E. coli* Xl1 Blue or Top10. The expression vectors used in this study were pMIH02 (Amp^R), pMGX-HK (Km^R), pMGX-K (Km^R), pMGX-HA (Amp^R), pET21b (Amp^R) and pMGX-A (Amp^R), all under the control of a T7 promoter with resistance markers for ampicillin or kanamycin.

Construction of, and cloning with pMIH02

pET21b, an Amp^R expression vector under T7 control and pET28b, a T7 controlled Km^R expression vector were digested with SphI/EcoRI. The pET21b backbone (5096 bp) was ligated with the pET28b insert (406 bp) using T4 DNA ligase (New England Biolabs). The ligated product

was transformed with the commercially available and chemically competent Top 10 *E. coli*. Successful ligation was confirmed by sequencing (Eurofins MWG Operon, Louisville, Kentucky, USA) and designated as pMIH02 (Amp^R, N- and C-terminal his-tag). *PseB* (pMIH04), *pseH (*pMIH03) and *pseG* (pMIH05) were digested with EcoRI/NdeI from PCR-blunt. The fragments were cloned into pMIH02 after digestion with EcoRI/NdeI using T4 DNA ligase (New England Biolabs). The ligated product was transformed with chemically competent Top 10. Successful ligations were screened and confirmed by sequencing (Eurofins MWG Operon, Louisville, Kentucky, USA)

Pseudaminic acid production in shake flasks

Starter cultures were grown in 1 mL of LB media, supplemented with the necessary antibiotics, at 37°C, 200 rpm, for 18 h. A 20 mL F2 minimal media consisting of per litre, 12.24 g of K_2HPO_4, 6.0 g of KH_2PO_4, 4.0 g $(NH_4)_2SO_4$ and 175.5mg of $MgSO_4$ (in 250-mL shake-flask), supplemented with the necessary antibiotics, 0.25% casitone and 1% (v/v) glycerol was inoculated with 0.2 mL of starter culture. Production cultures were grown at 37°C, 200 rpm, until an OD600 of 0.5 was reached. At this point, pseudaminic acid production was induced with 0.2 mM IPTG (final concentration), and the incubation temperature was lowered to 30°C. Production cultures were grown for 72 h and supplemented with 0.6% (v/v) glycerol, 0.2% GlcNAc, and necessary antibiotics at 0, 18 and 36 h post induction. Cells were removed from the sample by centrifugation. The supernatant was collected and pseudaminic acid was analyzed following DMB derivatization.

DMB derivitization and Analysis of Cell Free Culture Broth Using LC-ESI-MS/MS.

Authentic Pse5,7Ac_2 standard was provided by Dr. Ian C. Schoenhofen at the Institute for Biological Sciences, National Research Council (Ottawa, Ontario, Canada). A protocol for DMB derivitization was adopted from Martin *et al*. Pse5,7Ac_2 is mixed with a DMB solution (8 mM DMB (Sigma-Aldrich), 1.5 M acetic acid, 14 mM sodium hydrosulfite, 0.8 M 2-mercaptoethanol) in a 1:1 volume mixture. The sample is placed in a heat block at 50ºC for 2.5 h and shielded from light due to the sensitivity of DMB. The sample is placed in a Shimadzu API LC-ESI-MS/MS system. Conditions were: Thermo Scientific C_{18}, 3 u, 100 mm x 2.1 mm, mobile phase A: 95% water, 5.0% acetonitrile, 0.05% formic acid. Mobile phase B: 95% acetonitrile, 5.0% water, 0.05% formic acid. Gradient: Linear gradient of 0 % B to 20 % B over 15 min, followed by a linear gradient from 20% B to 100% B over 1 min until completion. Flow rate: 0.300 mL min^{-1}. MS (ESI); calculated for $C_{20}H_{265}N_4O_7$ ($M - OH^+$) is 433.1, observed 432.9. Calculated for $C_{20}H_{27}N_4O_8$ ($M + H^+$) is 451.2, observed 451.0; Calculated for $C_{20}H_{26}N_4O_8Na$ ($M + Na^+$) is 473.2, observed 473.0. HPLC analysis was performed on samples after a catastrophic fire destroyed the Shimadzu API LC-ESI-MS/MS. For HPLC analysis, the samples were placed in a 1260 Infinity High Performance Liquid Chomatography system (Agilent Technologies). The column conditions were as follows: Prontosil C18, 5 µm, 125 mm x 4 mm, flow rate: 1.00 mL min^{-1}, λ= 350 nm. Mobile phase: 84% water: 9% acetonitrile: 7 % methanol constant for 12 minutes, followed by a linear gradient of acetonitrile from 9% to100 % over 1 min until completion.

Western Blotting to Detect for His-tagged Pse5.7Ac$_2$ Proteins.

PseC (pNRC37.1) and *pseI* (pNRC36.3) were obtained from the NRC. *PseB, pseH and pseG* were cloned as described below. Plasmids encoding the *pseB*, *pseC*, *pseH, pseG* and *pseI* genes were each individually transformed in chemically competent BL21 *E. coli*. Protein expression was performed using the fermentation conditions previously described ("Pseudaminic acid production in shake flasks"). The only modification was that proteins were grown at 30^{o}C for 18 hours post induction without additional feedings. Production broths were centrifuged at 3000 x g for 20 minutes. Pellets were resuspended with 0.2 mL of Lysis buffer (100 mM sodium phosphate, 300 mM NaCl, 10% (v/v) glycerol, 1 mg/mL lysozyme, 1 µg/mL pepstatin A, 1 µg/mL leupeptin, pH 8.0) per 1 mL of culture. Cell lysis was done with sonication on ice (3 pulses, 1 minute each). The cellular suspension was centrifuged for 1 h at 9000 x g. The crude lysates were subject to Western blot analysis. Samples were run on Precise Tris-HEPES pre-cast gels (Thermo Scientific, 10 x 8.5 cm). Proteins were then transferred on to PVDF membrane (Immobilon, 0.45 µm, 10cm x 7cm). HP- conjugated Anti-his monoclonal antibody (Genescript) reacting specifically with N-terminal and C-terminal his-tags of proteins was used. Blocking, transfer and antibody dilution buffers were prepared as recommended by the antibody kit. Detection was performed with an Immobilon^{TM} Western Chemiluminescent HRP substrate (Fisher Scientific) following the manufacturer's instructions.

Mutagenesis of PseG via Splice by Overlap Extension (SOE)

Internal Primers containing the point mutation of interest were designed by IDT. M13 external primers were also used for cloning purposes (S3.2). Phusion High-Fidelity DNA Polymerase and Thermocycler Mastercycler were used for all PCR reactions. PCR conditions were as follows: one cycle of 95 °C for 5 min, 35 cycles of 30 s at 95 °C, 60 s at 60 °C, and 60 s at 72 °C; followed by one cycle of 10 min at 72 °C. Gel purification of the two amplified bands was performed using the Omega E.Z.N.A. Gel Extraction Kit following the manufacturer's instructions. A PCR reaction to fuse the two purified fragments was then performed. The annealed product was ligated into PCR-blunt to generate the plasmid pMIH06 (Km^R). Successful mutagenesis was confirmed by sequencing (Eurofins MWG Operon, Louisville, Kentucky, USA). The gene was then cloned into pMIH02, an Amp^R pET21b derivative under T7 control. The pseG fragment in pMIH06 was digested via EcoRI/NdeI and subsequently cloned into pMIH02 following EcoRI/NdeI digestion to generate the plasmid pMIH07 (Amp^R)

Golden Gate Assembly Procedure

The Golden Gate Enzyme Mix used for cloning was purchased from NEB. Primers containing BsaI cut sites and four complementary nucleotides for sequential gene fragments were automatically generated using Benchling's Golden Gate primer design program. Primers were then ordered from IDT. PCR amplification was done using Herculase II Fusion DNA polymerase (Agilent) or Q5 DNA polymerase (NEB). The PCR amplified DNA was either enzymatically purified or extracted after running on a 1 % agarose gel. The ratio of DNA inserts to vector was

performed as suggested by the manufacturer of the Golden Gate mix. pMGX-HA or pET21b were used as the vectors for assembly. The first thee genes of the five gene Pse5,7Ac_2 pathway were first cloned into pET21b using Golden Gate assembly (pse*BCH*, pMIH63). This plasmid was then used as the vector in addition to *pseG* and *pseI* for a 2nd round of Golden Gate assembly to generate pMIH64 (*pseBCHGI*). Transformation of the Golden Gate ligation mixture were performed using chemically competent XL1Blue and plated on LB + Ampicillin plates. 12 individual clones were selected and screened using NdeI or EcoRI. Successful clones were sent for sequencing (Nanuq, Genome Quebec).

3.3 References

(1) Varki, A.; Gagneux, P. *Ann. N. Y. Acad. Sci.* **2012**, *1253* (1), 16–36.

(2) Lundgren, B. R.; Boddy, C. N. *Org. Biomol. Chem.* **2007**, *5* (12), 1903.

(3) Hassan, M. I.; Lundgren, B. R.; Chaumun, M.; Whitfield, D. M.; Clark, B.; Schoenhofen, I. C.; Boddy, C. N. *Angew. Chemie Int. Ed.* **2016**, *55* (39), 12018–12021.

(4) Goon, S.; Kelly, J. F.; Logan, S. M.; Ewing, C. P.; Guerry, P. *Mol. Microbiol.* **2003**, *50*, 659–671.

(5) Tsvetkov, Y. E.; Shashkov, A. S.; Knirel, Y. A.; Zähringer, U. *Carbohydr. Res.* **2001**, *335* (4), 221–243.

(6) Zunk, M.; Williams, J.; Carter, J.; Kiefel, M. J. *Org. Biomol. Chem.* **2014**, *12* (18), 2918–2925.

(7) Lee, Y. J.; Kubota, A.; Ishiwata, A.; Ito, Y. *Tetrahedron Lett.* **2011**, *52* (3), 418–421.

(8) Dhakal, B.; Crich, D. *J. Am. Chem. Soc.* **2018**, *140* (44), 15008–15015.

(9) Liu, H.; Zhang, Y.; Wei, R.; Andolina, G.; Li, X. *J. Am. Chem. Soc.* **2017**, *139* (38), 13420–13428.

(10) Schoenhofen, I. C.; McNally, D. J.; Brisson, J. R.; Logan, S. M. *Glycobiology* **2006**, *16* (9), 8–14.

(11) Schoenhofen, I. C.; McNally, D. J.; Vinogradov, E.; Whitfield, D.; Young, N. M.; Dick, S.; Wakarchuk, W. W.; Brisson, J. R.; Logan, S. M. *J. Biol. Chem.* **2006**, *281*, 723–732.

(12) Plumbridge, J. A. *Mol. Microbiol.* **1989**, *3* (4), 505–515.

(13) Jackman, J. E.; Raetz, C. R. H.; Fierke, C. A. *Biochemistry* **1999**, *38* (6), 1902–1911.

(14) Martínez-gómez, K.; Flores, N.; Castañeda, H. M.; Martínez-batallar, G.; Hernández-chávez, G.; Ramírez, O. T.; Gosset, G.; Encarnación, S.; Bolivar, F. *Microb. Cell Fact.* **2012**, *11* (46), 1–21.

(15) Bearne, S. L.; Blouin, C. *J. Biol. Chem.* **2000**, *275* (1), 135–140.

(16) Mosberg, J. A.; Lajoie, M. J.; Church, G. M. *Genetics* **2010**, *186* (3), 791–799.

(17) Hofmann, M.; Boles, E.; Zimmermann, F. K. *Eur. J. Biochem.* **1994**, *221*, 741–747.

(18) Wendland, J.; Schaub, Y.; Walther, A. *Appl. Environ. Microbiol.* **2009**, *75* (18), 5840–5845.

(19) Mio, T.; Yabe, T.; Arisawa, M.; Yamada-okabe, H. **1998**, *273* (23), 14392–14397.

(20) Álvarez-Añorve, L. I.; Gaugué, I.; Link, H.; Marcos-Viquez, J.; Díaz-Jiménez, D. M.;

Zonszein, S.; Bustos-Jaimes, I.; Schmitz-Afonso, I.; Calcagno, M. L.; Plumbridge, J. *J. Bacteriol.* **2016**, *198* (11), 1610–1620.

(21) Liu, F.; Tanner, M. E. *J. Biol. Chem.* **2006**, *281* (30), 20902–20909.

(22) Gottesman, S. *Annu. Rev. Genet.* **1996**, *30*, 465–506.

(23) Murkin, A. S.; Chou, W. K.; Wakarchuk, W. W.; Tanner, M. E. *Biochemistry* **2004**, *43*, 14290–14298.

(24) Morrison, M. J.; Imperiali, B. *Biochemistry* **2014**, *53* (4), 624–638.

(25) Mauro, V. P.; Chappell, S. A. *Methods Mol. Biol.* **2018**, *1850*, Chapter 18.

(26) Hassan, M. I., McSorley, F. R., Hotta, K., Boddy, C. N. *J. Vis. Exp.* **2017**, *124* (e55187), doi:10.3791/55187.

(27) Gibson, D. G.; Benders, G. A.; Andrews-Pfannkoch, C.; Denisova, E. A.; Baden-Tillson, H.; Zaveri, J.; Stockwell, T. B.; Brownley, A.; Thomas, D. W.; Algire, M. A.; Merryman, C.; Young, L.; Noskov, V. N.; Glass, J. I.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. *Science* **2008**, *319* (5867), 1215–1220.

(28) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. *Nat. Methods* **2009**, *6* (5), 343–345.

(29) Engler, C.; Kandzia, R.; Marillonnet, S. *PLoS One* **2008**, *3* (11).

(30) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Slater, D. R.; Whalen, K. L. *Biochim. Biophys. Acta* **2015**, *1854* (8), 1019–1037.

(31) Khambhati, K.; Bhattacharjee, G.; Gohil, N.; Braddick, D.; Kulkarni, V.; Singh, V. *Front. Bioeng. Biotechnol.* **2019**, *7* (October), 1–16.

Chapter Four: *De novo* polyprotein design for the flexible *in vivo* heterologous expression of natural products.

4.1 Introduction

Heterologous expression of biosynthetic pathways as a tool of producing natural products is an extensively explored area of research, encompassing the production of simple molecules such as ethanol[1] to more complex intermediates such as artemisinic acid.[2] A wide range of microbial hosts have also been utilized for this purpose including prokaryotic organisms like *E. coli*[3], *Bacillus subtilis*[4] and the commonly used eukaryotic host *Saccharomyces cerevisae*.[5] Selection of an ideal host for the production of a desired natural product involves a number of factors, including identifying relevant primary and secondary metabolites present in a host organism that can be harnessed to optimize production levels of a specific compound. Different production hosts also utilize various promoters, selection markers, possess varying codon usage, transcript stability, and in the case of eukaryotic versus prokaryotic hosts, the manner and location whereby transcription and translation occurs also differ.[6] These factors contribute to the complexity of selecting a desirable host for the production of a specific natural product, and should be considered before moving natural product production between diverse hosts.

To overcome the barriers associated with moving biosynthetic pathways from one expression host to another, we sought to develop a flexible *in vivo* heterologous expression system that would enable rapid movement of natural product production between hosts with relative ease and minimal modifications. To achieve this goal, we took inspiration from the manner in which

viruses package, process and cleave their polyproteins.[7,8] Viruses arrange multiple genes in a single open reading frame, upon which transcription and translation generates one lengthy polyprotein that is then processed at specific sites by viral or cellular proteases.[9]

Before cloning and testing of a *de novo* polyprotein expression system, we first needed a natural product with a known biosynthetic pathway to use as a proof of principal. For this purpose, we used violacein, a bis-indole natural product with anti-tumor properties.[10] Violacein is a bright purple pigment that can be easily visualized when heterologously produced. The biosynthetic pathway of violacein was identified in *Chromobacterium violaceum*[11] and is well characterized, involving 5 enzymatic conversions starting from L-tryptophan (Figure 4.1).[12] The first step in the pathway is the oxidation of L-tryptophan[13] (**1**) to indole-3-pyruvic acid imine (IPA, **2**) by VioA, a tryptophan 2-monooxygenase. Then, VioB condenses two IPA moieties to produce an IPA imine dimer (**3**) VioE then converts **3** to protodeoxyviolacein (**4**) by formation of a lactam ring.[14] VioD oxidizes one indole ring to generate deoxyveolacein (**5**). VioC then hydroxylates a second indole ring, followed by spontaneous oxidative decarboxylation to produce violacein (**6**). The intermediates in the biosynthesis of violacein have been fully characterized[12], thus a lack of product formation from the polyprotein design due to non-functioning or truncated proteins could be readily identified. Additionally, because L-tryptophan is endogenously available in all heterologous hosts, supplementing production of this natural product with an exogenous starting substrate is not necessary.

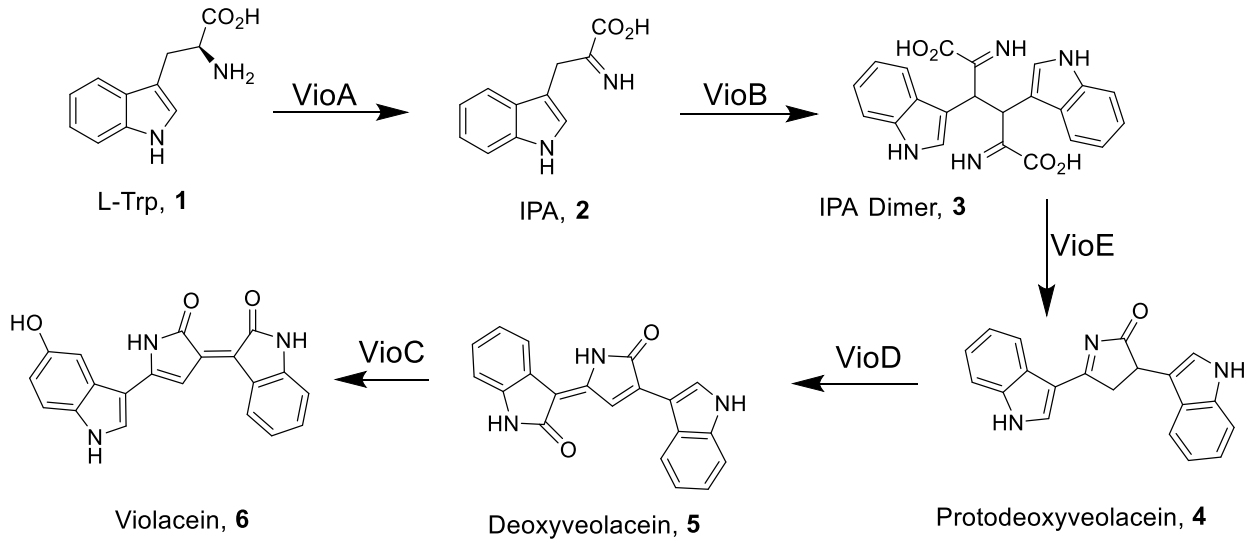


Figure 4.1: Biosynthetic pathway of violacein identified *in C. violaceum* starting from L-Tryptophan.

A key advantage to using violacein as the proof of concept natural product for this study is that the biosynthetic pathway from *C. violaceum* is relatively small (<8 Kb), thus minimizing cloning issues that could potentially arise when working with larger gene clusters. This allows us to readily study and optimize the polyprotein design as opposed to expending time and effort in cloning large pathways. Lastly, violacein is an attractive natural product for initial studies because the biosynthetic pathway has been shown to express as individual functional proteins in a number of diverse organisms including *E. coli*[15] and *S. cerevisae*.[16]

Construction of a *de novo* polyprotein was designed as shown in Figure 4.2. In addition to the violacein pathway, an N-terminal spacer sequence was to be cloned in front of each individual gene in the biosynthetic gene cluster. This sequence possessed a few features, the first of which is a linker designed to enable flexibility of the mature polyprotein. Variable linker lengths will be tested to strike a balance between linker length and violacein titers. Initially linkers lengths of 5

amino acids were cloned. We decided to use the Tobacco Etch Virus (TEV) protease, a 49 kDa protein that is used to remove affinity tags during protein expression, to cleave the polyprotein at specific recognition sites to liberate the mature individual proteins.[17] TEV protease has been shown to express in prokaryotes (*E. coli*)[18] and in eukaryotes (*S. cerevisiae*)[19], and is therefore an ideal protease to use for this study. The TEV protease recognition sequence was encoded downstream of the linker in the N-terminal spacer region between each violacein gene.[20] The last feature of the polyprotein construct is a FLAG tag sequence downstream of the TEV protease, which is to be probed with a FLAG tag antibody for western blotting. The FLAG tag is situated such that cleavage of the TEV protease at recognition sequence will leave the FLAG tag at the N-terminus of each protein in the violacein biosynthetic pathway.

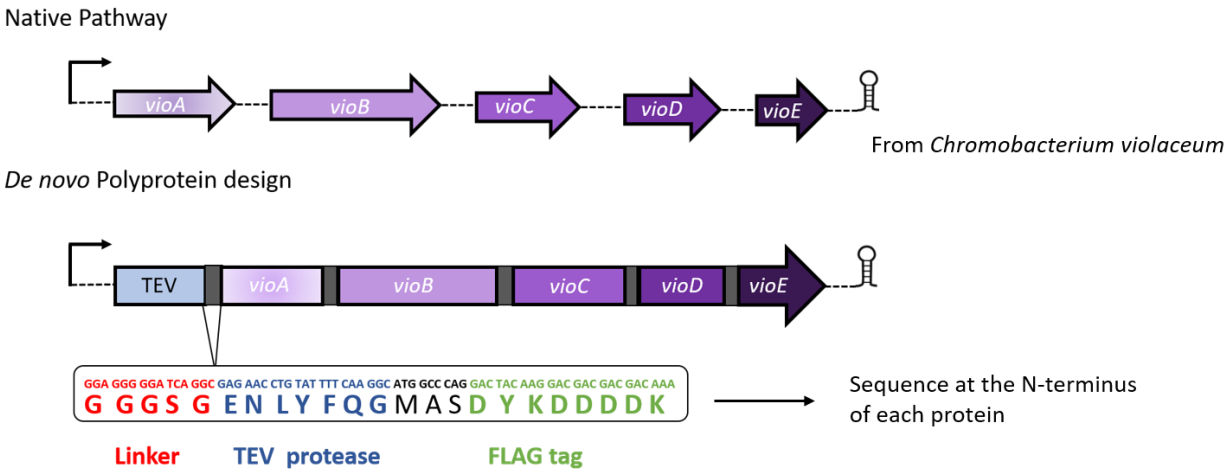


Figure 4.2 *De novo* polyprotein design for the heterologous expression of natural products. All five genes are fused together with a cleavable linker (red). TEV protease (blue) self-cleaves once expressed and can then process the individual proteins which can be detected by an N-terminal FLAG tag (green). 90 degree arrow indicates a promoter and stem loop represents terminator sequence.

Figure 4.3 highlights the predicted process for polyprotein cleavage and formation of fully mature proteins. The polyprotein was designed with multiple genes in a single open reading frame, with expression controlled by a promoter compatible with the host organism in which violacein is to be expressed. For example, if expression of the polyprotein in *E. coli* is desired, the full violacein polyprotein cluster (Figure 4.2) could be cloned under an inducible T7 promoter or a constitutive *trc* promoter. Upon initiation of transcription and completion of translation, TEV protease self-cleaves at the TEV recognition sequence located in the N-terminus of VioA. The free TEV protease can then proceed to cleaving each and every TEV protease recognition sequence to produce fully mature proteins required for violacein biosynthesis. Alternatively, as TEV protease is located at the N-terminus, it's conceivable that proteolysis starts before translation is complete. Differences in transcription and translation between prokaryotes and eukaryotes could account for the different mechanisms by which polyprotein processing and cleavage occurs.[6]

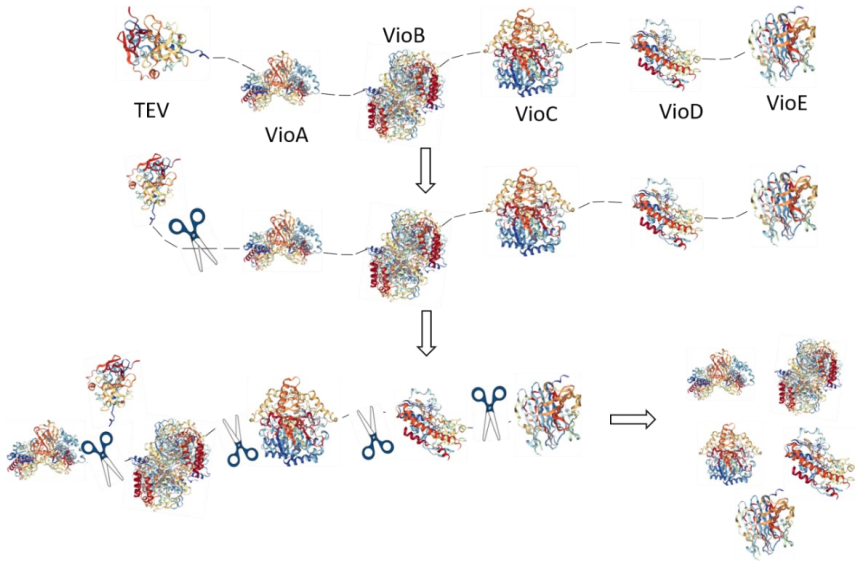


Figure 4.3 Processing of polyprotein is initiated by TEV protease self cleavage. TEV protease then cuts at N-terminal recognition sequences allowing mature proteins to fold.

