METABOLIC ENGINEERING AND OMICS ANALYSIS OF *AGROBACTERIUM* SP. ATCC 31749 FOR OLIGOSACCHARIDE SYNTHESIS

A Dissertation
Presented to
The Academic Faculty

by

Anne M. Ruffing

In Partial Fulfillment
of the Requirements for the Degree
Ph.D. in Chemical Engineering in the
School of Chemical & Biomolecular Engineering

Georgia Institute of Technology
May 2010

METABOLIC ENGINEERING AND OMICS ANALYSIS OF *AGROBACTERIUM* SP. ATCC 31749 FOR OLIGOSACCHARIDE SYNTHESIS

Approved by:

Dr. Rachel Ruizhen Chen Advisor
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. Jim Spain
School of Civil & Environmental Engineering
Georgia Institute of Technology

Dr. Martha Grover
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. Donald Doyle
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Mark Prausnitz
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Date Approved: January 12, 2010

ACKNOWLEDGEMENTS

I am deeply indebted to those who have supported me throughout my time here at Georgia Tech. First, I wish to thank my advisor, Dr. Rachel Chen, for her support and guidance. Her ideas and insight were essential in the progression of this thesis project. I would like to thank my other committee members Dr. Martha Grover, Dr. Mark Prausnitz, Dr. Jim Spain, and Dr. Donald Doyle for their time and suggestions. Several others have directly contributed to the work presented in this dissertation, deserving both my thanks and acknowledgment for their contribution. Dr. Zichao Mao, a former postdoctoral fellow of the Chen group, contributed to the development of the LacNAc-producing biocatalyst by constructing the *Agrobacterium* expression vector, pBQ, and the recombinant plasmid, pKEL. We collaborated with Dr. Wei-Shou Hu and Marlene Castro at the University of Minnesota for the genome sequencing of ATCC 31749. Marlene ran the program for sequence assembly and assisted with gene prediction. I am grateful to Dr. Hyun-Dong Shin for sharing his knowledge of molecular biology and for his helpful suggestions regarding this thesis work. I also wish to thank all members of the Chen group throughout my years at Georgia Tech, particularly Shara McClendon, my officemate for the last 5 years, and Manoj Agrawal.

I am fortunate to have an amazing group of people who have supported me throughout the many tribulations, setbacks, and triumphs of my graduate career. Thanks to my parents, Bob and Margie, who have driven to Atlanta numerous times over the years to help me move apartments, to celebrate birthdays, or just to make sure I take a break from the lab. I would also like to thank my siblings, Mark, Nick, and Kristen, for

their support. And last, but certainly not least, I wish to thank my fiancé, Carter Dietz, for his friendship and encouragement, for making me laugh, and for the many lunch-time discussions trouble-shooting my research problems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

6PG	6-phosphogluconate
α-Gal	terminal α1,3-galactose oligosaccharide
ACoA	acetyl coenzyme A
ADP	adenosine diphosphate
aKG	alpha-ketoglutarate
Ala	alanine
Arg	arginine
Asn	asparagine
Asp	aspartate
ATP	adenosine triphosphate
bp	base pair
c-di-GMP	cyclic-dimeric-guanosine monophosphate
CCR	carbon catabolite repression
CIT	citrate
cfu	colony forming unit
Crd	curdlan
Cys	cysteine
DCW	dry cell weight
DNA	deoxy-ribonucleic acid
ECF	extracytoplasmic function
ED	Entner-Doudoroff
EMP	Embden Meyerhof Parnas
EPS	exopolysaccharide

F6P	fructose-6-phosphate
FAD	flavin adenine dinucleotide (oxidized)
FADH2	flavin adenine dinucleotide (reduced)
Fru	fructose
G1P	glucose-1-phosphate
G6P	glucose-6-phosphate
Gal	galactose
GalNAc	*N*-acetylgalactosamine
GAP	glyceraldehyde-3-phosphate
gDNA	genomic DNA
Glc	glucose
GlcNAc	N-acetylglucosamine
Gln	glutamine
Glu	glutamate
Gly	glycine
Glycerol-3-P	glycerol-3-phosphate
GMP	guanosine monophosphate
GPI	glycosylphosphatidylinositol
GTP	guanosine triphosphate
Hib	*Haemophilus influenza* type B
His	histidine
ICT	isocitrate
Ile	isoleucine
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilobase pair

KDPG	2-keto-3-deoxygluconate-6-phosphate
KO	knockout
Lac	lactose
LacNAc	N-aceytllactosamine
LB	Luria Bertani/Luria Broth/Lysogeny Broth
Leu	leucine
Lys	lysine
MAL	malate
Man-1-P	mannose-1-phosphate
Man-6-P	mannose-6-phosphate
Met	methionine
NAD	nicotinamide adenine dinucleotide (oxidized)
NADH	nicotinamide adenine dinucleotide (reduced)
NDP	nucleotide diphosphate\
OAA	oxaloacetate
P3	product 3 (galactose-β1,4-mannose)
PAA	polyacrylamide
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
Phe	phenylalanine
Pi	phosphate
polyP	polyphosphate
(p)ppGpp	(guanosine pentaphosphate) guanosine tetraphosphate
PPi	pyrophosphate
PPP	pentose phosphate pathway

Pro	proline
PRPP	5-phosphoribosyl-1-pyrophosphate
PTS	phosphotransferase system
PYR	pyruvate
R5P	R5P: ribose-5-phosphate
Rif	rifampicin
RNA	ribonucleic acid
RNAP	RNA polymerase
rRNA	ribosomal ribonucleic acid
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	serine
sLe^x	sialyl Lewis x
Suc	sucrose
SUCC	succinate
SUCC-CoA	succinyl coenzyme A
TCA	tricarboxylic acid
Thr	threonine
tRNA	transfer RNA
Trp	tryptophan
Tyr	tyrosine
UDP	uridine diphosphate
UDP-GA	uridine diphosphate-glucuronic acid
UDP-Gal	uridine diphosphate galactose
UDP-GalNAc	uridine diphosphate-*N*-acetylgalactosamine
UDP-Glc	uridine diphosphate glucose

UDP-GlcNAc	uridine diphosphate-*N*-acetylglucosamine
UDPG	uridine diphosphate glucose
UMP	uridine monophosphate
UTP	uridine triphosphate
Val	valine

LIST OF GENE AND ENZYME NOMENCLATURE

α1,3-GalT	α1,3-galactosyltransferase
β1,4-GalT	β1,4-galactosyltransferase
Amp^R	ampicillin resistance
celD	required for cellulose synthesis
celE	required for cellulose synthesis
chvG	two component sensor kinase
chvI	two component response regulator
crdA	hypothetical gene required for curdlan synthesis
crdC	hypothetical gene required for curdlan synthesis
crdS/CrdS	curdlan synthase
crdR	predicted transcriptional regulator required for curdlan synthesis
fixK	nitrogen fixation regulatory protein
fnrN	Crp family transcriptional regulator
galE	UDP-galactose 4'-epimerase
galK	galactokinase
galT	galactose-1-phosphate uridylyltransferase
galU	UDPG pyrophosphorylase/UTP:G1P uridylyltransferase
glnB	nitrogen regulatory protein PII
glnD	PII uridylyltransferase
glnK	nitrogen regulatory protein PII
GM^R	gentamycin resistance
HK	hexokinase

KM^R	kanamycin resistance
lacY/LacY	lactose permease
lgtB	β1,4-galactosyltransferase
lgtC	α1,4-galactosyltransferase
laqIq	lactose reporessor
nifR	nitrogen regulation protein
NodC	chitin oligosaccharide synthase
nolR	ArsR family transcriptional regulator
ntrB	two component sensor kinase
ntrC	two component response regulator
ntrX	two component response regulator (nitrogen limitation)
ntrY	two component sensor kinase (nitrogen limitation)
pfp	pyrophosphate-dependent fructose-6-phosphate-1-phosphotransferase
pgi	phosphoglucose isomerase
pgm	phosphoglucomutase
phoB	two component response regulator (phosphate limitation)
ppa	inorganic pyrophosphatase
ppk	polyphosphate kinase
ppx	exopolyphosphatase
pss_{AG}	phosphatidylserine synthase
ptsN	nitrogen regulatory IIA protein
rep	replication protein
RpoD (σ^{70})	sigma factor-70
RpoE (σ^{24})	sigma factor-24 (stress response)
rpoN/RpoN (σ^{54})	sigma factor-54 (nitrogen limitation

RpoS (σ^{38})	sigma factor-38 (stationary phase)
rrpP	membrane-bound proton-translocating pyrophosphatase

SUMMARY

Oligosaccharides are important biomolecules that are targets and also components of many medical treatments, including treatments for cancer, HIV, and inflammation. While the demand for medically-relevant oligosaccharides is increasing, these compounds have proven difficult to synthesize. Whole-cell oligosaccharide synthesis is a promising method that requires relatively inexpensive substrates and can complete the synthesis in just one step. However, whole-cell oligosaccharide synthesis employing common microorganisms like *E. coli* have been plagued by low yields. This dissertation investigates an alternative microorganism for oligosaccharide production: *Agrobacterium* sp. ATCC 31749. This *Agrobacterium* strain produces high levels of curdlan polysaccharide, demonstrating its natural ability to produce the sugar nucleotide precursor for oligosaccharide production. The two main objectives of this dissertation are 1) to develop biocatalysts for oligosaccharide synthesis by engineering ATCC 31749 and 2) to determine what factors affect poly- and oligosaccharide production in this *Agrobacterium* strain.

ATCC 31749 was engineered to produce two oligosaccharides of medical importance: *N*-acetyllactosamine and galactose-α1,3-lactose. Oligosaccharide production in the biocatalyst was further improved with additional metabolic engineering. Substrate uptake was increased through expression of a lactose permease, and availability of the sugar nucleotide substrate improved with gene knockout of the curdlan synthase gene. Both of these engineering efforts led to increased oligosaccharide synthesis in the *Agrobacterium* biocatalyst. Overall, the engineered *Agrobacterium* strains synthesized

gram-scale quantities of the oligosaccharide products in just one step and requiring only a few inexpensive substrates and cofactors.

Additional improvement of the oligosaccharide-producing biocatalysts required further investigation of the factors influencing poly- and oligosaccharide production in ATCC 31749. In this dissertation, several environmental and intracellular factors are identified that affect both oligosaccharide and curdlan production. Sucrose was the preferred carbon source for oligosaccharide synthesis, and the addition of citrate to the synthesis reaction led to significant improvement in oligosaccharide production. To identify the genetic factors and possible mechanisms regulating curdlan production, the genome of ATCC 31749 was sequenced. The genome sequence was utilized for transcriptome analysis of ATCC 31749. In the transcriptome analysis, genes significantly up- and down-regulated during curdlan production were identified. Subsequent gene knockout experiments showed several factors to be important for curdlan synthesis, namely the nitrogen signaling cascade, polyphosphate, and the GTP-derived second messengers (p)ppGpp and c-di-GMP. In addition to the development of biocatalysts for oligosaccharide production, this investigation provides insight into the complex mechanisms regulating exopolysaccharide synthesis.

PREFACE

The work contained in this dissertation was written either with intent to publish, in submission currently, or already published in scientific journals. Previously published or accepted manuscripts were modified to include additional data and to adhere to the formatting requirements. This dissertation includes an introduction containing background information on oligosaccharides and *Agrobacterium* sp. ATCC 31749 (Chapter 1), the experimental findings of this research project along with discussion (Chapters 2-6), and a summary of the conclusions of this dissertation along with recommendations for future investigation (Chapter 7).

CHAPTER 1

INTRODUCTION

1.1 Oligosaccharides

The term ‘oligosaccharide’ comes from the Greek word for ‘few’ (*oligos*) and the term for sugar (saccharide). As suggested by its etymology, oligosaccharides are comprised of a few sugar residues, generally two to ten, linked by glycosidic bonds. Sugar residues have a general molecular formula of $(C{\cdot}H_2O)_n$, allowing oligosaccharides to also be classified as carbohydrates. Carbohydrates are one of the four main components of a cell, with lipids, nucleic acids, and amino acids as the other three. Despite its abundance in nature, the biological significance of carbohydrates and of oligosaccharides in particular has remained a mystery until the emergence of glycobiology in the late 1980’s [1].

1.1.1 Oligosaccharide structure

Oligosaccharides typically consist of a combination of the nine common monosaccharides: glucose, *N*-acetylglucosamine, galactose, *N*-acetylgalactosamine, xylose, glucuronic acid, mannose, fucose, and *N*-acetylneuraminic acid or sialic acid [2]. Despite this limited selection of monosaccharides, oligosaccharides display great structural diversity. The sugar residues of oligosaccharides can be linked at different positions around the sugar ring, and each glycosidic linkage can be either α or β with respect to its geometry (Figure 1.1). Furthermore, oligosaccharides are classified based on their linkage to a main scaffold [3]. O-linked glycosylation involves an

oligosaccharide attached to a serine or threonine residue of a protein (Figure 1.1a). In N-linked glycosylation, an oligosaccharide is attached to the asparagine residue of an Asn-X-Ser/Thr sequence in a protein, where X is any amino acid except for proline (Figure 1.1b). Oligosaccharides can be attached to lipids, forming glycolipids (Figure 1.1c), and glycosylphosphatidylinositol(GPI)-anchored proteins utilize glycolipids to anchor the protein to the cell membrane (Figure 1.1d). In addition to variations in the linkage and attachment to a main scaffold, oligosaccharide structure can also be modified by the addition of side groups to the monosaccharide components. With this great structural diversity, oligosaccharides also display a great range of functional diversity.

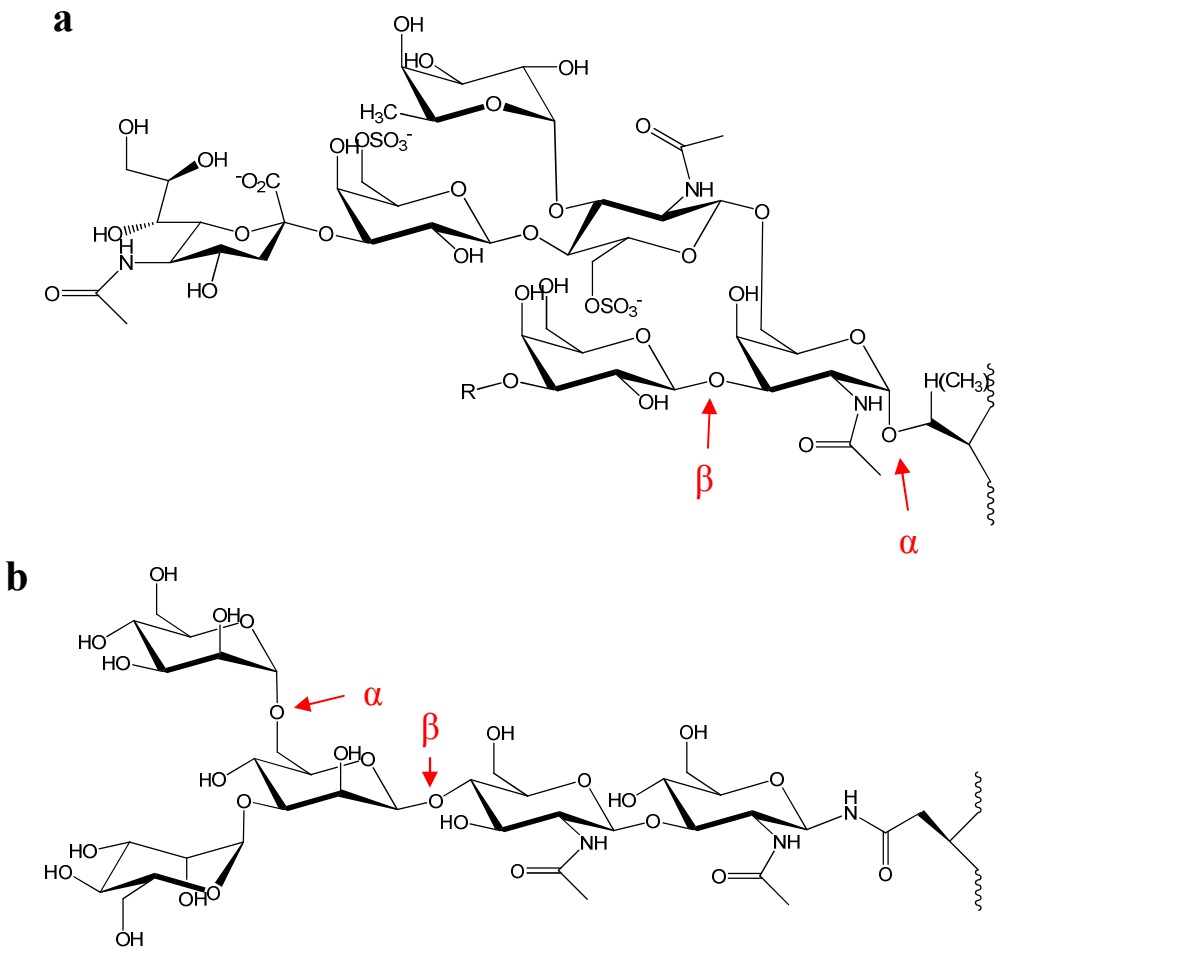


Figure 1.1 Examples of oligosaccharides linked to a main scaffold. a) O-linked glycosylation b) N-linked glycosylation c) glycolipid d) GPI anchor.

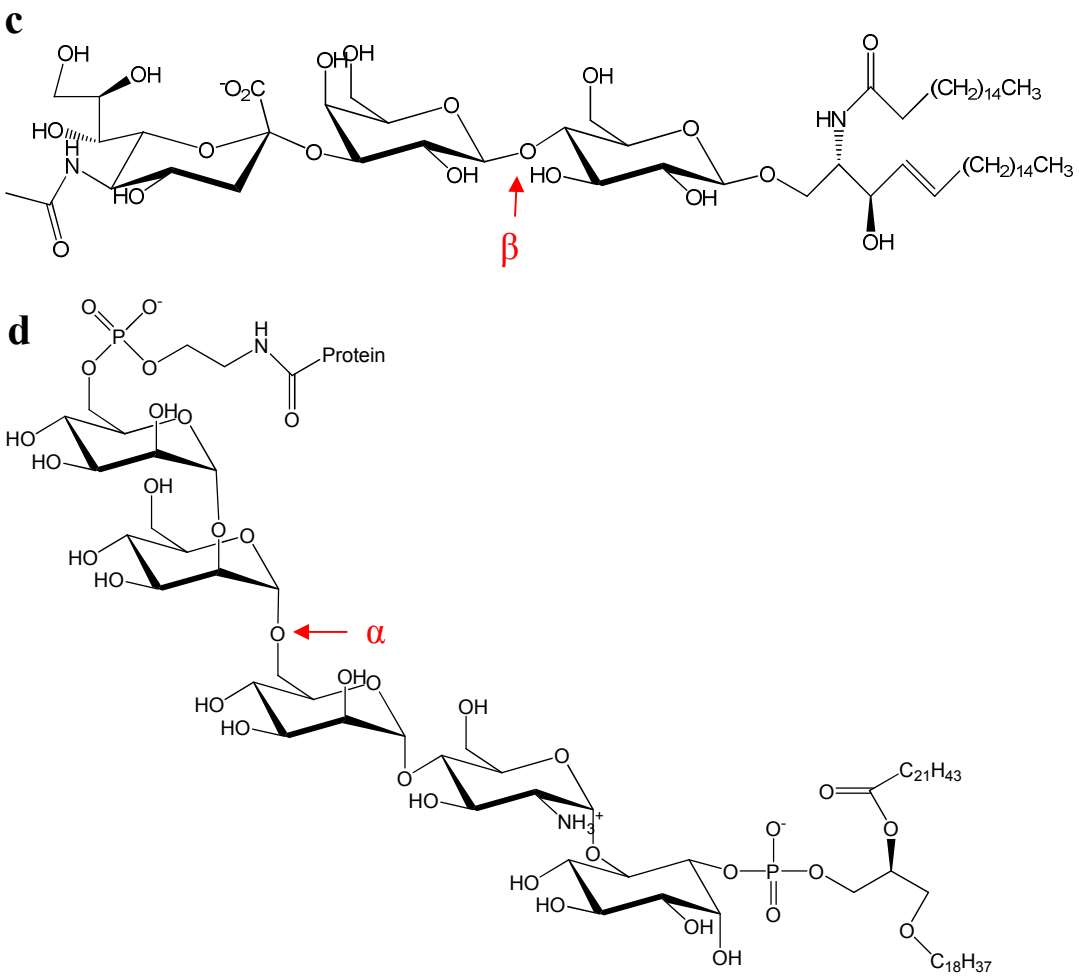


Figure 1.1 Continued.

1.1.2 Biological function

Carbohydrates have long been recognized for their roles as energy sources and carbon storage compounds. Carbohydrates are common nutrients in the human diet, and polysaccharides such as glycogen and starch provide a carbon storage mechanism for animal and plant cells respectively. Other polysaccharides contribute to cell structure. For example, cellulose is found in the cell walls of plants and other organisms, and chitin contributes to arthropod exoskeletons and some fungal cell walls. While these energy, storage, and structural functions are biologically important, a new role for carbohydrates has been identified in the past few decades. Cell surface carbohydrates were found to be

responsible for many interactions between the cell and its environment, contributing to cell-cell interactions, antibody attachment, hormone and toxin binding, and viral and bacterial infection [3]. Cell surface carbohydrates are also important in immune response. In fact, some bacteria and viruses can avoid detection by the immune system through coating their surface with exopolysaccharides. In addition, oligosaccharides can promote proper protein folding, and glycosylation can be essential for activating protein function. With these various biological roles, it is not surprising that many poly- and oligo-saccharides are involved in several disease states of the cell.

1.1.3 Medical applications

With recent discoveries in oligosaccharide function, medical researchers have a new class of candidates to consider for medical treatments. In fact, many oligosaccharide-based medical treatments are currently under development. Just as the structure and function of oligosaccharides vary, so do the types of diseases with potential for oligosaccharide-based treatment. Oligosaccharide-based treatments are under investigation for diseases ranging from cancer to infection to inflammation. The oldest and most well known carbohydrate-based drug is heparin. Heparin is a sulfated glycosaminoglycan which has used as an anticoagulant since 1937 [4]. Following the discovery of heparin, other carbohydrate-based medical treatments proved difficult to develop. In 1980, a carbohydrate-based vaccine was developed to fight against *Haemophilus influenza* type B (Hib), the causative agent of bacterial meningitis (group C) [5]. With the advent of this vaccine, Hib has been virtually eliminated from the U.S., most European countries, and other developing nations, saving thousands, if not millions of lives worldwide [1].

1.1.3.1 Oligosaccharide-based treatments for bacterial, parasitic, and viral infections

Since the advent of the Hib vaccine, carbohydrate-based treatments have been under investigation for many other bacterial and parasitic infections. A potential malaria vaccine is in development which contains a hexasaccharide found on the cell surface of the malaria-causing parasite *Plasmodium falciparum* [6]. Other notable carbohydrate-based vaccines under development stimulate immune defense against *Bacillus anthracis*, the causative agent of anthrax [7-9]. With the promise of these oligosaccharide vaccines, the threat of *B. anthracis* as a deadly bioweapon may soon be eliminated. In addition to the malaria and anthrax vaccines, numerous other bacterial and parasitic infections are being targeted by carbohydrate-based vaccines, including opportunistic staphylococcal infections, Shigella, cholera, typhoid, *Candida albicans*, leprosy, and leishmaniasis [5].