To implement the designed polyprotein system, we first sought to express the full construct containing the violacein biosynthetic pathway in *E. coli*. The polyprotein design encompassing the TEV protease, linker regions and two sets of violacein genes were cloned. One set of genes was from *C. violaceum* (pFM101) and was cloned by Dr. Fern McSorley. Another set of genes were codon optimized for expression in *S. cerevisae* (pMIH90) and ordered as five individual fragments (each 1-2 Kb) that were cloned with a combination of SOE and Gibson assembly. Gibson assembly is a cloning strategy that enables the assembly of multiple linear DNA fragments into a desired plasmid. This method employs small homologous overhangs installed by PCR amplification that typically consist of 15-40 bp.[21] During assembly of these fragments, an exonuclease exposes the 5' ends a DNA polymerase is responsible for gap filling. Finally, a ligase covalently joins consecutive segments with homologous overlapping regions.[22]

The fully assembled polyprotein design was cloned into a pET28b vector under the control of an inducible T7 promoter for expression in *E. coli*. LB plates containing viable clones in the *E. coli* strain BL21 turned purple after 5 days in the absence of the inducer (Figure 4.4A), indicating formation of violacein and thus successful expression of the polyprotein. Single clones were then selected and grown in nutrient rich liquid media and 12 h after induction with IPTG, the media turned a dark purple colour (Figure 4.4A). To confirm complete processing of the polyprotein into individual mature proteins, a western blot using the N-terminal FLAG tags that were installed in the linker regions was performed. All five violacein proteins were detectable (VioC and VioD at lower levels) and importantly, a large smear indicative of incomplete polyprotein processing in the region of the gel above the individual proteins was not observed (Figure 4.4B).

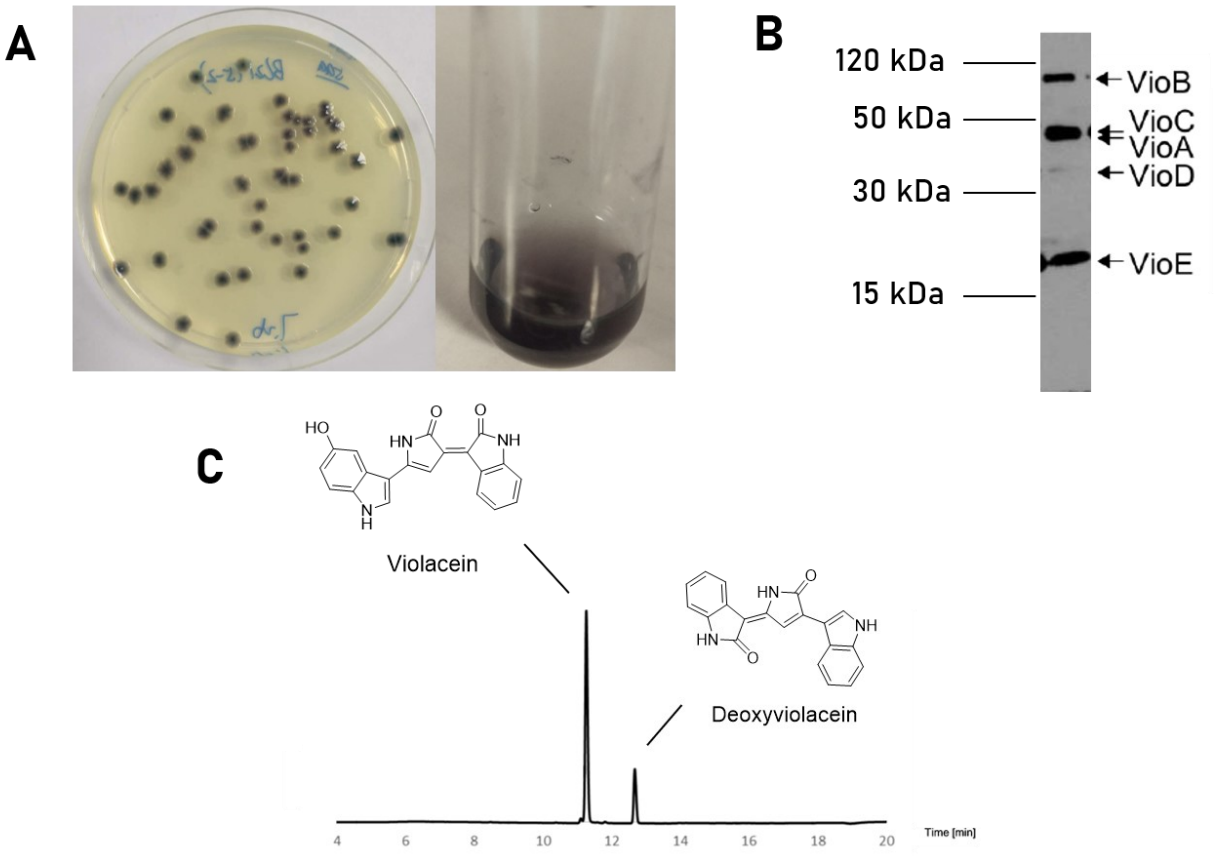


Figure 4.4: Violacein expression in *E. coli* **A**: Purple pigment production from violacein using the *de novo* polyprotein design is evident on LB plates as well as in liquid media. **B**: Western blotting of violacein proteins performed with pre-cast TGX using primary FLAG and secondary mouse IgG antibodies. Complete processing and cleavage of individual proteins from a polyprotein is observed. **C**: HPLC analysis (λ= 585 nm) of violacein produced in *E. coli*. The violacein (retention time 11.4 mins, 37 ± 1 mg L^{-1}) to deoxyviolacein (retention time 12.5 mins) ratio is 85:15.

HPLC analysis confirmed the presence of violacein and its intermediate deoxyviolacein in an 85:15 ratio (Figure 4.4C) in *E. coli*. Violacein is produced intracellularly and is thus present in the cell pellet. To purify violacein from *E. coli*, enzymatic lysis with lysis buffer was performed followed by sonication to release violacein. Separation of the desired natural product from cellular

debris was achieved by several rounds of ethyl acetate extractions. The aqueous layers were collected, evaporated weighed and the resulting material containing highly pure violacein is resuspended with methanol before HPLC analysis. The violacein titer in *E. coli* was also quantified using a standard curve that was generated using highly pure violacein. It was determined that a titer of 37 ± 1 mg L^{-1} was produced. The *S. cerevisiae* codon optimized violacein pathway produced 60 ± 13 mg L^{-1} when expressed in *E. coli*. Furthermore, mass spectral analysis was used to confirm that the product obtained was violacein.

Table 4.1: Violacein titers obtained after increasing linker lengths in the polyprotein design. Violacein titers were determined by HPLC quantification after ethyl acetate extraction from *E. coli.*

Linker Length	**Titer (mg L^{-1})**
5 amino acids	37 ± 1
10 amino acids	46 ± 1
15 amino acids	55 ± 2
18 amino acids	54 ± 1

After successfully producing violacein in *E. coli*, we sought to test certain parameters of the polyprotein design to enhance production levels. Initial polyprotein designs contained a linker with five amino acids. We sought to examine the effects of slightly larger amino acid linkers in

terms of violacein titers in *E. coli* (Table 4.1). Dr. Fern McSorley cloned three violacein polyprotein pathways (pFM101) from *C. violaceum* containing 10, 15 and 18 amino acid linkers in addition to the initial construct possessing a five amino acid linker. As the amino acid linkers were increased in length, there was slight increase in violacein titer. Increasing the amino acid linker length from five to 15 resulted in a 50% increase in violacein production. While this is a significant increase in terms of violacein titers, the five amino acid linker was used for violacein production in other organisms moving forward. While an increase in production levels is desired, the use of lengthy and repetitive linker lengths could introduce difficulties in terms of designing and cloning of the polyprotein constructs.

To test the robustness of this polyprotein design, we wanted to express violacein in a tumor targeting strain of *Salmonella typhimurium*.[23] The similarities between the genomes of *E. coli* and *S. typhimurium* should facilitate successful production of violacein in both of these gram negative bacterial strains.[24] The full polyprotein construct was cloned into pKK233, a low copy plasmid containing a constitutive *trc* promoter (pMIH92). Screening of successful clones was simplified due to the use of a constitutive promoter prompting immediate production of violacein production, which can be visualized on LB plates (Figure 4.5A). Large scale shake flask production was performed to verify that pigmentation is a result of violacein production, which was confirmed by HPLC (Figure 4.5B). Violacein was found to be the major product in *S. typhimurium*. Trace amounts of the dehydroxylated precursor deoxyviolacein, as well as the dideoxy precursor protodeoxyviolacein were isolated and detected by HPLC. The polyprotein construct under the control of a trc promoter should function and produce violacein in *E. coli* BL21 and was therefore

tested. As expected, successful biosynthesis of violacein was obtained. Quantification of this construct in *E. coli* showed that utilizing the trc promoter produced substantially lower titers of violacein (< 4 mg L^{-1}) in comparison to the T7 promoter (37 mg L^{-1}). The ability to manipulate *in vivo* production levels by simply altering the strength of the promoter that is used could provide a very useful level of control. For example, if this polyprotein design is used to produce proteins and/or natural products that are toxic to the host organism, their output can be minimized by using relatively weaker promoters such as *trc* and thus minimize cell death.

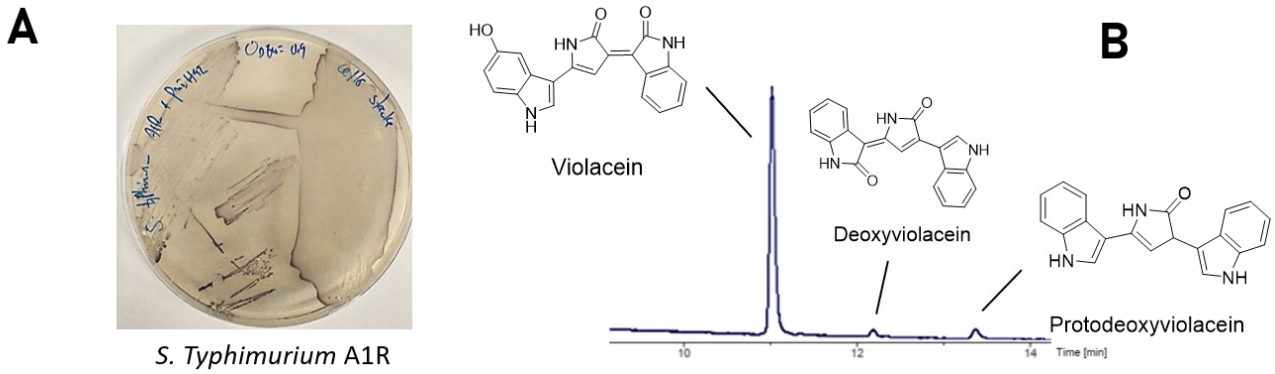


Figure 4.5: Violacein production *in S. typhimurium* A1R. A: Purple pigmentation is observed after overnight growth of streaked bacterial cultures containing transformed violacein polyprotein clones. B: HPLC analysis (λ= 585 nm) of violacein production in *S. typhimurium* A1R. Predominant product is violacein (retention time 11.1 mins) with a small amount of deoxyviolacein (retention time 12.2 mins) and protodeoxyviolacein (retention time 13.3 mins) intermediates detected.

After expression of the *de novo* polyprotein design in two gram-negative prokaryotic hosts, we wanted to expand the scope of this method and test the feasibility of utilizing this system in eukaryotic organisms. A well-established methodology is using HEK293T[25] and 786-O[26]

mammalian cell lines for protein expression. Post-translational modifications absent from bacterial organisms make these eukaryotes highly attractive cell lines for biologists. The full, unmodified polyprotein construct containing five amino acid linkers and the violacein biosynthetic pathway from *C. violaceum* was cloned into pCDNA3.1 (pFM110), a mammalian expression vector containing a constitutive cytomegalovirus immediate-early (CMV) promoter. This construct was then transfected into HEK293T, 786-O and HeLa cells by Nicholas Boileau, an undergraduate honour's student. After 72 h post transfection there was a lack of purple pigmentation in the liquid cultures, which typically indicates an absence of violacein production.

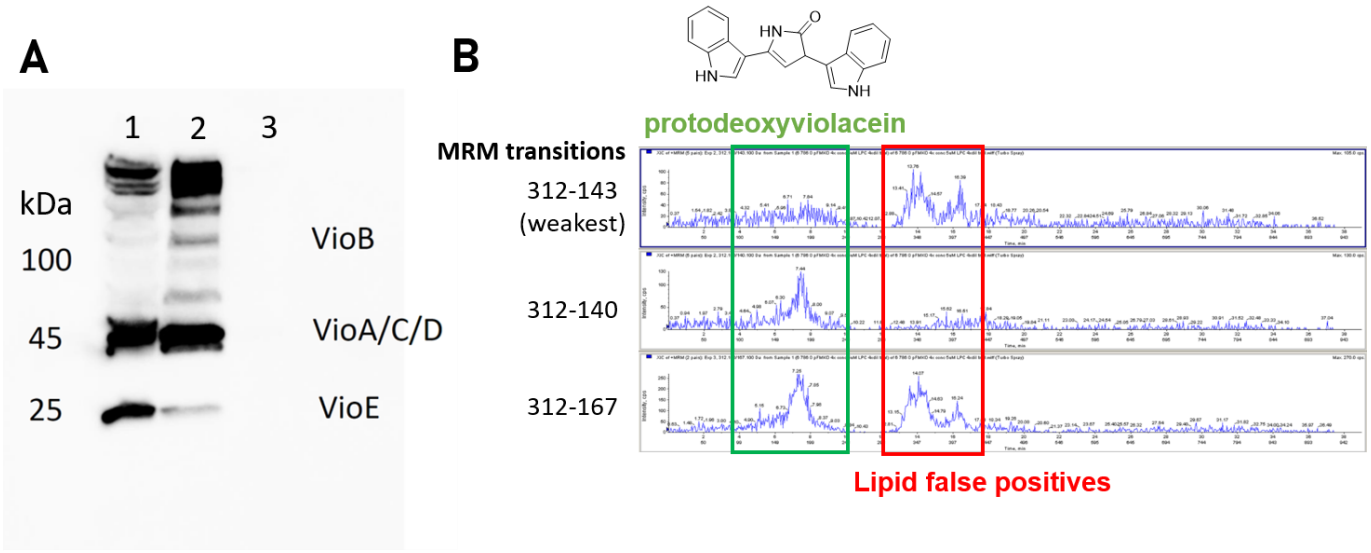


Figure 4.6: Unsuccessful violacein production in mammalian cells is observed despite polyprotein processing. **A:** Western blotting of HEK293T (1), 786-O (2) and HeLa (3) cells. Western blotting was performed with pre-cast TGX gels using primary FLAG, and secondary mouse IgG antibodies. All violacein biosynthetic proteins are present. Incomplete polyprotein processing is also detected. **B**: Metabolomics-type extraction with acidified Bligh and Dyer ($MeOH_{(aq)}$/$CHCl_3$ partition) detects protodeoxyviolacein intermediate. After separation, both positive and negative electron spray ionization (ESI) were tested for the most sensitive MRM transitions, with a positive ESI at 35 eV yielding the most sensitive transitions.

Western blotting of the mammalian cell lysates (Figure 4.6A) with the FLAG tag antibody showed that the polyprotein was expressed, with individual violacein proteins detected. Western blotting also shows a large mass of proteins above the largest violacein protein (VioB), which indicates that while the polyprotein is expressed in mammalian cell cultures, there is incomplete or inefficient processing by the TEV protease resulting in multiple large protein smears appearing on the gel. This is observed in both HEK293T and 786-O cell lines while protein expression was not observed in HeLa cells. Although a purple pigment was not observed from liquid cultures, partial polyprotein processing may result in the detection of violacein or its biosynthetic intermediates. After analysis by LCMS, violacein was not detected in cell cultures. To improve detection by LCMS, cell cultures were extracted by the classical acidified Bligh and Dyer, a method typically used for the isolation of lipids that is suitable for extraction with violacein due to its greasy properties.[27] After extraction using methanol and chloroform, samples were analyzed via multiple reaction monitoring (MRM) to detect for fragmentation of violacein and its intermediates (Figure 4.6B).[28] Both positive and negative electron spray ionization were tested for the most sensitive MRM transitions. It was found that product ion scans in positive ESI at 35 eV resulted in the most sensitive transitions. Although violacein was not detected, transitions for the dideoxy intermediate protodeoxyviolacein were observed.

The inability to produce violacein in mammalian cells raised a number of questions about the polyprotein design. One significant concern is that certain features may need to be modified for complete processing of the polyprotein to be observed. To determine whether a lack of violacein production was strictly a mammalian cell line issue or if there are broader concerns

regarding expression of this system in eukaryotic organisms, we decided to test whether the polyprotein design is functional in *S. cerevisiae*. The full polyprotein construct containing five amino acid linkers with a violacein pathway codon optimized for expression in yeast was cloned into p426, a vector containing a constitutive TEF1 promoter. Successful clones were screened by restriction enzyme digestion and transformed into *S. cerevisiae* BY4741 using the LiAc/SS carrier DNA/PEG method.[29,30] Individual clones were selected and grown in yeast synthetic dropout media with appropriate amino acids and supplemented with rich media. After several days the individual colonies developed an unusual green pigmentation (Appendix, Figure S4.1). Western blotting of the polyprotein system in *S. cerevisiae* with a FLAG tag antibody failed to show any detectable proteins (Appendix, Figure S4.2). Nevertheless, the green pigmentation observed from the yeast colonies prompted further analysis.

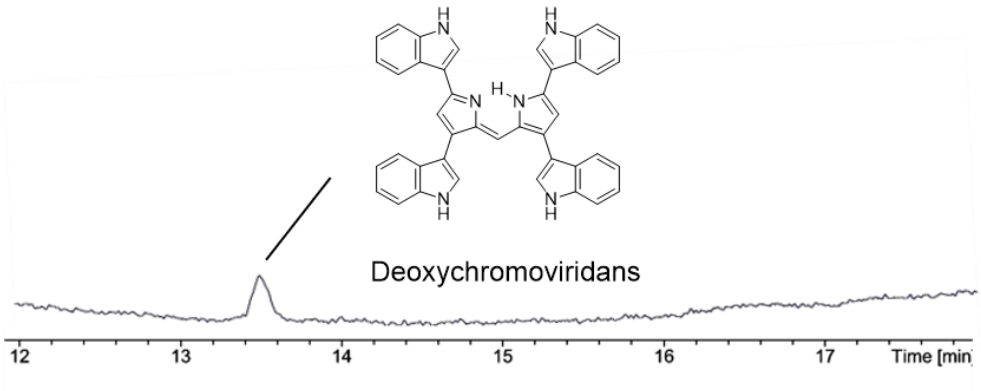


Figure 4.7: HPLC analysis (λ= 585 nm) of crude extracts from *S. cerevisiae* transformed with violacein polyprotein pathway. Retention time of deoxychromoviridans peak at 13.5 mins.

Large scale cultures of *S. cerevisiae* transformed with the plasmid harboring the polyprotein pathway were grown and cell pellets were extracted with ethyl acetate. The extracts were than analyzed by LCMS, where neither violacein nor deoxyviolacein were detected. Further analysis was performed on an HPLC with a diode array detector. Optimal signal for violacein is observed at 585 nm and after analyzing the crude extracts from *S. cerevisiae*, a small peak was observed (Figure 4.7). The retention time of this peak coupled with the green pigmentation observed from the plates points towards the production of deoxychromoviridans, a dimer that is formed by the condensation of two IPA intermediates of the violacein pathway.[12]

To investigate the lack of protein production and violacein formation in *S. cerevisiae*, we probed whether a lack of transcription and formation of mRNA is the cause of this observation. To test whether transcription is occuring, mRNA was extracted from *S. cerevisiae* and cDNA was generated using reverse transcriptase and TEV gene specific primers. [31] The cDNA was amplified with TEV specific PCR primers and a band on the DNA gel was observed, confirming transcription and formation of mRNA (Figure 4.8). A RT-PCR control sample without the addition of reverse transcriptase (Figure 4.8, lane 4) serves as a negative control to verify that the DNA bands obtained from the PCR reaction are from cDNA rather than contaminant genomic or plasmid DNA.

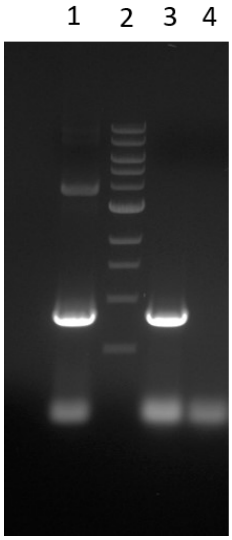


Figure 4.8: RT-PCR transcriptional analysis of *S. cerevisiae* polyprotein construct. TEV protease was converted to cDNA and detected on a DNA gel after PCR amplification. Lane 1: TEV Control PCR. Lane 2: DNA Ladder. Lane 3: S. cerevisiae TEV cDNA amplification. Lane 4: *S. cerevisiae* TEV cDNA amplification, no reverse transcriptase added (negative control).

A lack of protein expression in *S. cerevisiae* despite evidence that transcription and mRNA formation is occurring resulted in screening a variety of culture conditions in an attempt to supplement protein production. The starting substrate of the violacein biosynthetic pathway is tryptophan, which is also required for amino acid and protein production in *S. cerevisiae*. Producing proteins and a natural product that both share a substrate from primary metabolism may be taxing to the host organism and potentially result in insufficient protein production and as an added affect, limited violacein output. We predicted that supplementing the liquid medium that is used in growth and expression studies with tryptophan would circumvent this phenomenon if indeed intracellular levels of tryptophan are insufficient. Concentrations of L-tryptophan ranging from 50 μM to 1 mM were added to liquid cultures during shake flask production to no effect.

Neither a purple pigment nor a HPLC signal were detected for all concentrations that were screened. Additionally, we also tested various rich media, extending production times, and even tested the violacein pathway from *C. violaceum* as opposed to the codon optimized pathway with no success.

In conclusion we have developed a *de novo* polyprotein design for the heterologous expression of natural products. The polyprotein design consists of a TEV protease, a spacer region consisting of: a flexible five amino acid linker, a TEV protease recognition sequence and a FLAG tag at the N-terminus of each gene. Violacein was used as a proof of principle for this methodology. Production of a violacein polyprotein was achieved with great success in prokaryotes where in *E. coli*, titers of approximately 37 mg L^{-1} were achieved. Violacein was also produced in *S. typhimurium*, where the polyprotein pathway was expressed with a constitutive promoter. On the other hand, very limited success was achieved when expressing the violacein polyprotein design in eukaryotes. In mammalian cells, there was an absence of purple pigment, with incomplete processing of the polyprotein evident by western blotting. A violacein intermediate lacking the final two oxidation steps were observed, but violacein was not detected by LCMS. A similar trend was observed in *S. cerevisiae*, where purple pigmentation was once again lacking after expression of the polyprotein design. Instead, a green pigment was observed, and HPLC analysis points to the possibility that the IPA intermediate from the violacein biosynthesis has dimerized. Stronger evidence is required before we can definitively support this claim.

Future work regarding this project will focus primarily on efforts to successfully express this system in eukaryotic hosts. Transfection and growth of the mammalian cell lines were

performed at 30 °C due to the instability of VioB at higher temperatures, which prevented the typical 37 °C working temperatures for mammalian cells.[32] An alternative natural product biosynthetic pathway may be a suitable alternative if poor expression of VioB at 37 °C limits violacein production. In *S. cerevisiae*, an alternative expression vector with a strong inducible promoter will be tested.

It is possible that small modifications may prove insufficient in terms of successfully expressing violacein with a polyprotein system in eukaryotes. A more in-depth systematic modification of the polyprotein sequence and positioning of the protease may be necessary. In the TEV, a (+) ssRNA virus that infects plants, the TEV protease is part of the NIa sequence of the polyproteins with 7 viral cleavage sites.[34] In contrast to our *de novo* polyprotein design, where the TEV protease is situated at the N-terminus, the protease for TEV is located towards the C-terminus of the polyprotein. The positioning of the protease is especially relevant in mammalian systems where translation is significantly slower than in prokaryotes.[35] It is possible that the fully mature TEV protease is formed, self-cleaves and diffuses away from the rest of the polyprotein before translation and formation of the TEV cleavage sites can completed. Placing the TEV protease towards the C-terminus of the *de novo* polyprotein design will be tested to improve production of violacein in eukaryotes. This will enable protease activity to begin after more of the protein is translated, thus ensuring that the protease is around to cleave the latter parts of the polyprotein.

Another method that could be explored is implementing an alternative protease with improved expression in eukaryotes, such as the HIV-1 protease. This protease requires the formation of a dimer by folding together two polyproteins to produce an active dimer, may improve

polyprotein processing that was incomplete in mammalian cells (Figure 4.6A).[33] Alternatively, optimizing the codons of the TEV protease recognition sequence for expression in eukaryotes may be a simpler modification. Additionally, testing of various amino acid linker lengths was optimized in *E. coli,* therefore optimizations may be required to identify suitable linker lengths for expression in eukaryotes.

Ultimately these modifications may necessitate the use of two polyprotein designs to express natural products; one tailored for expression of natural products in prokaryotes and another for expression in eukaryotes. This would deviate from our initial goal of a single system to express across diverse hosts, but would nonetheless contribute a valuable tool to researchers in the field of natural products biosynthesis. Future work will also include utilizing this tool to express natural products with larger biosynthetic pathways than the relatively small violacein pathway tested in this work.

4.2 Materials and Methods

General methods

Reagents were purchased from Thermo Fisher Scientific, Bio-Rad and Sigma Aldrich and were used without further purification. Synthetic DNA fragments were codon optimized and purchased from GeneUniversal (Delaware, USA). Oligonucleotides used for Gibson assembly, SOE and PCR were purchased from IDT. Plasmid purification and Gel extraction kits were purchased from Omega Bio-Tek. PCR reactions for cloning were performed using Herculase II Fusion DNA polymerase Agilent following the manufacturer's recommended protocol. The Zero Blunt PCR Cloning Kit from Invitrogen was used for subcloning amplified DNA fragments. Restriction endonuclease cloning was performed using standard methods, unless otherwise stated. Plasmids were maintained in and replicated using *E. coli* Xl1Blue. The expression vectors used in this study were pET28b (T7, Km^R), pMGX-A (T7, Amp^R). pKK233 (*trc*, Amp^R), pRS426 (TEF1, Amp^R) and p426 (TEF1, URA3, Amp^R).

Cloning of violacein pathway with splice by overlap extension (SOE) and Gibson Assembly

The violacein pathway as a polyprotein was designed, codon optimized for expression in *S. cerevisiae* and ordered as synthetic DNA fragments. Gibson assembly primers were designed using Benchling's automated primer design feature and ordered from IDT. The synthetic fragments were amplified by PCR using standard conditions. After gel extraction and purification, Gibson assembly was unsuccessfully attempted using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) using manufacturer's recommendations. An alternative strategy of employing

SOE to ligate adjacent amplified fragments was necessary to mask the N-terminal FLAG tags present in every gene. *VioAB* and *vioCDE* were assembled via SOE of individual gene fragments. Next, the forward primer of Vio*A* and the reverse primer of vio*E* were used in a SOE PCR to generate *vioABCDE*. One large 8 Kb fragment was successfully amplified and Gibson assembly using the NEBuilder HiFi DNA Assembly Master Mix was used to clone this fragment into pFM82, an expression vector under the control of a T7 promoter with an installed TEV protease to generate pMIH90 (*vioABCDE*, T7 promoter, Km^R)

Expression of pMIH90 in *E. coli* to produce violacein

pMIH90 was transformed into *E. coli* for the production of violacein. Starter cultures were grown in 1 mL of LB media, supplemented with the necessary antibiotics, at 37°C, 200 rpm, for 18 h. A 250 mL flask containing 20 mL LB media supplemented with the necessary antibiotics was inoculated with 0.1 mL of starter culture. Cultures were grown in a shake flask at 37 °C, 200 rpm, until an OD_{600} of 0.5 was reached. At this point, violacein production was induced with a final concentration of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) and the incubation temperature was lowered to 30 °C due to the instability of VioB at higher temperatures. Production cultures were grown for 24 h upon which cell pellets were harvested by centrifugation.

Cloning and expression of pMIH92 in *S. typhimurium*

The violacein polyprotein design was cloned from pMIH90 into a plasmid pKK233-hCAM via SpeI/PacI restriction sites following standard endonuclease cloning to produce pMIH92 ((*vioABCDE*, *trc* promoter, Amp^R). To transform this plasmid, electrocompetent *S. typhimurium*

A1R was first prepared. 100 mL LB was inoculated with a fresh O/N culture and grown at 37°C, 200 rpm until an OD_{640} of 0.75 was reached. Cells were chilled on ice for 15 mins, followed by centrifugation at 4000 x g for 10 mins. Cells were then washed with a culture volume of HEPES then once with 10% glycerol (1% culture volume). Cells were then resuspended with 10% glycerol and stored at -80ºC. Next, 0.5 ng of DNA was added to 40 µl of electrocompetent with gentle mixing. This mixture was then transferred into a prechilled cuvette and placed in an Eppendorf Eporator electroporator. Electroporation was initiated with the following settings: Voltage: 2500 V; time constant: 5 ms. 1 mL of SOC medium was immediately added and the mixture was transferred to a sterile culture tube. After 1.5 h of recovery at 37 ºC, 200 RPM the cells were placed on LB + ampicillin plates. Successful clones were grown in liquid media is described in "Expression of pMIH90 in *E. coli* to produce violacein".

Extraction, detection and quantification of violacein

Cell pellets from production experiments in *E. coli, S. typhimurium* and *S. cerevisiae* were resuspended in 0.2 mL of Lysis buffer (100 mM sodium phosphate, 300 mM NaCl, 10% (v/v) glycerol, 1 mg mL^{-1} lysozyme, 1 µg mL^{-1} pepstatin A, 1 µg mL^{-1} leupeptin, pH adjusted to 8.0) per 1 mL of culture. The resulting cellular debris was sonicated on ice (3 times, 30 s on, 30 s off). The cellular suspension was then placed in a separatory funnel and extracted with 3x ethyl acetate. The organic phase was discarded and the aqueous phase was dried with magnesium sulphate, filtered of solid residues and evaporated with a rotary evaporator. The resulting crude product was resuspended in MeOH prior to analysis by HPLC. The samples were placed in a 1260 Infinity

High Performance Liquid Chomatography system (Agilent Technologies). The column conditions were as follows: Prontosil C18, 5 µm, 125 mm x 4 mm, flow rate: 0.3 mL min^{-1}, λ= 585 nm. Mobile phase A: 99.95% water, 0.05% formic acid. Mobile phase B: 99.95% acetonitrile, 5.0% water, 0.05% formic acid. Gradient: Linear gradient of 0% B to 100 % B over 25 min, followed by a linear gradient from 100% B to 0 % B over 5 min until completion.

Transfection and expression of pFM110 in mammalian cell lines

pFM110, a plasmid containing the violacein pathway from *C. violaceum* expressed as a polyprotein in pCDNA3.1, was cloned by Dr. Fern Mcsorley. Transfection of mammalian cells with pFM110 was performed by Nicholas Boileau. Lipofectamine 2000 (Thermo Fisher Scientific) was used as the transfection reagent in all transfections detailed herein. First, an ideal ratio of pFM110 pDNA (µg):lipofectamine (µL) was determined by testing; 2.5:5, 2.5:10, 5:10, 5:15, with the best ratio determined to be 5:15. HEK293T cells were 80% confluent at the time of the transfection. They were transfected with 5 µg of pDNA and 15 µL of Lipofectamine. Cells were left to grow for 72 h at 30 °C and 5% CO_2 at which point the cells were lysed using RIPA+ buffer and samples were collected for western blotting.

Western blotting of mammalian cell lines expressing the violacein polyprotein.

The protein content of the cell lysates was determined using the DC assay kit (Bio-Rad). An equal amount of protein for each sample was mixed with 4X solubilizing buffer and water to a volume of 25 µL. These samples were boiled and loaded on TGX electrophoresis gels (Bio-Rad). Gels were run at 180 V until the loading front reached the end of the gel. The gel was then inserted into

the transfer apparatus with a 0.22 µm nitrocellulose membrane, and 15% MeOH transfer buffer. Transfer was for 1 h at 125 V in a cold room. The membrane was blocked with 5% skimmed milk in TBS-T for 30 mins at RT with agitation. The membrane was then rinsed 3x for 5 mins each with TBS-T and then probed with 1° α-FLAG antibody O/N at 4 °C with agitation. The membrane was then rinsed with TBS-T and probed for 1 h at RT with 2° α-mouse IgG antibody. The membrane was washed 3x before bands were detected with the Clarity Western ECL (Bio-Rad) solutions (per manufacturer's directions) added to the membrane for 5 mins. The membrane was then placed into the film cassette in a dark room. Film was exposed for intervals between 5 s and 30 mins, and developed until observable bands were generated.

MRM of violacein and intermediates from mammalian cells

This MS protocol was developed and performed by Carlos Canez at the Jeff Smith Lab, Carleton University. Solvent gradient for LCMS was A: 30% methanol, 70% water, 10 mM ammonium acetate and B: 100% methanol, 10 mM ammonium acetate. Crude violacein extracts were dissolved in 75% water and 25% methanol before injection. After separation, both positive and negative electron spray ionization (ESI) were tested for the most sensitive MRM transitions, with a positive ESI at 35 eV producing the most sensitive transitions. The limit of detection using this method was 4000 x dilution of the violacein crude.

Cloning and expression of pMIH91 in *S. cerevisae.*

The violacein polyprotein design was cloned from pMIH90 into a plasmid p426 via SpeI/PacI restriction sites following standard endonuclease cloning to produce pMIH91 ((Vio*ABCDE*, TEFI promoter, URA3, Amp^R). Transformation of pMIH91 into S. cerevisiae BY4741 was performed following the LiAc/SS carrier DNA/PEG method.[29] Growth experiments were carried out synthetic minimal media (Sigma Aldrich) containing 1.46 g L^{-1} yeast synthetic dropout supplement, 6.7 g L^{-1} yeast nitrogen base (w/o amino acids), 2% glucose (w/v) and supplemented with necessary amino acids (leucine 380 mg L^{-1}, tryptophan 76 mg L^{-1}). Experiments were performed in 250 mL shake flasks at 200 RPM, 30°C with medium comprising 10% of the total flask volume. Individual colonies were transferred to liquid media and pre-cultured for 14 hours. Inoculation of experimental cultures were grown to an OD_{660} of 0.6 prior to supplementing with 1 volume of YPD medium (YPD medium containing 1% (w/v) yeast extract, 2% peptone and 2% glucose).

Transcriptome analysis of violacein in *S. cerevisae*

20 mL cultures of *S. cerevisiae* transformed with pMIH91 were grown as described above. Cells were spun down by centrifugation at 4000 x g for 10 mins. RNA was extracted using RNeasy Mini Kit (Qiagen) following the manufacturer's recommended protocol with 1 modification. Samples were treated with RNAprotect (Qiagen) prior to elution of RNA. Once RNA was isolated, conversion to cDNA was performed in a sterile microcentrifuge tube by adding 0.5 µg RNA, 40 µM Oligo-dT primer, 10 mM dNTP and nuclease-free H_2O to a total volume of 16 µL. Samples

were heated for 5 minutes at 80°C and placed on ice. 1X RT buffer (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT), 10 U of RNAse inhibitor (NEB) and 200 U of Reverse Transcriptase (NEB) were was added to the microcentrifuge tube to a final volume of 20 µL and incubated at 42°C for 1 h, followed by an enzyme inactivation step at 90°C for 10 mins. Control reactions did not contain Reverse Transcriptase. cDNA was then amplified by TEV protease primers using standard PCR conditions.

4.3 References

(1) Yang, S.; Jia, N.; Li, M.; Wang, J. *Mol. Biol. Rep.* **2011**, *38* (1), 59–64.

(2) Ro, D.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C. Y.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D. *Nature* **2006**, *440* (April), 3–6.

(3) Rosano, G. L.; Ceccarelli, E. A. *Front. Microbiol.* **2014**, *5*, 1–17.

(4) Cui, W.; Han, L.; Suo, F.; Liu, Z.; Zhou, L.; Zhou, Z. *World J. Microbiol. Biotechnol.* **2018**, *34* (10), 0.

(5) Balbás, P.; Lorence, A. *Methods Mol. Biol.* **2004**, *824*, 329–347.

(6) Clancy, S. *Nat. Educ.* **2008**, *1* (1), 125.

(7) Bartenschlager, R.; Ahlborn-Laake, L.; Mous, J.; Jacobsen, H. *J. Virol.* **1994**, *68* (8), 5045–5055.

(8) Tian, Y. C.; Shih, D. S. *J. Virol.* **1986**, *57* (2), 547–551.

(9) Marcotrigiano, S. A. Y. and J. *Curr. Opin. Virol.* **2008**, *6* (9), 2166–2171.

(10) Hoshino, T.; Kondo, T.; Uchiyama, T.; Ogasawara, N. *Agric. Biol. Chem.* **1987**, *51* (3), 965–968.

(11) Federal, U.; Catarina, D. S. *Genet. Mol. Res.* **2004**, *3* (1), 85–91.

(12) Balibar, C. J.; Walsh, C. T. *Biochemistry* **2006**, *45* (51), 15444–15457.

(13) DEMOSS, R. D.; EVANS, N. R. *J. Bacteriol.* **1960**, *79* (1), 729–733.

(14) Ryan, K. S.; Balibar, C. J.; Turo, K. E.; Walsh, C. T.; Drennan, C. L. *J. Biol. Chem.* **2008**, *283* (10), 6467–6475.

(15) Fang, M. Y.; Zhang, C.; Yang, S.; Cui, J. Y.; Jiang, P. X.; Lou, K.; Wachi, M.; Xing, X. H.

Microb. Cell Fact. **2015**, *14* (1), 1–13.

(16) Mitchell, L. A.; Chuang, J.; Agmon, N.; Khunsriraksakul, C.; Phillips, N. A.; Cai, Y.; Truong, D. M.; Veerakumar, A.; Wang, Y.; Mayorga, M.; Blomquist, P.; Sadda, P.; Trueheart, J.; Boeke, J. D. *Nucleic Acids Res.* **2015**, *43* (13), 6620–6630.

(17) Kapust, R. B.; Waugh, D. S. *Protein Expr. Purif.* **2000**, *19* (2), 312–318.

(18) Van Den Berg, S.; Löfdahl, P. Å.; Härd, T.; Berglund, H. *J. Biotechnol.* **2006**, *121* (3), 291–298.

(19) Yi, L.; Gebhard, M. C.; Li, Q.; Taft, J. M.; Georgiou, G.; Iverson, B. L. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (18), 7229–7234.

(20) Carrington, J. C.; Dougherty, W. G. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (10), 3391–3395.

(21) Gibson, D. G.; Benders, G. A.; Andrews-Pfannkoch, C.; Denisova, E. A.; Baden-Tillson, H.; Zaveri, J.; Stockwell, T. B.; Brownley, A.; Thomas, D. W.; Algire, M. A.; Merryman, C.; Young, L.; Noskov, V. N.; Glass, J. I.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. *Science* **2008**, *319* (5867), 1215–1220.

(22) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison Iii, C. A.; Smith, H. O. *Nat. Methods* **2009**, *6* (5), 343–345.

(23) Walker, J. M. *Methods Mol. Biol.* **2016**, *1409*.

(24) Sargo, C. R.; Campani, G.; Silva, G. G.; Giordano, R. C.; Silva, A. J. Da; Zangirolami, T. C.; Lu, R. W.; Correia, D. M.; Lu, R. W. *Biotechnol. Prog.* **2015**, *31* (5), 1217–1225.

(25) Thomas, P.; T, T. G. S. *J. Pharmacol. Toxicol. Methods* **2005**, *51*, 187–200.

(26) Ou, Y.; Li, J.; Wang, J.; Chang, C.; Wu, C.; Chen, W.; Kuan, Y.; Liao, S.; Lu, H.; Chen, C. *Int. J. Mol. Sci.* **2019**, *20*, 2792–2808.

(27) E.G. Bligh, W. J. D. *Can. J. Biochem. Physiol.* **1959**, *37* (8), 911–917.

(28) Kitteringham, N. R.; Jenkins, R. E.; Lane, C. S.; Elliott, V. L.; Park, B. K. *J. Chromatogr. B* **2009**, *877*, 1229–1239.

(29) Gietz, R. D.; Schiestl, R. H. *Nat. Protoc.* **2008**, *2* (1), 31–35.

(30) Gietzt, R. D.; Schiestls, R. H.; Willems, A. R.; Woods, R. A. *Yeast* **1995**, *11*, 355–360.

(31) Shiao, Y. *BMC Biotechnol.* **2003**, *3*, 1–12.

(32) Duran, R. R. M. H. N. *Brazilian Hournal Med. Biol. Res.* **1989**, *22* (5), 569–577.

(33) Wlodawer, A.; Miller, M.; Jasklski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kentt, S. B. H. *Science (80-.).* **1989**, *245*.

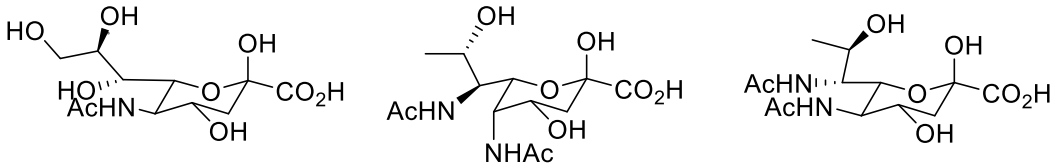
(34) Carrington, J.C *et al*, *Journal of Virology,* **1993**, 67, 12, 6995-7000

(35) Jackson, R.; Hellen, C.; Pestova, T. *Nat. Rev. Mol. Cell Biology,* **2010**, 11(2), 113–127.

Chapter Five: Concluding Statements and Future Directions

5.1 Summary: *In vivo* production of complex carbohydrates

The primary objective of this thesis was to harness microbial biosynthetic pathways to access a group of complex, bacterial nine carbon sialic acid analogues that are referred to as NulO's, which are extensively summarized in Chapter 1. Of particular interest to us were Leg5,7Ac$_2$ and Pse5,7Ac$_2$, two stereoisomers that to date have been difficult to produce by synthetic or chemoenzymatic methods (Figure 5.1). Previous successful implementation of a heterologous expression strategy to produce the eukaryotic and prokaryotic NulO sialic acid in *E. coli* by the Boddy group laid the groundwork for expanding the scope of that work and enabled the generation of additional nine carbon analogues.[1]



Sialic Acid Pse5,7Ac$_2$ Leg5,7Ac$_2$

Figure 5.1: Structure of bacterial nonulosonic acids.

Chapter 2 discusses the development of a cell-based production strategy for Leg5,7Ac$_2$ with excellent yields.[2] A *de novo* biosynthetic pathway from *S. cerevisiae, C. jejuni* and *L. pneumophila* produced over 100 mg L^{-1} of Leg5,7Ac$_2$ in an *E. coli* strain lacking the catabolic pathways capable

of NulO degradation. An essential aspect of this design was the overexpression of UDP-GlcNAc, the starting substrate of the *de novo* biosynthetic pathway of Leg5,7Ac$_2$. A purification strategy resulted in highly pure Leg5,7Ac$_2$ that was further converted to produce biologically relevant analogues, such as the activated CMP-Leg5,7Ac$_2$. This sugar was obtained by enzymatic modifications via a synthetase and is currently being tested as a novel antibiotic against *N. gonorrhoeae*. Leg5,7Ac$_2$ was also chemically activated as the thioglycoside, providing an intermediate for designing complex glycoconjugates. Finally, large scale production using a bioreactor generated over 5 g L^{-1}, thus satisfying our initial goal of providing a reliable and relatively cheap method to obtain sufficient quantities of this complex carbohydrate.

In chapter 3 we focused our efforts to produce Pse5,7Ac$_2$, the naturally occurring bacterial stereoisomer of Leg5,7Ac$_2$.[3] We utilized the framework that provided good yields for the related NulO's, sialic acid and Leg5,7Ac$_2$ and assumed due to their stereochemical similarities, high yields would also be obtained. A *de novo* biosynthetic pathway constructed from the genes of *H. pylori* and *C. jejuni* was initially tested using the microbial production strategy, yet titers >20 mg L^{-1} of Pse5,7Ac$_2$ were not achieved. A number of modifications including the substitution of genes from various organisms, modification of 6x-Histidine tags and codon optimizations did not result in improved yields. The key distinction that results in a stereochemical difference between Leg5,7Ac$_2$ and Pse5,7Ac$_2$ is the initial dehydration of the latter by PseB also involves an additional epimerization. Screening of multiple dehydratase/epimerase identify a stable pair that result in improved titers will be the next immediate step. Methods to decrease screening times using cell-free protein synthesis will be explored.[4]

5.2 Summary: *De novo* polyprotein design for the heterologous expression of natural products

In chapter 4 we proposed a novel method for rapidly moving natural product biosynthesis across diverse hosts. We took inspiration from the way viruses package and processes their polyproteins[5] and adopted a similar strategy by attempting to produce violacein, a dark purple pigment with anti-tumour properties, in a number of heterologous hosts. A key feature of our polyprotein design was a spacer sequence at the N-terminus of each gene that contained a flexible linker, a TEV protease recognition sequence and a FLAG tag for protein detection. Polyprotein cleavage was controlled by an N-terminal TEV protease.[6]

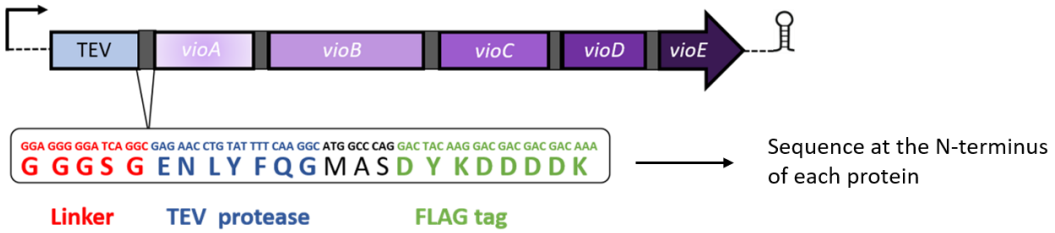


Figure 5.2: *De novo* polyprotein design for the heterologous production of natural products inspired by viral polyprotein processing. Key features include linker (red), TEV protease recognition sequence (blue) and FLAG tag (green).

Violacein production by the *de novo* polyprotein design was successfully observed in prokaryotic hosts including *E. coli* and *S. typhimurium*. The flexibility of this system was demonstrated by utilizing different promoters, linkers and nucleotides for violacein genes. When production was moved to eukaryotic organisms, violacein production was no longer observed. Incomplete polyprotein cleavage is observed in mammalian cell lines, although a dideoxy

intermediate of violacein is detected by MS. Production of violacein in *S. cerevisae* is also faced with a lack of a purple pigment. Protein production is not observed in yeast, but transcriptome analysis confirms production of mRNA. Future work for this project will focus on extensively optimizing production of violacein in eukaryotic hosts. This will then be followed up by demonstrating the flexibility of this system by expressing natural products with large biosynthetic pathways to highlight that pathway length is not a limiting factor when using this sytem.

5.3 Future applications for the *in vivo* production of various carbohydrates.

In addition to the NulO's described above, we will seek to generate additional sugars and generate structural diversity in two ways. First, harnessing known biosynthetic pathways for the production of related nine carbon analogues such as acinetaminic acid and 8-epi-Leg5,7Ac$_2$ will be explored.[7] Secondly, utilizing modified starting substrates to produce azido derivatives of NulO's will be tested. We will attempt to produce 5-azidoacetyl-Leg5,7Ac$_2$ and 7-azidoacetyl-Leg5,7Ac$_2$ (Figure 5.3). The synthesis of the starting substrates to be used for this work (2-azidoacetyl-AltNac or 4-azidoacetyl-AltNAc) have been established for Pse5,7Ac$_2$, and will be synthesized for Leg5,7Ac$_2$ as well.[8] Protein engineering to enable the conversion of substrates with bulky functional groups by Agm1 and Uap1, the two yeast derived enzymes that are a key component of the microbial production strategy, may be required.[9] This would generate Leg5,7Ac$_2$ and Pse5,7Ac$_2$ analogues containing residues that would enable click chemistry and thus provide an alternative method to further the role of these complex sugars.

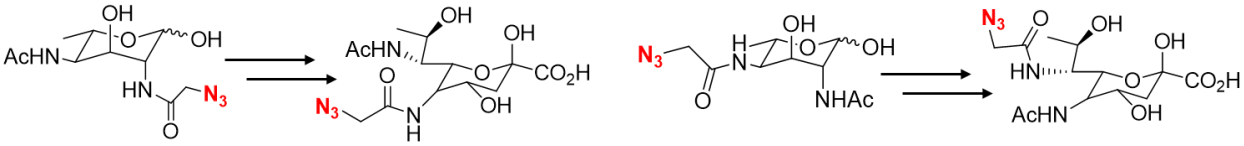


Figure 5.3: Azidoacetyl NulO derivatives to be produced *in vivo* by using azidoacetyl-AltNAc as a starting substrate.

To date, the microbial production strategy highlighted in this thesis has been limited to accessing complex nine carbon carbohydrates. Looking forward, we will seek to use this established strategy to generate other sugars that utilize UDP-GlcNAc as a starting substrate, thus widening the scope of the work described herein. Future work will include developing an *E. coli*-based system for the production of N-acetyl-D-galactosamine (GalNAc). GalNAc is currently produced for commercial uses by isolation from bovine trachea and avian parts. The animal origin of GalNAc is limiting its use as feedstock for pharmaceutical applications, particularly in the development of synthetic cancer vaccines. We will look to develop an *E. coli*-based, non-animal origin GalNAc manufacturing method which would provide a key reagent and alleviate a major GalNAc supply problem. An important asset for this work is BRL04, a modified *E. coli* strain that has been genetically engineered to lack *nagA*, a gene involved in the C-2 deacetylation of UDP-GlcNAc, the starting substrate for the biosynthesis of GalNAc. BRL04 should therefore generate a high intracellular concentration of GlcNAc, thus improving GalNAc production. Additionally, amino sugars such as GalNAc can act as excellent carbon sources for *E. coli* therefore catabolism of GalNAc must be prevented. We hypothesize that deletion of this *E. coli* transporter will prevent

the re-uptake of GalNAc, therefore further strain optimization and engineering will be required for this project.