Carbohydrates are also important components of treatments for several viruses. Viral coats commonly contain glycoproteins which help the virus to evade the immune system of the host; the carbohydrate components of glycoproteins offer new targets for treatment of viral infections [1]. Numerous research efforts are underway to develop an HIV vaccine by targeting the gp120 viral envelope glycoprotein. Previous research has shown that infected individuals who generate antibodies against the gp120 glycoprotein can survive longer than those who lack the antibody [10]. HIV vaccines which promote gp120 antibody production may be able to extend the lifetime of the estimated 36 million people currently infected with HIV. Carbohydrate targets may also lead to the development of an antiviral treatment for influenza. The influenza virus enters mammalian cells when glycoproteins on the viral coat bind to *N*-acetylneuraminic acid (sialic acid) residues on the host cell [2]. A sugar that mimics *N*-acetylneuraminic acid

may inhibit the binding of the influenza virus and the subsequent infection of the host cell. As knowledge grows regarding the mechanisms of viral infection, carbohydrates have emerged as attractive targets for preventing virus transmission and dissemination.

1.1.3.2 Oligosaccharide-based treatments for cancer

One disease has proven difficult to treat due to the fact that the diseased cells are nearly identical to healthy cells; this disease is cancer. Fortunately, cancerous cells are *nearly* identical to healthy cells, not *exactly*. One differentiating feature is the cell surface oligosaccharides. Cancerous cells may over- or under-express certain natural oligosaccharides or they may express oligosaccharides which are typically expressed only in embryonic tissues [11]. The change in oligosaccharide expression is not universal for all cancers; instead, it is specific for each type of cancer. Table 1.1 lists the known changes in oligosaccharide expression for various types of malignant tissue. These differences in carbohydrate expression allow for differentiation of cancerous cells from healthy cells and thus provide an avenue for targeted treatment of malignant tissue. Vaccines under development for breast cancer and small cell lung cancer link the cancer-associated oligosaccharide, Globo-H and Fucosyl GM1, with a highly immunogenic protein to recruit the immune system to attack malignant cells [1]. Other carbohydrate-based cancer vaccines under development include treatments for malignant melanoma, prostate cancer, non-small-cell lung carcinoma, small-cell lung carcinoma, and ovarian cancer [5]. Additional research and improvements are required before oligosaccharide-based cancer vaccines can be used in clinical application, yet these vaccines may become important weapons in the fight against cancer.

Table 1.1 Changes in oligosaccharide expression in malignant tissues [11]

Oligosaccharide	*Malignant Tissue*							
	Ovary	Pancreas	Breast	Colon	Brain	Prostate	Skin	Lung
sLe^x		X	X	X				X
sLe^a		X	X	X				X
sTn	X	X	X	X		X		X
TF	X		X	X		X		
Le^y	X	X	X	X		X		X
Globo H	X	X	X	X		X		X
PSA		X	X		X			X
GD2					X		X	
GD3					X		X	
Fucosyl GM1								X
GM2	X	X	X	X	X	X	X	X

With the promising research efforts on carbohydrates described above, companies have invested in carbohydrate ventures to capitalize on this new pharmaceutical tool. Large pharmaceutical companies like Amgen, Merck, and Wyeth have developed carbohydrate-based products that are now commercially available, and many new start-up companies have focused on carbohydrate science, including Oxford Glycosciences, GlycoTech, Oncothyreon (formerly Biomira), TheraCarb, and Synsorb Biotech [1, 5]. Unfortunately, several carbohydrate-based treatments, including Biomira's Theratope for breast cancer treatment and Oxford Glycosciences' Vevesca for treatment of Gaucher's disease, have fallen short of expectations during final stages of clinical testing. Despite these setbacks, several carbohydrate-based treatments are currently in market, including those listed in Table 1.2. Before the 1980's, very little was known about carbohydrate function, yet in just a few decades, carbohydrate-based medical treatments have advanced

for diseases including malaria, anthrax, HIV, cancer, and inflammation. As we continue to learn more about how carbohydrates contribute to biological processses, additional medical treatments and diagnostics will undoubtedly emerge as novel solutions to these complex medical problems.

Table 1.2 Commercially available carbohydrate-based treatments

Treatment	Function	Company
EPOGEN®	Stimulates red blood cell production	Amgen
Prevnar®	Treats bacteremia and meningitis	Wyeth
Menjugate®	Treats meningitis C	Chiron
NeisVac-C	Treats meningitis C	Baxter
Hib vaccine	Treats meningitis and pneumonia caused by *H. influenza* type B	Merck

1.2 Oligosaccharide Synthesis

1.2.1 Chemical synthesis

Oligosaccharides have been synthesized via chemical mechanisms since the development of the Koenigs-Knorr glycosylation in 1901 [12]. The Koenigs-Knorr reaction utilizes substituted sugars with a good leaving group at the C1 position and suitable protection groups added to the remaining hydroxyl groups. This sugar derivative is reacted with another sugar containing a free hydroxyl group to form a glycosidic bond (Figure 1.2). The leaving group is typically a halide, with bromide being the most preferred. The protecting groups can vary, yet acetate, benzoate, ester, and ether groups are commonly used [13]. Although it has been more than century since the advent of the Koenigs-Knorr reaction, this reaction remains the general method for chemical synthesis of oligosaccharides.

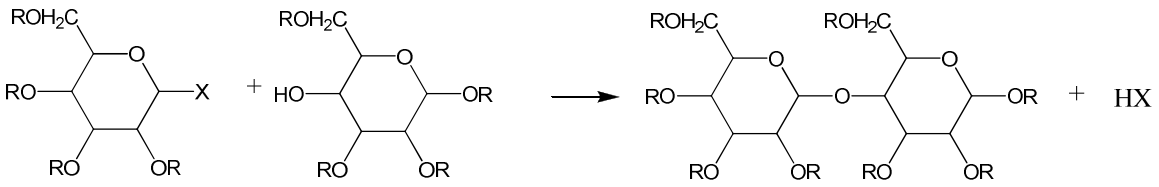


Figure 1.2 Koenigs-Knorr glycosylation reaction. X: leaving group; R: protection group.

There are many challenges in the chemical synthesis of oligosaccharides. Four main challenges are (1) numerous reactive hydroxyl groups (2) reactivity of the nucleophilic hydroxyl group (3) stereochemistry of glycosidic linkage and (4) yield [13]. From the chemical structure, it is clear that sugars have multiple hydroxyl groups with similar reactivity. This complicates the formation of a specific glycosidic bond, generating many byproducts. To dictate the regiochemistry of the glycosidic bond, protection groups are added to the hydroxyl groups that do not participate in the desired bond formation. To address the second challenge, a leaving group must be added to the substrate sugar to promote good reactivity. The Koenigs-Knorr reaction illustrates these solutions to the first two challenges. When glycosidic bonds are formed, there are two possible sterochemistries: α and β. To control stereochemistry, different protection groups are added, producing steric hindrance which favors either the α or β conformation [12]. As with any reaction, one important factor is the product yield. The individual steps for glycosidic bond formation often have high yields (50-90%), yet the additional steps of adding and removing the protecting and leaving groups leads to low overall yields. For example, chemical synthesis of an α-Gal epitope is shown in Figure 1.3. While this process reduces the number of steps by using substrates with protection groups already added, it still requires 12 steps to produce one α1,3-glycosidic linkage. Furthermore, the

overall yield of the product is less than 10%. The complex synthesis steps and low overall yields contribute to the high cost of the oligosaccharide product, and these issues remain as major challenges in chemical oligosaccharide synthesis. Furthermore, if multiple glycosidic bonds are required for producing larger oligosaccharides, these problems are compounded.

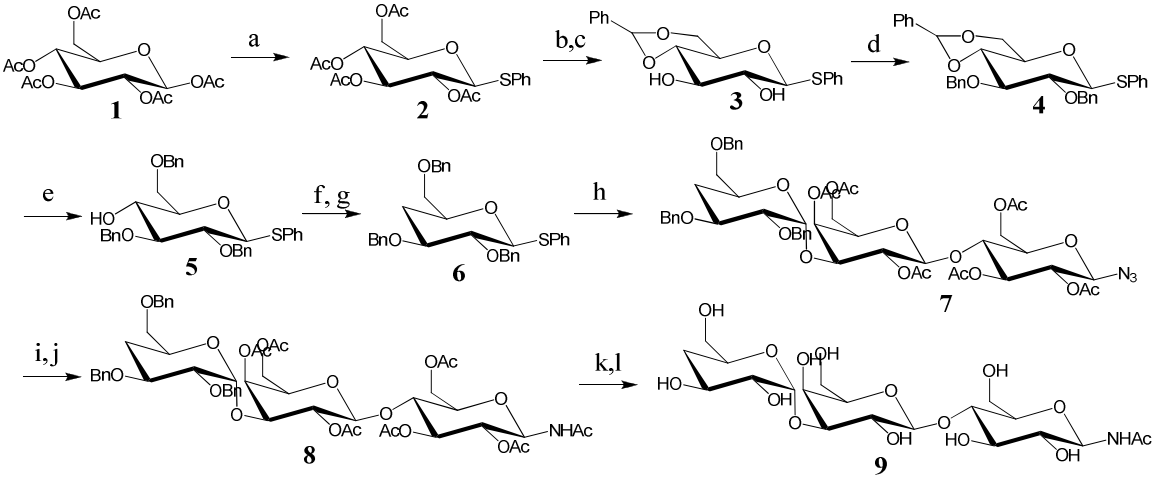


Figure 1.3 Reaction scheme for α-Gal epitope synthesis. Adapted from [14]. Yields: (a) 80%; (b,c) 86%; (d) 80%; (e) 53%; (f, g) 80%; (h) 60%; (i, j) 70%; (k, l) >99%.

1.2.2 Biological synthesis

1.2.2.1 Enzymatic synthesis

The challenges of chemical oligosaccharide synthesis have prompted scientists to turn towards another catalyst source: Mother Nature. To synthesize cell surface carbohydrates, why not use the same mechanism that naturally produces these compounds, enzymes. Enzymatic synthesis of oligosaccharides has many distinct advantages over chemical synthesis. Enzymes form glycosidic bonds in only one step. Hence, synthesis of the α-Gal epitope requires only one enzymatic step, compared to the twelve steps shown for chemical synthesis (Figure 1.3). This greatly improves the

possibility for industrial-scale synthesis. Enzymes can also form glycosidic bonds that are difficult or impossible to form chemically. For example, synthesis of α-sialyl and β-mannosyl linkages is kinetically unfavorable, complicating chemical synthesis [15]. Oligosaccharides containing these glycosyl linkages are biologically important and are needed for biological studies and applications. The natural enzymes, α2,3-sialyltranserase and β1,4-mannosyltransferase, can easily form these unfavorable bonds and produce the oligosaccharides required for advancing biological and medical applications.

The first challenge facing enzymatic oligosaccharide synthesis was obtaining a significant amount of enzyme. From the 1930's to 1970's, chemical synthesis was greatly favored over enzymatic synthesis. During this time, the enzymes for oligosaccharide synthesis were acquired by large-scale purification from their natural sources [16]. The cellular production of a particular cell surface oligosaccharide is relatively low, and consequently, there is a naturally low abundance of the requisite enzyme. This translates into low enzyme quantity after purification. The development of recombinant DNA technology in the 1980's allowed for large-quantity enzyme production. As a result, the interest in enzymatic methods for oligosaccharide synthesis drastically increased. Two types of enzymes have been studied and utilized for oligosaccharide synthesis: glycosidases and glycosyltransferases.

1.2.2.1.1 Glycosidases

Glycosidases have long been studied for their ability to break down oligo- and polysaccharides [17], yet application of glycosidases for oligosaccharide synthesis was not common until the 1950's when glycosidases such as levansucrase were used to produce carbohydrates like α-lactosyl-β-fructofuranoside and α-galactopyranosyl-β-

fructofuranoside [18, 19]. Since then, glycosidases from various sources have been employed to synthesize numerous oligosaccharides. Recently, a β-galactosidase from *Bacillus circulans* enabled synthesis of the medically-relevant diasaccharide, *N*-acetyllactosamine, with a 42% yield [20]. Other glycosidases used for oligosaccharide synthesis include β-glucosidase, α-galactosidase, α-fucosidase, β-mannosidase, and β-*N*-hexosaminidase [15, 21].

Glycosidases are enzymes that naturally catalyze oligosaccharide hydrolysis, but by altering conditions to shift the equilibrium to favor the reverse reaction, glycosidases can form glycosidic bonds [22]. A typical glycosidase-catalyzed reaction is shown in Figure 1.4. Several techniques are employed to shift the equilibrium in favor of the reverse reaction. Adding excess substrate, i.e. excess galactose or glucose in the glycosidase reaction in Figure 1.4, will push the equilibrium towards formation of the oligosaccharide. This is the most common approach; however, removal of the oligosaccharide product will also favor glycosidic bond formation [22]. Moreover, many glycosidases are solvent tolerant, allowing the enzymatic reaction to occur in a solvent instead of water. Eliminating or reducing the amount of water also shifts the equilibrium of the reaction towards oligosaccharide formation [15].

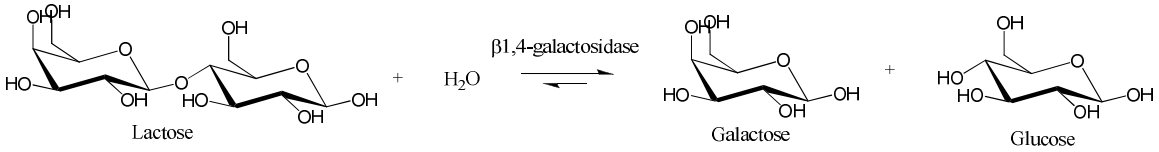


Figure 1.4 Glycosidase-catalyzed reaction.

Glycosidases have several advantages over other enzymes capable of producing glycosidic bonds, namely glycosyltransferases. Glycosidases are abundant in nature and are therefore readily available for oligosaccharide synthesis. The glycosyl donors for glycosidases are simple sugars which are inexpensive and do not require extensive modification, reducing the cost of synthesis [15, 23]. Furthermore, many glycosidases are thermally stable, allowing for higher synthesis temperatures [22]. While these advantages are appealing, there are also many drawbacks to glycosidase-catalyzed oligosaccharide synthesis. An inherent disadvantage of glycosidases is that the enzymes naturally favor oligosaccharide hydrolysis. This requires additional expense such as excess substrate or product removal to shift the equilibrium towards oligosaccharide synthesis, and despite these extra efforts, overall yields for glycosidase-catalyzed synthesis range from 20-60%, significantly lower than the yields achieved with glycosyltransferases [21]. The low yields are also caused by the lack of glycosidase regiospecificity. While glycosidases have high stereospecificity, the glycosidic bond is often formed at several different carbons on the acceptor, leading not only to lower yields but also a mixture of products [15, 23]. Thus, the lack of regiospecificity also complicates downstream oligosaccharide purification. Despite these disadvantages, glycosidases remain a viable enzymatic method for synthesizing oligosaccharides.

1.2.2.1.2 Glycosyltransferases

Glycosyltransferases are enzymes that naturally catalyze cellular glycosidic bond formation *in vivo*. A typical glycosyltransferase-catalyzed reaction is shown in Figure 1.5. Glycosyltransferases require sugar nucleotides as substrates for oligosaccharide synthesis. In 1950, Luis Leloir and colleagues discovered the first sugar nucleotide, UDP-

glucose [24]. Based on this discovery, glycosyltransferases requiring sugar nucleotide substrates are known as Leloir glycosyltransferases. Non-Leloir glycosyltransferases require sugar-1-phosphates as glycosyl donors, and these glycosyltransferases suffer from low yields and poor regioselectivity. Consequently, non-Leloir glycoysltransferases are not frequently used for oligosaccharide synthesis [25]. Therefore, this section will focus only on Leloir glycosyltransferases and any subsequent mention of a glycosyltransferase refers to the Leloir type.

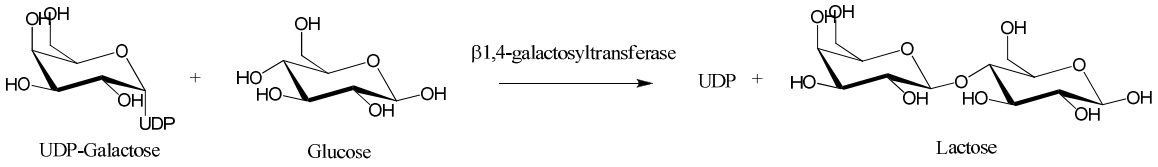


Figure 1.5 Glycosyltransferase-catalyzed reaction.

Although Leloir first discovered sugar nucleotides in 1950, the role of sugar nucleotides as glycosyl donors was not demonstrated until 1953 with the identification of UDP-glucuronic acid as donor for phenyl-β-D-glucopyranosiduronic acid synthesis and UDP-glucose as substrate for biosynthesis of trehalose phosphate [26, 27]. A few years later, Glaser and Brown provided evidence for the use of sugar nucleotide donors in the biosynthesis of polysaccharides such as hyaluronic acid [28]. Since the 1950's, glycosyltransferases have been pursued for their synthesis capabilities, yet their use in oligosaccharide synthesis has been limited by their availability. The pioneering studies on glycosyltransferases focused on enzymes from mammalian sources [29]. With the development of recombinant DNA technology, however, it soon became apparent that mammalian glycosyltransferases were not optimal for large-scale production.

Mammalian glycosyltransferases often require detergents for solubilization and retention of activity *in vitro*. They can also be difficult to purify and are often glycoproteins themselves, requiring a eukaryotic host organism for production [30]. Eukaryotic production of a glycosyltransferase typically has lower yields and higher cost compared to production in a prokaryotic host. Due to these complications, bacterial sources of glycosyltransferases were investigated. A major challenge in identifying bacterial glycosyltransferases arises from the fact that most oligosaccharides of interest are those related to human health. Therefore, one must identify bacterial glycosyltransferases that can synthesize the glycosidic linkages found in human oligosaccharides. This led many scientists to consider pathogenic bacteria as sources of glycosyltransferases. Many bacterial pathogens coat their cell surfaces with oligosaccharides that mimic those of their host in order to avoid detection by the immune system, and as a result, glycosyltransferases from pathogenic bacteria are often capable of producing the oligosaccharides found on human cell surfaces [30]. These bacterial glycosyltransferases are optimal for large-scale synthesis, as they can be produced in large quantities using fast growing bacterial hosts like *Escherichia coli*. A list of bacterial glycosyltransferases used for oligosaccharide production can be found in Table 1.3. All of the bacterial glycosyltransferases in Table 1.3 are derived from human pathogens or invasive bacteria with the exception of the β1,4-*N*-acetylglucosaminyltransferase from *R. leguminosarum*, which is a plant symbiont.

Table 1.3 Bacterial glycosyltransferases. Adapted from [25].

Glycosyltransferase	Stereo-chemistry	Gene	Source*
Galactosyltransferase			
	β1,4	*lgtB*	*N. meningitidis, N. gonorrhoeae*
	β1,4	*cps14J*	*S. pneumoniae*
	β1,4	*hpgalT*	*H. pylori*
	β1,3	*cgtB*	*C. jejuni*
	α1,4	*lgtC*	*N. meningitidis, N. gonorrhoeae*
N-acetylglucosaminyltransferase			
	β1,3	*lgtA*	*N. meningitidis, N. gonorrhoeae*
	β1,4	*nodC*	*R. leguminosarum*
N-acetylgalactosaminyltransferase			
	β1,3	*lgtD*	*N. meningitidis, N. gonorrhoeae*
	β1,4	*cgtA*	*C. jejuni*
Fucosyltransferase			
	α1,3	*HpfucT*	*H. pylori*
	α1,3	*fucT*	*H. pylori*
	α1,2	*fucT2*	*H. pylori*
	α1,3/4	*fucT*	*H. pylori*
Sialyltransferase			
	α2,3	*lst*	*N. meningitidis, N. gonorrhoeae*
	α2,3	*cst-I*	*C. jejuni*
	α2,3	*cst-II*	*C. jejuni*
	α2,6	*bst*	*P. damsel*
	α2,8/9	*neuS*	*E. coli*

* *C. – Campylobacter; E. – Escherichia; H. – Helicobacter; N.- Neisseria; P. – Photobacterium; R. – Rhizobium; S. – Streptococcus.*

As nature's chosen catalyst for oligosaccharide synthesis, glycosyltransferases have high stereo- and regio-selectivity, generating very few byproducts, if any. Glycosyltransferases also exhibit high yields generally ranging from 70-100% [23]. Moreover, unnatural oligosaccharides can be synthesized using glycosyltransferases. Mammalian glycosyltransferases, for example, have been shown to tolerate a range of

glycosyl acceptors and donors, with corresponding yields ranging from 30 to 99% [31]. Bacterial glycosyltransferases have demonstrated even greater flexibility, utilizing a broad range of acceptors, sometimes with only a slight decrease in reaction rate [30]. One main problem deters the application of glycosyltransferases for industrial production: the high cost associated with the sugar nucleotide substrate. Sugar nucleotides themselves are difficult to synthesize. Chemical synthesis of the requisite sugar nucleotides is a laborious process and one that is not feasible for industrial production. Enzymatic sugar nucleotide synthesis is also difficult, in that it requires additional purified enzymes and other compounds such as sugar phosphates, phosphoenolpyruvate, and nucleoside 5'triphosphates which are not readily available [25]. The high cost of substrates and cofactors required for enzymatic oligosaccharide synthesis often prevent large-scale oligosaccharide production utilizing this method.

1.2.2.2 Whole-cell synthesis

Both enzymatic and whole-cell biocatalysts have been employed to synthesize oligosaccharide structures. However, whole-cell strategies offer several advantages. They avoid the difficulties of enzyme isolation and purification and do not require extensive downstream separation processes to isolate the desired product, as whole-cells can be easily removed by centrifugation. Additionally, enzymatic synthesis requires expensive sugar nucleotides as substrate while whole-cell techniques generate the sugar nucleotide *in vivo* from a less expensive carbon source [32]. With these advantages, whole-cell biocatalysis offers enormous potential for improving oligosaccharide synthesis.

1.2.2.2.1 Metabolic Engineering

Whole-cell oligosaccharide synthesis is made feasible through the application of metabolic engineering. Metabolic engineering is a relatively new method of strain improvement, emerging from the development of recombinant DNA technology in the 1970's and 1980's. The term 'metabolic engineering' was first coined by James Bailey, Gregory Stephanopoulos, and Joseph Vallino in the June publication of *Science* in 1991 [33, 34]. Metabolic engineering is the application of recombinant DNA technology to rationally alter cellular genetics for targeted improvement of cell function(s). Metabolic engineering strategies rely on previously acquired knowledge of an organism's genetics, metabolism, and gene-function relationships. With this knowledge, the cell's metabolism can be genetically engineered to bring about a desired phenotype, typically the improved production of a target metabolite.

There are two main resources which must be supplied for whole-cell oligosaccharide synthesis: carbon and energy. Oligosaccharides are composed of monosaccharide subunits, and as a result, the production of these carbohydrates is carbon-intensive, competing with the cell's metabolism for available carbon sources. Therefore, any metabolic engineering effort must find a balance between the cell's metabolic needs and the carbon requirement for oligosaccharide synthesis. Moreover, oligosaccharide synthesis requires sugar nucleotides as precursor. These sugar nucleotides are also major components in cell wall biosynthesis, and so, carbohydrate production directly competes with cell growth [35]. When a cell is forced to choose between carbohydrate production and cell growth, oligosaccharide synthesis is often minimized, producing only the quantities necessary for cell survival and growth. As a

result, there remains significant potential for improving oligosaccharide synthesis through the application of metabolic engineering.

Production of the sugar nucleotide precursors required for oligosaccharide synthesis interfaces with the cell's primary metabolism. Figure 1.6 depicts the metabolic pathway in *E. coli* for production of two sugar nucleotides, UDP-glucose and UDP-galactose. Sugar nucleotide biosynthesis clearly effects the glycolysis pathway, which generates energy for the cell and provides precursors for other metabolites produced by the cell. An optimal flux distribution must be determined to find a balance between the carbon supply for glycolysis and carbohydrate synthesis. Through natural evolution, the carbon flux distribution has been optimized with the objective of promoting cell growth and survival. The flux distribution for maximum oligosaccharide synthesis, however, may be significantly different from the naturally evolved fluxes. Metabolic engineering provides the tools necessary for manipulating the carbon flux distribution to optimize oligosaccharide and polysaccharide production.

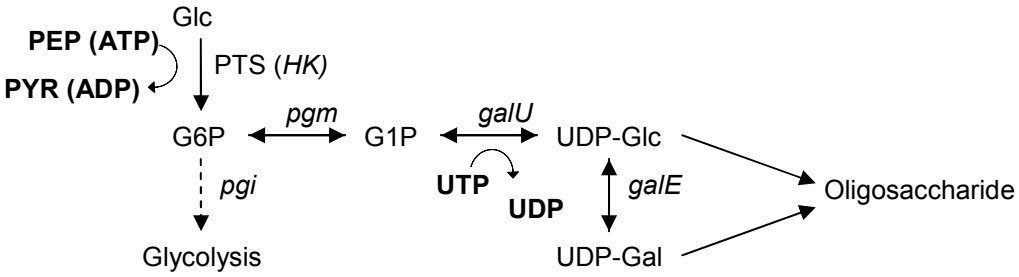


Figure 1.6 UDP-glucose and UDP-galactose synthesis pathway in *E. coli.*

Oligosaccharide synthesis is not only carbon-intensive; it is also an energy-intensive process. As shown in Figure 1.6, two high-energy compounds are consumed for biosynthesis of each sugar nucleotide. One high-energy compound, either ATP or

phosphoenolpyruvate (PEP) depending upon the sugar transport system, is used to activate the sugar component. This step is illustrated by the conversion of glucose to glucose-6-phosphate in Figure 1.6. A second high-energy compound, UTP or another nucleoside triphosphate, is required for synthesis of the sugar nucleotide precursor. This energy requirement makes oligosaccharide and polysaccharide synthesis particularly costly for the cell. Additionally, the sugar component of the sugar nucleotide represents an unutilized carbon source, which has the potential to yield up to 38 ATP molecules through glycolysis, the TCA cycle, and oxidative phosphorylation [36]. Taking into account both the energy requirement of sugar nucleotide biosynthesis and the potential energy loss from an unutilized carbon source, a cell will lose up to 40 high-energy compounds for each sugar nucleotide synthesized in oligosaccharide production. This energy loss impacts many metabolic reactions throughout the cell's metabolism. Just as the carbon flux distribution must be optimized, the cell's energy production may also need modification for optimal oligosaccharide synthesis. Again, metabolic engineering offers a means of achieving this goal.

1.2.2.2.2 Single-cell synthesis

Single-cell synthesis of oligosaccharides requires the use of glycosyltransferases for glycosidic bond synthesis as glycosidases display hydrolytic activity *in vivo*. As a result, the main concerns in whole cell oligosaccharide synthesis are similar to those of glycosyltransferase-catalyzed synthesis, namely provision of the sugar nucleotide donor and acceptor.

An important consideration in single-cell oligosaccharide production is the selection of the host strain. The most common microorganism employed for single-cell

oligosaccharide synthesis is *Escherichia coli*. *E. coli* is a fast-growing organism commonly which is commonly used for the industrial production of many compounds. Furthermore, the tools for genetic manipulation of *E. coli* are well-established and commercially available. *E. coli* has been successfully engineered to produce a variety of oligosaccharides including *N*-acetllactosamine (LacNAc), globotriose, α-Gal epitope, the carbohydrate moieties of gangliosides GM1 and GM2, and several other higher order oligosaccharides [32, 37-40]. While recombinant *E. coli* have been successful in synthesizing the desired oligosaccharide, the product concentrations are quite low, typically less than 1 g/L. Due to the low yields in *E. coli*, others have chosen to engineer microorganisms that naturally produce the oligosaccharide product. For example, *Corynebacterium glutamicum* has been metabolically engineered to improve production of the trisaccharide, trehalose. Trehalose synthesis in *C. glutamicum* was increased nearly 10-fold using metabolic engineering, resulting in nearly 10 g/L of trehalose [41]. While trehalose synthesis was successfully enhanced by engineering its natural host, many other oligosaccharides cannot be synthesized in significant quantities from their natural hosts. Many microorganisms cannot be grown in the laboratory using available growth media, and many also have slow growth rates. In addition, the tools available for genetic manipulation may not be applicable to the natural host strain. The difficulty of finding a host strain for efficient oligosaccharide synthesis remains a major obstacle in single-cell oligosaccharide production.

1.2.2.2.3 Bacterial coupling

Despite successful oligosaccharide synthesis with single-cell engineered strains, a multiple-strain approach was developed to meet the industrial demand for high product

concentrations. This approach recognizes the need for an abundant supply of high-energy compounds, in particular nucleoside triphosphates, for production of sugar nucleotide precursors. A bacterial strain with efficient nucleoside triphosphate synthesis is employed to maximize the energy supply. Pairing this strain with recombinant strains possessing the enzymes for sugar nucleotide synthesis and a glycosyltransferase allows for efficient oligosaccharide production. This strategy of combining bacterial strains of specialized function is known as bacterial coupling. Bacterial coupling is analogous to an assembly line, where each bacterial strain is selected and engineered to perform one task to the best of its ability, and the product of this task is transferred to the next strain which then performs its specific task. With this strategy, bacterial coupling maximizes each step of oligosaccharide synthesis by engineering each strain for optimal production of an essential component. It also reduces the metabolic burden imposed by metabolic engineering through using several strains for recombinant protein expression.

Bacterial coupling was first applied to oligosaccharide synthesis in 1998 for production of the trisaccharide, globotriose [42]. This application uses three bacterial strains: *Corynebacterium ammoniagenes* for production of UTP from orotic acid, recombinant *E. coli* containing genes for UDP-galactose synthesis, and another recombinant *E. coli* strain possessing an α1,4-galactosyltransferase (Figure 1.7). Coupling of these three strains enabled production of 266 mM (134 g/L) of globotriose with lactose supplied as the acceptor. Similar strategies was used to produce the disaccharide, *N*-acetyllactosamine, sialylated oligosaccharides, and a Lewis x trisaccharide [43-45]. Overall product concentrations with bacterial coupling were high, ranging from 20 to 140 g/L.

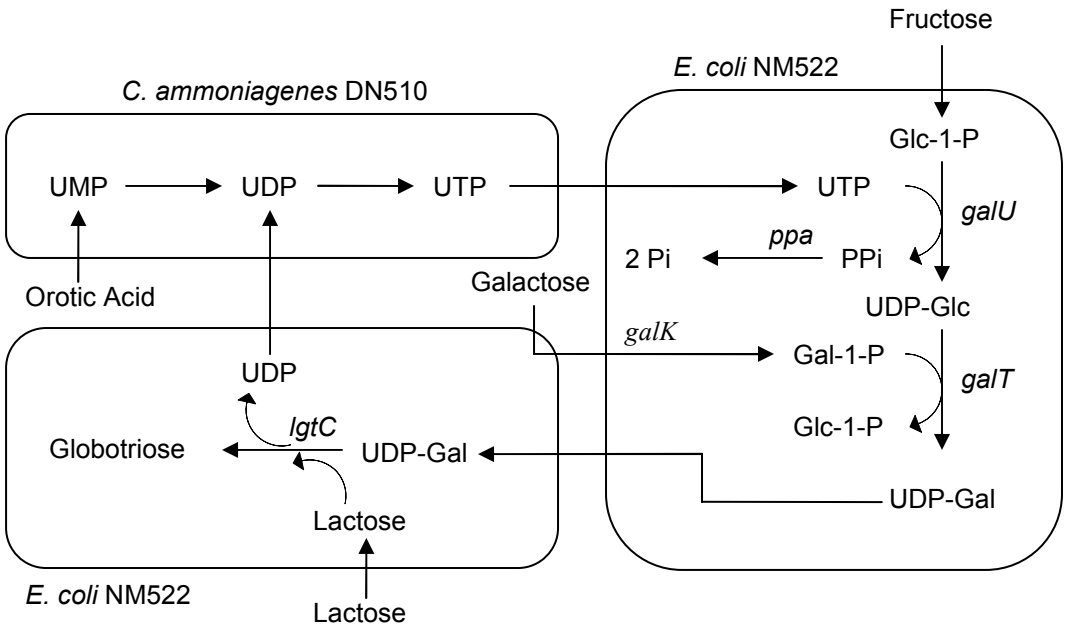


Figure 1.7 Bacterial coupling system for globotriose production. Genes named in this figure represent recombinant enzymes while all unidentified genes are native enzymes.

While bacterial coupling yields high concentrations of the oligosaccharide product, there are several drawbacks to this approach. Cell permeability is a major issue. In bacterial coupling, the product of one strain must be transported out of the producing cell and into the next strain where it will be used as substrate. With three to five bacterial strains involved in each system, cell membrane permeability posses a significant limitation, particularly when large, highly charged molecules such as nucleoside triphosphates must be transported. To address this problem, a surfactant, polyoxyethylene octadecylamine, and xylene were added to the reaction mixture [42-45]. Chemically-induced cell membrane permeability is accompanied by a dissipation of the membrane potential which will likely reduce the energy capacity of the cell. Surprisingly, permeabilization in the reported bacterial coupling examples did not seem to significantly impair the high-energy requiring oligosaccharide production. Bacterial coupling is also

inherently expensive from an industrial perspective. Each system requires three to five bacterial strains which must be grown in separate fermentations prior to use in oligosaccharide synthesis. So for each glycosidic linkage, at least four fermentations are required: one for growth of each microorganism and one for oligosaccharide production. Another aspect that may prevent the industrial application of bacterial coupling is the high cost of raw materials. Expensive cofactors such as orotic acid, GMP, and sialic acid as well as multiple carbon sources are required for the reported bacterial coupling strategies. These disadvantages make bacterial coupling less than ideal, yet the high levels of oligosaccharide synthesis are unmatched by any other whole-cell synthesis technique.

1.3 *Agrobacterium* sp. ATCC 31749

1.3.1 Phylogeny of the *Agrobacterium* genus

Members of the genus *Agrobacterium* are α-proteobacteria of the Rhizobiales order and Rhizobiaceae family. Rhizobiaceae are a diverse group of aerobic, rod-shaped bacteria classified based on their 16S rRNA gene sequence. Genera of the Rhizobiaceae family include *Rhizobium, Carbophilus, Chelatobacter, Ensifer,* and *Sinorhizobium* in addition to *Agrobacterium* [46]. Of these genra, *Rhizobium* and *Sinorhizobium* appear to be the most closely related to *Agrobacterium*. *Rhizobium, Sinorhizobium,* and *Agrobacterium* include microorganisms commonly found in soil which are known for their association with plants, although not all species are plant-associated. *Rhizobium* and *Sinorhizobium* have similar phenotypes and are known to invade the root hairs of leguminous plants, forming root nodules. This relationship is symbiotic, where the plant provides the *Rhizobium* or *Sinorhizobium* with nutrients and the bacterium fixes nitrogen

for the plant. On the other hand, *Agrobacterium* species are well-known as plant pathogens. Some *Agrobacterium* species infect plants, causing the formation of tumors known as gall. While famous for their plant interactions, it has long been known that the genes responsible for *Rhizobium* symbiosis and *Agrobacterium* pathogenesis reside on transferrable megaplasmids, not chromosomal DNA [47, 48]. This has sparked a great debate over the classification of *Rhizobium, Sinorhizobium,* and *Agrobacterium* based on their relation to plants [49-51].

Within the *Agrobacterium* genus, there is also confusion regarding the taxonomy and phylogeny of species. There are five species of *Agrobacterium* dictated by the disease phenotype of the organism. *A. tumefaciens* is the most well-known species, and it causes crown gall in over 140 species of dicotyledonous plants. *A. rubi* causes crown gall in raspberries, while *A. vitis* causes gall in grapevines. *A. rhizogenes* is the only phytopathogenic species that does not cause gall, but instead, causes a condition known as hairy root. *A. radiobacter* is the only non-phytopathogenic species of *Agrobacterium* [52]. An alternative classification for *Agrobacteria* was proposed based on physiological and biochemical properties, with no influence from the disease phenotype. This classification segregates *Agrobacteria* into three biovars [53]. Both the species and biovar classifications are employed throughout the literature, yet with the increase in sequenced genomes from *Agrobacterium*, there have been recent efforts to reorganize the species classifications of this genus.

Perhaps the most well known *Agrobacterium* is *A. tumefaciens*. *A. tumefaciens* has agricultural significance as the causative agent for crown gall disease in dicotyledonous plants. This species is also exploited for the genetic engineering of plants.

On the tumor inducing (Ti) plasmid of *A. tumefaciens,* there are genes known as T-DNA which allow the transfer of genetic material from *A. tumefaciens* to the infected plant. By replacing the transferred genes of the T-DNA with a chosen gene sequence, *A. tumefaciens* will transfer the chosen gene to the plant, allowing for the generation of transgenic crops. With its significance as a plant pathogen and tool for gene transfer, *A. tumefaciens* was the first species sequenced from the *Agrobacterium* genus [54, 55].

As the *Agrobacterium* genus is primarily known for its phytopathogenicity, very few studies have focused on the metabolism of this genus. Radiorespirometric and enzymatic activity studies performed in the 1970's revealed that *Agrobacteria* primarily use the Entner-Doudoroff (ED) pathway for glucose metabolism, not the energetically favorable Embden Meyerhof Parnas (EMP) pathway [56, 57]. Two enzymes in the EMP pathway were found to have very low activity, phosphofructokinase and fructose-1,6-diphosphate aldolase. In addition to the ED pathway, the pentose phosphate pathway (PPP) was found to be important in glucose catabolism. The percentage of glucose flux through the ED pathway ranged from 55-88% while the remaining 45-12% was diverted through the PPP [57]. Furthermore, *A. tumefaciens* was shown to have functional tricarboxylic acid (TCA) and glyoxylate cycles [56]. The primary metabolism of microorganisms in the *Agrobacterium* genus consists of the following pathways: ED, PPP, TCA cycle, and glyoxylate cycle.

1.3.2 History of ATCC 31749

The natural parent strain of *Agrobacterium* sp. ATCC 31749 was first isolated from soil in 1964 by Harada and Yoshimura [58]. This bacterium was classified as *Alcaligenes faecalis* var. *myxogenes* 10C3, and it was initially of interest for its

production of a polysaccharide containing succinic acid, which was subsequently named succinoglycan. A spontaneous mutant derived from *Alcaligenes faecalis* var. *myxogenes* 10C3 was found to produce a β1,3-glucan polysaccharide which forms a gel when heated in an aqueous suspension [59]. This polysaccharide was called curdlan [59, 60]. In 1977, another spontaneous mutant was obtained which produced only curdlan and no succinoglycan. This strain was named 10C3K, and it appeared to be genetically stable [61]. Shortly thereafter, a uracil auxotroph was derived from strain 10C3K and designated as *Alcaligenes faecalis* var. *myxogenes* IFO 13140 (ATCC 21680). In 1983, Phillips, Lawford, and coworkers isolated a mutant from ATCC 21680 which is not a uracil auxotroph, is unable to hydrolyze starch, and can grow using citric acid as the sole carbon source; this strain was submitted as ATCC 31749 [62, 63]. ATCC 31750, a spontaneous mutant of ATCC 31749, produces only water-insoluble curdlan as opposed to ATCC 31749 which produces both water-soluble and water-insoluble curdlan. Lawford obtained a patent for a continuous process for the production of curdlan from ATCC 31749 and ATCC 31750 in 1982 [64]. In 1991, all of these strains, originally classified as *Alcaligenes*, were reclassified as *Agrobacterium* based on 16S rRNA sequence [63]. Prior to the work detailed in this dissertation, ATCC 31749 has only been utilized for curdlan production.

1.3.2.1 Curdlan polysaccharide

In addition to ATCC 31749 and its parent strains, other microorganisms have been shown to produce curdlan. Detectable amounts of curdlan were purified from other Rhizobia, including *A. radiobacter* (IFO 12607, 12665, 13127, and 13256), *A. rhizogenes* (IFO 13259), *Rhizobium trifolii* J60, and *Rhizobium* sp. TISTR 64B. In addition to these

Gram-negative bacteria, several Gram-positive strains are also capable of producing curdlan, namely *Cellumonas spp.* and *C. flavigena* KU [65]. Of all the curdlan-producing strains identified so far, ATCC 31749 is the most studied and has been shown to produce the highest amount of curdlan.

As mentioned above, curdlan is a linear β1,3-glucose polymer containing up to 12,000 glucose monomers. Curdlan and other β1,3-glucans can be detected using an Aniline Blue staining which produces a blue color when the dye binds to the β1,3-glucan [66]. Most forms of curdlan are insoluble in water but readily dissolve in basic solutions. This property is commonly used to isolate and purify curdlan. Curdlan is well-known for its ability to form gels upon heating. By controlling the temperature of the heating process, two different types of gel can be formed from aqueous curdlan suspensions. High-set, thermo-irreversible gels are formed by heating above 80°C and then cooling, while low-set, thermo-reversible gels are produced by heating to 55°C, followed by cooling. High-set gels differ structurally from low-set gels in that they contain triple-stranded helices of curdlan. Low-set gels primarily contain single-stranded helices [65]. This unique gel-forming ability distinguishes curdlan from other glucan polysaccharides.

Initial applications of curdlan focused on this gel-forming ability. Curdlan is extensively employed in the food industry as a gelling agent, thickener, extender, binding agent, aroma masker, water-holding agent, and coating agent. With these functions, curdlan has been used as an additive in many foods including jelly, pudding, tofu, noodles, hamburger, sausage, frozen egg products, boiled rice, and salad dressings. Curdlan is not easily degraded by human digestive enzymes, allowing for applications in calorie-reduced and diabetic food products. Moreover, curdlan gels resist degradation due

to freeze and thaw cycles, suggesting potential for frozen food applications. Curdlan also has applications in the agricultural industry. Curdlan has been introduced to animal and fish feed to inhibit intestinal putrefying bacteria that cause animals to age and to improve immune activity [63]. In addition, curdlan has many medical applications. As a gel, curdlan can serve as a drug release vehicle and coating. In fact, a patent was filed for the use of curdlan as a sustained release suppository for drugs such as indomethacin, diclofenac sodium, and ibuprofen [67]. Additional patents were also filed using curdlan as an inhibitor of blood coagulation and as a treatment for dementia [68, 69]. Extensive research has been conducted to investigate curdlan sulfate as an anti-HIV treatment, and in 1994, phase I clinical trials suggested that curdlan sulfate may have therapeutic effects [70]. Additionally, sulfoalkyl derivatives of curdlan have been shown to have antitumor activity [71]. There are many more applications for curdlan; in fact, 171 patents have been filed involving the use of curdlan. These numerous applications have led to a high demand for curdlan, and consequently, Takeda Chemical Industries Ltd. produces 600 to 700 tons of curdlan each year [63].

1.3.2.2 Curdlan production in ATCC 31749

ATCC 31749 has been studied exclusively for its natural production of curdlan polysaccharide. In the 1980s, initial studies determined that curdlan production was triggered by nitrogen limitation in ATCC 31749 [72]. Since then, many parameters have been determined to influence curdlan production in this strain. Numerous carbon sources were utilized by ATCC 31749, with overall results indicating that sucrose and maltose yield the highest curdlan production [73]. As with many fermentations, the pH and level of dissolved oxygen were important factors for the production of curdlan by

Agrobacterium sp. The optimum pH for growth was determined to be pH 7 while pH 5.5 led to higher amounts of curdlan during nitrogen-limited conditions [74]. High levels of dissolved oxygen resulted in high curdlan production [75]. This result is not surprising as high levels of dissolved oxygen allow for increased energy production through the TCA cycle. In turn, the additional energy is used for the energy-intensive process of polysaccharide synthesis. The concentration of residual phosphate also influences curdlan production. Both high and very low phosphate concentrations yielded low curdlan production, but a residual phosphate concentration of 0.5g/L led to curdlan production as high as 65g/L [76]. The analysis of intracellular nucleotides during curdlan production showed high concentrations of UMP, and UMP or uracil supplementation of curdlan fermentations showed increased curdlan production [77, 78]. Again, this is not surprising as both uracil and UMP are precursors for the production of UDP-glucose, the sugar nucleotide required for curdlan production. Overall, these studies show that curdlan production is regulated by nitrogen limitation, pH, and phosphate concentration, and curdlan synthesis in ATCC 31749 is limited by energy and nucleotide availability, not the carbon flux or sugar moiety.

Since most studies have focused on the optimization of conditions for the industrial production of curdlan, little is known regarding the genetics and molecular regulation of curdlan production. Four genes have been identified as being essential for curdlan production in ATCC 31749. A curdlan synthase gene (*crdS*) was characterized as the enzyme responsible for transferring glucose from the sugar nucleotide, UDP-glucose, to the growing β1,3-glucan polymer chain [79]. In the same putative operon, there are two other genes believed to contribute to curdlan synthesis; these were designated as

crdA and *crdC*. These genes have no significant homology to known protein sequences, and their role in curdlan production is still unknown [80]. A phosphatidylserine synthase gene (pss_{AG}) was shown to be required for the production of high molecular weight curdlan, indicating the importance of membrane composition in curdlan synthesis [81]. A putative regulatory gene, *crdR*, was also necessary for curdlan production, yet this gene still remains to be characterized. In addition to identifying the genes involved in curdlan production, there is also an effort to uncover the regulatory mechanisms controlling curdlan production. As nitrogen limitation is necessary for curdlan production, the genes involved in nitrogen regulatory systems were investigated. Mutants of *ntrBC* failed to produce curdlan, suggesting that the nitrogen signaling cascade may regulate transcription of the curdlan synthesis operon. However, no RpoN-dependent promoter has been found for the *crd* operon [65]. The mechanism for enhanced curdlan production at pH 5.5 is also an area of interest. A proteomic study of *Agrobacterium* sp. comparing protein levels at pH 7 and pH 5.5 showed higher amounts of the proteins involved in the curdlan synthesis pathway, including curdlan synthase, UTP-glucose-1-phosphate uridylyltransferase, and phosphoglucomutase [82]. While these recent genetic and proteomic studies have identified genes important for curdlan production, much work still remains in characterizing these genes and elucidating the regulation of curdlan synthesis in *Agrobacterium* sp.

1.4 Project Objectives

The work described in this dissertation focuses on two main objectives: 1) to develop a whole-cell biocatalyst for oligosaccharide synthesis utilizing ATCC 31749 as host and 2) to determine the factors influencing curdlan production in ATCC 31749.