GalNAc biosynthesis has been characterized in *Pseudomonas aeruginosa*.[10] It involves the epimerization of the sugar moiety of UDP-GlcNAc at the C-4 position by *WbpP*. To convert the UDP-GlcNAc into UDP-GalNAc, *wbpP*, encoding the UDP-N-acetylglucosamine 4-epimerase, will be cloned and expressed *in vivo*. To enable over-production of GalNAc from UDP-GalNAc, we will engineer a novel GalNAc hydrolase by designing chimeric proteins that can both recognize and hydrolyze UDP-GalNAc. Detection and quantification of GalNAc is essential for development of an over-production strain. Differentiating GalNAc from GlcNAc, its stereoisomer, presents a major challenge. Initial experiments show that GlcNAc and GalNAc can be differentiated by 400 MHz H-NMR, and an alternative will be to phosphorylate GlcNAc with NagK and ATP. NagK is highly selective for GlcNAc over GalNAc. By chemically modifying GlcNAc we can easily quantify GalNAc levels by LC-ESI-MS/MS.

5.4 Concluding remarks

The overarching goal of this thesis was to display the effectiveness of metabolic engineering and synthetic biology at generating difficult to access natural products, with a particular focus on complex carbohydrates. While chemical synthesis has an important role in the production of natural products and associated analogues, the expansion of bioinformatics, microbial engineering strategies and advanced cloning methods will ensure synthetic biology is an important tool to access novel, difficult to make molecules.

5.5 References

(1) Lundgren, B. R.; Boddy, C. N. *Org. Biomol. Chem.* **2007**, *5* (12), 1903.

(2) Hassan, M. I.; Lundgren, B. R.; Chaumun, M.; Whitfield, D. M.; Clark, B.; Schoenhofen, I. C.; Boddy, C. N. *Angew. Chemie Int. Ed.* **2016**.

(3) Goon, S.; Kelly, J. F.; Logan, S. M.; Ewing, C. P.; Guerry, P. *Mol. Microbiol.* **2003**, *50*, 659–671.

(4) Khambhati, K.; Bhattacharjee, G.; Gohil, N.; Braddick, D.; Kulkarni, V.; Singh, V. *Front. Bioeng. Biotechnol.* **2019**, *7* (October), 1–16.

(5) Marcotrigiano, S. A. Y. and J. *Curr. Opin. Virol.* **2008**, *6* (9), 2166–2171.

(6) Carrington, J. C.; Dougherty, W. G. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (10), 3391–3395.

(7) Kenyon, J. J.; Notaro, A.; Hsu, L. Y.; Castro, C. De; Hall, R. M. *Sci. Rep.* **2017**, *7* (11357), 6–11.

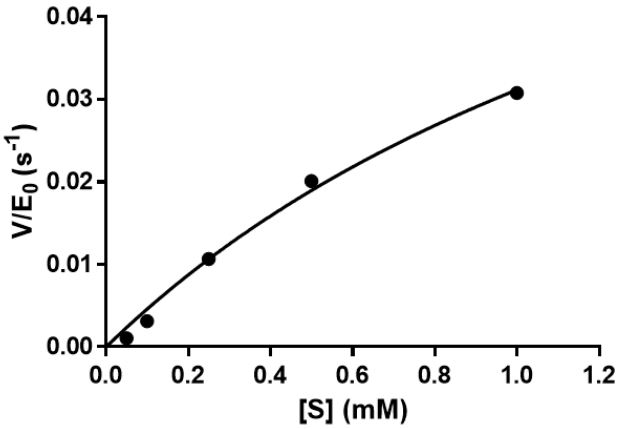
(8) Andolina, G.; Wei, R.; Liu, H.; Zhang, Q.; Yang, X.; Cao, H.; Chen, S.; Yan, A.; Li, X. D.; Li, X. *ACS Chem. Biol.* **2018**, *13* (10), 3030–3037.

(9) Horsman, M. E.; Lundgren, B. R.; Boddy, C. N. *Chem. Eng. Commun.* **2016**, *203* (10), 1326–1335.

(10) Creuzenet, C.; Belanger, M.; Wakarchuk, W. W.; Lam, J. S. *J. Biol. Chem.* **2000**, *275* (25), 19060–19067.

Appendix I: Supplemental Data

Chapter 2: Total biosynthesis of Legionaminic acid, a bacterial sialic acid analogue



	Value	Error
R^2	0.9817	N/A
K_{cat}	0.52	0.10
K_M	1.80	0.51

Figure S2.1. Initial reaction rates (s^{-1}) plotted against substrate concentration (mM) for the degradation of legionaminic acid by NanA. The data were fit to the Michaelis-Menten model using Graphpad Prism 6.

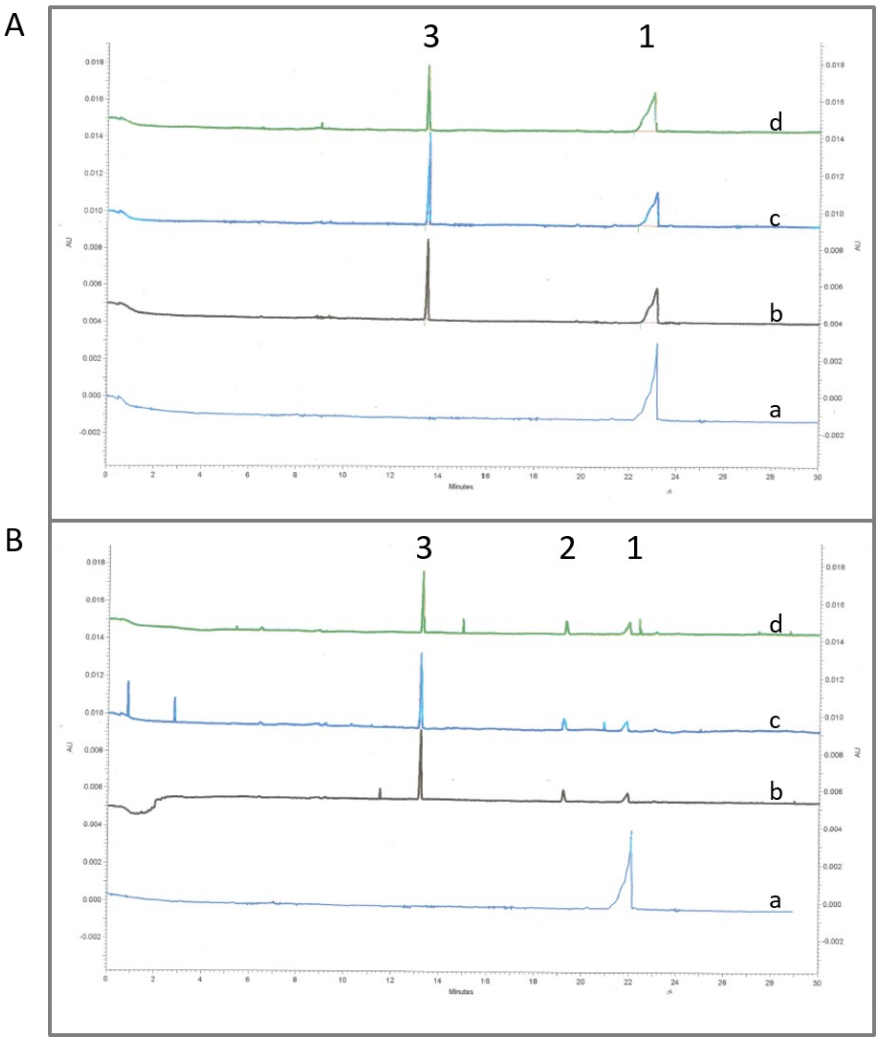
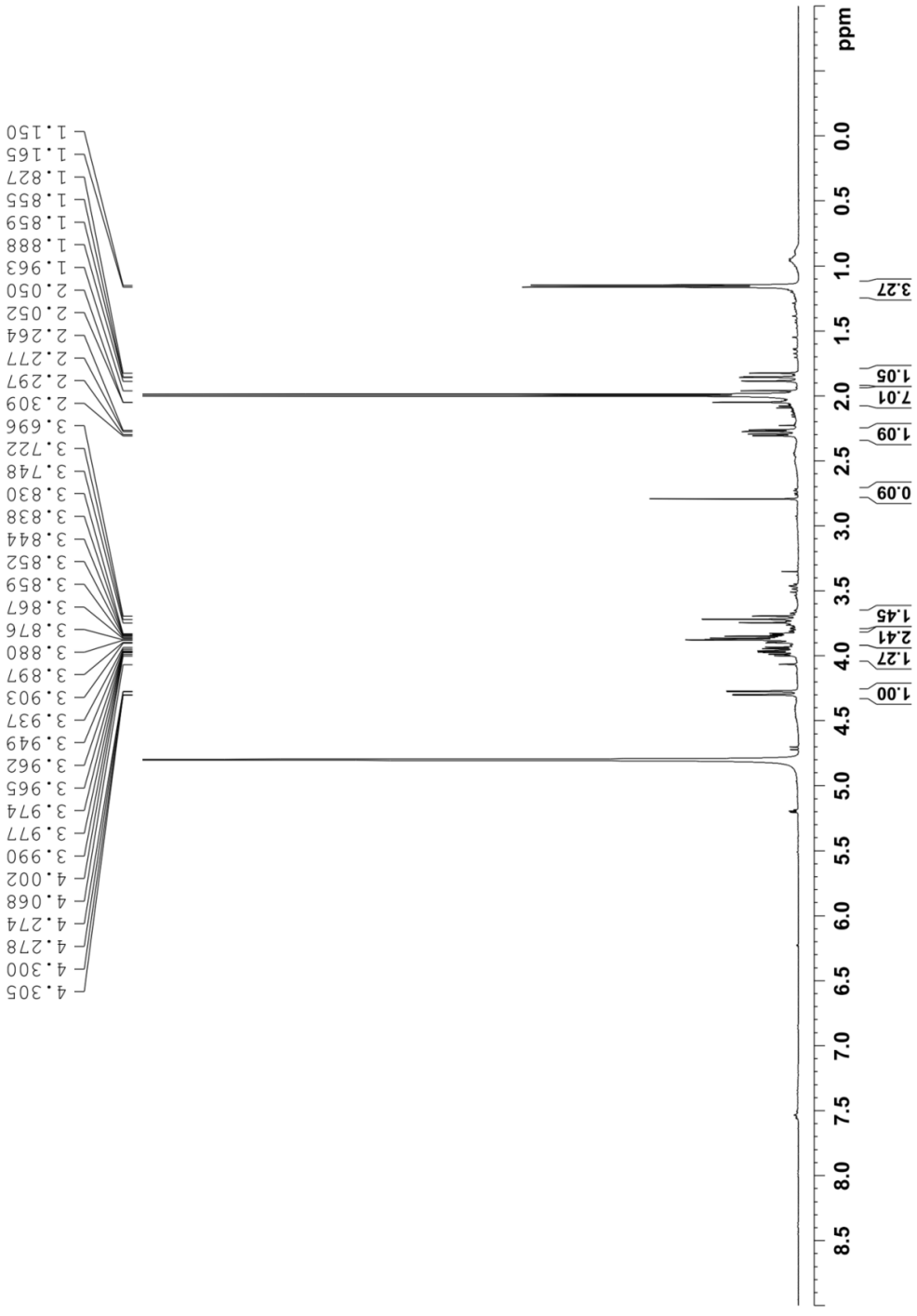


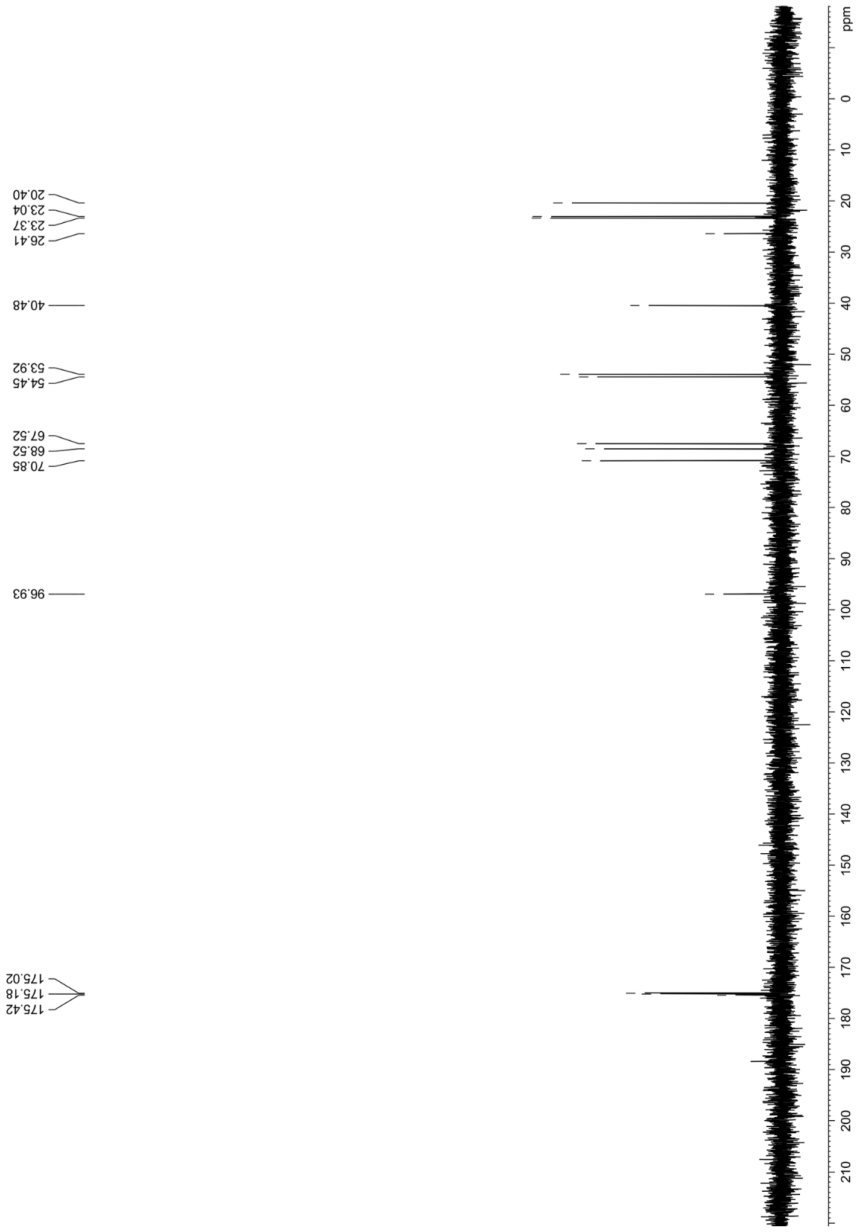
Figure S2.2. Near-complete enzymatic conversion of Leg5,7Ac$_2$ to CMP-Leg5,7Ac$_2$: conversion efficiency is similar for semi-pure and pure preparations of Leg5,7Ac$_2$. CMP-activation was performed as described in Methods section using pure (>95% purity; trace b), and semi-pure (70-85% purity; traces c and d) Leg5,7Ac$_2$ preparations for either 3.5 h (panel A) or overnight (panel B). Trace a in both panels is a non-enzyme control. The positions of CTP (1), CMP (2), and CMP-Leg5,7Ac$_2$ (3) are labeled. To note, the small accumulation of CMP found in panel B is typical of overnight reactions and is derived from CMP-Leg5,7Ac$_2$ material that has become labile. Samples were analyzed by capillary electrophoresis as described in reference 10.

S.2.3 Selected NMR Spectra

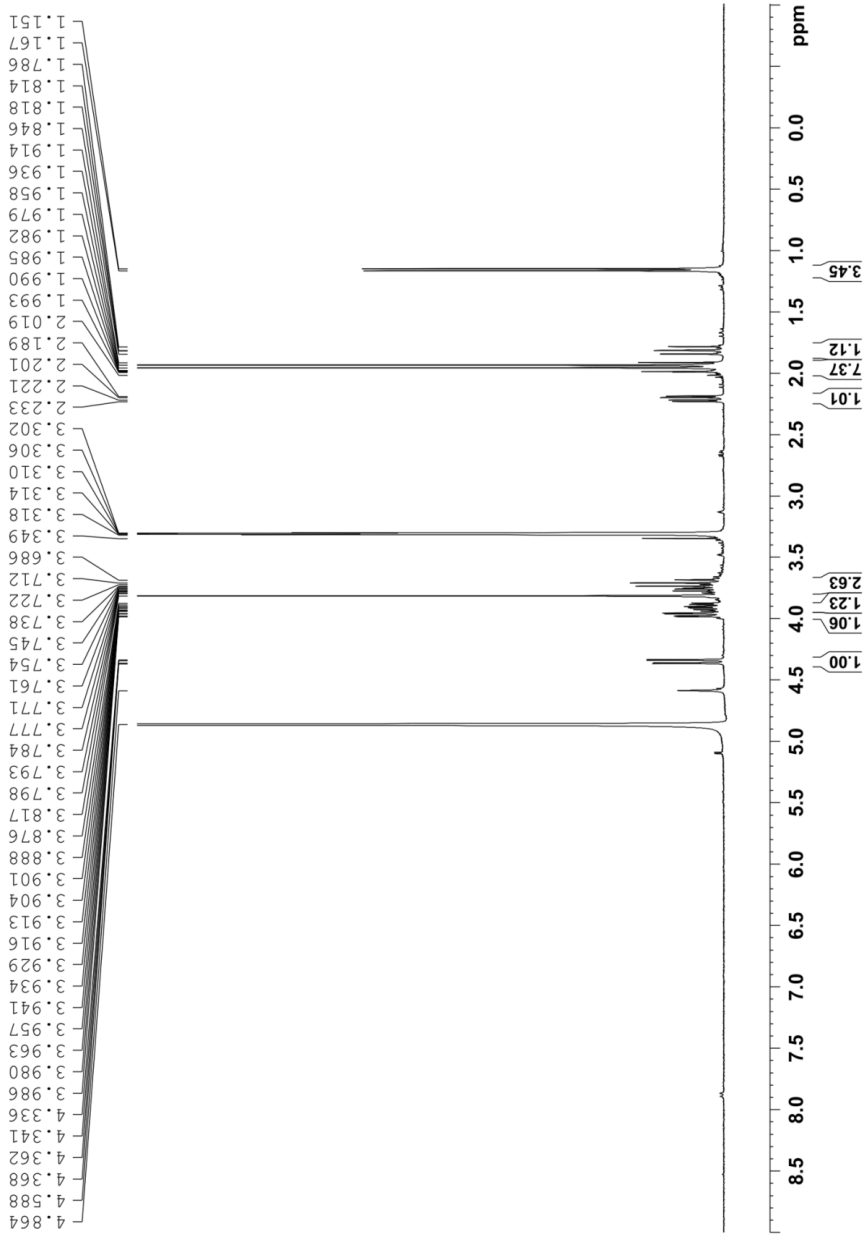
[1]H-NMR of 5,7-diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulopyranosiduronate, **1**



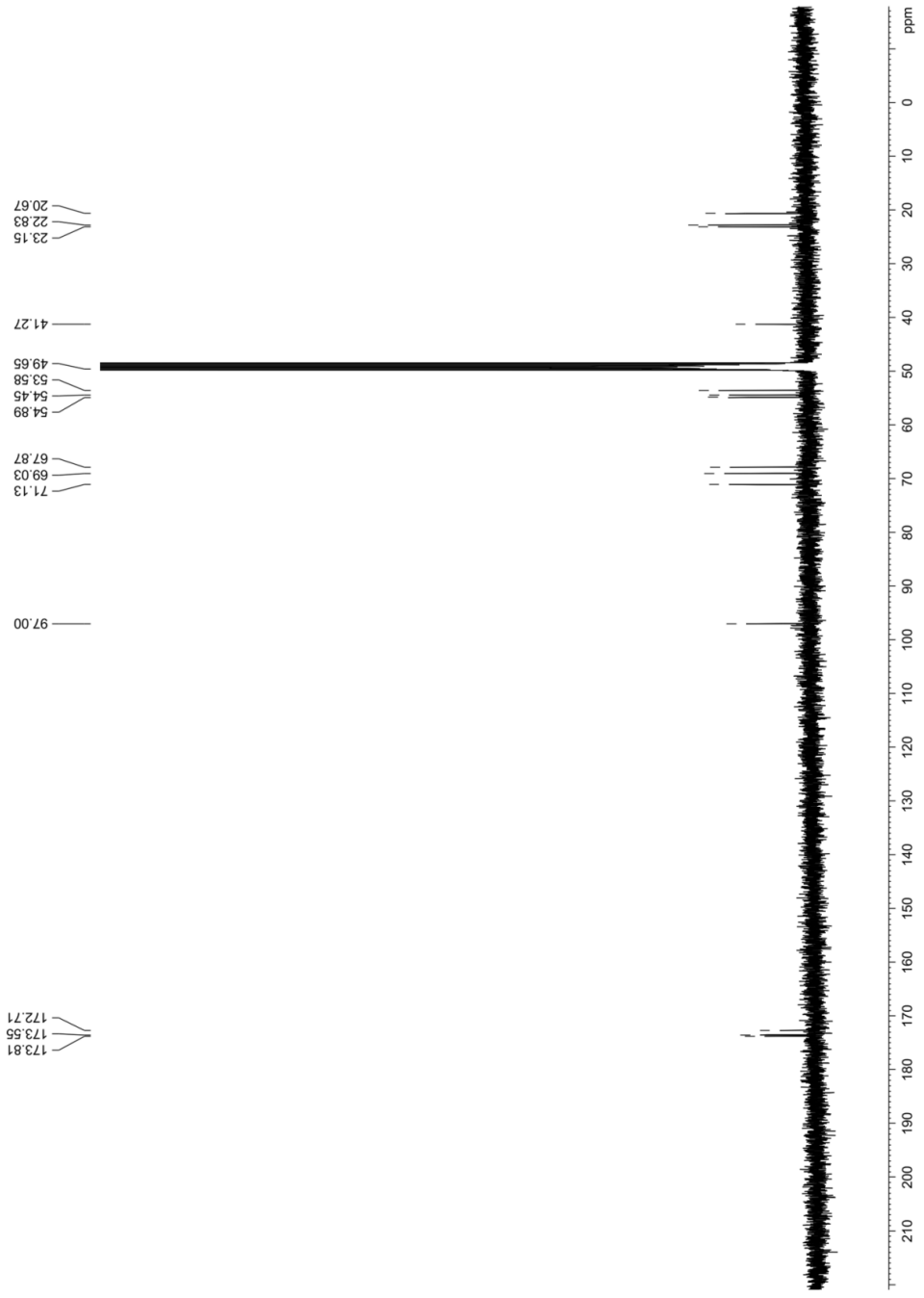
^{13}C-NMR of 5,7-diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulopyranosiduronate, **1**



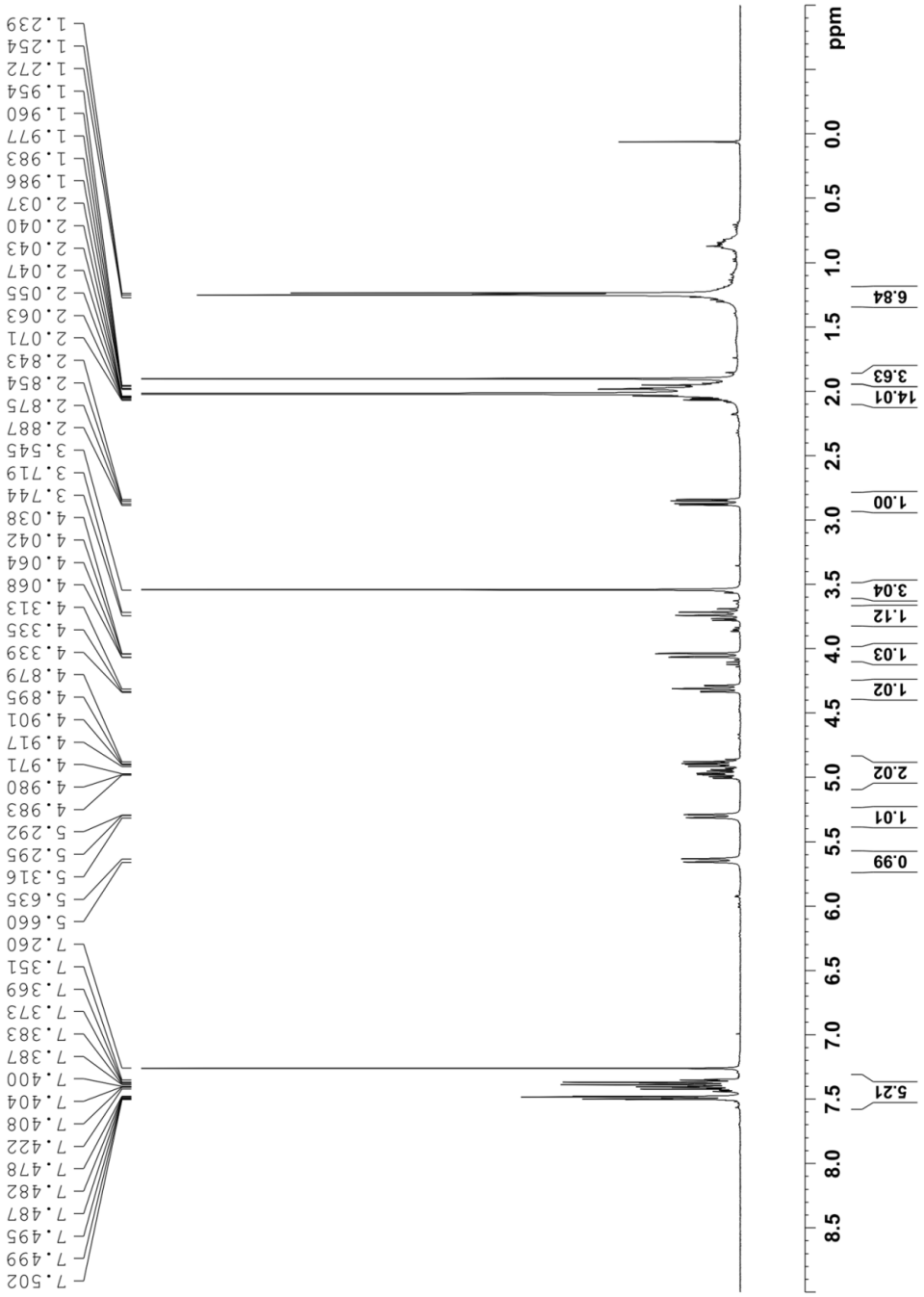
^{1}H-NMR of methyl-5,7-diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulopyranosiduronate, **15**