1.4.1 Metabolic engineering of ATCC 31749 for oligosaccharide synthesis

The production of large amounts of curdlan polysaccharide make *Agrobacterium* sp. strain ATCC 31749 an ideal host for oligosaccharide synthesis. ATCC 31749 has been shown to produce up to 93 g/L of curdlan [78]; this requires the microorganism to produce 292 g/L of the sugar nucleotide substrate, UDP-glucose. UDP-glucose is easily converted into other sugar nucleotides, such as UDP-galactose, with the introduction of just one additional enzyme. In turn, a glycosyltransferase can use UDP-galactose to produce galactose-containing oligosaccharides. Thus, by introducing only two additional enzymes into ATCC 31749, the microorganism can be transformed into a whole-cell biocatalyst for oligosaccharide synthesis. With the high amount of sugar nucleotide available in the *Agrobacterium* host, there is potential for producing oligosaccharides at concentrations greater than 100 g/L. Since ATCC 31749 utilizes a cheap carbon source like sucrose, the cost of producing oligosaccharides can be greatly reduced compared to the current chemical and enzymatic methods of synthesis. Furthermore, the engineered ATCC 31749 would only require one fermentation as opposed to the 3 to 5 fermentations required by bacterial coupling, making the *Agrobacterium* biocatalyst ideal for industrial production. By increasing oligosaccharide availability and reducing the cost associated with oligosaccharide synthesis, the *Agrobacterium* biocatalyst will advance the development of oligosaccharide-based medical treatments and make current oligosaccharide treatments more affordable.

1.4.2 What factors affect production of the curdlan polysaccharide in ATCC 31749?

Oligosaccharide synthesis in the *Agrobacterium* biocatalyst will employ the same metabolic pathways as curdlan production, and therefore, the pathways and

corresponding regulatory systems of curdlan production will also influence oligosaccharide synthesis. Information regarding curdlan production can thereby be exploited to enhance oligosaccharide synthesis. As discussed in section 1.3.2.2, very little information is available about how ATCC 31749 produces such large quantities of curdlan polysaccharide. By identifying the factors which allow for high polysaccharide production, I will also identify metabolic engineering targets for improving oligosaccharide synthesis in ATCC 31749. For example, if a transcriptional activator governs curdlan production under nitrogen-limited conditions, this activator can be overexpressed during nitrogen-rich conditions, allowing for curdlan and oligosaccharide production during the growth phase. Unique energy-generating genes or pathways may also contribute to curdlan production, and once identified, the corresponding genes can be manipulated to enhance oligosaccharide synthesis in the engineered *Agrobacterium* biocatalyst. In addition to these potential benefits for oligosaccharide production, this investigation will contribute to the fundamental knowledge and understanding of exopolysaccharide production and regulation, an area that is not well understood.

1.5 References

1. Alper, J., R.F. Service, and M. Balter, *Searching for medicine's sweet spot.* Science, 2001. **291**(5512): p. 2338-2343.

2. Weiss, A.A. and S.S. Lyer, *Glycomics aims to interpret the third molecular language of cells.* Microbe, 2007. **2**(10): p. 489-497.

3. Saxon, E. and C.R. Bertozzi, *Chemical and biological strategies for engineering cell surface glycosylation.* Annual Review of Cell and Developmental Biology, 2001. **17**: p. 1-23.

4. Linhardt, R.J., *Heparin: An important drug enters its seventh decade.* Chemistry & Industry, 1991. **2**: p. 45-50.

5. Borman, S., *Carbohydrate vaccines: Novel chemical and enzymatic oligosaccharide synthesis techniques could lead to a new generation of carbohydrate-based vaccine agents.* Chemical and Engineering News, 2004. **August, 9**: p. 31-35.

6. Seeberger, P.H. and D.B. Werz, *Synthesis and medical applications of oligosaccharides.* Nature, 2007. **446**: p. 1046-1051.

7. Daubenspeck, J.M., et al., *Novel oligosaccharide side chains of the collagen-like region of BclA, the major glycoprotein of the Bacillus anthracis exosporium.* Jounral of Biological Chemistry, 2004. **279**(30): p. 30945-30953.

8. Tarasenko, O., et al., *Polymeric glycoconjugates protect and activate macrophages to promote killing of Bacillus cereus spores during phagocytosis.* Glycoconjugate Journal, 2008. **25**: p. 473-480.

9. Werz, D.B. and P.H. Seeberger, *Total synthesis of antigen Bacillus anthracis tetrasaccharide - Creation of an anthrax vaccine candidate.* Angewandte Chemie (International ed.), 2005. **44**(39): p. 6315-6318.

10. Scanlan, C.N., et al., *Exploiting the defensive sugars of HIV-1 for drug and vaccine design.* Nature, 2007. **446**: p. 1038-1045.

11. Dube, D.H. and C.R. Bertozzi, *Glycans in cancer and inflammation - potential for therapeutics and diagnostics.* Nature Reviews, 2005. **4**: p. 477-488.

12. Bertozzi, C.R. and L.L. Kiessling, *Chemical glycobiology.* Science, 2001. **291**: p. 2357-2364.

13. Flowers, H.M., *Chemical synthesis of oligosaccharides.* Methods in Enzymology, 1978. **50**: p. 93-121.

14. Janczuk, A.J., et al., *The synthesis of deoxy-α-Gal epitope derivatives for the evolution of an anti-α-Gal antibody binding.* Carbohydrate Research, 2002. **337**: p. 1247-1259.

15. Palcic, M.M., *Biocatalytic synthesis of oligosaccharides.* Current Opinion in Biotechnology, 1999. **10**: p. 616-624.

16. Ichikawa, Y., G.C. Look, and C.H. Wong, *Enzyme-catalyzed oligosaccharide synthesis.* Analytical Biochemistry, 1992. **202**(2): p. 215-238.

17. Coley, N.G. and M. Keynes, *History of Biochemistry*, in *Encyclopedia of Life Sciences*. 2002, John Wiley & Sons.

18. Avigad, G., *Enzymatic synthesis and characterization of a new trisaccharide, α-lactosyl-β-fructofuranoside* The Journal of Biological Chemistry, 1957. **1**: p. 121-129.

19. Feingold, D.S., G. Avigad, and S. Hestrin, *Enzymatic synthesis and reactions of a sucrose isomer α-D-galactopyranosyl-β-D-fructofuranoside* The Journal of Biological Chemistry, 1957. **1**: p. 295-307.

20. Vetere, A. and S. Paoletti, *High-yield synthesis of N-aceytllactosamine by regioselective transglycosylation.* Biochemical and Biophysical Research Communications, 1996. **219**: p. 6-13.

21. Crout, D.H. and G. Vic, *Glycosidases and glycosyl transferases in glycoside and oligosaccharide synthesis.* Current Opinion in Chemical Biology, 1998. **2**(1): p. 98-111.

22. Monsan, P. and F. Paul, *Enzymatic synthesis of oligosaccharides.* FEMS Microbiology Reviews, 1995. **16**: p. 187-192.

23. Watt, G.M., P.A.S. Lowden, and S.L. Flitsch, *Enzyme-catalyzed formation of glycosidic linkages.* Current Opinion in Structural Biology, 1997. **7**: p. 652-660.

24. Caputto, R., et al., *Isolation of the coenzyme of the galactose phosphate-glucose phosphate transformation.* The Journal of Biological Chemistry, 1950. **184**: p. 333-350.

25. Endo, T. and S. Koizumi, *Large-scale production of oligosaccharides using engineered bacteria.* Current Opinion in Structural Biology, 2000. **10**: p. 536-541.

26. Dutton, G.J. and I.D.E. Storey, *The isolation of a compound of uridine diphopshate and glucuronic acid from liver.* Biochemical Journal, 1953. **53**: p. xxxvii-xxxviii.

27. Leloir, L.F. and E. Cabib, *The enzymatic synthesis of trehalos phosphate.* Journal of American Chemical Society, 1953. **75**(21): p. 5445-5446.

28. Glaser, L. and D.H. Brown, *The enzymatic synthesis in vitro of hyaluronic acid chains.* Proceedings of the National Academy of Sciences of the United States of America, 1955. **41**: p. 253-260.

29. Spiro, R.G., *Glycoproteins.* Annual Review of Biochemistry, 1970. **39**: p. 599-638.

30. Johnson, K.F., *Synthesis of oligosaccharides by bacterial enzymes.* Glycoconjugate Journal, 1999. **16**: p. 141-146.

31. Ernst, B. and R. Oehrlein, *Substrate and donor specificity of glycosyl transferases.* Glycoconjugate Journal, 1999. **16**: p. 161-170.

32. Bettler, E., et al., *The living factory: In vivo production of N-acetyllactosamine containing carbohydrates in E. coli.* Glycoconjugate Journal, 1999. **16**: p. 205-212.

33. Bailey, J.E., *Toward a science of metabolic engineering.* Science, 1991. **252**: p. 1668-1675.

34. Stephanopoulos, G. and J.J. Vallino, *Network rigidity and metabolic engineering in metabolite overproduction.* Science, 1991. **252**: p. 1675-1681.

35. Ramos, A., et al., *Relationship between glycolysis and exopolysaccharide biosynthesis in Lactococcus lactis.* Applied and Environmental Microbiology, 2001. **67**(1): p. 33-41.

36. Voet, D. and J.G. Voet, *Biochemistry*. 3rd ed. Vol. 1. 2004: John Wiley & Sons, Inc. 1137.

37. Antoine, T., et al., *Large-Scale in vivo synthesis of the carbohydrate moieties of gangliosides GM1 and GM2 by metabolically engineered Escherichia coli.* ChemBioChem, 2003. **4**: p. 406-412.

38. Bettler, E., et al., *Production of recombinant xenotransplantation antigen in Escherichia coli.* Biochemical and Biophysical Research Communications, 2003. **302**: p. 620-624.

39. Dumon, C., et al., *In vivo fucosylation of lacto-N-neotetraose and lacto-N-neohexaose by heterologous expression of Helicobacter pylori α-1,3 fucosyltransferase in engineered Escherichia coli.* Glycoconjugate Journal, 2001. **18**: p. 465-474.

40. Zhang, J., et al., *Large-scale synthesis of globotriose derivatives through recombinant E. coli.* Org. Biomol. Chem., 2003. **1**: p. 3048-3053.

41. Carpinelli, J., R. Kramer, and E. Agosin, *Metabolic engineering of Corynebacterium glutamicum for trehalose overproduction: Role of the TreYZ trehalose biosynthetic pathway.* Applied and Environmental Microbiology, 2006. **72**(3): p. 1949-1955.

42. Koizumi, S., et al., *Large-scale production of UDP-galactose and globotriose by coupling metabolically engineered bacteria.* Nature Biotechnology, 1998. **16**: p. 847-850.

43. Endo, T., et al., *Large-scale production of N-acetyllactosamine through bacterial coupling.* Carbohydrate Research, 1999. **316**: p. 179-183.

44. Endo, T., et al., *Large-scale production of CMP-NeuAc and sialylated oligosaccharides through bacterial coupling.* Applied Microbiology & Biotechnology, 2000. **53**: p. 257-261.

45. Koizumi, S., et al., *Large-scale production of GDP-fucose and Lewis X by bacterial coupling.* Journal of Industrial Microbiology & Biotechnology, 2000. **25**: p. 213-217.

46. Boone, D.R., R.W. Castenholz, and G.M. Garrity, eds. *Bergey's Manual of Systematic Bacteriology*. 2001, Springer: New York. 324-361.

47. Drummond, M.H. and M.D. Chilton, *Tumor-inducing (Ti) plasmids of Agrobacterium share extensive regions of DNA homology.* Jounral of Bacteriology, 1978. **136**(3): p. 1178-1183.

48. Rosenberg, C., et al., *Genes controlling early and late functions in symbiosis are located on a megaplasmid in Rhizobium meliloti.* Molecular Genetics and Genomics, 1981. **184**(2): p. 326-333.

49. Young, J.M., et al., *A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of Agrobacterium Conn 1942 and Allorhizobium undicola de Lajudie et al. 1998 as new combinations: Rhizobium radiobacter, R. rhizogenes, R. rubi, R. undicola, R. vitis.* International Journal of Systematic and Evolutionary Microbiology, 2001. **51**: p. 89-103.

50. Farrand, S.K., P.B. van Berkum, and P. Oger, *Agrobacterium is a definable genus of the family Rhizobiaceae.* International Journal of Systematic and Evolutionary Microbiology, 2003. **53**: p. 1681-1687.

51. Young, J.M., et al., *Classification and nomenclature of Agrobacterium and Rhizobium - a reply to Farrand et al. (2003).* International Journal of Systematic and Evolutionary Microbiology, 2003. **53**: p. 1689-1695.

52. Slater, S.C., et al., *Genome sequence of three Agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria.* Jounral of Bacteriology, 2009. **191**(8): p. 2501-2511.

53. Matthysse, A.G., *The genus Agrobacterium*, in *The Prokaroytes, Third Edition*, M. Dworkin, et al., Editors. 2006, Springer. p. 91-107.

54. Wood, D.W., et al., *The genome of the natural genetic engineer Agrobacterium tumefaciens C58.* Science, 2001. **294**: p. 2317-2323.

55. Goodner, B., et al., *Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58.* Science, 2001. **294**: p. 2323-2328.

56. Arthur, L.O., et al., *Carbohydrate metabolism in Agrobacterium tumefaciens.* Jounral of Bacteriology, 1973. **116**(1): p. 304-313.

57. Arthur, L.O., et al., *Carbohydrate catabolism of selected strains in the genus Agrobacterium.* Applied Microbiology, 1975. **30**(5): p. 731-737.

58. Harada, T. and T. Yoshimura, *Production of a new acidic polysaccharide containing succinic acid by a soil bacterium.* Biochimica et biophysica acta, 1964. **83**: p. 374-376.

59. Harada, T., et al., *Production of firm, resilient gel-forming polysaccharide by a mutant of Alcaligenes faecalis var. myxogenes 10C3.* Agricultural Biological Chemistry, 1966. **30**: p. 196-198.

60. Harada, T., A. Misaki, and H. Saito, *Curdlan: A bacterial gel-forming beta-1,3-glucan.* Archives of Biochemistry and Biophysics, 1968. **124**(1): p. 292-298.

61. Amemura, A., M. Hisamatsu, and T. Harada, *Spontaneous mutation of polysaccharide production in Alcaligenes faecalis var. myxogenes 10C3.* Applied and Environmental Microbiology, 1977. **34**(6): p. 617-620.

62. Phillips, K.R., et al., *Production of curdlan-type polysaccharide by Alcaligenes faecalis in batch and continuous culture.* Canadian Journal of Microbiology, 1983. **29**(10): p. 1331-1338.

63. Lee, I.-Y., *Curdlan*, in *Biotechnology of Biopolymers: From Synthesis to Patents*, A. Steinbuchel and Y. Doi, Editors. 2005, Wiley-VCH: Weinheim, Germany. p. 457-480.

64. Lawford, H.G., *Continuous process for the production of gelable exopolysaccharide*, U.S.P. Office, Editor. 1982, George Weston Limited: United States. p. 5.

65. McIntosh, M., B.A. Stone, and V.A. Stanisich, *Curdlan and other bacterial (1→3)-β-D-glucans.* Applied Microbiology and Biotechnology, 2005. **68**: p. 163-173.

66. Nakanishi, I., et al., *Demonstration of curdlan-type polysaccharide and some other β-1,3-glucan in microorganisms with aniline blue.* The Journal of General and Applied Microbiology, 1976. **22**: p. 1-11.

67. Toshiko, S., *Sustained release suppository*, J.P. Office, Editor. 1992: Japan.

68. Tsuneo, A., K. Koichi, and K. Junji, *Blood coagulation inhibitor*, J.P. Office, Editor. 1990: Japan.

69. Hiroshi, Y., et al., *Medicine for treating dementia*, J.P. Office, Editor. 1997: Japan.

70. Gordon, M., et al., *A phase I study of curdlan sulfate--an HIV inhibitor. Tolerance, pharmacokinetics and effects on coagulation and on CD4 lymphocytes.* Journal of Medicine, 1994. **25**(3-4): p. 163-180.

71. Demleitner, S., J. Kraus, and G. Franz, *Synthesis and antitumour activity of sulfoalkyl derivatives of curdlan and lichenan.* Carbohydrate Research, 1992. **226**(2): p. 247-252.

72. Lawford, H.G. and K.R. Phillips, *A two stage continuous process for the production of thermogellable curdlan-type exopolysaccharide.* Biotechnology Letters, 1982. **4**: p. 689-694.

73. Lee, I.-Y., et al., *Production of curdlan using sucrose or sugar cane molasses by two-step fed-batch cultivation of Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1997. **18**: p. 255-259.

74. Lee, J.-H., et al., *Optimal pH control of batch processes for production of curdlan by Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1999. **23**: p. 143-148.

75. Lee, I.Y., et al., *Influence of agitation speed on production of curdlan by Agrobacterium species.* Bioprocess Engineering, 1999. **20**: p. 283-287.

76. Kim, M.K., et al., *Residual phosphate concentration under nitrogen-limiting conditions regulates curdlan production in Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 2000. **25**: p. 180-183.

77. Kim, M.K., et al., *Higher intracellular levels of uridinemonophosphate under nitrogen-limited conditions enhance metabolic flux of curdlan synthesis in Agrobacterium species.* Biotechnology and Bioengineering, 1999. **62**(3): p. 317-323.

78. Lee, J.H. and I.Y. Lee, *Optimization of uracil addition for curdlan (β-1→3-glucan) production by Agrobacterium sp.* Biotechnology Letters, 2001. **23**: p. 1131-1134.

79. Karnezis, T., et al., *Topological characterization of an inner membrane (1→3)-β-D-glucan (curdlan) synthase from Agrobacterium sp. strain ATCC 31749.* Glycobiology, 2003. **13**(10): p. 693-706.

80. Stasinopoulos, S.J., et al., *Detection of two loci involved in (1→3)-β-glucan (curdlan) biosynthesis by Agrobacterium sp. ATCC31749, and comparative sequence analysis of the putative curdlan synthase gene.* Glycobiology, 1999. **9**(1): p. 31-41.

81. Karnezis, T., et al., *Cloning and characterization of the phosphatidylserine synthase gene of Agrobacterium sp. strain ATCC 31749 and effect of its inactivation on production of high-molecular-mass (1→3)-β-D-glucan (curdlan).* Journal of Bacteriology, 2002. **184**(15): p. 4114 - 4123.

82. Jin, L.H., et al., *Proteomic analysis of curdlan-producing Agrobacterium sp. in response to pH downshift.* Journal of Biotechnology, 2008. **138**(3-4): p. 87-87.

CHAPTER 2

METABOLIC ENGINEERING OF *AGROBACTERIUM* SP. FOR UDP-GALACTOSE REGENERATION AND LACNAC SYNTHESIS[1]

2.1 Abstract

Curdlan-producing *Agrobacterium* sp. is unique in possessing a highly efficient UDP-glucose regeneration system. A broad-host-range expression strategy was successfully developed to exploit the unique metabolic capability for UDP-galactose regeneration during oligosaccharide synthesis. The engineered *Agrobacterium* cells functioned as a UDP-galactose regeneration system, allowing galactose-containing disaccharides to be synthesized from glucose or other simple sugars. Unexpectedly, a lag period of 24 hours, which could be eliminated with rifampicin, preceded the active synthesis. An intracellular nucleotide profiling revealed that the UMP level was elevated by 3.8 fold in the presence of rifampicin, suggesting that rifampicin simulated a nitrogen-limitation condition that triggered the metabolic change. Product selectivity was improved nearly 40-fold by using high acceptor concentration and restricting glucose supply. *N*-acetyllactosamine concentration near 20 mM (7 g/L) was obtained, demonstrating the effectiveness of the engineered strain in UDP-galactose regeneration. This organism could be engineered to regenerate other UDP-sugar nucleotides using the same strategy as illustrated here.

1. Portions of this chapter were previously published: Ruffing A., Mao Z., Chen, R. *Metabolic Engineering.* 2006. 8(5): 465-473.

2.2 Introduction

Over the past decade, metabolic engineering has emerged as an effective tool for developing whole-cell biocatalysts for oligosaccharide synthesis [1-5]. Whole-cell catalysts provide facile regeneration of cofactors including sugar nucleotides, which are products of multiple interacting metabolic pathways (Figure 2.1). Unlike nicotinamide cofactors, sugar nucleotides are not easily regenerated using isolated enzymes. Not surprisingly, *Escherichia coli* was often used for whole-cell cofactor regeneration as it is one of most well-studied organisms and numerous genetic tools are well established [6, 7]. Although *E. coli* is capable of synthesizing UDP-glucose, a precursor to other UDP-sugars and cell wall synthesis, a high regeneration rate of UDP-glucose does not normally occur in wild type cells without significant metabolic engineering efforts. Furthermore, sugar nucleotide synthesis interfaces with the primary metabolism, and therefore, any improvement in sugar nucleotide regeneration through metabolic engineering will involve or affect not only sugar nucleotide synthesis but also the primary metabolism. Importantly, this will also include other interacting pathways and regulations of all these pathways that often resist any carbon flux redistribution not optimal for cell growth [8]. Biosynthesis of sugar nucleotides directly involves high-energy compounds such as UTP, bringing additional complications from energy production, regulation, and the inter-conversion of high-energy compounds. Hence, like any other metabolic engineering research, multiple cycles of trial and error may be necessary to achieve the desired outcome [9]. Judicious choice of organism is important to maximize the chance of success for such endeavors.