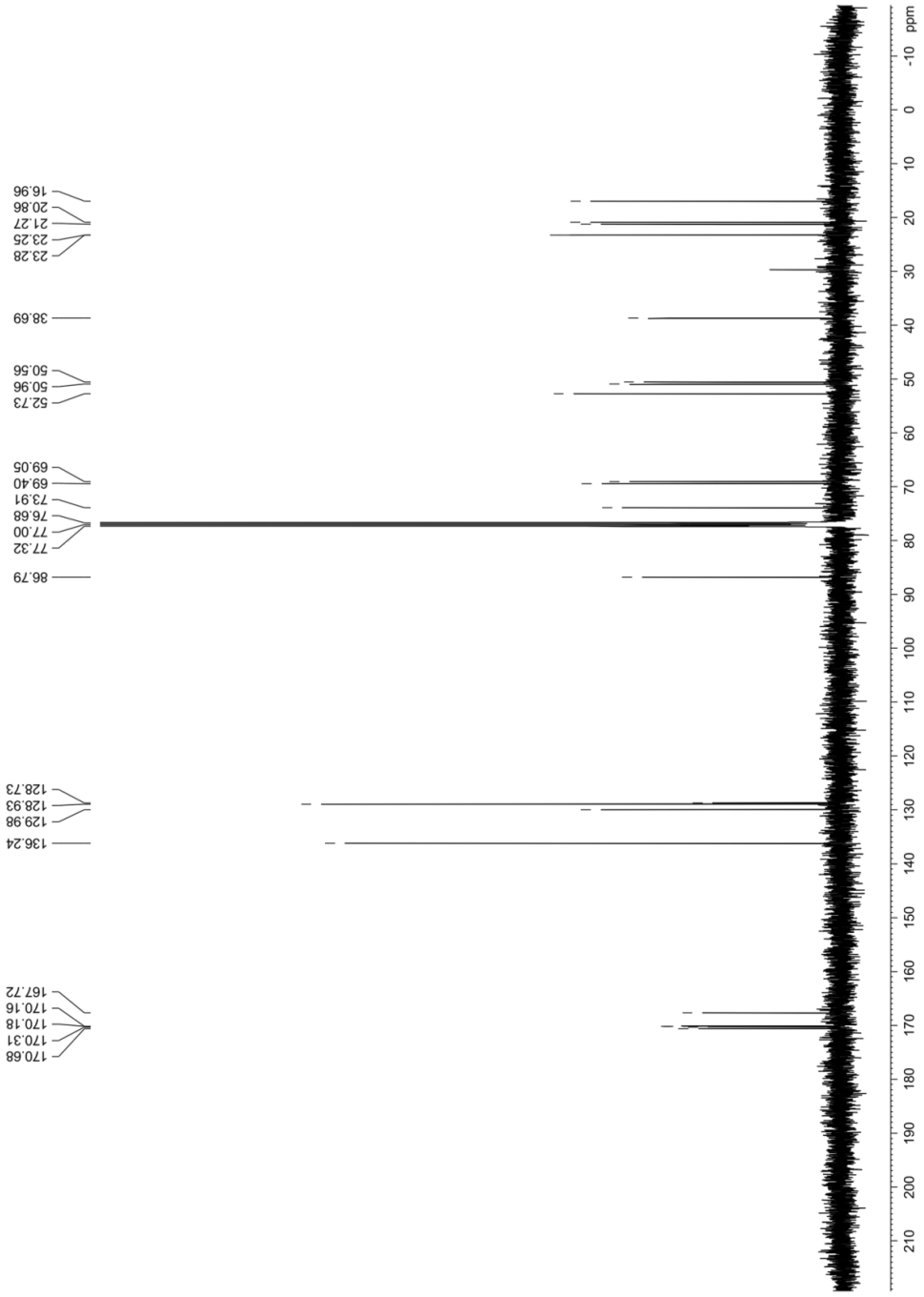
^{13}C-NMR of methyl-5,7-diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-D-*galacto*-nonulopyranosiduronate, **15**



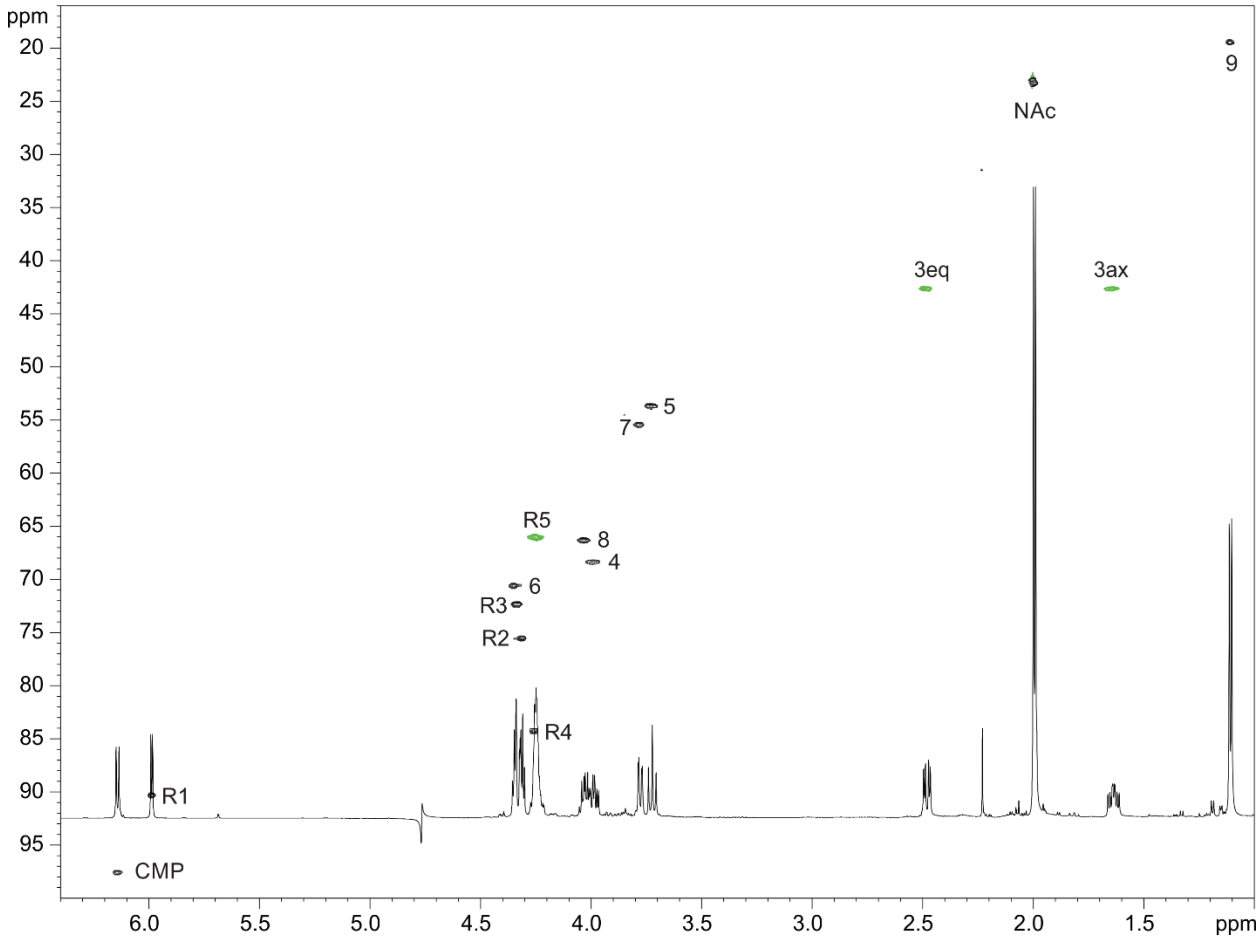
[1]H-NMR of Phenyl methyl (5,7-diacetamido 4,8-di-*O*-acetyl-3,5,7,9-tetradeoxy-D-*glycero*-α-D-*galacto*-2-thio-nonulopyranosyl)onate, **16**



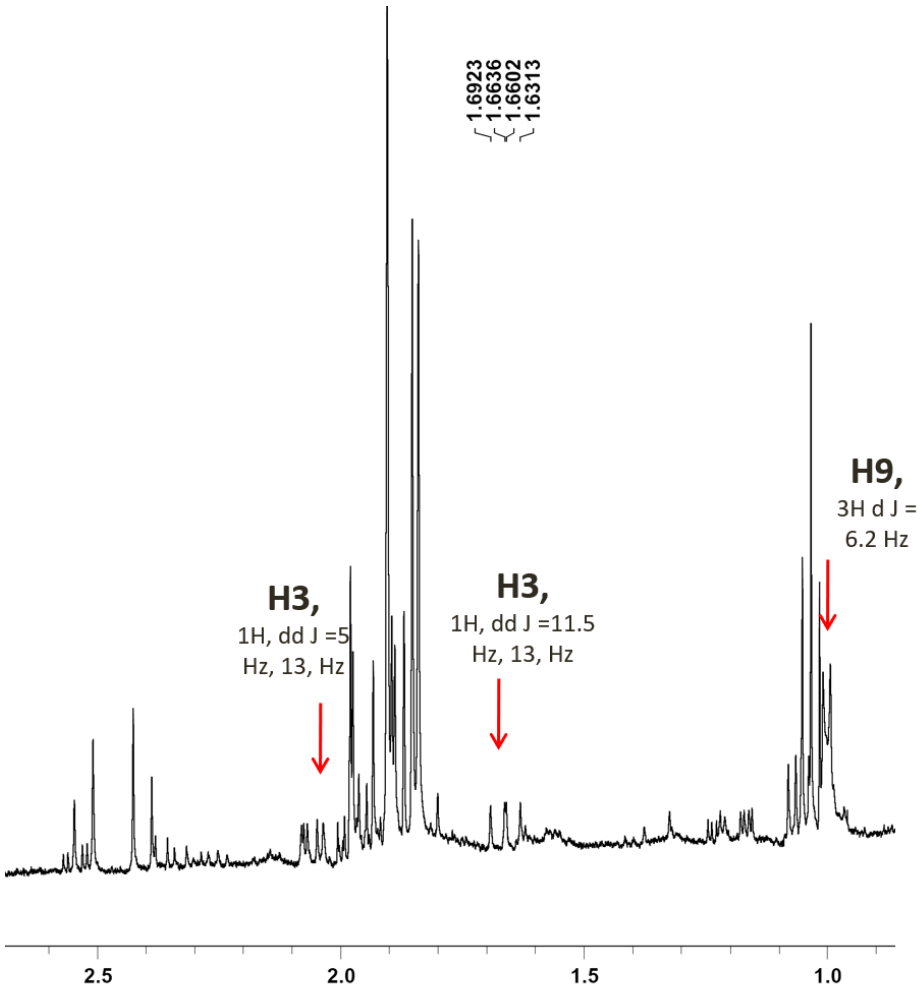
^{13}C-NMR of Phenyl methyl (5,7-diacetamido 4,8-di-*O*-acetyl-3,5,7,9-tetradeoxy-D-*glycero*-α-D-*galacto*-2-thio-nonulopyranosyl)onate, **16**



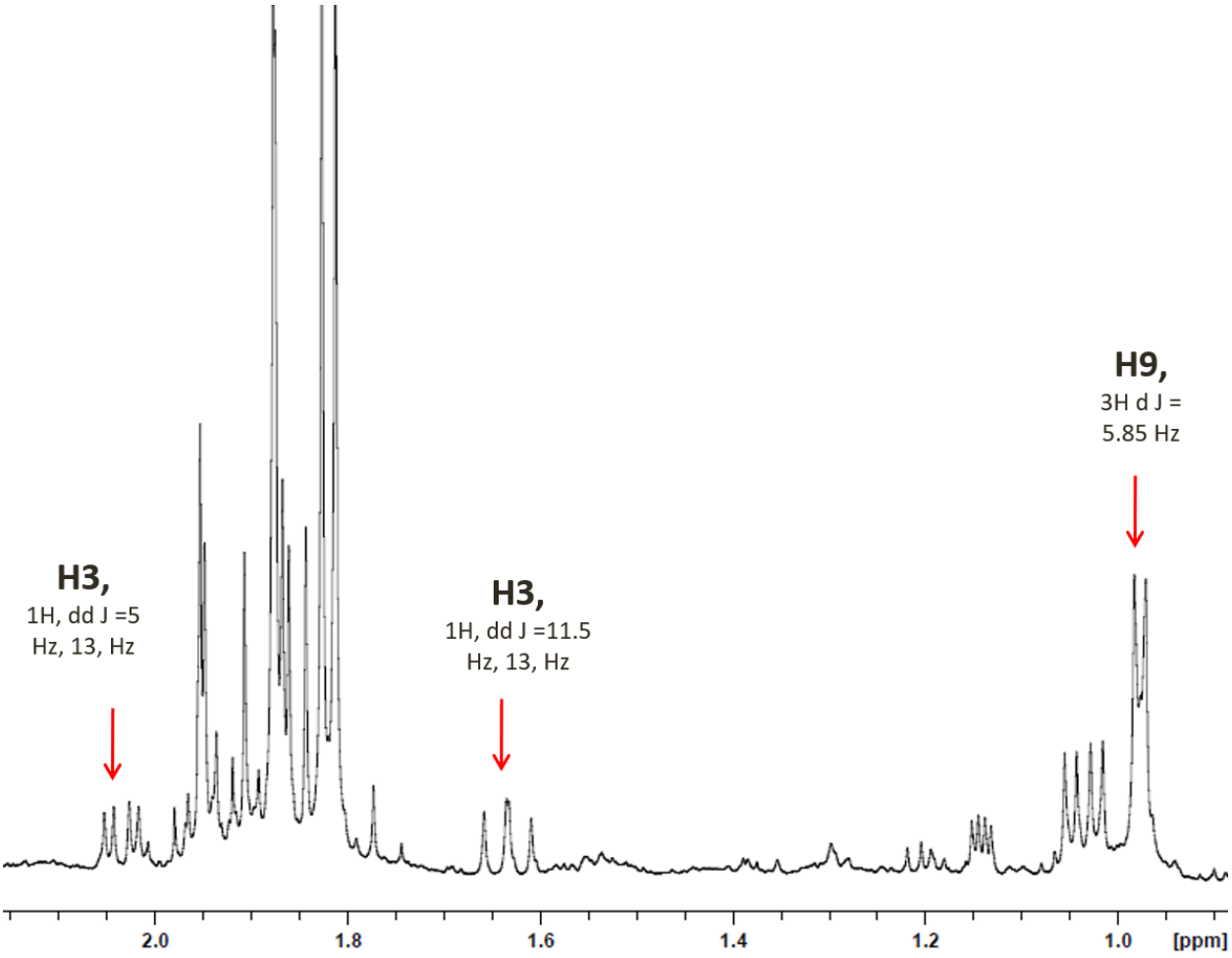
Portion of the ^{1}H-NMR of Disodium 5'-phosphoryl -cytidyl (5,7-diacetamido-3,5,7,9-tetradeoxy-D-*glycero*-α-D-*galacto* nonulopyranosyl)onate with ^{13}C-^{1}H-HSQC spectrum superimposed



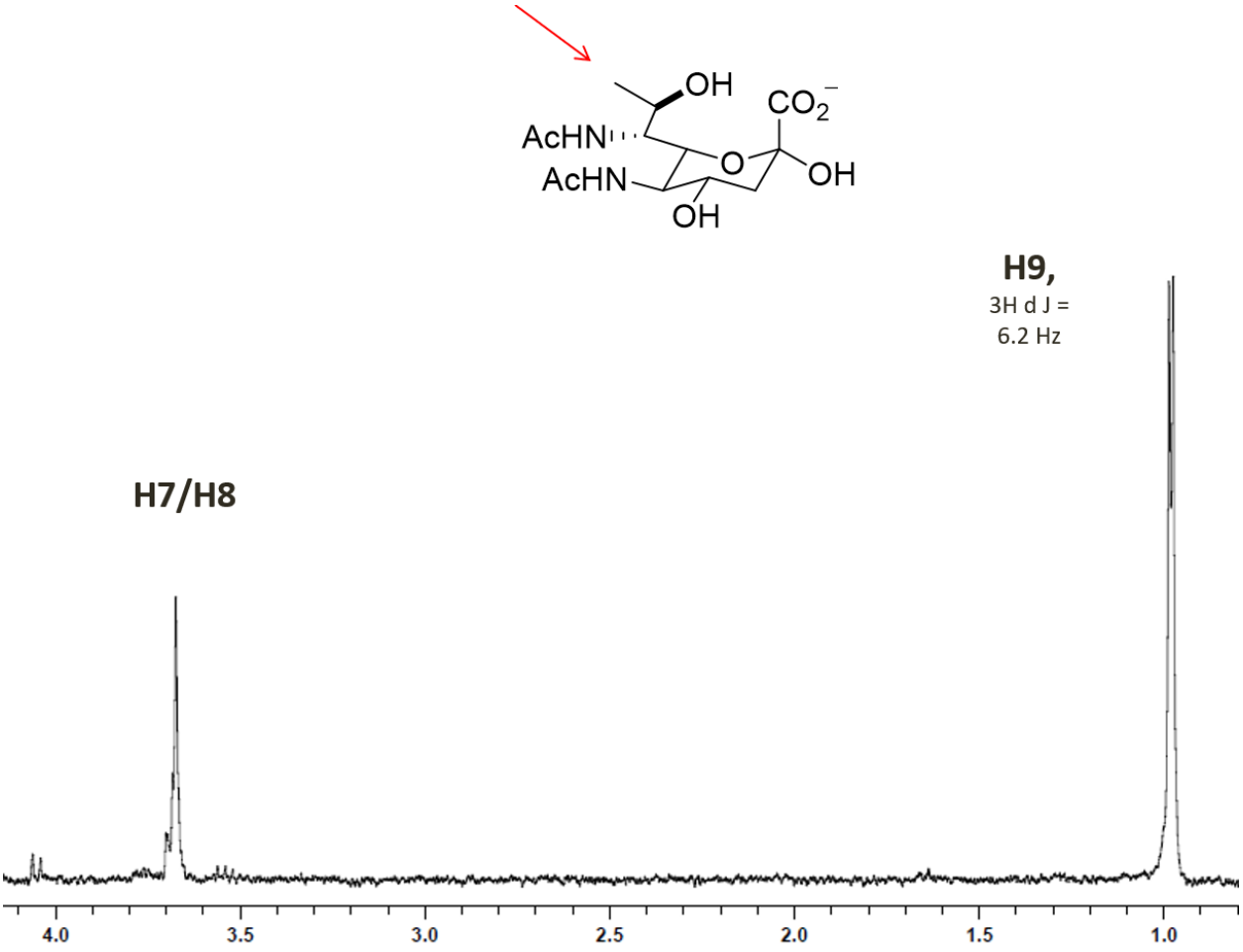
^{1}H-NMR of 0.5 mL clarified Leg5,7Ac$_2$ production broth at 162hr with 0.5 mL D_2O



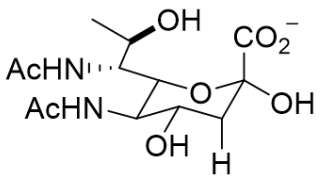
^{1}H-NMR of 40mL of Leg5,7Ac$_2$ production broth after lyophilization and extraction with methanol. 0.5 mL of D_2O was added to evaporated product before analysis.

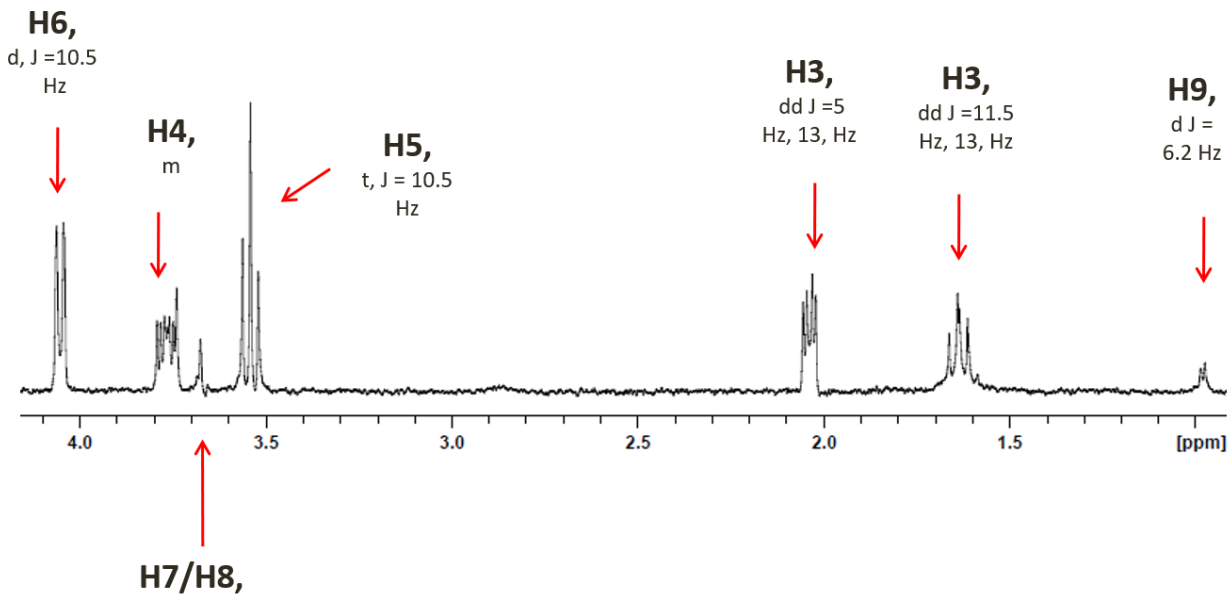


^{1}H-NMR of 1-D TOCSY analysis of Leg5,7Ac$_2$ directly from culture broth selectively irradiated at H9



^{1}H-NMR of 1-D TOCSY analysis of Leg5,7Ac$_2$ directly from culture broth selectively irradiated at H3 Axial





S2.4: Codon Optimized Genes Sequences Used in pMIH37

>PglF, *Campylobacter jejuni.*

GACGCATATGGGTAGCAGCCATCATCATCACCATCATGGTAGCATGCTGGTTGATTT
TAAACCGAGCCGTATGAAAGAAGAAGAAACCCCTTGTATTGTTGTTGGTGCAACCA
GCAAAGCACTGCATCTGCTGAAAGGTGCAAAAGAAGGTAGCCTGGGTCTGTTTCCG
GTTGGTGTTGTTGATGCACGTAAAGAACTGATTGGCACCTATTGCGATAAATTCATC
GTGGAAGAAAAAGAAAAAATCAAAAGCTACGTGGAACAGGGTGTTAAAACCGCAA
TTATTGCACTGCGTCTGGAACAAGAAGAACTGAAAAAACTGTTTGAAGAACTGGTG
GCCTATGGCATTTGTGATGTTAAAATCTTTAGCTTCACCCGCAATGAAGCACGCGAT
ATTAGCATTGAAGATCTGCTGGCACGTAAACCGAAAGATCTGGATGATAGCGCAGT
TGCAGCATTCCTGAAAGATAAAGTTGTTCTGGTTAGCGGTGCCGGTGGCACCATTGG
TAGCGAACTGTGTAAACAGTGTATCAAATTTGGTGCCAAACACCTGATTATGGTGGA
TCATAGCGAATATAACCTGTACAAAATCAACGATGACCTGAACCTGTATAAAGAGA
AAATTACCCCGATTCTGCTGAGCATTCTGGATAAACAGAGCCTGGATGAAGTTCTGA
AAACCTATAAACCGGAACTGATTCTGCATGCCGCAGCATATAAACATGTTCCGCTGT
GTGAACAGAATCCGCATAGTGCAGTTATTAACAATATTCTGGGCACCAAAATTCTGT
GCGATAGCGCCAAAGAAAACAAAGTTGCCAAATTCGTTATGATCAGCACCGATAAA
GCAGTTCGTCCGACCAATATTATGGGTTGTACCAAACGTGTTTGCGAACTGTATACC
CTGAGCATGAGTGATGAAAATTTTGAAGTTGCCTGCGTGCGTTTTGGTAATGTTCTG
GGTAGCTCAGGTAGCGTTATTCCGAAATTCAAAGCACAGATTGCCAATAATGAACC
GCTGACCCTGACCCATCCGGATATTGTTCGTTACTTTATGCTGGTAGCAGAAGCAGT
TCAGCTGGTTCTGCAGGCAGGCGCAATTGCAAAAGGTGGTGAACTGTTTGTTCTGGA
TATGGGTAAACCGGTGAAAATTATCGATCTGGCCAAAAAAATGCTGCTGCTGAGTA
ATCGTAATGATCTGGAAATCAAAATCACCGGTCTGCGCAAAGGTGAAAAACTGTAC
GAGGAACTGCTGATCGATGAAAATGATGCAAAAACCCAGTATGAGAGCATCTTTGT
GGCGAAAAATGAAAAAGTTGATCTGGATTGGCTGAACAAAGAAATCGAAAACCTGC
AGATTTGCGAAGATATTAGCGAAGCACTGCTGAAAATTGTGCCGGAATTTAAACAT
AACAAAGAAGGCGTGTGAGAATTCGCAG

>PglE, *Campylobacter jejuni.*
GACGCATATGGGTAGCAGCCATCATCATCACCATCATGGTAGCATGCGTTTTTTTCT
GAGCCCTCCGCACATGGGTGGTAATGAACTGAAATATATCGAAGAAGTGTTTAAAA
GCAACTATATTGCACCGCTGGGCGAATTTGTTAATCGTTTTGAACAGAGCGTTAAAG
CCTATAGCAAAAGCGAAAATGCACTGGCACTGAATAGCGCAACCGCAGCACTGCAT
CTGGCCCTGCGTGTTGCCGGTGTTAAACAGGATGATATTGTTCTGGCAAGCAGCTTT
ACCTTTATTGCAAGCGTTGCACCGATTTGTTATCTGAAAGCAAAACCGGTGTTTATC
GATTGTGATGAAACCTATAACATCGACGTGGATCTGCTGAAACTGGCCATTAAAGA
ATGTGAAAAAAAACCGAAAGCCCTGATTCTGACCCATCTGTATGGTAATGCAGCAA
AAATGGATGAGATTGTGGAAATCTGCAAAGAAAACGAAATCGTGCTGATTGAAGAT
GCAGCAGAAGCACTGGGTAGCTTCTATAAAAACAAAGCCCTGGGCACCTTTGGTGA
ATTTGGTGCATATTCATATAACGGCAACAAAATCATTACAACCAGCGGTGGTGGTAT
GCTGATTGGCAAAAACAAAGAGAAAATCGAGAAAGCCCGTTTCTATAGCACCCAGG
CACGTGAAAATTGTCTGCATTATGAACATCTGGATTATGGCTATAACTATCGCCTGA
GCAATGTTCTGGGTGCAATTGGTGTTGCACAGATGGAAGTTCTGGAACAGCGTGTTC
TGAAAAAACGCGAAATCTATGAGTGGTATAAAGAATTTCTGGGCGAGTGCTTTAGCT
TTCTGGATGAACTGGAAAATAGCCGTAGCAATCGTTGGCTGAGCACCGCACTGATTG
ATTTTGATAAAAATGAGCTGAACAGCTGCCAGAAAGATATTAACATTAGCCAGAAA
AACATCACCCTGCATCCGAAAATTAGCAAACTGATCGAAGATCTGAAAAACGAGCA
GATTGAAACCCGTCCGCTGTGGAAAGCAATGCATGCACAAGAGGTTTTTAAAGGTG
CAAAAGCCTATCTGAATGGCAACAGCGAACTGTTTTTCCAGAAAGGTATTTGTCTGC
CGAGCGGCACCGCAATGAGCAAAGATGATGTTTATGAAATCTCCAAACTGATCCTG
AAAAGCATTAAAGCCTGAGAATTCGCAG

>pglD, *Campylobacter jejuni.*
GACGCATATGGCACGCACCGAAAAAATCTATATCTATGGTGCAAGCGGTCATGGTCT
GGTTTGTGAAGATGTTGCCAAAAACATGGGCTACAAAGAATGCATCTTTCTGGATGA
TTTTAAAGGCATGAAATTCGAAAGCACCCTGCCGAAATATGATTTCTTTATTGCCAT
TGGCAACAACGAGATCCGCAAAAAAATCTACCAGAAAATTTCCGAGAACGGCTTCA
AAATTGTGAACCTGATTCATAAAAGCGCACTGATTAGCCCGAGCGCAATTGTTGAAG
AAAATGCAGGTATTCTGATCATGCCGTATGTTGTGATTAATGCCAAAGCCAAAATTG
AAAAAGGCGTGATTCTGAATACCAGCAGCGTTATTGAACATGAATGCGTGATTGGT
GAATTTAGCCATGTTAGCGTTGGTGCAAAATGTGCCGGTAATGTGAAAATTGGCAAA
AATTGCTTTCTGGGCATTAATAGCTGTGTTCTGCCGAATCTGAGCCTGGCAGATGAT
AGCATTCTGGGTGGTGGTGCAACCCTGGTTAAAAATCAGGATGAAAAAGGTGTGTTT
GTTGGCGTTCCGGCAAAACGTATGCTGGAACATCATCATCACCATCATTGAGAATTC
GCAG

>LegG, *Legionella pneumophila.*
GACGCATATGGGTAGCAGCCATCATCATCATCATCACGGTAGCATTCGCAAAATTAT
CTATGTTACCGGCACCCGTGCAGATTATGGTCTGATGCGTGAAGTTCTGAAACGTCT
GCATCAGAGCGAAGATATTGATCTGAGCATTTGTGTTACCGGTATGCATCTGGATGC
ACTGTATGGTAATACCGTGAATGAAATTAAAGCCGACCAGTTTAGCATTTGCGGTAT
TATTCCGGTTGATCTGGCAAATGCACAGCATAGCAGCATGGCAAAAGCAATTGGTC
ATGAACTGCTGGGTTTTACCGAAGTTTTTGAAAGCGAAACACCGGATGTTGTTCTGC
TGCTGGGTGATCGTGGTGAAATGCTGGCAGCAGCAATTGCAGCCATTCATCTGAATA
TTCCGGTGGTTCATCTGCATGGTGGTGAACGTAGCGGCACCGTTGATGAAATGGTTC
GTCATGCAATTAGCAAACTGAGCCATTATCATTTTGTTGCAACCGAAGCAAGCAAAC
AGCGTCTGATTCGTATGGGTGAAAAAGAAGAAACCATTTTTCAGGTTGGTGCACCGG
GTCTGGATGAAATTATGCAGTATAAAACCAGCACCCGTGATGTGTTTAATCAGCGTT
ATGGTTTTGACCCGGACAAAAAAATCTGCCTGCTGATTTATCATCCGGTTGTTCAAG
AAGTGGACAGCATCAAAATTCAGTTTCAGAGCGTTATTCAGGCAGCACTGGCAACC
AACCTGCAGATCATTTGTCTGGAACCGAATAGCGATACCGGTGGCCATCTGATTCGC
GAAGTTATTCAAGAATATATCGATCACCCGGATGTGCGCATTATCAAACATCTGCAT
CGTCCGGAATTTATTGATTGTCTGGCCAATTCAGATGTGATGCTGGGTAATAGCAGC
AGCGGCATTATTGAAGCAGCATCCTTTAATCTGAATGTTGTGAATGTGGGTAGCCGT
CAGAATCTGCGTGAACGCAGCGATAATGTTATTGATGTTGATGTTACCTATGACGCC
ATTCTGACCGGTCTGCGCGAAGCACTGAATAAACCGAAAATCAAATACAGCAACTG
CTATGGTGATGGCAAAACCAGCGAACGTTGTTATCAGCTGCTGAAAACCATTCCGCT
GCATAGCCAGATCCTGAATAAATGTAATGCCTATTGAGAATTCGCAG

>LegI, *Legionella pneumophila*
GACGCATATGGGTAGCAGCCATCATCATCACCATCATGGTAGCCTGGAAGGTAGCA
ATCGTAAAATCAATGGTATTAAACCGCGTGGTAGCAGCATGACCTGTTTTATCATTG
CCGAAGCCGGTGTTAATCATAATGGTGATCTGCAGCTGGCAAAAGAACTGGTTTATG
CAGCAAAAGAAAGCGGTGCAGATGCAGTTAAATTTCAGACCTTTAAAGCAGATACC
CTGGTGAATAAAACCGTGGAAAAAGCCGAATACCAGAAAAATAACGCACCGGAAA
GCAGCACCCAGTATGAAATGCTGAAAGCACTGGAACTGAGCGAAGAAGATCATTAT
CTGCTGAGCGAACTGGCAAATAGCCTGGGTATTGAATTTATGAGCACCGGTTTTGAT
GAACAGAGCATCGATTTTCTGATTAGCCTGGGCGTGAAACGTCTGAAAATTCCGAGC
GGTGAAATTACCAATGTTCCGTATCTGCAGCATTGTGCAAGCAAAAAACTGCCGCTG
ATTATTAGTACCGGTATGTGCGATCTGCAAGAAGTTCGCGTTGCAATTGATACCGTT
AAACCGTATTATGGTAATAGCCTGAGCGATTATCTGGTTCTGCTGCATTGTACCAGC
AATTATCCGGCAAGCTATCAGGATGTTAATCTGAAAGCAATGCAGACCCTGGCAGA
TGAATTTCAGCTGCCGGTTGGTTATAGCGATCATACCCTGGGCATTCTGGTTCCGAC
CCTGGCCGTTGGTATGGGTGCATGTGTTATTGAAAAACACTTCACCATGGATAAAAG
CCTGCCTGGTCCGGATCATCTGGCAAGCATGGATCCGGAAGAAATGAAAAATCTGG
TTCAGAGCATTCGTGATGCAGAAACCGTTCTGGGTAGTGGTGAAAAAAAACCGAGC
GATAATGAACTGCCGATTCGTGCACTGGTTCGTCGTAGCATTACCCTGCGTCGTGAT
CTGGTTAAAGGTGCACAGATTAGCAAAGAAGATCTGATCCTGCTGCGTCCGGGTAC
AGGTATTGCACCGAGCGAAATTAGCAATATTGTTGGTAGCCGTCTGAGCATGAATCT
GAGCGCAGGCACCACCCTGCTGTGGGAACATATTGAAGCATGAGAATTCGCAG

Chapter 3: Heterologous expression of pseudaminic acid using a genetically engineered strain optimized for complex sugar production

```
Query  421  TTAATCAGGGAAGAATTTTATCAAGAGGCCAAAGAAAATAGAGaaaaaaaaaTATGATTT  480
            ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Sbjct  421  TTAATCAGGGAAGAATTTTATCAAGAGGCCAAAGAAAATAGAG-AAAAAAAATATGATTT  479
```

Figure S3.1: Basic Local Alignment Search Tool (BLAST) results showing the site of pseG mutation in pMIH05 (highlighted in red). An extra base is present 464 bp downstream of the start codon. Query: Mutant *pseG* from pMIH05. Sbjct: Wild type *pseG* from *C. jejuni*.

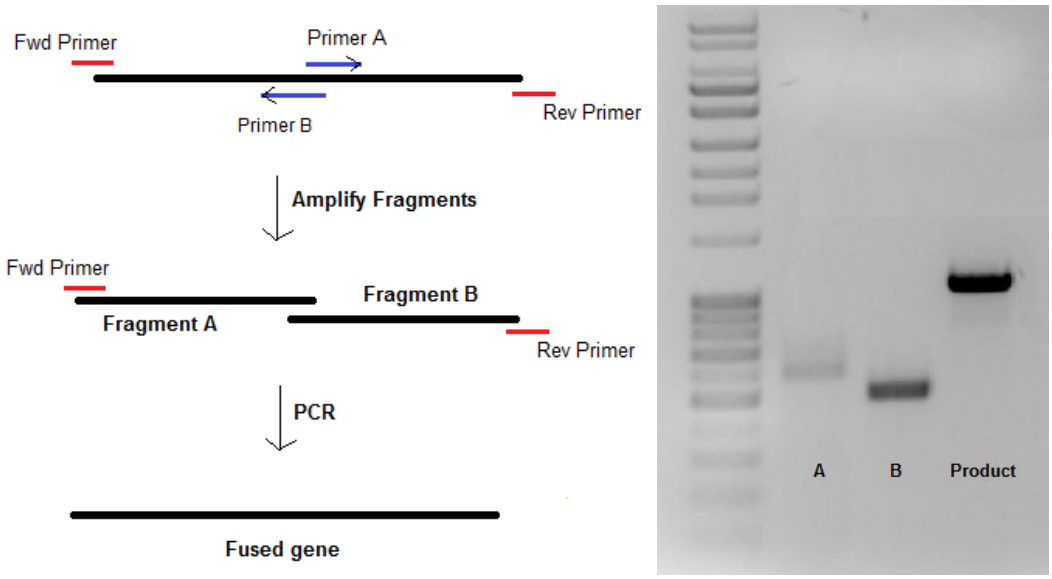


Figure S3.2: Left: Overview of SOE procedure, as described in the "Materials and Methods" section. **Right**: Image of purified PCR products used in the mutagenesis of pseG. Fragment A is from 1 to 480 bp, Fragment B is from 460 to 821 bp, and Product is the annealed gene that is obtained after PCR of the two fragments.

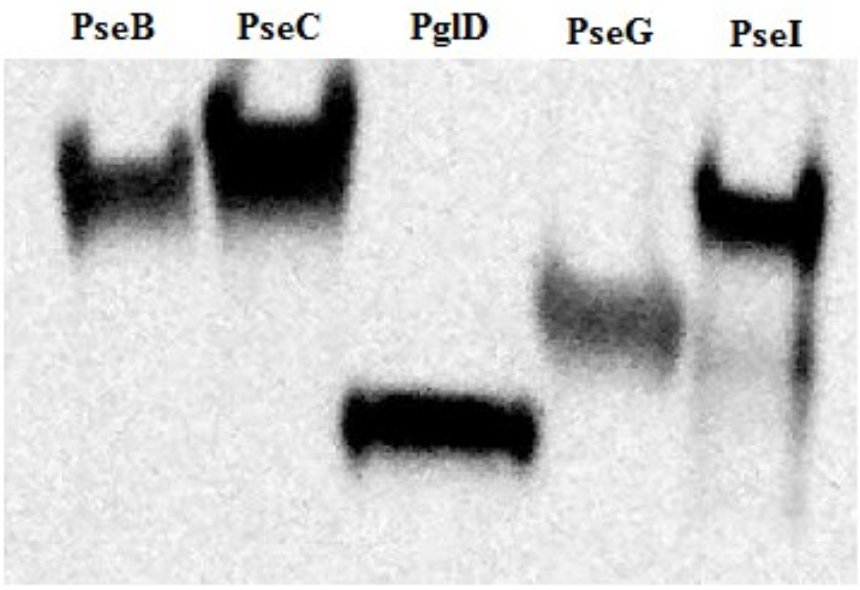


Figure S3.3 Expression of Pse5,7Ac$_2$ proteins in *E. coli* BL21 analyzed by western blotting. Expected size of proteins: pseB: 37 kDa, pseC: 42 kDa, pglD: 20 kDa, pseG: 33 kDa, pseI: 28 kDa

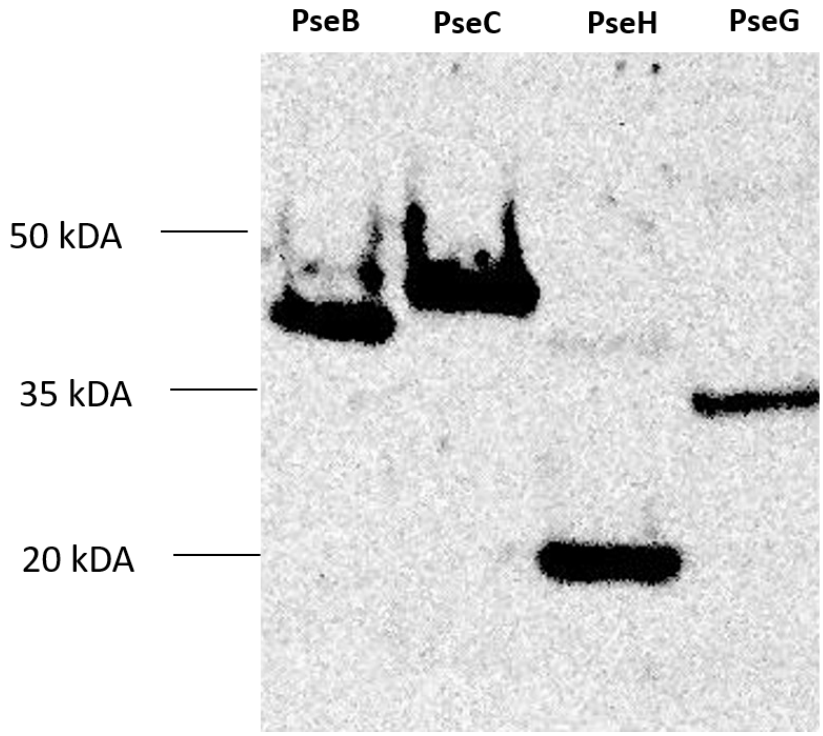


Figure S3.4 Expression of the codon optimized first four genes Pse5,7Ac$_2$ pathway from *H. pylori* in *E. coli* BL21. Proteins were assessed via western blotting using an Anti-6xHis HP-conjugated antibody. Expected size of proteins: pseB: 37 kDa, pseC: 42 kDa, pseH: 20 kDa, pseG: 33 kDa,

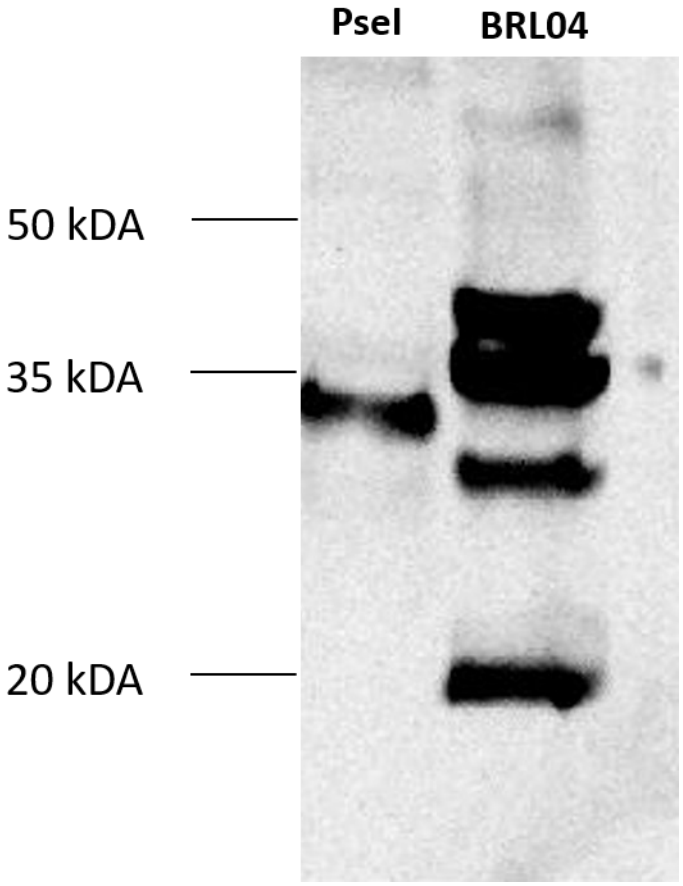


Figure S3.5 Expression of codon optimized PseI (28 kDa) in the *E. coli* strain BL21 and the fully cloned pseBCHGI pathway in the E. coli strain BRL04. Expression was analyzed via western blotting with an anti-6x-Histidine tag.

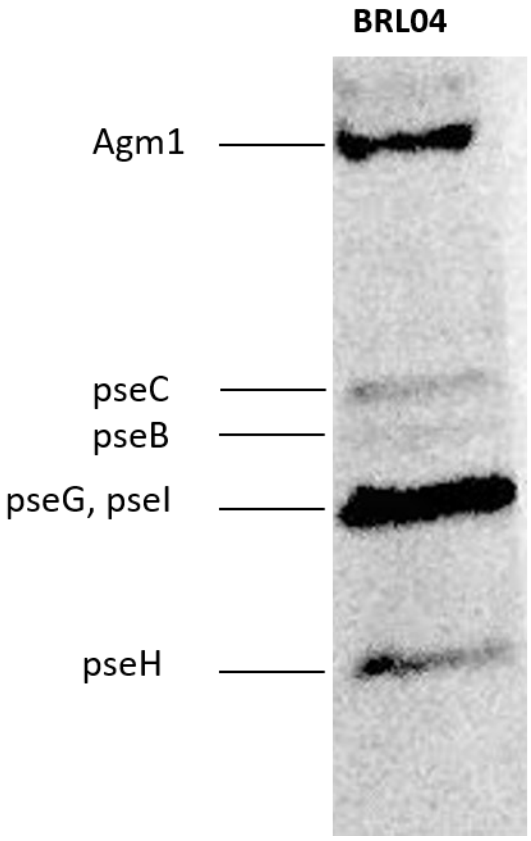


Figure S3.6 Expression in E. coli BRL04 of the codon optimized pentacistronic plasmid possessing the Pse5,7Ac$_2$ pathway from *H. pylori*. Proteins were assessed via western blotting using an Anti-6xHis HP-conjugated antibody. Expected size of proteins: pseB: 37 kDa, pseC: 42 kDa, pseH: 20 kDa, pseG: 33 kDa, agm1: 60 kDa

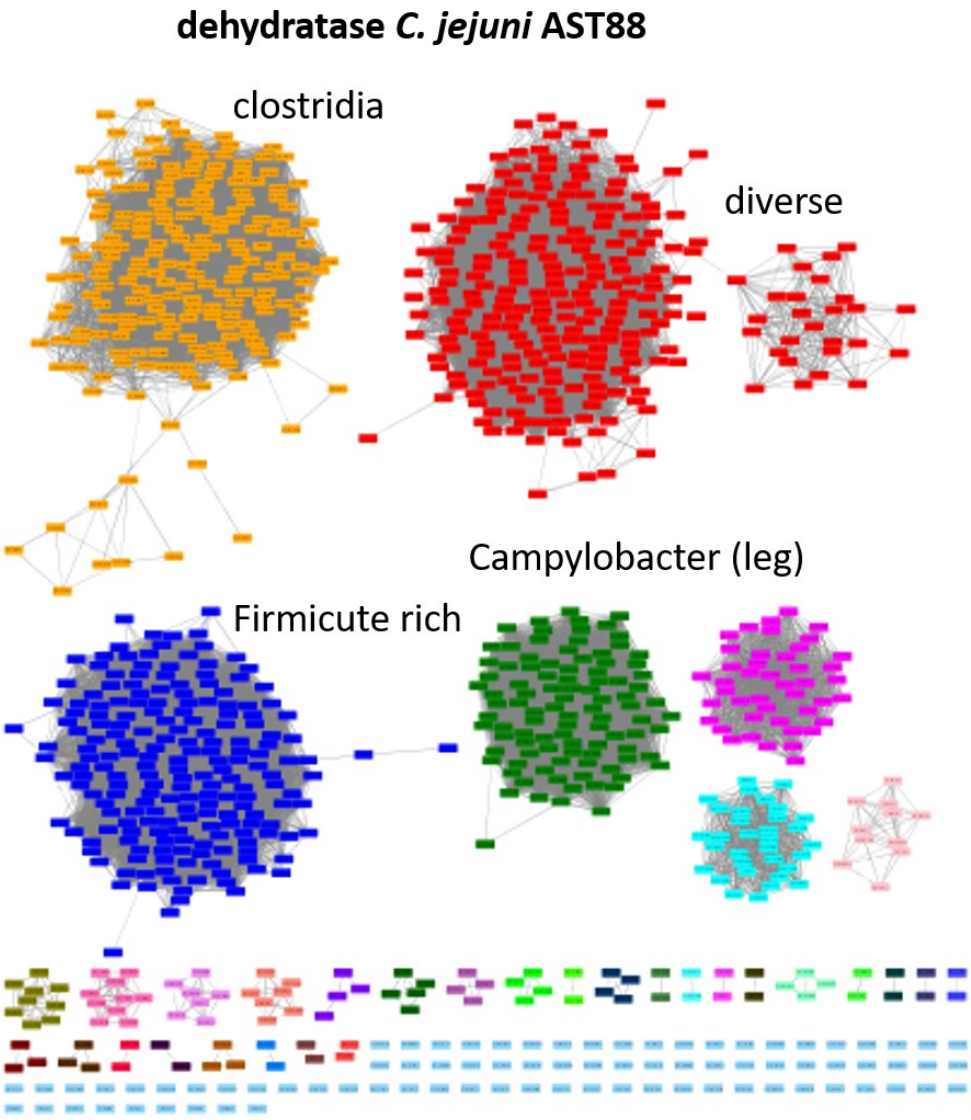


Figure S3.7 Clustering network for the genes associated with Pse5,7Ac$_2$ dehydratase *pseB*. Main groups are *Clostridia, Campylobacter*, and a firmicute rich region. Clusters were generated by querying *pseB* using EFI-EST, a tool that can generate a sequence similarity network for clusters of a protein sequence from the UniProt database. Results above were queried with an AST score of 88.