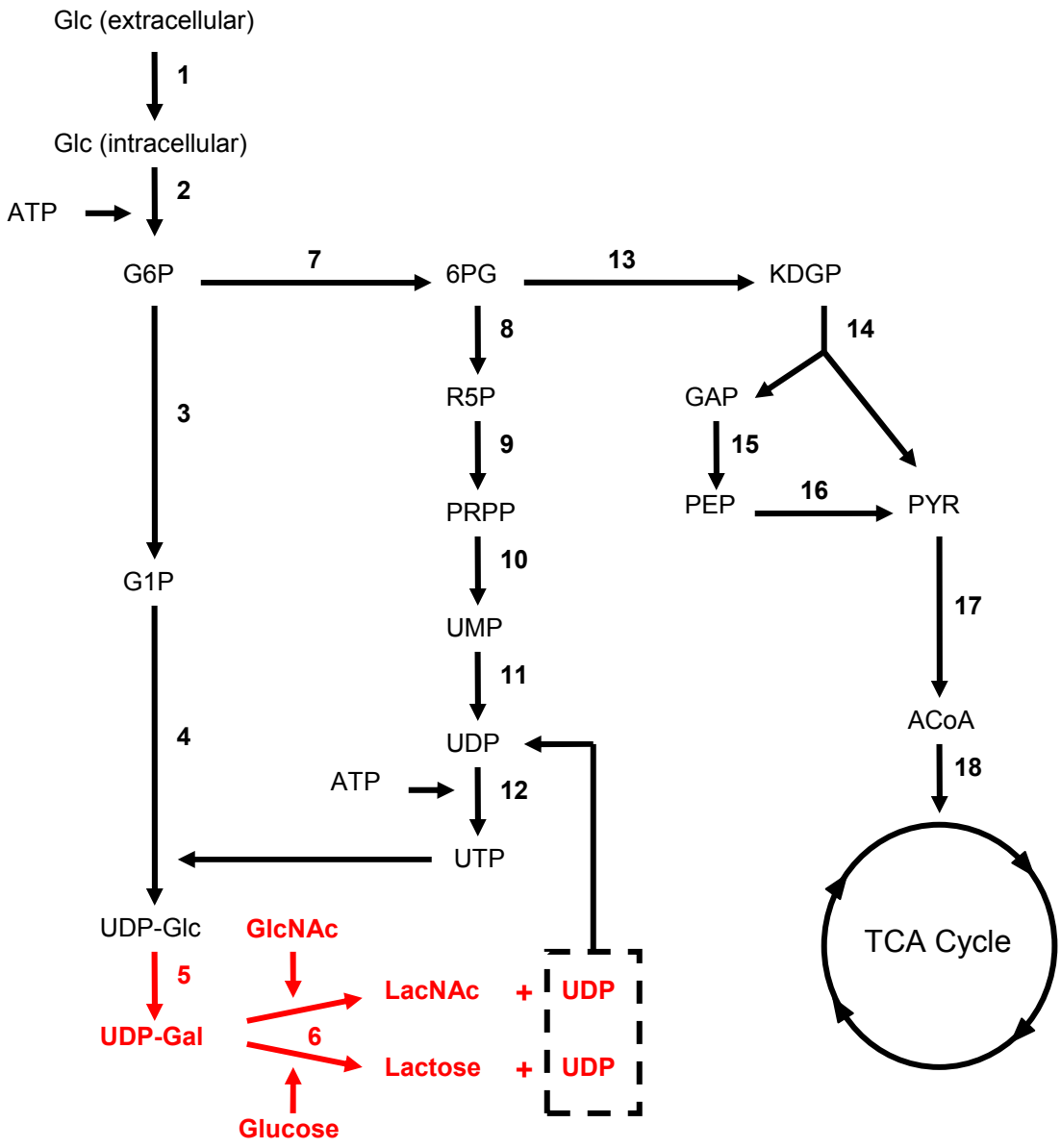


Figure 2.1 Proposed metabolic network of UDP-glucose synthesis in *Agrobacterium* sp. The engineered metabolic pathway is indicated in red. Enzymes/Proteins: 1: glucose-binding-protein for glucose uptake; 2: hexokinase; 3: phosphoglucomutase; 4: UDP-glucose phosphorylase; 5: UDP-galactose-4'-epimerase; 6: β-1,4-galactosyl--transferase; 7: glucose-6-phosphate dehydrogenase; 8: 6-phosphogluconate dehydrogenase; 9: ribose-phosphate diphosphokinase; 10: orotate phosphoribosyltransferase and orotidine-5'-phosphate decarboxylase; 11: uridylate kinase; 12: UDP kinase; 13: 6-phosphogluconate dehydrase; 14: 2-keto-3-deoxygluconate-6-phosphate aldolase; 15: glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate mutase, and enolase; 16: pyruvate kinase; 17: pyruvate dehydrogenase multienzyme complex; 18: citrate synthase.

Nature has evolved mechanisms for the regeneration of sugar nucleotides with efficiency unparalleled by any artificially assembled system reported thus far. In fact, these mechanisms are common and seem to operate in many species. Of particular interest is one such mechanism that operates in curdlan-producing *Agrobacterium* sp. strains. Curdlan, a β-1,3-glucose homopolymer, is synthesized in certain *Agrobacterium* strains under non-growing conditions. The synthesis is triggered by limitation of the nitrogen source [10]. During curdlan synthesis, glucose is transported and subsequently phosphorylated to glucose-6-phosphate, which is then converted to glucose-1-phosphate by phosphoglucomutase (Figure 2.1). The polymer precursor, UDP-glucose, is synthesized from UTP and glucose-1-phosphate, catalyzed by UDP-glucose pyrophosphorylase. The subsequent polymerization is catalyzed by curdlan synthase, which transfers a glucose molecule from UDP-glucose to the nascent polymer chain and releases a molecule of UDP. In the cytosol, UDP is converted back to UTP by UDP kinase using ATP derived from glycolysis or the TCA cycle. The glucose polymer synthesis is very efficient, with a productivity of 1.0 g-curdlan/hr/g-dry cell and a yield of 0.5 g curdlan per gram of glucose consumed [11]. Final product could be accumulated to 90 g/L [12]. This high efficiency suggests that the regeneration of the precursor molecule, UDP-glucose, must also be very efficient. When this high-capacity sugar nucleotide regeneration is compared to the best system reported [3], it surpasses the latter system by 32-fold. The higher flux of UDP-glucose regeneration in this organism can be attributed to its exquisite mechanism that couples energy production, carbon metabolism, and UDP-glucose synthesis, giving rise to an optimal rate of UDP-glucose regeneration during polymerization.

In this work, we sought a metabolic engineering strategy to exploit this highly efficient sugar-nucleotide regeneration system for oligosaccharide synthesis. A curdlan-deficient strain generated by mutagenesis with NTG (kindly provided by Dr. Stone and Dr. Stanisich, La Trobe University) ([13] and personal communication) was further engineered into a UDP-galactose regeneration system. We demonstrate its utility through the synthesis of a model disaccharide, *N*-acetyllactosamine (LacNAc).

2.3 Results

2.3.1 Engineering a UDP-galactose regeneration system for disaccharide synthesis

To exploit the efficient UDP-glucose regeneration system in this organism for UDP-galactose regeneration and oligosaccharide synthesis, it is necessary to introduce new enzymes into this organism. The two required enzymes for synthesis of LacNAc are UDP-galactose 4' epimerase and β-1,4-galactosyltransferase (Figure 2.1). Therefore, our first task is to construct a functional expression vector for this organism. A fusion expression strategy was adopted and a six-residue peptide linker (Ser-Ala-Ala-Gly-Gly-Ser) was designed to place between the two enzymes (detailed in section 2.4). Figure 2.2 illustrates the strategy used to construct the broad-range host expression vector, pKEL. The fused gene was first cloned into an *E. coli* T5-driven expression vector, pQE80L (Qiagen), resulting in pQEL. Successful expression of the fusion enzyme in *E. coli* was verified with SDS-PAGE gel analysis (data not shown) and activity assays (Table 2.1). To construct an expression vector functional in *Agrobacterium* sp., the replication region and kanamycin marker of a broad-host-range cloning vector pBBR122 (MoBiTec, Germany) was PCR-cloned into pQEL. The resulting plasmid, pKEL (Figure 2.2), is expected to be functional in a wide range of gram-negative bacteria, including

Agrobacterium and *E. coli*. To assess the effectiveness of this expression strategy in *Agrobacterium*, enzyme activities for the galactosyltransferase and fusion enzymes were measured and compared to those in *E. coli* (Table 2.1). Using the broad-host-range vector (pKEL), the enzyme activities were reduced 2 to 3-fold in *E. coli,* compared to the expression level from the smaller vector pQEL. The fusion enzyme was also active in *Agrobacterium* strains. The activity level obtained was about 50-70% of the level achieved in *E. coli* with the same plasmid. This is a significant improvement from our earlier work, where the expression level was only about 10% of that in *E. coli* [13]. A control plasmid containing no fusion genes, pBQ, was constructed in otherwise identical fashion for comparison studies.

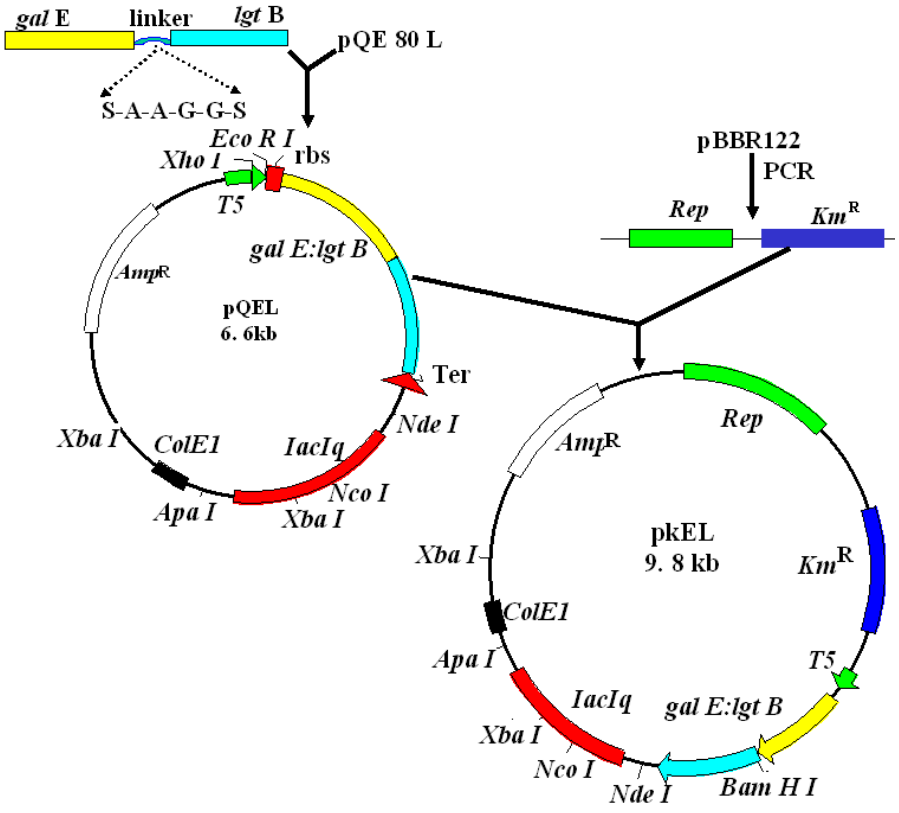


Figure 2.2 Construction of pKEL.

Table 2.1 Activities of β-1,4-galactosyltransferase and fusion enzyme in *E. coli* and *Agrobacterium* sp.[2]

Vector	*E. coli* AD202 (U/L)		*Agrobacterium* sp. ATCC 31749 (U/L)	
	Fusion	LgtB	Fusion	LgtB
pQEL	22 ± 2.2	25 ± 3.1	N/A	N/A
pKEL	7.5 ± 0.67	10.5 ± 0.12	5.4 ± 0.51	5.0 ± 0.38

2. Data are averages of three independent determinations. A unit of activity is defined as the amount of enzyme to synthesize 1 μmol of product per minute. NA: not applicable.

To demonstrate that the above metabolic engineering strategy resulted in a functional UDP-galactose regeneration system in this organism, LacNAc synthesis was carried out with the curdlan-deficient strain transformed with pKEL (LTU265/pKEL). The synthesis was compared to the same *Agrobacterium* strain, LTU265, transformed with the control plasmid pBQ. The β-1,4-galactosyltransferase, as it turned out, was promiscuous, accepting both glucose and GlcNAc as substrates and resulting in the synthesis of both lactose and LacNAc simultaneously (Figure 2.1). As shown in Figure 2.3, significant products were synthesized from glucose and GlcNAc with LTU265/pKEL, 0.83 mM LacNAc and 7.6 mM lactose (or a total of 8.4 mM UDP-galactose derives disaccharides), while no product formation was detected in LTU265/pBQ. These results indicate the successful construction of pKEL, which allows the plasmid-bearing cells to function as a UDP-galactose regeneration system.

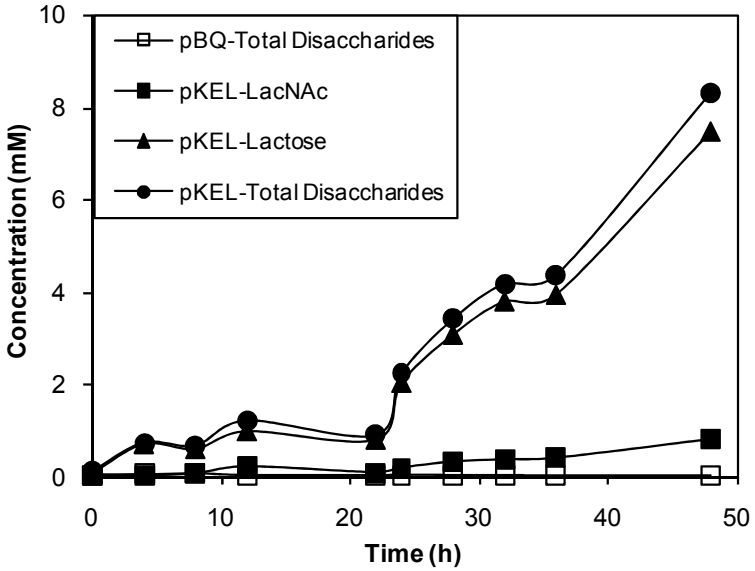


Figure 2.3 Disaccharides synthesized with LTU265/pKEL and LTU265/pBQ over a period of 48 hours under standard reaction conditions. LTU265/pBQ, LacNAc and lactose (□); LTU265/pKEL, LacNAc (■); LTU265/pKEL, lactose (▲); LTU265/pKEL, LacNAc and lactose (●). Data are averages of three independent determinations.

2.3.2 Rifampicin effect and intracellular nucleotide concentrations

While the construction allowed the engineered strain to produce UDP-galactose derived sugars, a rather long lag period preceded the onset of synthesis (Figure 2.3), typically 12-24 hours. It is known that curdlan synthesis is initiated only after the nitrogen source is depleted, and there appears to be no lag phase with respect to curdlan production [10, 19]. The engineered *Agrobacterium* cells were grown in LB medium, and prior to synthesis, they were washed twice with 10 % glycerol before suspension in the reaction buffer. This process should remove any residual nitrogen source prior to transfer to the reaction buffer. Thus, the long lag phase preceding the synthesis was puzzling, as it cannot be explained by the time required to exhaust the residual nitrogen source. In the course of this work, we discovered that a potent transcriptional inhibitor, rifampicin, could eliminate the lag phase (Figure 2.4). Consequently, a significant increase in

synthesis (1.6 to 4-fold) was observed in the presence of the inhibitor. Rifampicin diffuses freely into bacteria and inhibits the DNA-dependent RNA synthesis by blocking the elongation of RNA when the transcript becomes 2-3 nt in length [20]. The mechanism of rifampicin implicates the involvement of nucleotides in the metabolic change accompanying the transition from lag phase to active synthesis. We therefore undertook an intracellular nucleotide profiling to probe the underlying mechanism for the observed rifampicin effect. The intracellular nucleotides were analyzed using an HPLC ion-pairing method [18]. Intracellular concentrations of UMP, UDP, UTP, ATP, and two sugar nucleotides, UDP-glucose and UDP-galactose, were measured during the 48-hour synthesis reaction. Data are tabulated in Table 2.2. Concentrations at 12 hr represent values during the lag phase for reactions without rifampicin and during active synthesis for reactions with rifampicin. Comparison of the six compounds at 12 hr shows that only the UMP concentration was significantly different from samples taken with and without rifampicin. The UMP value in the active synthesis phase was about 3.8-fold as high as that in the lag phase, and the UMP concentration remained high in subsequent hours, corresponding to a period with continued oligosaccharide synthesis. No appreciable concentration differences in ATP, UTP, and UDP were observed (Table 2.2). Neither were there significant differences in the sugar nucleotides UDP-glucose and UDP-galactose. This result is reminiscent of the data reported by Kim et al. (1999). As shown in their study, active curdlan synthesis took place under nitrogen-limited conditions, and under these conditions, the UMP concentration was about 3.5 times as high as that during nitrogen-sufficient conditions with no curdlan synthesis. Their data also indicated negligible influence of other nucleotide concentrations on nitrogen-limited curdlan

production. Taken together, the data suggest that the addition of rifampicin simulates a nitrogen-limited condition, which triggers the synthesis of curdlan, and in our case, the synthesis of disaccharides. It is of interest to note that the rifampicin effect seemed to be limited to the elimination of lag phase. There were no significant differences in nucleotide concentrations measured at 24 and 36 hours from samples taken with and without rifampicin.

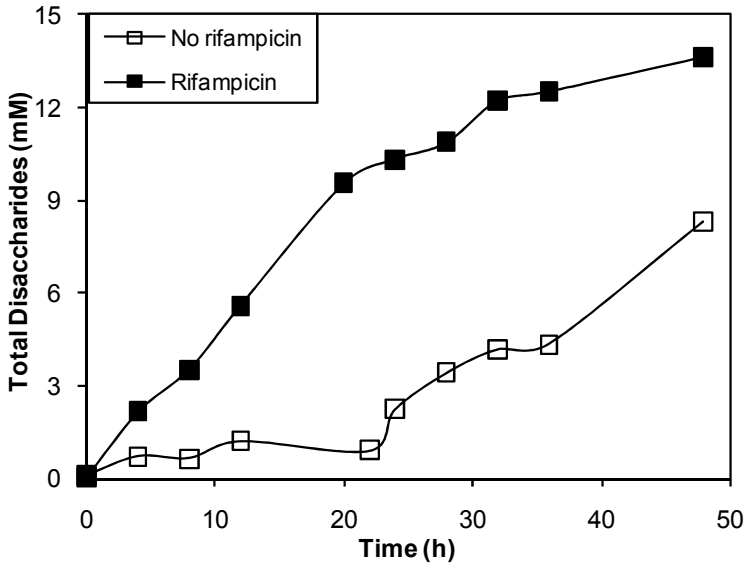


Figure 2.4 Total disaccharides (LacNAc and lactose) synthesized with LTU265/pKEL over a period of 48 hours under the following conditions: standard reaction conditions (□) and standard reaction conditions with 50µg/mL rifampicin (■). Data are averages of three independent experiments.

Table 2.2 Comparison of cellular nucleotide levels under conditions of no rifampicin addition and 50μg/ml of rifampicin[3]

Concentration (μmol / g DCW)	0 h No Rif	Rif	12 h No Rif	Rif	36 h No Rif	Rif
UMP	0.23 ± 0.09	0.32 ± 0.14	0.23 ± 0.09	0.87 ± 0.22	0.91 ± 0.06	0.69 ± 0.23
UDP	0.17 ± 0.03	0.22 ± 0.07	0.50 ± 0.13	0.61 ± 0.22	0.61 ± 0.03	0.76 ± 0.13
UTP	0.04 ± 0.02	0.05 ± 0.03	0.09 ± 0.02	0.10 ± 0.04	0.11 ± 0.003	0.14 ± 0.03
UDPG	0.28 ± 0.14	0.33 ± 0.25	0.27 ± 0.18	0.39 ± 0.36	0.44 ± 0.14	0.60 ± 0.15
UDP-gal	1.37 ± 0.54	0.82 ± 0.54	1.85 ± 0.70	1.34 ± 0.35	0.98 ± 0.37	1.31 ± 0.97
ATP	0.09 ± 0.02	0.13 ± 0.05	0.11 ± 0.03	0.11 ± 0.02	0.10 ± 0.01	0.09 ± 0.01

3. Nucleotide concentrations are measured in μmol per gram of dry cell weight. Data are averages of three independent determinations and are given for four reaction times: 0, 12, 24, and 36 hours.

2.3.3 Product selectivity

β-1,4-galactosyltransferase (*lgtB*) has been shown to have a wide range of acceptor specificity, including both GlcNAc and glucose. In *in-vitro* studies, the enzyme was shown to have a slightly higher specificity for GlcNAc than glucose as the acceptor; the activity with glucose as the acceptor was approximately 60-70% of that with GlcNAc [17, 21]. In *Agrobacterium* sp., glucose is directly transported into the cell through glucose-binding proteins without phosphorylation [22, 23]. The presence of intracellular glucose in high concentration could compete with GlcNAc in the synthesis reaction to form the undesired byproduct, lactose. This was observed in this study (Figure 2.3, Table 2.3). In fact, the initial conditions, 280 mM of glucose and 20 mM of GlcNAc, generated an unfavorable condition for LacNAc synthesis. Consequently, lactose was accumulated

as the predominant product. Reduction of intracellular glucose concentration using a fed-batch addition led to a more than 2-fold increase in the ratio of product (LacNAc) to byproduct (lactose) (0.13 vs. 0.29, Table 2.3).

Table 2.3 LacNAc and lactose synthesized after 48 hours reaction time with various carbon sources[4]

Carbon Source	LacNAc (mM)	Lactose (mM)	Total (mM)	LacNAc: Lactose
280 mM glucose	1.9 ± 1.3	14.6 ± 5.9	16.5 ± 7.2	0.13
10 mM glucose (fed batch)	5.1 ± 0.2	17.9 ± 8.9	23 ± 9.1	0.29
140 mM sucrose	6.4 ± 3.7	14.9 ± 4.8	21.3 ± 8.5	0.43
70 mM sucrose	6.4 ± 0.2	7.9 ± 3.1	14.3 ± 3.3	0.81
200 mM GlcNAc[5]	3.4 ± 0.07	0.0 ± 0.0	3.4 ± 0.07	N/A

4. All data are averages of two or three independent experiments.
5. The data point for 44.2g/L GlcNAc is after 22 hours reaction time (the maximum product synthesized).

To further increase synthesis of the desired product, alternative carbon sources were investigated, including fructose, glycerol, and sucrose. Both fructose and glycerol led to significant decreases in the formation of both products (data not shown). In contrast, sucrose (140 mM) yielded a significant increase in LacNAc production compared to an equivalent monomer amount of glucose (280 mM), improving the product ratio to 0.43. Further enhancement in LacNAc selectivity was achieved by using a lower sucrose concentration (Table 2.4), increasing the LacNAc to lactose ratio to 0.81. The lower sucrose concentration apparently suppressed the formation of lactose, favoring synthesis of the desired product. The enzymatic breakdown of sucrose into its constituent components may have led to a low intracellular glucose concentration, similar to that observed with the fed-batch addition of 10 mM glucose, favoring LacNAc formation.

However, the increase observed with sucrose was higher than that seen with the fed-batch glucose addition. Sucrose, therefore, is a preferred carbon and energy source for this synthesis.

In an effort to completely eliminate lactose synthesis, 200 mM GlcNAc was used as the sole carbon source. This led to the formation of 3.4 mM LacNAc with no byproduct synthesis. This suggests that eliminating glucose as the carbon and energy source could eliminate byproduct formation.

The product ratio could also be influenced by increasing GlcNAc concentrations. A combination of limiting intracellular glucose concentration (by using a relatively low concentration of sucrose) and a higher extracellular concentration of GlcNAc (up to 100 mM) led to dramatic improvement in product ratio (Figure 2.5). These favorable conditions yielded LacNAc concentrations of up to 19.1 mM (7.3 g/L) with concomitant reduction of lactose formation to below 5 mM (Figure 2.5). A product ratio of 4.0, in favor of LacNAc, was obtained, representing a near 40-fold improvement over the initial value. This indicates the possibility to improve synthesis and product selectivity through judicious selection of process conditions. Further increase of product selectivity is possible, yet an alternative galactosyltransferase with better substrate specificity will be a better choice as it avoids the complication completely.

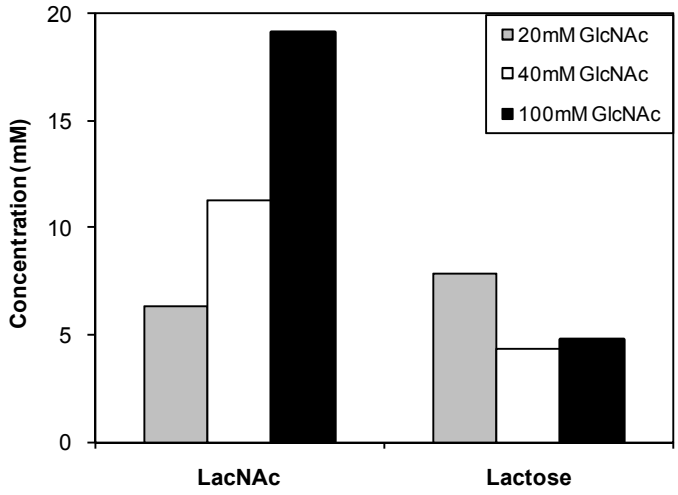


Figure 2.5 Comparison of LacNAc and lactose produced after 48 hours reaction time with increasing substrate concentrations of 20mM GlcNAc (gray), 40mM GlcNAc (white), and 100mM GlcNAc (black).

2.3.4 Identification of oligosaccharide byproducts

In addition to the desired product, LacNAc, several other byproducts were formed during the oligosaccharide synthesis reaction (Figure 2.6). Based on retention time, the byproducts are likely disaccharides, and since the recombinant β1,4-galactosyltransferase, LgtB, is known to allow several sugars to serve as acceptor, we suspect that the promiscuity of this enzyme led to the observed byproduct peaks. As evidence, β-galactosidase was found to hydrolyze all three products into their respective monosaccharide components. The only sugars initially present in the reaction medium are sucrose and GlcNAc, but the metabolism of ATCC 31749 can convert these sugars into other forms which may serve as acceptors to generate the β-galactose disaccharides. The most likely candidates are glucose and fructose, as these sugars are formed from sucrose hydrolysis. An enzymatic assay using the GalE:LgtB fusion enzyme revealed that glucose is readily used as an acceptor to produce lactose, with a retention time of 19.3 min. The

other byproduct (retention time 21.5 min) proved more difficult to identify. Replacing sucrose with other carbon sources provided some insight into the composition of this byproduct: the byproduct was formed with sucrose or fructose as the carbon source, but not with glucose. This suggests that the byproduct is formed from fructose metabolism. Since little is known about the metabolism of ATCC 31749, we considered the metabolism of a well-studied relative, *Agrobacterium tumefaciens* C58. From the predicted metabolic pathways of *Agrobacterium tumefaciens* C58, several potential acceptor sugars were identified: mannose, fructose, sorbitol, and mannitol. Of these candidates, only mannose yielded a product peak from the GalE:LgtB enzyme assay with a retention time matching that of the unidentified byproduct. Thus, the third product is likely galactose-β1,4-mannose. As this disaccharide is not commercially available, the concentration of this byproduct could not be determined and was therefore not shown in the analysis above.

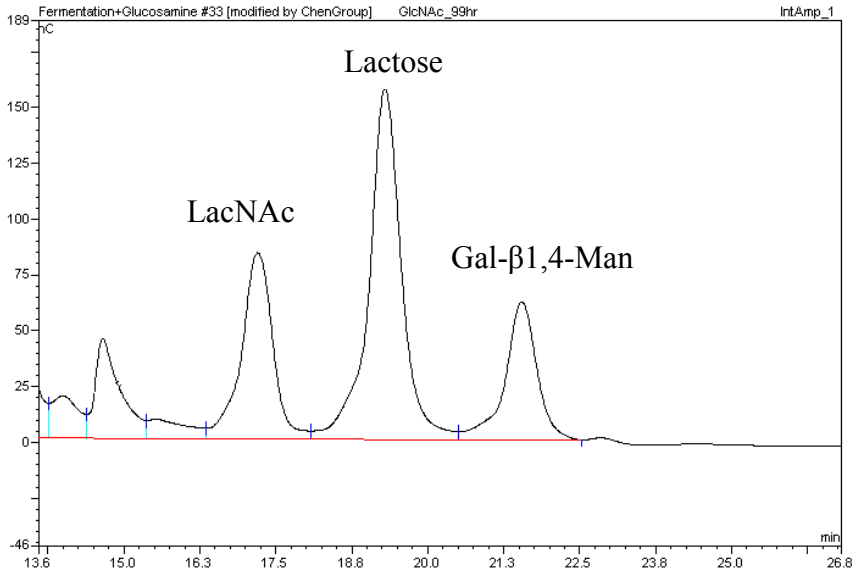


Figure 2.6 Chromatogram of oligosaccharide products

2.4 Discussion and Conclusions

The high amount of product synthesized with the engineered *Agrobacterium* sp. demonstrates the advantage of utilizing the cell's naturally efficient UDP-glucose regeneration system for oligosaccharide synthesis. This advantage is most clearly seen by comparing LTU265/pKEL to AD202/pQEL, an *E. coli* host containing a plasmid with the same fusion protein (*galE*:*lgtB*) [24]. The engineered *E. coli* produced a total of approximately 2.5 mM of UDP-galactose derived disaccharides (Table 2.4). Under similar conditions, the engineered *Agrobacterium* sp. yielded about 15.7 mM of disaccharides, indicating that its natural regeneration system is indeed superior to other organisms.

Table 2.4 Maximum LacNAc and lactose synthesized with *E. coli* AD202/pQEL and *Agrobacterium* sp. LTU265/pKEL[6]

	E. coli* with *galE:lgtB	***Agrobacterium* sp. with *galE:lgtB***
LacNAc (mM)	1.0	11.3
Lactose (mM)	1.5	4.4

6. The following reaction conditions were used for product synthesis: *E. coli* – 200mM glucose, 50mM GlcNAc, and 16 hr reaction time; *Agrobacterium* sp. – 70 mM sucrose, 40mM GlcNAc, and 48 hr reaction time.

The successful development of broad-host-range expression vectors allows exploitation of the efficient UDP-glucose regeneration system in this organism for oligosaccharide synthesis. In this paper, the naturally efficient UDP-glucose regeneration system in *Agrobacterium* sp. was metabolically engineered into an effective UDP-galactose regeneration system. With one additional enzyme, UDP-glucose and UDP-galactose can be converted into UDP-glucuronic acid and UDP-galacuronic acid,

respectively. Thus, *Agrobacterium* sp. could be similarly engineered to form effective UDP-galacuronic acid and UDP-glucuronic acid regeneration systems.

The use of sucrose as the carbon source leads to higher product concentrations than other sources including glucose, fructose, and glycerol. This result is in agreement with previous research on curdlan production with *Agrobacterium* sp. [19, 25]. Lee et al. not only found that sucrose led to a higher production of curdlan, they also found that more curdlan was produced with sucrose than with a comparable mixture of glucose and fructose. Sucrase activity was not detected in the culture broth or cell suspensions of *Agrobacterium* sp., suggesting that sucrose is directly transported into the cell [19]. The subsequent phosphorylation could be catalyzed by sucrose phosphorylase. Sucrose phosphorylase uses inorganic phosphate to convert sucrose to fructose and glucose-1-phosphate. If this is true, the advantage it confers is two fold. (1) It eliminates ATP consumption for glucose phosphosphorylation. (2) As glucose-1-phosphate is directly formed by this enzyme, it eliminates the conversion of glucose-6-phosphate to glucose-1-phosphate for UDP-glucose formation (Figure 2.1), avoiding the competition occurring at the glucose-6-phosphate node. The molecular basis for the superiority of sucrose as carbon source remains to be elucidated as our initial experiments failed to detect sucrose phosphorylase activity in a crude lysate of the engineered *Agrobacterium* sp.

The finding with rifampicin is intriguing, as it seems to relate to the regulation of polysaccharides synthesis. In a study where uracil was fed to ammonia-depleted cells, the synthesis of curdlan was initiated only when the uracil feeding occurred 24 hours after exhaustion of the nitrogen source [12]. Feeding uracil immediately after ammonia depletion resulted in more cell growth but no curdlan synthesis. Apparently, a waiting

period of 24 hours is necessary to prepare cells for the transition between the two metabolic states. Despite an 80% reduction in curdlan synthesis, the mutant (LTU265) used in this study retains the characteristic regulation observed with curdlan synthesis. It will be interesting to generate regulatory mutants that allow synthesis of oligosaccharides without the 24 hour lag phase. Recent advances in understanding curdlan synthesis in this organism have identified genes involved in the regulation [13, 26], paving the way for further metabolic engineering of the strain for practical applications.

2.5 Materials and Experimental Methods

2.5.1 Materials

Molecular biology reagents, such as A addition kit, plasmid purification kit, gel DNA recovery kit, Taq and pfu DNA polymerase, PCR primers, restriction enzymes, T4 DNA ligase, and SDS-PAGE ready gel were purchased from commercial sources, including Qiagen, Invitrogen, Promega, and BioRad. The chemicals used in this study were obtained from Sigma-Aldrich (lactose, LacNAc, UMP, UDP, UTP, UDP-glucose, UDP-galactose, ATP), Mallinckrodt (glucose), MP Biomedicals Inc.(GlcNAc), Fisher, and VWR (all other chemicals).

2.5.2 Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are listed in Table 2.5.

Table 2.5 Bacterial strains and plasmids used in this study

Strains/Plasmids	Description	Source
pQE80L	Commercial expression vector containing T5 promoter and *lacIq*	Qiagen
pBBR122	Commercial broad-host-range vector	Mobitec
pCW-lgtB	β-1,4-galactosyltransferase gene from *Nesseria meningitidis*	[14]
pLysPTE7	T7 expression vector containing the *E. coli* UDP-galactose epimerase gene	[15]
Teasy(-)GalE	pGEM-T Easy vector containing *galE* gene with *Bgl*II and *Bam*HI sites	This study
T-ElgtBD	pGEM-T Easy vector containing fused *galE:lgtB* With *Bgl*II and *Xho*I sites	This study
pQEL	pQE80L containing T5 driven *galE:lgtB* fusion expression vector	This study
pKEL	T5 directed *galE:lgtB* fusion gene for *Agrobacterium* sp. expression	This study
pBQ	Empty vector for pKEL, constructed from pQE80L and pBBR122	This study
Jm109	*E. coli*, F' *traD36 proA*$^{+}$*B*$^{+}$ *lacI*^{q} Δ*(lacZ) M15/*Δ*(lac-proAB) glnV44 e14*$^{-}$ *gyrA96 recA1 relA1 endA1 thi hsdR17*	[16]
AD202	*E. coli* protease free host	W.W. Wakarchuk
ATCC 31749	*Agrobacterium* sp., wild type	ATCC
LTU265	locus II mutant of *Agrobacterium* sp. ATCC 31749	B.A. Stone V.A. Stanisich

2.5.3 Construction of *Agrobacterium* sp. expression vector, pKEL and analysis of GalE-LgtB fusion enzyme activity

The gene encoding *E. coli* UDP-galactose 4'-epimerase (*galE*) was amplified from pLysPTE7 [15]. The 5' primer for *galE*, (5'-TGG***AGATCT***ATGAGAGTTCTGG

TTACCGGTGGT-3'), contains a *Bgl*II site (in bold italics) and start codon (underlined). The 3' primer for *galE*, (5'-TA***GGATCC***GCCAGCGC*TGA*ATCGGGATATCCCTGTG GAT G-3'), contains a *Bam*HI site (in bold italics) with the original stop codon, TAA, mutated to Ser codon, TCA (reverse complementary sequence underlined), and a mini-linker sequence coding for a linker peptide, Ala-Ala-Gly-Gly-Ser. The PCR fragment (about 1 kb) was ligated into the pGEM-T easy vector to give the plasmid pT-GalE. The gene for β-1,4-galactosyltransferase from *Nesseria meningitidis* was cloned from plasmid pCW-lgtB [14]. The 5' primer for *lgtB*, (5'-TA***GGATCC***ATGCAAAACCACGTTATCA GCTTAG-3'), contains a *Bam*HI site (in bold italics) and start codon (underlined); the 3' primer for *lgtB*, (5'-TA***CTCGAG***TTATTGGAAAGGCACAATGAACTG-3'), contains an *Xho*I site (in bold italics). The amplified PCR fragment (about 0.8 kb) was inserted into the pGEM-T easy vector to give the plasmid pT-lgtB. The *lgtB* gene fragment, obtained from pT-LgtB by the double digestion by *Bam*HI and *Sac*II, was inserted into pT-GalE downstream from *galE* to give pT-EL. The fusion gene fragment, obtained by *Bgl*II and *Xho*I digestion, was fused into the *Bam*HI and *Sal*I sites of pQE80L to give pQEL. To construct the expression vector for *Agrobacterium* sp. strains, a fragment from pBBR122 (MoBiTec, Germany), containing its replica region and kanamycin marker, was amplified with the primers derived from the pBBR122 sequence (GenBank accession Y14439). Both the 5' primer (5'-TGGT***GTCGAC***CTTGCCAGCCCGTGGATATGTGG-3') and the 3' primer (5'-AGGT***GTCGAC***TCTGTGATGGCTTCCATGTCGGCAG-3') contain a *Sal*I site (in bold italics). After digestion with *Sal*I, the 3.2 kb PCR fragment was ligated into the *Xho*I site of pQEL to obtain pKEL, a plasmid which can be

maintained in most gram negative bacteria with both ampicillin and kanamycin markers. The organization of the plasmid pKEL is shown in Figure 2.2.

E. coli strain AD202, *Agrobaterium* sp. ATCC 31749 and its mutant strain LTU265 were transformed with the expression vector pKEL by electroporation. The overnight inoculums of all transformants were prepared at 30^{o}C in a culture tube with 3 mL of LB containing 100 μg/L ampicillin and 100 μg/L kanamycin. The cells were diluted 100 times and grown in freshly prepared LB medium supplemented with antibiotics until OD_{600} reached 0.3-0.4, upon which IPTG was added to a final concentration of 1 mM, and cells were incubated for another 6 hours. The induced cells were collected by centrifugation and used for the enzyme activity assay.

The enzyme activity assay was performed as described by Blixt *et al.* with modification [17]. The cell pellets were resuspended to 1/10 of the culture volume and washed in buffer containing Tris-HCl (25 mM, pH 7.5), $MnCl_2$ (10 mM), and Triton X (0.25% v/v). After an additional 15 min of centrifugation at 5,000 g, the cell pellets were resuspended in the same buffer to 1/20 of the culture volume. The cells were lysed by Branson Mode 250 Sonifer (6 x 15 sec sonication, 1min ice cooling), and the lysates were centrifuged for 15 min at 10,000 g. The supernatant was used as crude enzyme for the activity analysis. The β-1,4-galactosyltransferase assay contained the crude enzyme (10% w/v), *N*-acetylglucosamine (10 mM), and UDP-galactose (2 mM) in 25 mM Tris-HCl (pH 7.5 containing 100 mM NaCl, 10 mM MnCl 1% bovine serum albumin) with a final reaction volume of 100 μl. The assay was conducted at 30°C for 30 min, and the reaction was stopped by heating in boiling water for 10 min. The acceptor (*N*-acetylglucosamine) consumption was limited to 20% by the low donor (UDP-galactose) concentration. After

centrifugation at 13,000 rpm for 5 min, the product (LacNAc) was analyzed by the method described below (carbohydrate analysis). The fusion enzyme activity assay is identical to the β-1,4-galactosyltransferase assay with a donor substrate of UDP-glucose in place of UDP-galactose. The results of both enzyme activity assays are shown in Table 2.1.

2.5.4 LacNAc and lactose synthesis with LTU265 transformants

Agrobacterium sp. transformants ATCC 31749/pKEL and LTU265/pKEL were prepared as described above. The 10 mL reaction mixture contained 50 mM Tris-HCl (pH 7.5), 1 g/L $K_2HPO_4 \cdot 3H_2O$, 5 g/L $MgSO_4 \cdot 7H_2O$, 5 g/L sodium citrate, 5 mM $MnCl_2$, 50 g/L glucose, and 20 mM GlcNAc. The cell concentration was 10% (wet weight), and the reaction vessel was a 50 mL flask. The reaction was carried out at 30ºC and 250 rpm in a biological shaker. Samples were heated at 97°C for 10 min. The supernatant was obtained by centrifugation and analyzed with a Dionex BioLC system (details below). Any changes in these standard reaction conditions are explicitly stated in the results section.

2.5.5 Analytical methods

2.5.5.1 Carbohydrate analysis

A Dionex BioLC system with a CarboPac PA20 analytical column was used for carbohydrate analysis. The Dionex ED50 electrochemical detector measured carbohydrate concentrations through pulsed amperometry (waveform: t = 0.41 sec, p = -2.00 V; t = 0.42 sec, p = -2.00 V; t = 0.43 sec, p = 0.60 V; t = 0.44 sec, p = -0.10 V; t = 0.50 sec, p = -0.10 V). Lactose and LacNAc concentrations were determined using calibration curves prepared from standards (Sigma-Aldrich). The mobile phase consisted

of degassed 200 mM sodium hydroxide (A) and 18 MΩ-cm water (B), pressurized with inert gas (He) at a flow rate of 0.5 mL/min. The following linear gradient was used: t = 0 min, 5:95 (A:B); t = 5 min, 5:95; t = 10 min, 20:80; t = 20 min, 20:80; t = 21 min, 100:0; t = 30 min, 100:0; t = 35 min, 5:95; t = 50 min, 5:95.

2.5.5.2 Nucleotide extraction and assay

The intracellular nucleotides were extracted using a previously established method [18]. Nucleotide analysis was performed using an Agilent 1100 series HPLC system with a reverse-phase ODS column (Waters Jsphere ODS H80, 150 x 4.6 mm i.d., 4 μm-particle size). The buffers and gradient used in HPLC analysis were modified from the nucleotide assay used by Kim et al. A two-buffer system was used: (1) Buffer A - 100 mM potassium phosphate monobasic (KH_2PO_4) and 8 mM tetrabutylammonium hydrogen sulfate with pH adjusted to 5.3 with 2M KOH (2) Buffer B - 70% buffer A and 30% methanol (pH 5.9). Both buffers were filtered with a 0.22 μm cellulose acetate filter. The gradient described by Kim et al. was used, with an extended equilibrium phase of 100% buffer A for 14 min instead of 4 min. The injection volume was 20 μL, and the flow rate was 1 mL/min. Using this method, the nucleotides had the following retention times: UMP 4.1 min, UDP 9.8 min, UTP 22.7 min, UDP-glucose 8.9 min, UDP-galactose 8.4 min, and ATP 28.8 min. Nucleotide concentrations were determined using calibration curves prepared from standards (Sigma-Aldrich).

2.5.5.3 β-Galacotsidase assay

β-Galacotsidase was used to aid in the identification of disaccharide byproducts. The assay included 1x PBS buffer (pH 7.4), 30 units/mL of β-galacotsidase, and reaction sample with a total volume of 100 μL. The enzyme mixture was incubated at 37°C for 30

min, followed by heating at 97°C for 10 min to deactivate the β-galacotsidase. After centrifugation at 13,200 rpm for 2 min, the sample was diluted and analyzed using the method described in section 2.4.5.1 (above).

2.6 References

1. Koizumi, S., et al., *Large-scale production of UDP-galactose and globotriose by coupling metabolically engineered bacteria.* Nature Biotechnology, 1998. 16: p. 847-850.

2. Endo, T., et al., *Large-scale production of CMP-NeuAc and sialylated oligosaccharides through bacterial coupling.* Applied Microbiology and Biotechnology, 2000. 53: p. 257-261.

3. Endo, T., et al., *Large-scale production of N-acetyllactosamine through bacterial coupling.* Carbohydrate Research, 1999. 316: p. 179-183.

4. Endo, T. and S. Koizumi, *Large-scale production of oligosaccharides using engineered bacteria.* Current Opinion in Structural Biology, 2000. 10: p. 536-541.

5. Bettler, E., et al., *The living factory: In vivo production of N-acetyllactosamine containing carbohydrates in E. coli.* Glycoconjugate Journal, 1999. 16: p. 205-212.

6. Sorensen, H.P. and K.K. Mortenson, *Advanced genetic stratgies for recombinant protein expression in Escherichia coli.* Journal of Biotechnology, 2005. 115(2): p. 113-128.

7. Jana, S. and J.K. Deb, *Strategies for efficient production of heterologous proteins in Escherichia coli.* Applied Microbiology and Biotechnology, 2005. 67(3): p. 289-298.

8. Hua, Q., C. Yang, and K. Shimizu, *Metabolic flux analysis for efficient pyruvate fermentation using vitamin-auxotrophic yeast of Torulopsis glabrata.* Journal of Bioscience and Bioengineering, 1999. 87(2): p. 206-213.

9. Bailey, J.E., *Toward a science of metabolic engineering.* Science, 1991. 252: p. 1668-1675.

10. Phillips, K.R., J. Pik, and H.G. Lawford, *Production of curdlan-type polysaccharide by Alcaligenes faecalis in batch and continuous culture.* Canadian Journal of Microbiology, 1983. 29: p. 1331-1338.

11. Phillips, K.R. and H.G. Lawford, *Theoretical maximum and observed product yields associated with curdlan production by Alcaligenes faecalis.* Canadian Journal of Microbiology, 1983. 29: p. 1270-1276.

12. Lee, J.-h. and I.-Y. Lee, *Optimization of uracil addition for curdlan (β-1→3-glucan) production by Agrobacterium sp.* Biotechnology Letters, 2001. 23: p. 1131-1134.

13. Karnezis, T., et al., *Cloning and characterization of the phosphatidylserine synthase gene of Agrobacterium sp. strain ATCC 31749 and effect of its inactivation on production of high-molecular-mass (1→3)-β-D-glucan (curdlan).* Journal of Bacteriology, 2002. 184(15): p. 4114-4123.

14. Wakarchuk, W.W., et al., *Role of paired basic residues in the expression of active recombinant galactosyltransferases from bacterial pathogen Neisseria meningitidis.* Protein Engineering, 1998. 11(4): p. 295-302.

15. Oberpichler, I., et al., *Phytochromes from Agrobacterium tumefaciens: Difference spectroscopy with extracts of wild type and knockout mutants.* FEBS Letters, 2006. 580: p. 437-442.

16. Yanisch-Perron, C., J. Vieira, and J. Messing, *Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors.* Gene, 1985. 33: p. 103-119.

17. Blixt, O., et al., *Efficient preparation of natural and synthetic galactosides with a recombinant β-1,4-galactosyltransferase/UDP-4'-gal epimerase fusion protein.* Journal of Organic Chemistry, 2001. 66: p. 2442-2448.

18. Kim, M.-K., et al., *Higher intracellular levels of uridine monophosphate under nitrogen-limited conditions enhance metabolic flux of curdlan synthesis in Agrobacterium species.* Biotechnology and Bioengineering, 1999. 62(3): p. 317-323.

19. Lee, I.-Y., et al., *Production of curdlan using sucrose or sugar cane molasses by two-step fed-batch cultivation of Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1997. 18: p. 255-259.

20. Campbell, E.A., et al., *Structural mechanism for rifampicin inhibition of bacterial RNA polymerase.* Cell, 2001. 104: p. 901-912.

21. Lee, J.W., et al., *Exopolymers from curdlan production: Incorporation of glucose-related sugars by Agrobacterium sp. strain ATCC 31749.* Canadian Journal of Microbiology, 1997. 43: p. 149-156.

22. Cornish, A., J.A. Greenwood, and C.W. Jones, *Binding-protein-dependent sugar transport by Agrobacterium radiobacter and A. tumefaciens grown in continuous culture.* Journal of General Microbiology, 1989. 135: p. 3001-3013.