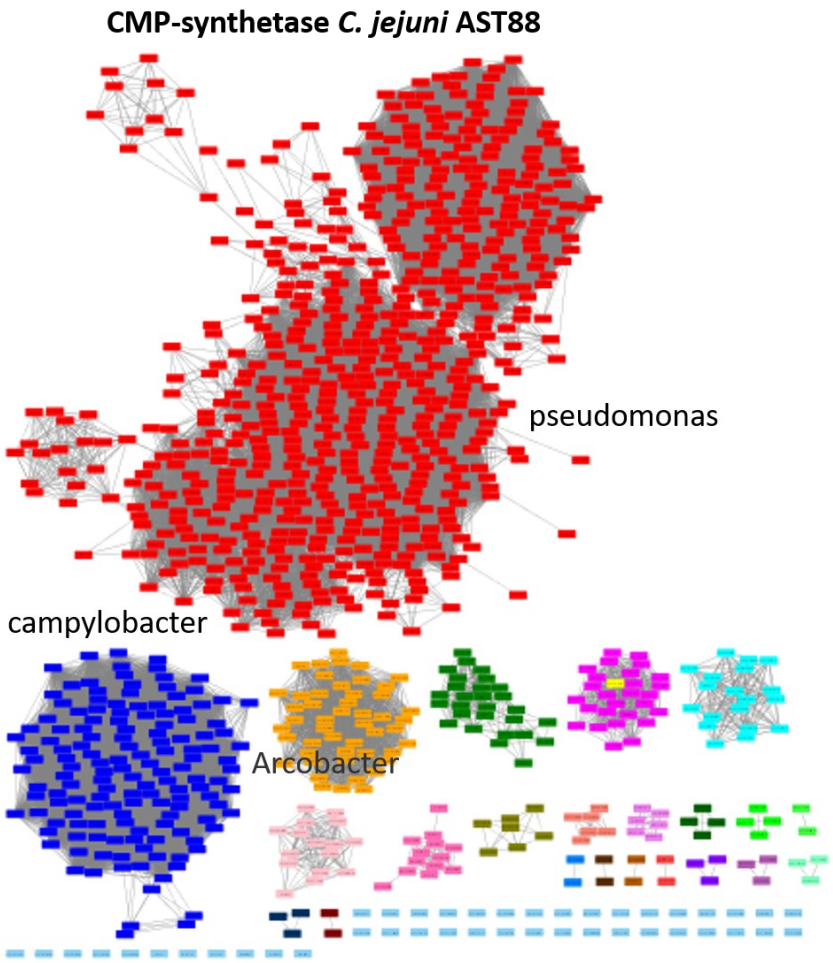


Figure S3.8: Clustering network for the Pse5,7Ac$_2$ CMP-synthetase *pseF*. Main groups are *Pseudamonas, Campylobacter*, and *Arcobacter*. Clusters were generated by querying *pseF* using EFI-EST, a tool that can generate a sequence similarity network for clusters of a protein sequence from the UniProt database. Results above were queried with an AST score of 88.

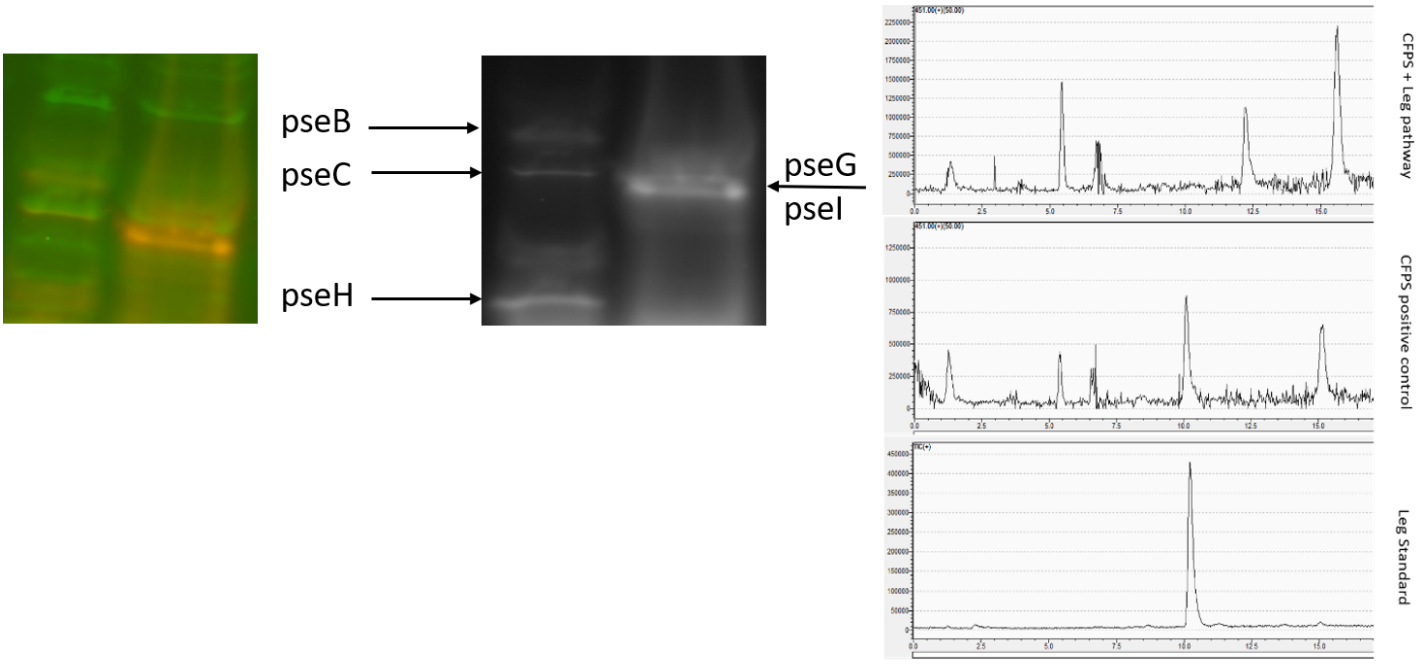


Figure S3.9: *In vitro* protein synthesis of Leg5,7Ac$_2$ pathway using CFPS kit from NEB. Protein expression is observed (left). LCMS analysis failed to detect Leg5,7Ac$_2$ production.

Chapter 4: De novo polyprotein design for the flexible in vivo heterologous expression of natural products

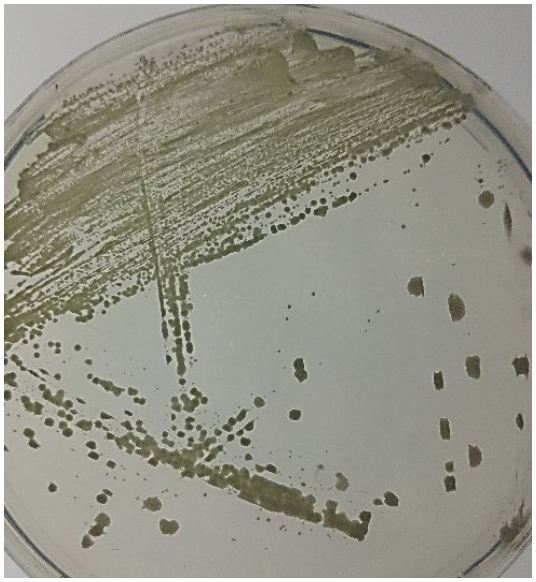


Figure S4.1: *S. cerevisiae* BY4741 transformed with pMI91 grown on YPD develops a green pigmentation after 5 days.

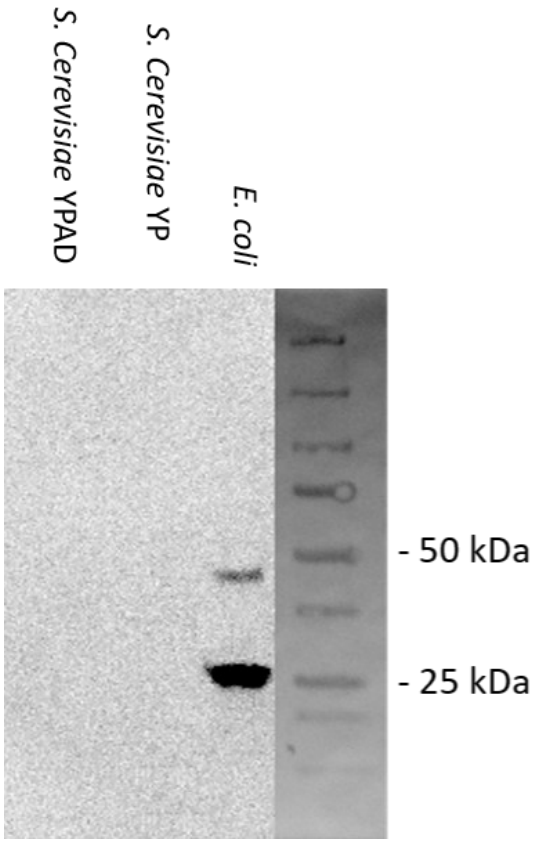
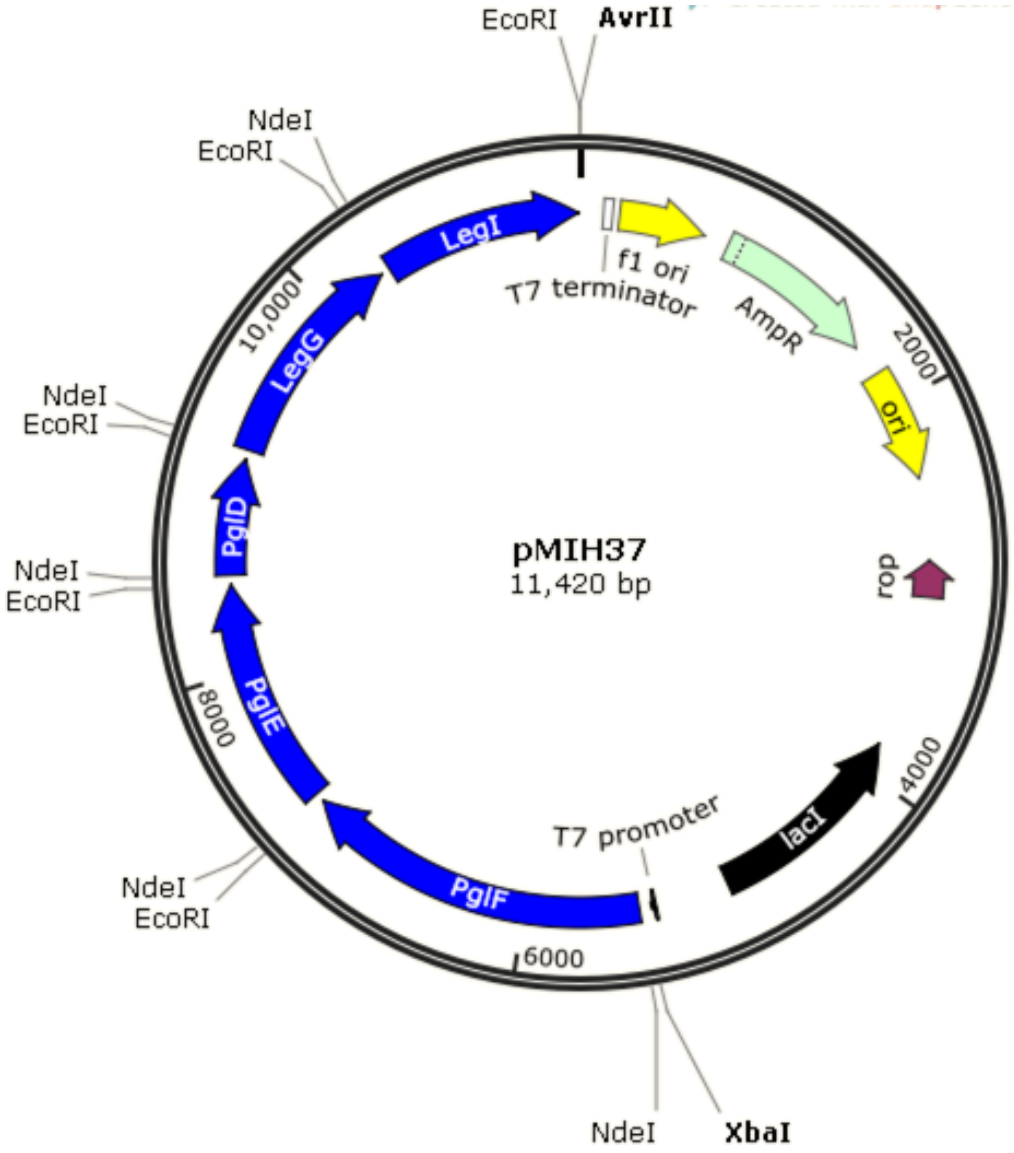
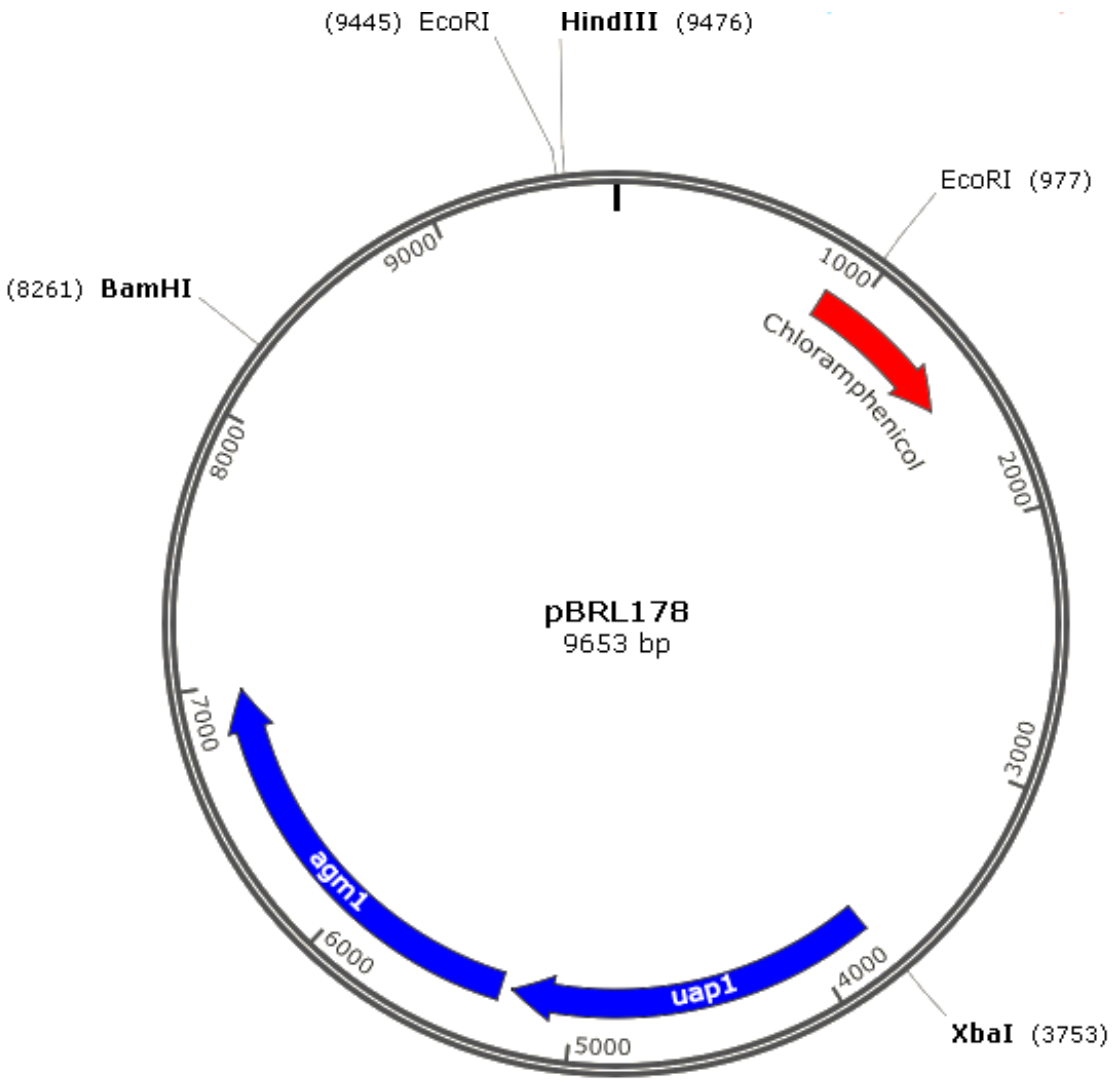


Figure S4.2: Western blotting of violacein polyprotein pathway in *S. cerevisae* BY4741 supplemented with YP and YPAD rich media. *E. coli* expressing the polyprotein design was used as a positive control. Western blotting was performed with pre-cast unstained gels and western blotting membrane was probed an HRP-conjugated primary FLAG antibody. Detection of bands was performed with chemiluminescence.

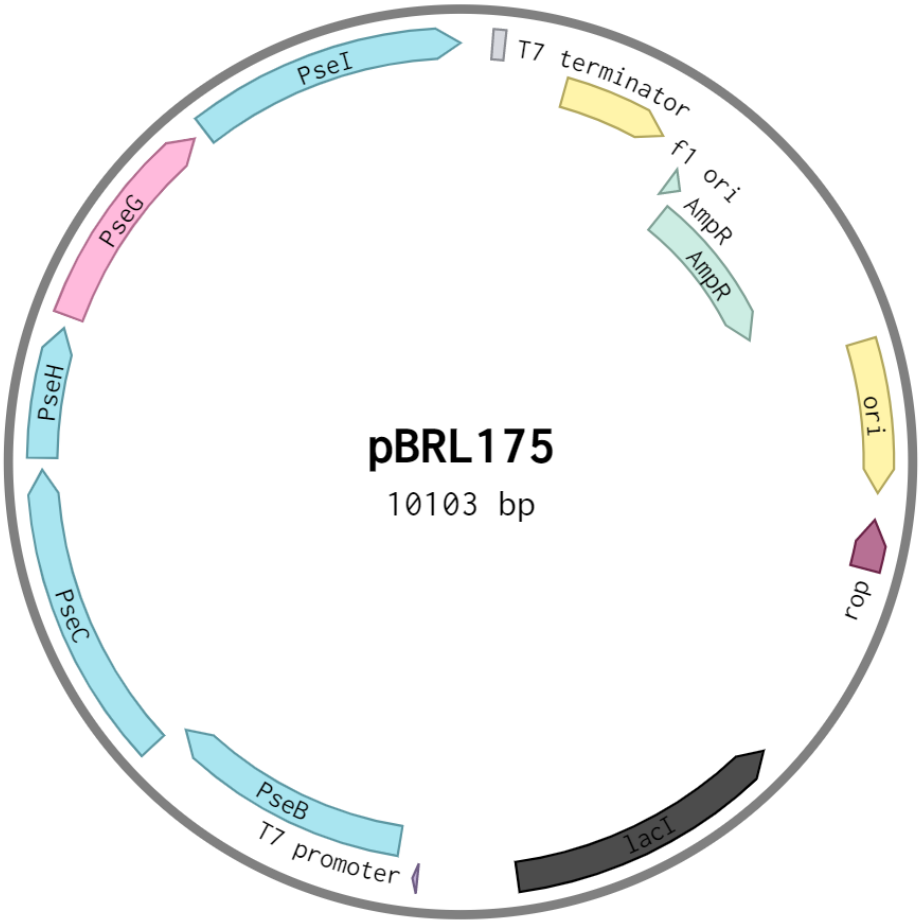
APPENDIX II: Plasmid Maps



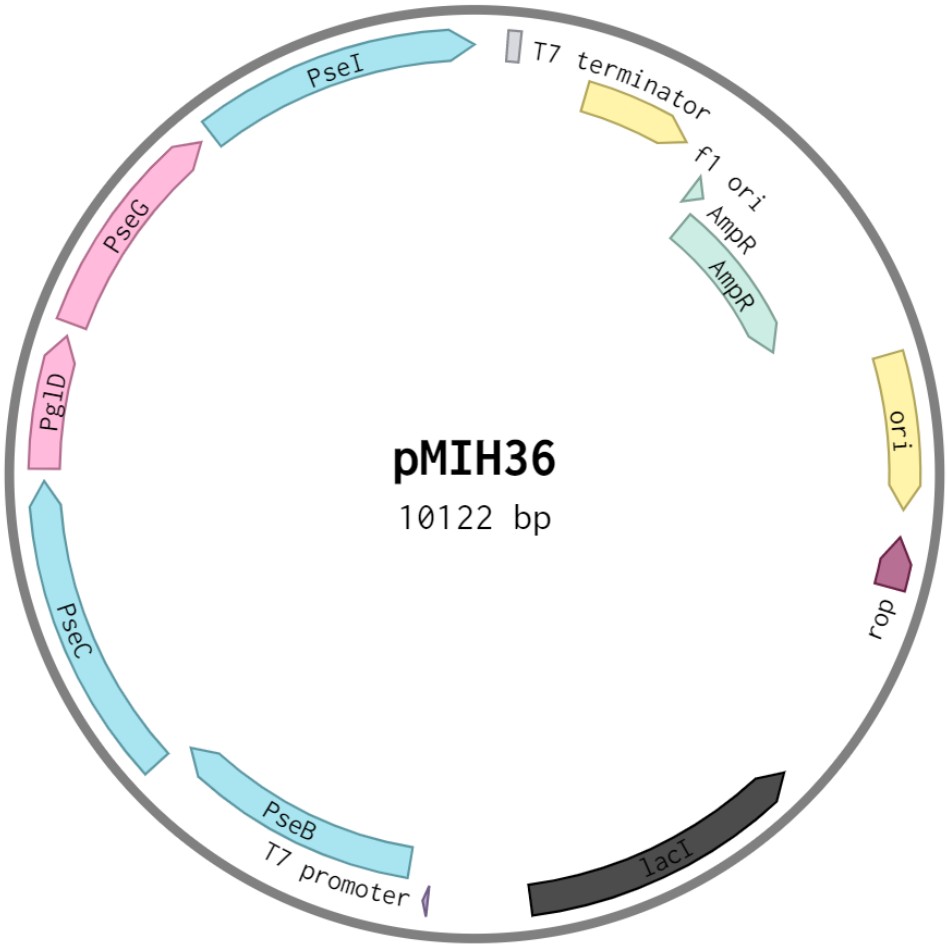
Chapter 2: Plasmid map of Leg5,7Ac$_2$ biosynthetic plasmid pMIH37.



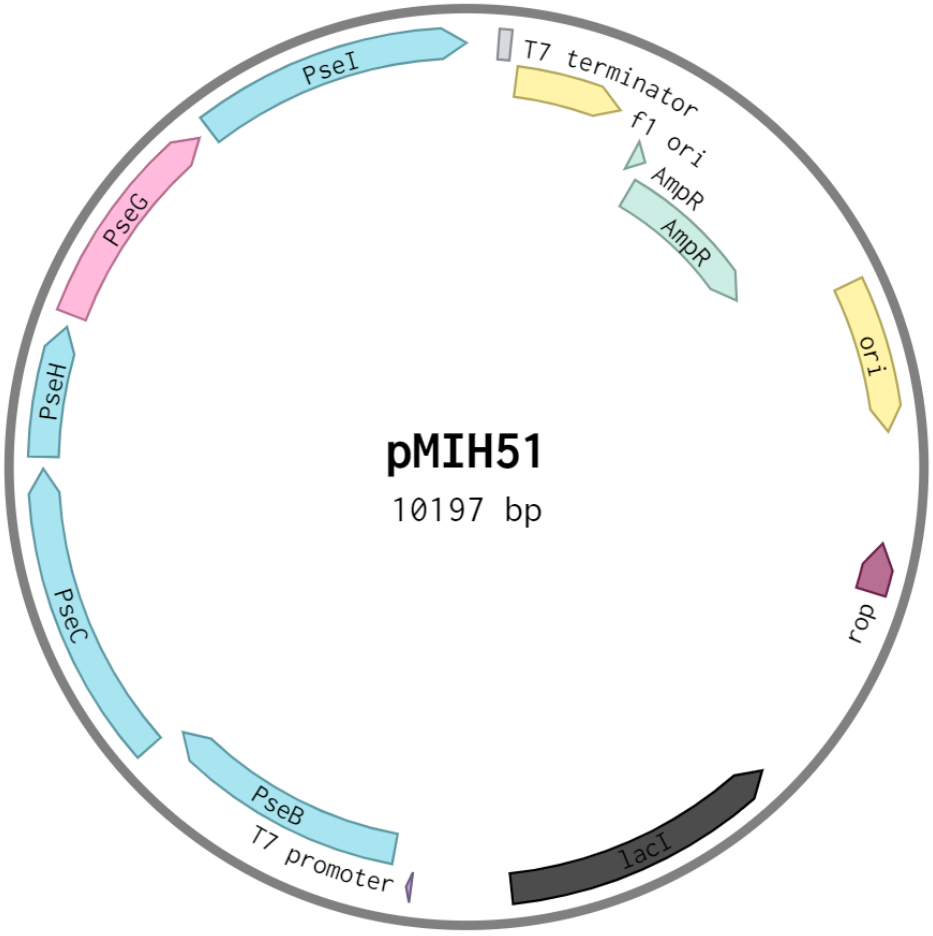
Chapter 2: Plasmid map of pBRL178.