23. Cornish, A., J.A. Greenwood, and C.W. Jones, *Binding-protein-dependent glucose transport by Agrobacterium radiobacter grown in continuous culture.* Journal of General Microbiology, 1988. 134: p. 3099-3110.

24. Mao, Z., H.-D. Shin, and R.R. Chen, *Engineering E. coli UDP-glucose synthesis pathway for oligosaccharide synthesis.* Biotechnology Progress, 2006.

25. Saudagar, P.S. and R.S. Singhal, *Fermentative production of curdlan.* Applied Biochemistry and Biotechnology, 2004. 118: p. 21-31.

26. Karnezis, T., et al., Topological characterization of an inner membrane (1→3)-β-glucan (curdlan) synthase from Agrobacterium sp. strain ATCC31749. Glycobiology, 2003. 13(10): p. 693-706.

CHAPTER 3

METABOLIC ENGINEERING OF *AGROBACTERIUM* SP. STRAIN ATCC 31749 FOR PRODUCTION OF AN α-GAL EPITOPE[1]

3.1 Abstract

Oligosaccharides containing a terminal Gal-α1,3-Gal moiety are collectively known as α-Gal epitopes. α-Gal epitopes are integral components of several medical treatments under development, including flu and HIV vaccines as well as cancer treatments. The difficulty associated with synthesizing the α-Gal epitope hinders the development and application of these treatments due to the limited availability and high cost of the α-Gal epitope. In this work, *Agrobacterium* sp. ATCC 31749 was engineered to produce Gal-α1,3-Lac by the introduction of a UDP-galactose 4'-epimerase:α1,3-galactosyltransferase fusion enzyme. The engineered *Agrobacterium* synthesized 0.4 g/L of the α-Gal epitope. Additional metabolic engineering efforts addressed the factors limiting α-Gal epitope production, namely the availability of the two substrates, lactose and UDP-glucose. Through expression of a lactose permease, the intracellular lactose concentration increased by 60 to 110%, subsequently leading to an improvement in Gal-α1,3-Lac production. Knockout of the curdlan synthase gene increased UDP-glucose availability by eliminating the consumption of UDP-glucose for synthesis of the curdlan polysaccharide. With these additional engineering efforts, the final engineered strain synthesized approximately 1 g/L of Gal-α1,3-Lac. The *Agrobacterium* biocatalyst

1. Portions of this chapter were submitted for publication: Ruffing, A. and Chen, R. *Microbial Cell Factories.* 2009.

developed in this work synthesizes gram-scale quantities of α-Gal epitope and does not require expensive cofactors or permeabilization, making it a useful biocatalyst for industrial production of the α-Gal epitope.

3.2 Introduction

α-Gal epitopes are oligosaccharides containing terminal Gal-α1,3-Gal residues. In nature, three main α-Gal epitopes are produced: two trisaccharides (Gal-α1,3-Gal-β1,4-GlcNAc and Gal-α1,3-Lac) and a pentasaccharide (Gal-α1,3-Gal-β1,4-GlcNAc-β1,3-Gal-β1,4-Glc). These epitopes are components of glycolipids and glycoproteins displayed on the cell surface of non-primate mammals and New World monkeys via expression of an α1,3-galactosyltransferase (α1,3-GalT). The α1,3-GalT was inactivated in ancestral Old World primates approximately 20-28 million years ago, resulting in the absence of α-Gal epitopes in humans, apes, and Old World monkeys today [1, 2]. These evolutionary descendents of Old World primates produce an antibody to Gal-α1,3-Gal-containing oligosaccharides known as anti-Gal. Anti-Gal is the most abundant natural antibody in humans, and as a result, exposure to α-Gal epitopes generates a strong immune response [3]. Many current research efforts exploit the human immune response to α-Gal epitopes. The efficacy of a vaccine is often determined by uptake of the vaccine by antigen presenting cells. Uptake can be greatly enhanced by the presence of an IgG antibody, such as anti-Gal, bound to its associated antigen. Based on this principle, several vaccines have been modified with α-Gal epitopes in an effort to improve vaccine uptake and efficacy. This strategy was applied to flu and HIV vaccines and was found to be more effective than the non-modified vaccine in animal studies [4, 5]. In addition to enhancing vaccine efficacy, the immunogenicity of α-Gal epitopes has been applied to improve

cancer treatments. Autologous tumor vaccines with α-Gal epitopes on the tumor cells and injections of α-Gal-containing glycolipids were shown to generate an immune response against malignant tumors in mice [6, 7]. The promising results of these α-Gal-based treatments have stimulated the demand for α-Gal epitope production.

The increasing interest in α-Gal epitopes for various medical applications necessitates an efficient and economical means of synthesizing α-Gal epitopes. Traditional chemical synthesis requires numerous reaction steps, leading to low overall yields, a high cost, and a process that is not applicable for large-scale production. Enzymatic production of α-Gal epitopes can be achieved in just one step through the use of an α1,3-GalT; however, enzymatic synthesis requires provision of an expensive sugar nucleotide, UDP-galactose. To reduce cost, enzymatic synthesis schemes often employ a UDP-galactose 4'-epimerase to provide the UDP-galactose from a less expensive sugar nucleotide, UDP-glucose [8, 9]. As UDP-glucose is still quite expensive, other enzymatic synthesis schemes have been developed to regenerate UDP-galactose through the use of additional enzymes [10, 11]. While these synthesis schemes reduce the cost of sugar nucleotide provision, they require production and purification of multiple enzymes, generally 4 to 6, and may also require other high energy compounds such as PEP which can still lead to high synthesis cost.

Alternatively, whole cell biocatalysts can synthesize α-Gal epitopes in just one step without enzyme purification. Different hosts and engineering strategies were explored by Wang and coworkers for whole-cell Gal-α1,3-Lac synthesis. An engineered *E. coli* was constructed by overexpressing five enzymes: three enzymes of the Gal operon (GalK, GalT, GalU) for UDP-galactose synthesis, a pyruvate kinase for energy

production, and an α1,3-GalT. In this strategy, glucose and catalytic amounts of other cofactors (ATP, UDPG, and G1P) were supplied to the engineered *E. coli* to provide the energy required for Gal-α1,3-Lac synthesis [12]. Alternatively, an engineered *Pichia pastoris* expressed a sucrose synthase which directly converts sucrose and UDP to UDP-glucose with fructose as coproduct. With two additional enzymes (UDP-glucose 4'-epimerase and α1,3-GalT), the modified *P. pastoris* required only sucrose, lactose, a catalytic amount of UDP, and a few essential nutrients to produce Gal-α1,3-Lac. The three recombinant enzymes in *P. pastoris* constituted an artificial pathway, whose operation was independent of cellular metabolism [13]. Both the engineered *E. coli* and *P. pastoris* were capable of synthesizing gram-scale amounts of Gal-α1,3-Lac.

One key challenge with whole-cell catalysts is uptake of the acceptor sugar, lactose, along with the primary sugar (i.e. sucrose or glucose). In both the engineered *E. coli* and *P. pastoris* biocatalysts, permeabilization was found to be necessary [12, 13]. While effective at improving lactose uptake, permeabilization also reduces cell viability and dissipates the proton gradient used for energy generation. This may not be a problem when cellular metabolism is not required to generate UDP-glucose, such is the case with sucrose synthase, but otherwise, it is detrimental to both cellular metabolism and α-Gal epitope synthesis. To avoid permeabilization, an *E. coli* biocatalyst was engineered to synthesize the acceptor sugar *in vivo*. Expressing a chitin oligosaccharide synthase (NodC) and several glycosyltransferases, the recombinant *E. coli* produced a heptasaccharide α-Gal epitope without permeabilization of the cell membrane [14]. While this method avoids the complications imposed by multiple sugar uptake, the diversion of cell resources for *in vivo* substrate synthesis may hinder α-Gal epitope production.

We have previously demonstrated that *Agrobacterium* sp. strain ATCC 31749 is a good host for oligosaccharide production through the synthesis of β1,4-Gal disaccharides [15]. ATCC 31749 naturally produces high amounts of a β1,3-glucan polysaccharide known as curdlan. High curdlan production in ATCC 31749 implies an efficient mechanism for regeneration of the requisite sugar nucleotide, UDP-glucose. In our previous work, ATCC 31749 was engineered to convert the UDP-glucose regeneration system into a UDP-galactose regeneration system. This current study utilizes the efficient UDP-galactose production for synthesis of the medically-relevant α-Gal epitope, Gal-α1,3-Lac. Instead of using permeabilization or *in vivo* substrate synthesis, an alternative strategy will be employed to address insufficient uptake of the substrate sugar, lactose. A lactose permease (LacY) from *E. coli* will be introduced in the *Agrobacterium* host to facilitate lactose transport across the cell membrane. In addition, sugar nucleotide availability will be improved by eliminating curdlan production, a competing pathway for utilization of UDP-glucose (Figure 3.1).

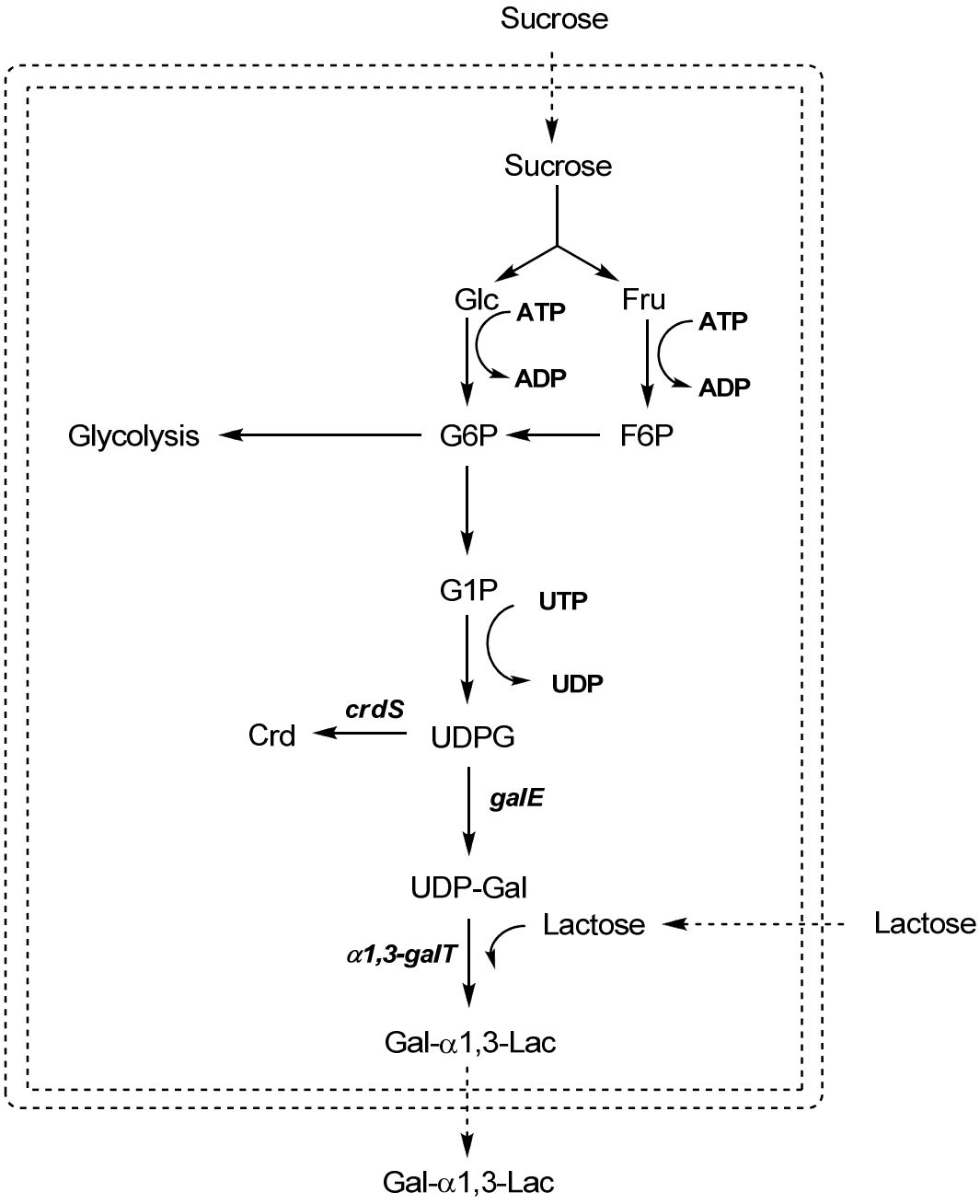


Figure 3.1 Metabolic pathway for Gal-α1,3-Lac synthesis in ATCC 31749. Cofactors are shown in bold and enzymes are in bold italics. The double dashed lines indicate the cell membrane, and the dashed arrows represent transport reactions.

3.3 Results

3.3.1 α-Gal epitope synthesis in ATCC 31749/pBQET

ATCC 31749 naturally produces high amounts of curdlan polysaccharide, suggesting an abundant supply of the sugar nucleotide precursor, UDP-glucose. In order to utilize UDP-glucose for synthesis of Gal-α1,3-Lac, two additional enzymes are required: a UDP-galactose 4'-epimerase and an α1,3-galactosyltransferase (Figure 3.1). In previous work, we demonstrated that the expression of a *galE:lgtB* fusion gene in ATCC 31749 can efficiently produce β1,4-Gal oligosaccharides [15]. Adopting a similar strategy for α-Gal epitope synthesis, a *galE:α1,3-galT* fusion gene was constructed using a UDP-galactose 4'-epimerase (*galE*) from *E.coli* and a truncated bovine α1,3-galactosyltransferase (*α1,3-galT*) (Section 3.4 Figure 3.2A). Inserting the fusion gene into the *Agrobacterium* expression vector, pBQ, yielded pBQET, which was transformed into ATCC 31749 for synthesis of the α-Gal epitope. Successful expression of the GalE:α1,3-GalT fusion enzyme was confirmed with an enzyme activity assay using ATCC 31749/pBQET. Concentrations of IPTG ranging from 0 to 1 mM were investigated to maximize activity of the fusion enzyme in the ATCC 31749 host. From 0 to 0.5 mM IPTG, the fusion enzyme activity increased with increasing IPTG concentration and reached a plateau from 0.5 to 1 mM IPTG (3.1). A concentration of 1 mM IPTG was selected for all subsequent α-Gal epitope synthesis reactions.

Table 3.1 Enzyme activity of GalE:α1,3-GalT fusion enzyme with varying IPTG concentration[2]

Strain	**IPTG (mM)**	**Fusion enzyme activity [µmol/(min•L)]**
ATCC 31749/pBQET	0	1.2 ± 0.76
	0.05	10.7 ± 1.1
	0.1	28.0 ± 0.52
	0.5	53.4 ± 16.3
	1	54.1 ± 5.3
ATCC 31749/pBQETY	0	6.4 ± 0.96
	0.05	13.2 ± 0.058
	0.1	15.6 ± 1.6
	0.5	14.3 ± 0.72
	1	11.6 ± 1.2
ATCC 31749Δ*crdS*/pBQET	1	38.2 ± 1.7
ATCC 31749Δ*crdS*/pBQET	1	12.6 ± 0.79

2. Data are reported as averages of at least two independent experiments with the associated standard deviation serving as an estimation of error.

The engineered *Agrobacterium,* ATCC 31749/pBQET, was utilized for small-scale α-Gal epitope synthesis. The synthesis process includes two phases. In the first phase, the engineered *Agrobacterium* is grown and induced. After recombinant protein production, the cells are transferred to a nitrogen-limited minimal media for synthesis of Gal-α1,3-Lac. Nitrogen-limited conditions are employed for α-Gal epitope synthesis as curdlan production, and hence UDP-glucose production, is activateded by this environmental signal. In the α-Gal epitope synthesis reaction, the engineered strain

synthesized 0.39 g/L of the desired product, Gal-α1,3-Lac (Figure 3.3). While this result confirms the successful engineering of ATCC 31749 for α-Gal epitope synthesis, the concentration of Gal-α1,3-Lac produced is lower than expected. In our previous work, the engineered *Agrobacterium* synthesized up to 7.5 g/L of β1,4-Gal oligosaccharides, nearly 20-fold greater than the amount of Gal-α1,3-Lac produced by ATCC 31749/pBQET [15]. Since ATCC 31749 has the potential to produce much higher levels of the α-Gal epitope, additional metabolic engineering strategies were investigated, focusing mainly on increased acceptor uptake and UDP-glucose availability.

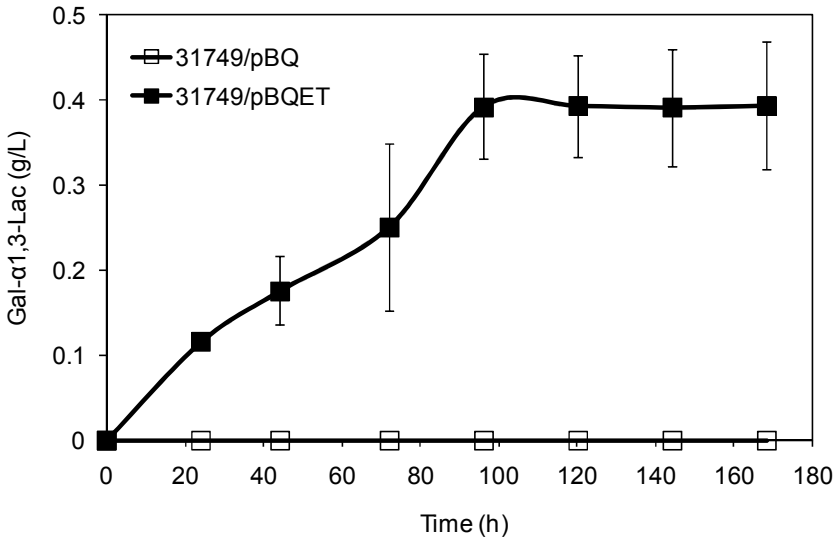


Figure 3.3 Synthesis of Gal-α1,3-Lac by ATCC 31749/pBQ (□) and ATCC 31749/pBQET (■). Data points are averages of three independent experiments with the standard deviation indicated by error bars.

3.3.2 Increased uptake of the acceptor, lactose

A major challenge of whole-cell synthesis is the efficient transport of reactants across the cell membrane. This task is particularly difficult for oligosaccharide synthesis reactions such as α-Gal epitope synthesis. The synthesis reaction requires uptake of two

sugars: the acceptor sugar (lactose) and a carbon source for production of the sugar nucleotide and cellular energy (sucrose) (Figure 3.1). Simultaneous uptake of multiple sugars is often prohibited by catabolite repression systems which allow for uptake of only a preferred carbon source [16]. Nominal uptake of the acceptor, lactose, was suspected to limit synthesis of the α-Gal epitope in ATCC 31749/pBQET. Measurement of lactose uptake by ATCC 31749 supported this hypothesis, as the rate of lactose consumption was more than 45-fold lower than the rate of sucrose consumption. To confirm that lactose is limiting for Gal-α1,3-Lac synthesis, the extracellular concentration of lactose was increased from 25 g/L to 50 g/L in an effort to increase the intracellular lactose concentration by means of diffusion. The higher lactose concentration led to a 28% increase in Gal-α1,3-Lac synthesis. An estimate of the intracellular lactose concentration indicates that roughly 1 mM of lactose is present intracellularly, which is more than 7-fold lower than the reported K_m for the GalE:LgtB fusion enzyme (K_m = 8.5 mM) [17]. These results indicate that lactose availability may limit α-Gal epitope synthesis.

While doubling the concentration of lactose was successful at increasing α-Gal epitope production, diffusion-mediated transport across the cell membrane is limited, requiring a large increase in lactose concentration to bring about a small improvement in synthesis. In this work, an alternative strategy is explored: introducing a heterologous lactose transporter to increase the availability of intracellular lactose. A lactose permease gene (*lacY*) from an *E. coli* K12 strain was expressed along with the fusion enzyme in the engineered *Agrobacterium* (Figure 3.2B). LacY is a lactose/proton symporter that transports lactose across the cytoplasmic membrane [18]. As expression of a transmembrane protein is generally not as straightforward as other soluble proteins,

additional steps were taken to provide evidence of successful expression of *lacY*. To analyze the expression and activity of LacY in the engineered *Agrobacterium*, uptake of lactose by ATCC 31749/pBQETY was compared to the lactose uptake of ATCC 31749/pBQET through analysis of residual lactose concentration in the extracellular medium. The rate of lactose uptake in ATCC 31749/pBQETY was approximately 50% greater than that of ATCC 31749/pBQET, indicating successful expression and activity of LacY. This measurement relies on extracellular lactose measurement, however, which assumes that decreasing lactose concentration is due solely to uptake by the cell, with no degradation of lactose in the extracellular medium. To confirm that LacY expression does indeed enhance lactose uptake, the amount of intracellular lactose was directly measured throughout the time course of the α-Gal synthesis reaction. These measurements show a 60 to 110% increase in intracellular lactose concentration for ATCC 31749/pBQETY compared to ATCC 31749/pBQET. Combining these results with the previous extracellular lactose measurements suggests successful expression of LacY and functional insertion of the protein into the cytoplasmic membrane.

ATCC 31749/pBQETY was employed to determine the effect of increased lactose availability on α-Gal epitope synthesis. The LacY-expressing strain synthesized 0.65 g/L of Gal-α1,3-Lac, a 67% improvement over the initial engineered strain, ATCC 31749/pBQET (Figure 3.4). Surprisingly, the activity of the GalE:α1,3-GalT fusion enzyme was 4.7-fold lower in ATCC 31749/pBQETY compared to ATCC 31749/pBQET. The lower activity in the LacY-expressing strain is presumably due to lower expression of the fusion enzyme. Varying IPTG concentration revealed that the high fusion enzyme activity of ATCC 31749/pBQETY was achieved with low

concentrations of IPTG (Table 3.1). To determine if the lower fusion enzyme activity of ATCC 31749/pBQETY restricts Gal-α1,3-Lac production, 0.05 mM of IPTG was used for α-Gal epitope synthesis, as both ATCC 31749/pBQET and ATCC 31749/pBQETY have similar fusion enzyme activities at this IPTG concentration. The amount of Gal-α1,3-Lac synthesized by the LacY-expressing strain was similar to that produced with 1 mM IPTG; this was expected as the enzyme activity levels are very similar at both IPTG concentrations. Unexpectedly, the amount of Gal-α1,3-Lac produced by the ATCC 31749/pBQET was similar for both 0.05 mM and 1 mM IPTG despite over a 5-fold reduction in fusion enzyme activity (Table 3.1). This suggests that the activity of the fusion enzyme is not a limiting factor in Gal-α1,3-Lac synthesis. With the higher concentration of intracellular lactose in ATCC 31749/pBQETY, the availability of UDP-glucose remains a potential limiting factor in α-Gal epitope production.