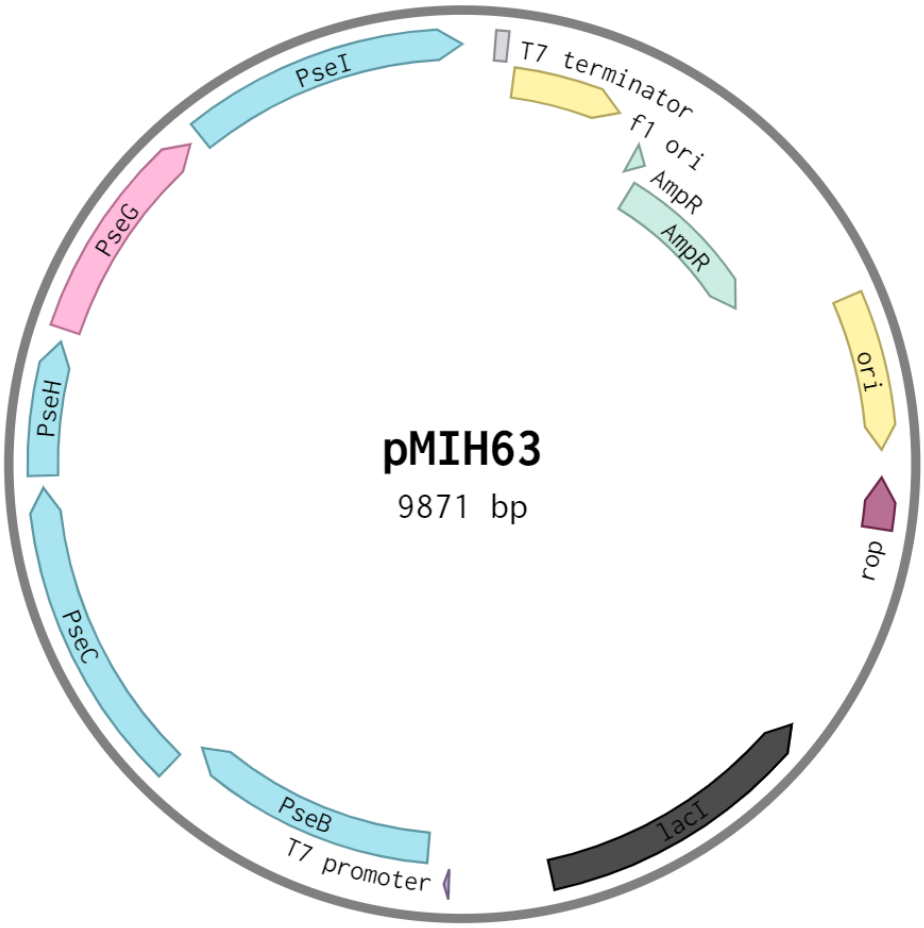
Chapter 3: Plasmid map of pBRL175



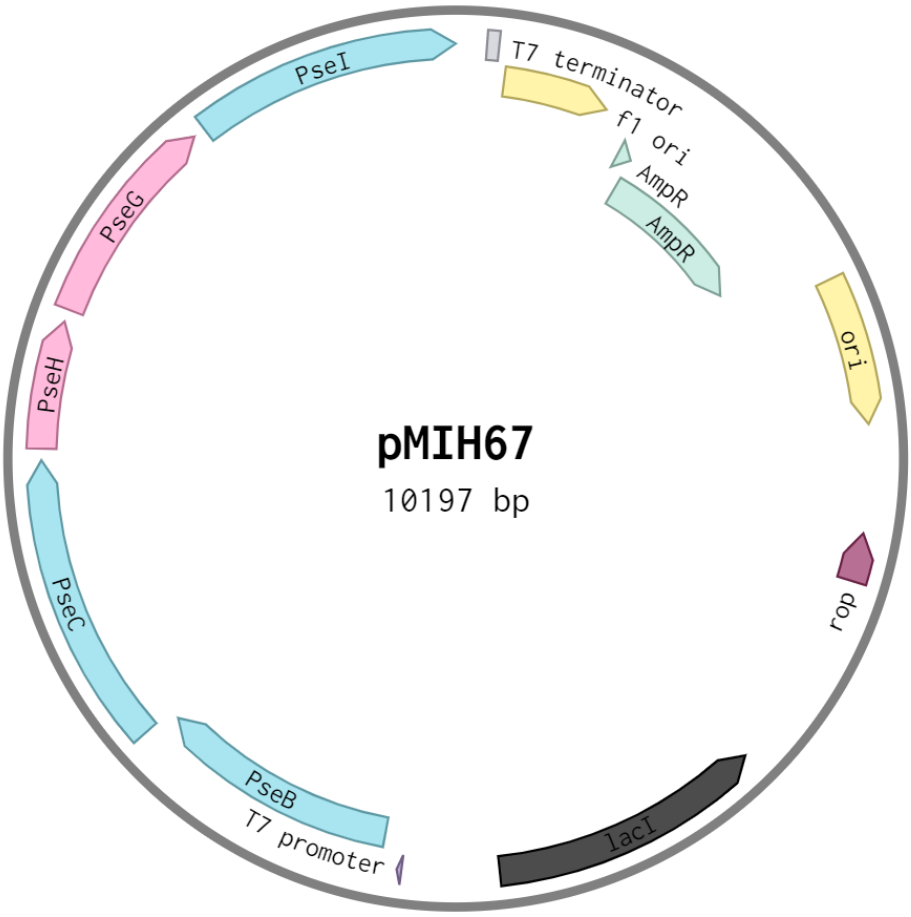
Chapter 3: Plasmid map of pMIH36



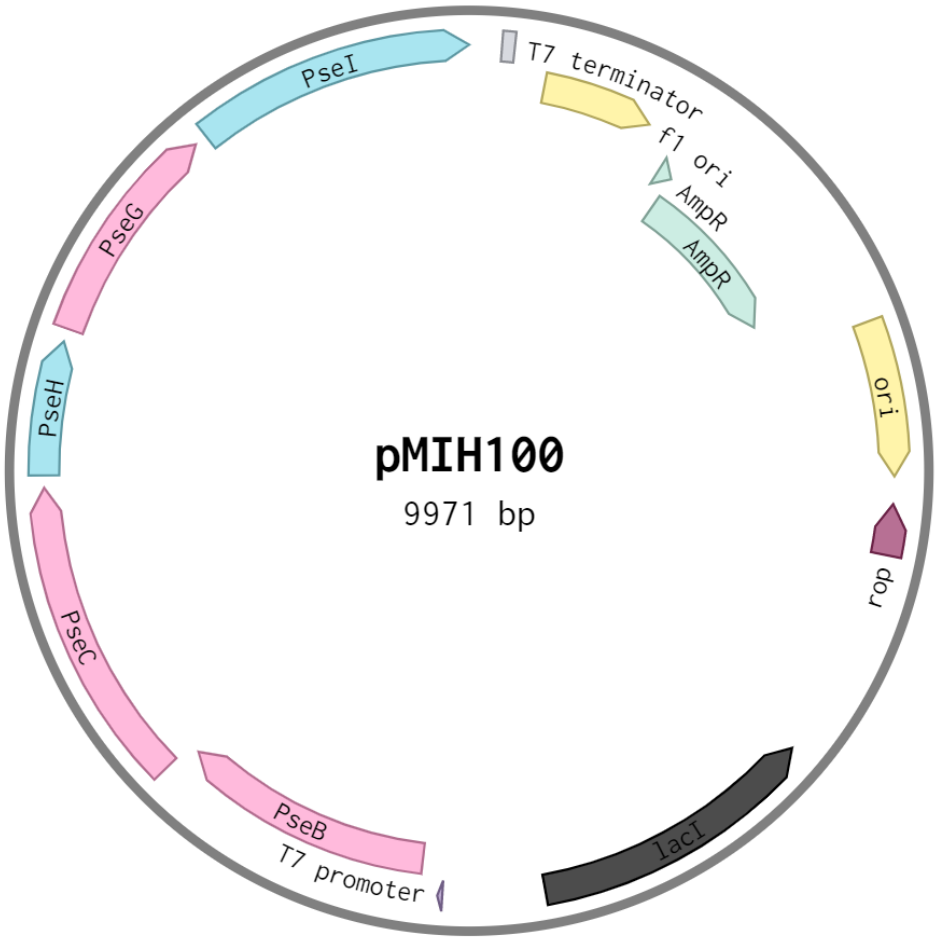
Chapter 3: Plasmid map of pMIH51



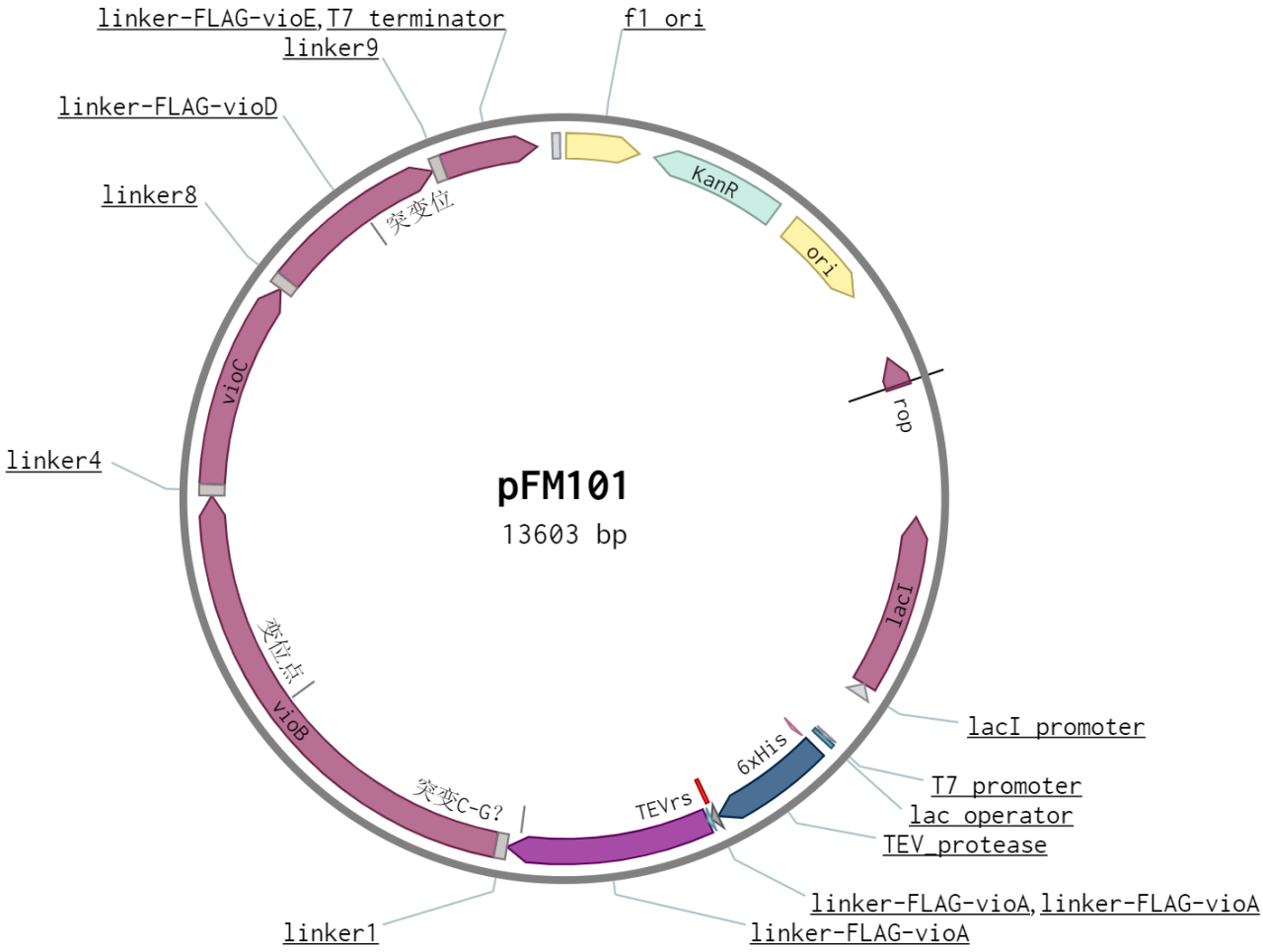
Chapter 3: Plasmid map of pMIH63



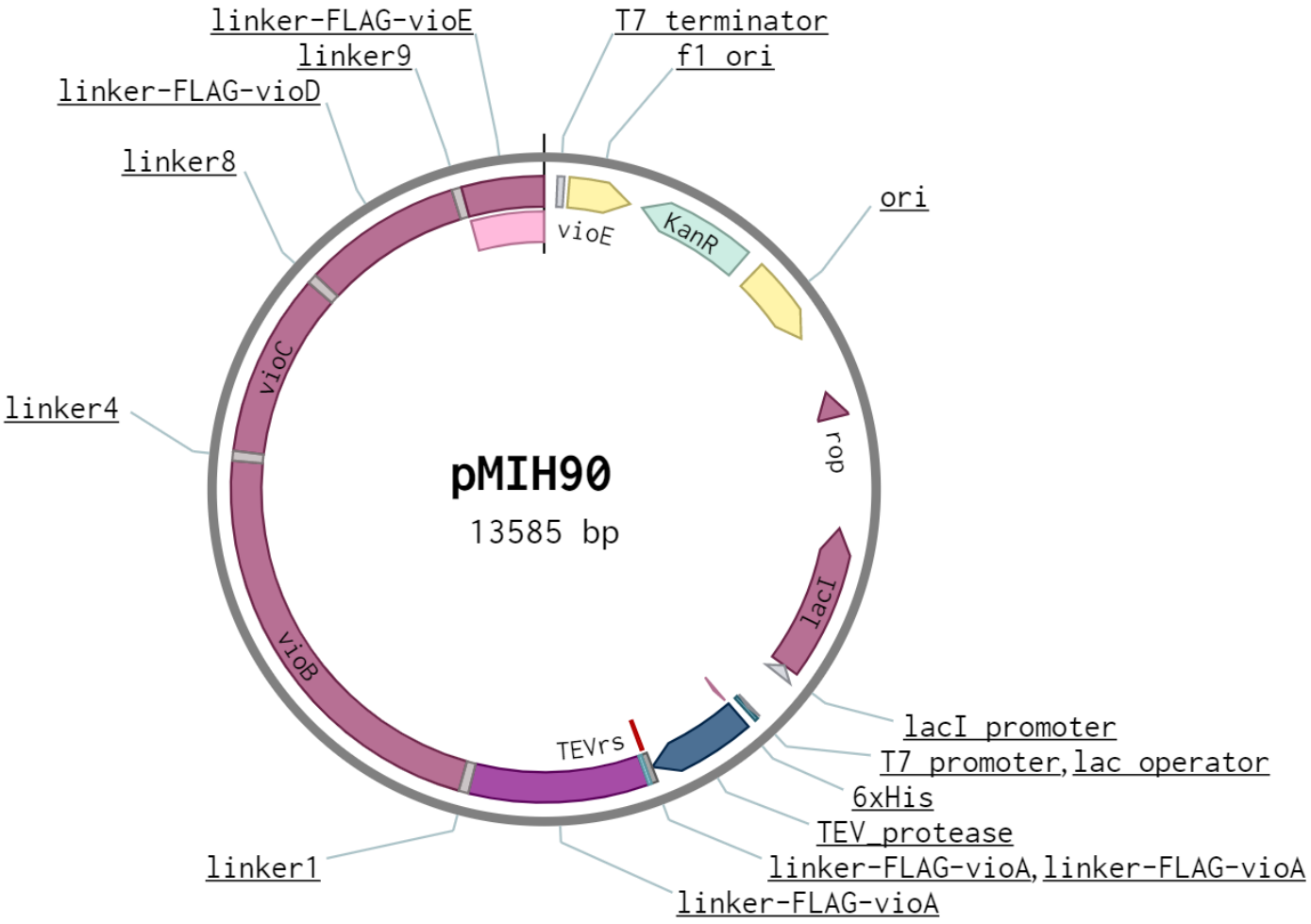
Chapter 3: Plasmid map of pMIH67



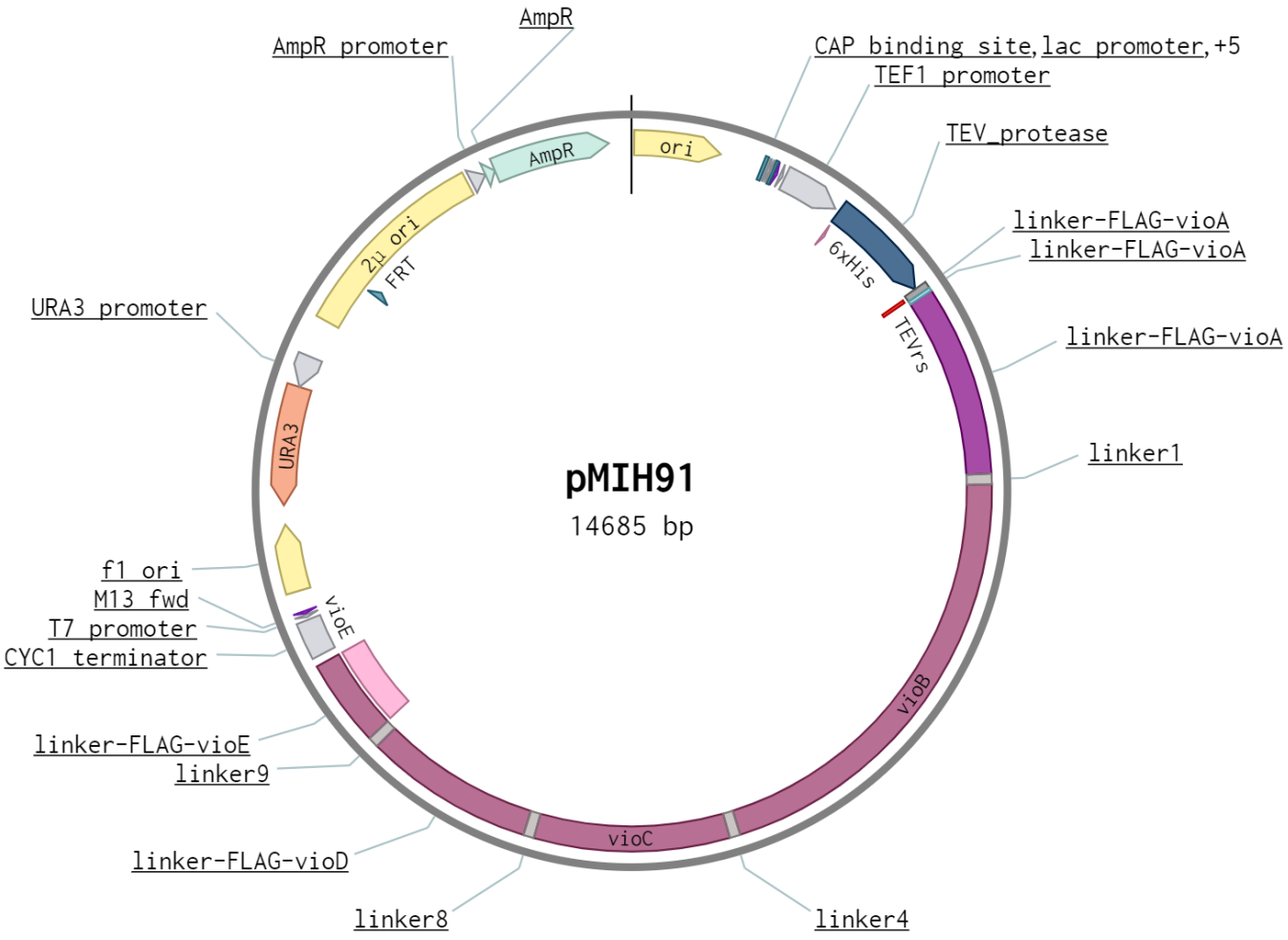
Chapter 3: Plasmid map of pMIH100



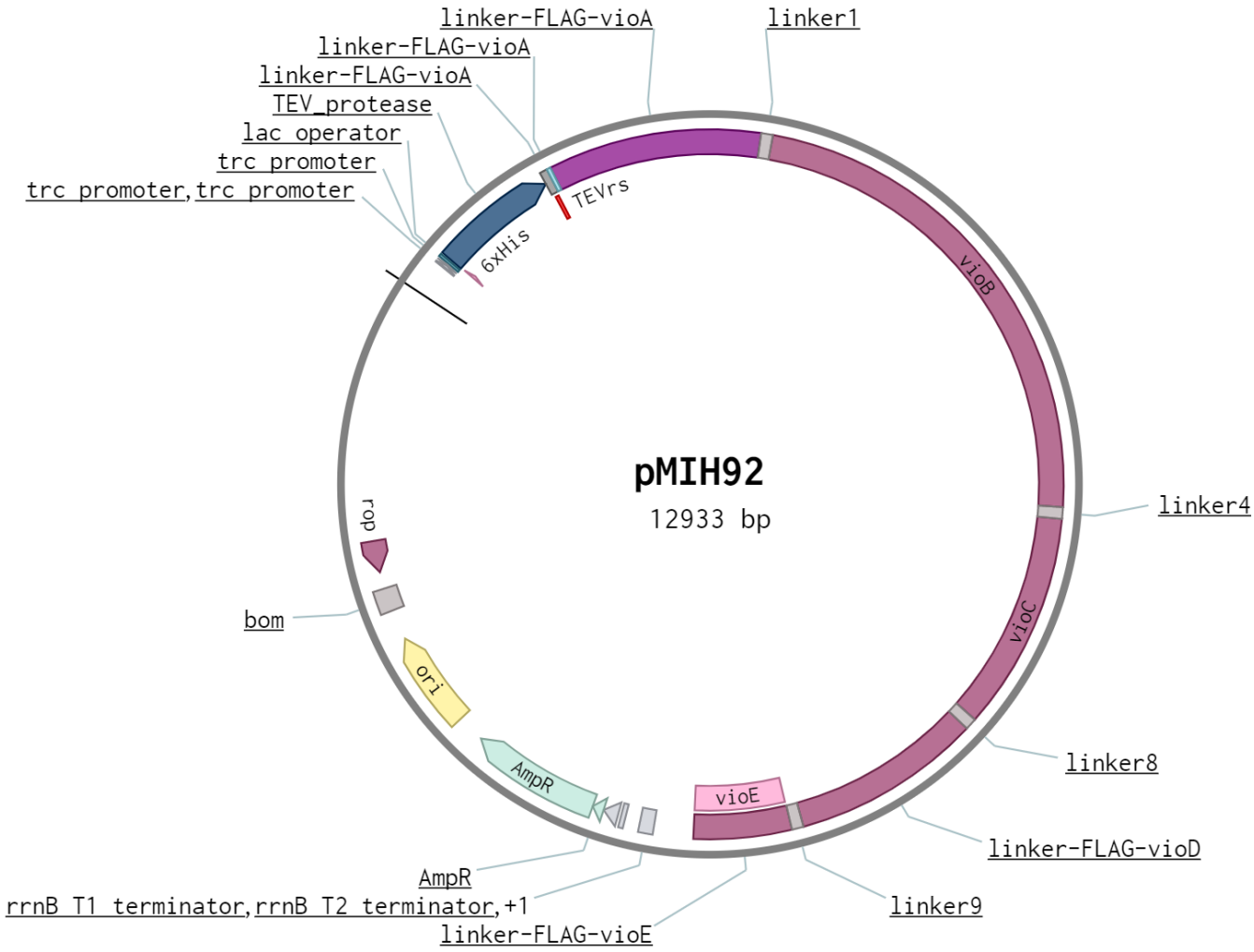
Chapter 4: Plasmid map of pFM101



Chapter 4: Plasmid map of pMIH90



Chapter 4: Plasmid map of pMIH91



Chapter 4: Plasmid map of pMIH92

APPENDIX III: Copyright Information

Total Biosynthesis of Legionaminic Acid, a Bacterial Sialic Acid Analogue

Author: Mohamed I. Hassan, Benjamin R. Lundgren, Michael Chaumun, et al
Publication: Angewandte Chemie International Edition
Publisher: John Wiley and Sons
Date: Aug 19, 2016

Order Completed

Thank you for your order.

This Agreement between Mohamed I Hassan ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4744851314655
License date	Jan 09, 2020

Printable Details

Licensed Content

Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Angewandte Chemie International Edition
Licensed Content Title	Total Biosynthesis of Legionaminic Acid, a Bacterial Sialic Acid Analogue
Licensed Content Author	Mohamed I. Hassan, Benjamin R. Lundgren, Michael Chaumun, et al
Licensed Content Date	Aug 19, 2016
Licensed Content Volume	55
Licensed Content Issue	39
Licensed Content Pages	4

Order Details

Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No

About Your Work

Title of your thesis / dissertation	Harnessing Microbial Biosynthetic Pathways for the Production of Complex Molecules
Expected completion date	Feb 2020
Expected size (number of pages)	250

Additional Data