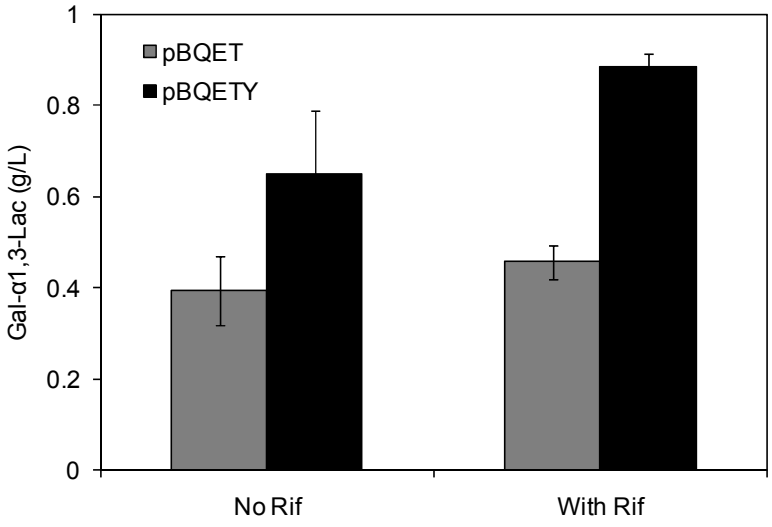


Figure 3.4 Synthesis of Gal-α1,3-Lac after 150 hours by ATCC 31749/pBQET and ATCC 31749/pBQETY without and with rifampicin. Data are averages of three independent experiments with the standard deviation indicated by error bars.

3.3.3 Curdlan synthase knockout for improved UDP-glucose availability

In addition to limitations imposed by low acceptor (lactose) availability, the amount of the sugar nucleotide, UDP-glucose, may also restrict α-Gal epitope synthesis by the GalE:α1,3-GalT fusion enzyme. Under the nitrogen-limited conditions of the synthesis reaction, the production of the α-Gal epitope competes with curdlan synthesis for the available UDP-glucose (Figure 3.1). To determine if UDP-glucose availability limits Gal-α1,3-Lac production, rifampicin was added to the synthesis reaction to prevent curdlan production. Since curdlan is only produced under nitrogen-limitation [19], it is suspected that the curdlan synthesis operon is only transcribed after exhaustion of the nitrogen source. Therefore, rifampicin, a transcription inhibitor, should prevent transcription of the genes responsible for curdlan synthesis during the production of Gal-α1,3-Lac. The addition of rifampicin led to a 37% increase in Gal-α1,3-Lac synthesis in ATCC 31749/pBQETY (Figure 3.4), indicating that UDP-glucose availability may limit α-Gal epitope production.

Curdlan, a β1,3-glucan polysaccharide, is not known to perform any essential or beneficial function for ATCC 31749. Therefore, by eliminating curdlan production, the amount of UDP-glucose available for α-Gal epitope synthesis will be increased without any detrimental impact on the cell. The gene responsible for the transfer of glucose from UDP-glucose to the growing curdlan polymer chain was previously determined to be curdlan synthase, c*rdS* [20]. CrdS has been characterized and the nucleotide sequence reported, providing the necessary information required for gene knockout.

This is the first reported attempt at gene knockout in *Agrobacterium* sp. strain ATCC 31749; however, successful gene knockout has been reported for a relative

organism, *Agrobacterium tumefaciens*. Gene knockout in *A. tumefaciens* employed the standard method of insertional mutagenesis via homologous recombination of a disruption cassette [21]. To determine an effective antibiotic resistance cassette for gene knockout, ATCC 31749 was tested for resistance against several antibiotics. Two antibiotics, gentamicin and tetracycline, were found to be effective at preventing ATCC 31749 growth at concentrations greater than 50 μg/mL. Since the genome of *A. tumefaciens* contains several tetracycline resistance genes, gentamicin was selected for the antibiotic resistance cassette in the *crdS* knockout plasmid. Sequences homologous to the 5' and 3' ends of *crdS* (500 bp) were added to each side of the gentamicin resistance cassette to allow for homologous recombination. The resulting fragment was inserted into the pBluescript II (KS^-) phagemid to give the *crdS* knockout plasmid, pBScrdShG (Figure 2C). After transformation of the *crdS* knockout plasmid into ATCC 31749, gentamicin-resistant colonies were screened using PCR for the presence of a 2236 kb fragment containing the interrupted *crdS* with gentamicin resistance cassette and the absence of the 1965 kb fragment corresponding to the intact *crdS*. A functional screening was also performed to confirm successful disruption of *crdS*, using aniline blue staining to detect curdlan production. The aniline blue dye binds to β1,3-glucan linkages in curdlan and produces a bright blue color [22]. In Figure 3.5, two candidate *crdS* mutants are compared to the wild type ATCC 31749 strain and a curdlan-deficient mutant produced via NTG mutagenesis, LTU265. The two *crdS* mutants show only a faint blue color, similar to LTU265, indicating curdlan synthesis was drastically reduced. As additional evidence for successful c*rdS* knockout, ATCC 31749Δ*crdS* was tested for the formation of curdlan gel. Heating an aqueous curdlan solution above 80°C and cooling

will lead to the formation of a curdlan gel [23]. The *crdS* mutant strain was cultivated in nitrogen-limited media to initiate curdlan production, yet the mutant did not form any visible curdlan gel after the requisite heating and cooling (data not shown). The cell growth profile for ATCC 31749Δ*crdS* in LB media showed no deviation from the wild-type strain, indicating that the *crdS* knockout did not effect cell growth (Figure 3.6). Since curdlan is not produced during cell growth, but rather, under nitrogen-limited conditions, the cell viability of ATCC 31749Δ*crdS* was studied in minimal, nitrogen-free media. The cell viability of the *crdS* mutant under curdlan-producing conditions showed no significant difference from the wild-type strain (Figure 3.7). Therefore, curdlan production does not appear to contribute to cell survival under nitrogen-limited conditions and should not have a detrimental effect on α-Gal epitope synthesis.

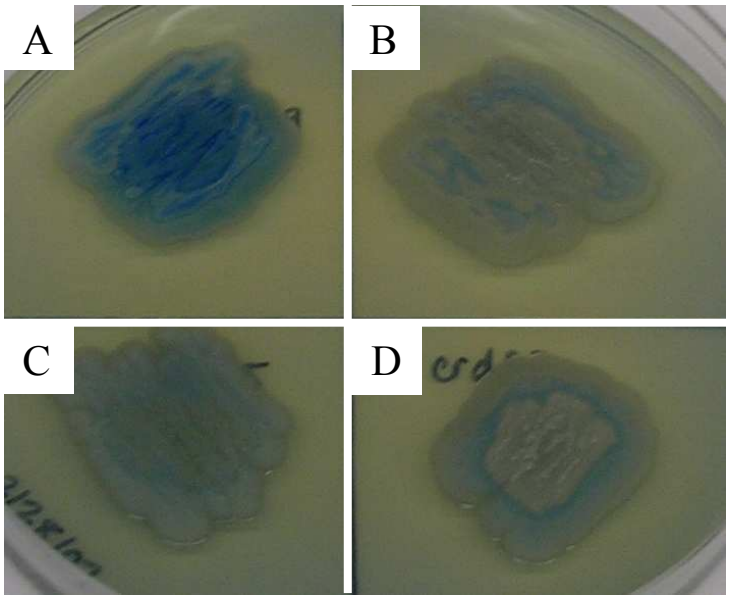


Figure 3.5 Aniline blue staining of curdlan. A: ATCC 31749; B: ATCC 31749ΔcrdS colony 8; C: LTU265; D: ATCC 31749ΔcrdS colony 14.

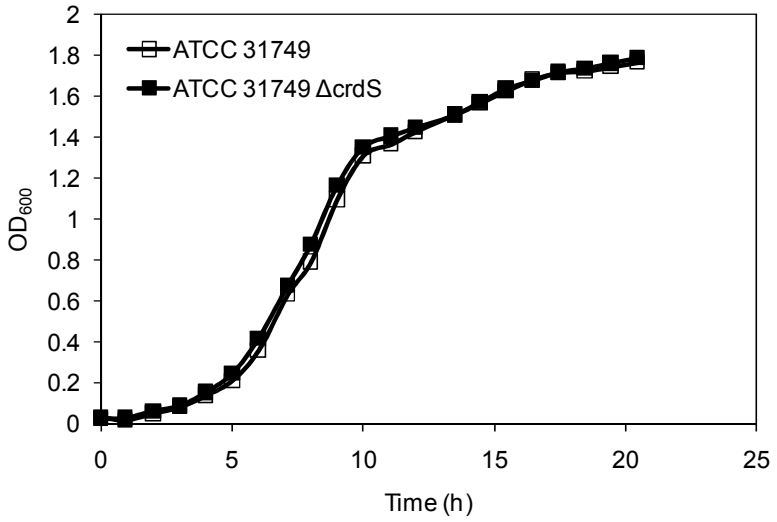


Figure 3.6 Growth profiles of ATCC 31749 and ATCC 31749Δc*rdS* in LB media.

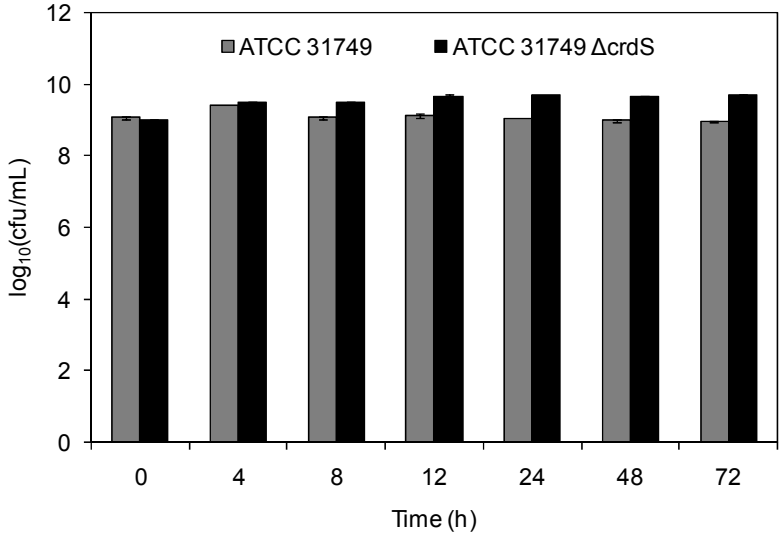


Figure 3.7 Cell viability of ATCC 31749 and ATCC 31749Δc*rdS* in minimal, nitrogen-free media. Data are averages of two independent experiments with the standard deviation indicated by error bars.

To determine the effect of the curdlan synthase knockout on α-Gal epitope synthesis, ATCC 31749Δ*crdS* was transformed with pBQET and pBQETY. ATCC 31749Δ*crdS*/pBQETY produced 0.96 g/L of Gal-α1,3-Lac (Figure 3.8), a concentration similar to that achieved with the addition of rifampicin (Figure 3.4). On the other hand,

ATCC 31749Δ*crdS*/pBQET synthesized only 0.40 g/L of the α-Gal epitope, indicating that synthesis remains limited by lactose availability. Overall, both lactose and UDP-glucose availability contributed to the low levels of Gal-α1,3-Lac produced by the initial engineered strain, ATCC 31749/pBQET. The construction of ATCC 31749Δ*crdS*/pBQETY, with increased levels of intracellular lactose and UDP-glucose, demonstrates the potential of ATCC 31749 as host for production of the α-Gal epitope. Requiring only a carbon source, the acceptor sugar, and a few essential nutrients (a phosphate source, buffer, and metal cofactors), the engineered *Agrobacterium* is capable of producing gram-scale quantities of the α-Gal epitope for medical research and applications.

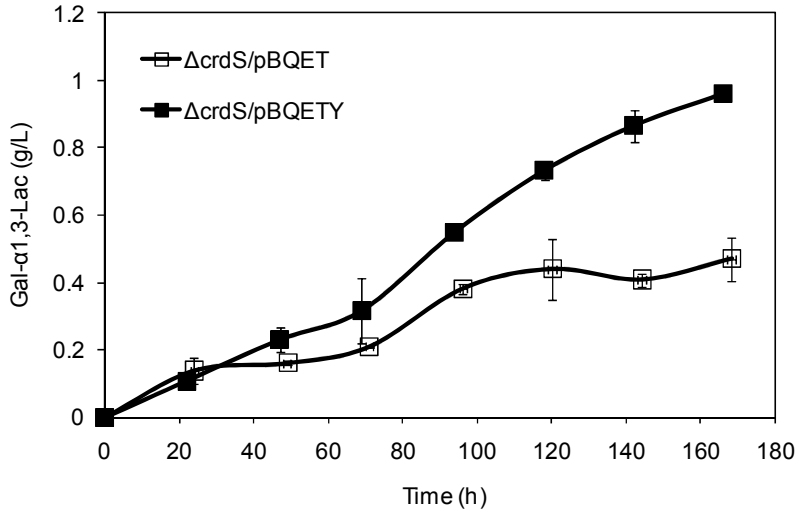


Figure 3.8 Synthesis of Gal-α1,3-Lac by ATCC 31749Δ*crdS*/pBQET and ATCC 31749Δ*crdS*/pBQETY. Data points are averages of three independent experiments with the standard deviation indicated by error bars.

3.4 Discussion and Conclusions

The whole-cell biocatalysts developed in this study provide a basis for an efficient and cost-effective means for large-scale production of the Gal-α1,3-Lac epitope. Enzymatic and whole-cell methods developed for α-Gal epitope synthesis face many obstacles including (1) low synthesis levels, (2) inefficient substrate transport across the cell membrane, and (3) expensive cofactors (Table 3.2). This work addresses all three of these issues. Host selection is a critical factor in determining product synthesis levels. To provide an adequate host for synthesis of the oligosaccharide, Gal-α1,3-Lac, a polysaccharide-producing microorganism was selected, *Agrobacterium* sp. ATCC 31749. This microorganism can produce up to 93 g/L of curdlan polysaccharide [24], demonstrating a natural proclivity for synthesizing the sugar nucleotide precursor, UDP-glucose. With ATCC 31749 as the host, gram-scale quantities of Gal-α1,3-Lac were produced, demonstrating an advantage over traditional hosts such as *E. coli*. Insufficient uptake of substrate and acceptor sugars is yet another obstacle in whole-cell oligosaccharide synthesis. To enhance sugar uptake, other strategies have used permeablization techniques to weaken the cell membrane or have expressed additional enzymes to synthesize the acceptor sugar *in vivo* [12-14]. In this work, a lactose permease was introduced into the *Agrobacterium* host to increase uptake of the acceptor, lactose. This strategy increased lactose uptake without the undesirable consequences of cell permeablization techniques, leading to a 67% improvement in Gal-α1,3-Lac synthesis. The LacY-expressing *Agrobacterium* may also be used to synthesize other lactose-containing oligosaccharides such as Globo-H, the major component of a vaccine for metastatic breast cancer [25]. Lastly, expensive cofactors are often added to enhance

production of the sugar nucleotide, including UDP, ATP, PEP, or even UDP-glucose itself. By deleting the gene responsible for curdlan synthesis, the efficient UDP-glucose synthesis pathway of the *Agrobacterium* host was exploited for α-Gal epitope synthesis, negating a need for expensive cofactors. Without the requirements of cell permeablization or cofactors, the engineered *Agrobacterium* developed in this work is a suitable biocatalyst for industrial-scale production of the α-Gal epitope, and with only slight modification, it may be utilized for the production of other medically-relevant oligosaccharides.

Table 3.2 Comparison of enzymatic and whole-cell methods for α-Gal epitope synthesis

Method	Enzymes/Host	α-Gal Epitope	Cell concentration	Cofactors	Permeabilization technique	Carbon source(s)	Vol. (mL)	Ref.
	Purified Enzymes							
Enzymatic	GalE; α1,3-GalT	8	-	UDPG, ATP	-	-	5	[8]
	GalE:α1,3-GalT	11.3	-	UDPG	-	-	-	[9]
	Sucrose synthase; GalE; β1,4-GalT; α1,3-GalT	1.1	-	UDPG, TDP-6-deoxy-4-keto-glucose	-	-	5	[10]
	UGPase; PPase; GalE; α1,3-GalT, β1,4-GalT, PK	3.3	-	G1P, PEP, UTP	-	-	10	[11]
	Host (recombinant enzyme)							
Whole-cell	*E. coli* (NodC, LgtB, GstA)	0.68	OD_{600}: 100	-	-	glycerol	1,000	[14]
	E. coli (GalK, GalT, GalU, PykF, α1,3-GalT)	7.2	13%	ATP, G1P, UDPG	freeze/thaw and permeablization (Triton X-100)	Gal and Glc	500	[12]
	Pichia pastoris (sucrose synthase, GalE, α1,3-GalT)	14	25%	UDP	freeze/thaw	Suc	200	[13]
	Agrobacterium sp. (GalE:α1,3-GalT)	0.96	10%	-	-	Suc	15	This study

While the *Agrobacterium* biocatalysts developed in this work have some advantages over other methods of α-Gal epitope synthesis, the amount of Gal-α1,3-Lac produced is lower than expected. The metabolic engineering strategies of *lacY* expression and *crdS* knockout resulted in only moderate improvements in α-Gal epitope synthesis. Based on the high levels of curdlan produced by ATCC 31749, the *Agrobacterium* host has the theoretical potential to produce over 270 g/L of the α-Gal epitope. Unfortunately, little is known regarding the mechanism and regulation of curdlan synthesis in ATCC 31749. Perhaps, once this regulatory mechanism is elucidated, additional metabolic engineering could reveal the full potential of the *Agrobacterium* host for α-Gal epitope and oligosaccharide synthesis. Until this information is available, several other strategies may help improve Gal-α1,3-Lac production. Other reported methods of whole-cell α-Gal epitope synthesis utilize fermenters for either high cell density growth or both growth and the epitope synthesis reaction [12-14]. Large-scale fermentation may be particularly beneficial with the engineered *Agrobacterium* as high curdlan synthesis requires high levels of dissolved oxygen and pH control at pH 5.5 [26, 27]. Under these conditions, the engineered *crdS* mutant should produce high levels of UDP-glucose for Gal-α1,3-Lac synthesis. Unfortunately, initial fermentation attempts with the engineered *Agrobacterium* were unsuccessful due to low recombinant protein production. Chromosomal integration of the *galE:α1,3galT* fusion enzyme and expression using a natural host promoter may overcome the limited recombinant protein production to allow large-scale fermentation. Chromosomal integration of *lacY* may also improve α-Gal epitope production. As shown in Table 3.1, the addition of *lacY* to pBQET leads to nearly a 5-fold reduction in GalE:α1,3-GalT fusion enzyme activity. Through chromosomal

integration of *lacY*, lactose uptake will be improved while maintaining the high fusion enzyme activity of the pBQET strain. Lastly, expression of a sucrose synthase may increase UDP-glucose production. Sucrose synthase conserves cellular energy by directly converting sucrose and UDP to fructose and UDP-glucose, saving two energy equivalents for each molecule of UDP-glucose synthesized compared to the traditional UDP-glucose synthesis pathway (Figure 3.1). Sucrose synthase has been successfully used in both enzymatic and whole-cell production of α-Gal epitopes [10, 13]. Implementation of the additional strategies outlined in this discussion should generate a biocatalyst for economical, large-scale production of Gal-α1,3-Lac.

3.5 Materials and Experimental Methods

3.5.1 Materials

The chemicals used in this study were obtained from Sigma-Aldrich (lactose, $K_2HPO_4 \cdot 3H_2O$, and $MnCl_2 \cdot 4H_2O$); Acros organics (aniline blue); V-labs, Inc. (Gal-α1,3-Gal-β1,4-Glc); and Fisher (all other chemicals).

3.5.2 Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are described in Table 3.3.

Table 3.3 Bacterial strains and plasmids used in this study

Strain or plasmid	Description	Source
ATCC 3179	Curdlan-producing *Agrobacterium* sp.	ATCC
LTU265	ATCC 31749 locus II mutant with decreased curdlan production	B.A. Stone V.A. Stanisich
JM109	*E. coli* K12 strain used for *lacY* cloning	Promega
ATCC 31749Δ*crdS*	ATCC 31749 with interrupted curdlan synthase gene (*crdS*)	This study
pGEM-T easy	Vector used for gene cloning	Promega
pBQ	Broad-host-range expression vector constructed for gene expression in ATCC 31749	[15]
pET15b-αGalT	Plasmid containing truncated bovine α1,3-galactosyltransferase gene	[11]
pT-α1,3-galT	pGEM-T easy vector containing truncated bovine α1,3-galactosyltransferase with *BamHI* and *SacI* restriction sites	This study
pT-GalE	pGEM-T easy vector containing UDP-galactose 4'-epimerase (*galE*) from *E. coli* with a shor linker sequence for fusion enzyme construction	[15]
pTET	pGEM-T easy vector containing the *galE:α1,3galT* fusion gene	This study
pBQET	pBQ with *galE:α1,3galT* for fusion enzyme expression in ATCC 31749 and ATCC 31749Δ*crdS*	This study
pT-lacY	pGEM-T easy vector containing lactose permease gene (*lacY*) from JM109 with *SacI* and *XhoI* restriction sites	This study
pBQETY	pBQ with *galE:α1,3galT* and *lacY* for expression of the fusion enzyme and *E. coli* lactose permease in ATCC 31749 and ATCC 31749Δ*crdS*	This study
pT-crdS	pGEM-T easy vector containing curdlan synthase gene (c*rdS*) with *SacI* and *KpnI* restriction sites	This study
pT-crdSh	pGEM-T easy vector containing two 500 bp regions of homology to *crdS* with *SacI/KpnI* and *EcoRI* /*HindIII* restriction sites	This study
pBluescript II (KS^-)	Phagemid for gene knockout in ATCC 31749	Stratagene

Table 3.3 Continued

Strain or plasmid	Description	Source
pBScrdSh	pBluescript II (KS^-) containing two 500 bp regions of homology to *crdS* with *EcoRI* and *HindIII* restriction sites	This study
pYanni2	Plasmid containing a gentamicin resistance cassette	[28]
pT-GmR	pGEM-T easy vector containing gentamicin resistance cassette with *EcoRI* and *HindIII* restriction sites	This study
pBScrdShG	*crdS* knockout plasmid; pBluescript II (KS^-) containing a gentamicin resistance cassette flanked by two 500 bp regions of homology to c*rdS*	This study

3.5.3 Construction of pBQET and pBQETY for α-Gal epitope synthesis

A fusion enzyme containing the *E. coli* UDP-galactose 4'-epimerase (*galE*) and truncated bovine α1,3-galactosyltransferase (*α1,3-galT*) was constructed and inserted into the *Agrobacterium* expression vector, pBQ, to form pBQET for α-Gal epitope synthesis. The truncated bovine *α1,3-galT* was amplified from pET15b-αGalT [11]. The 5' primer for *α1,3-galT* (5'-TC***GGATCC***ATGGAAAGCAAGCTTAAGCTATC-3') contains a *BamHI* site (in bold italics) and a start codon (underlined). The 3' primer for *α1,3-galT* (5'-TC***GAGCTC***TCAGACATTATTTCTAACCACATT-3') contains a *SacI* site (in bold italics) and a stop codon (underlined). The amplified *α1,3-galT* was inserted into the pGEM-T easy vector to form pT-α1,3-galT. The α1,3-galT fragment, obtained by double digestion with *BamHI* and *SacI*, was inserted into the respective restriction sites of the pT-GalE plasmid containing *galE* with a linker sequence [15]. Successful ligation of the α1,3-galT fragment to the digested pT-GalE yielded pTET, containing the *galE:α1,3-galT* fusion gene for α-Gal epitope synthesis. The fusion gene fragment, obtained from

BglII and *SacI* digestions, was fused to the *BamHI* and *SacI* sites of pBQ. The resulting plasmid, pBQET, is shown in Figure 3.2A.

The lactose permease gene (*lacY*) was cloned from the genomic DNA of *E. coli* K12 strain JM109. The 5' primer for *lacY* (5'-AC***GAGCTC***AAAGAGGAGAAAT TAACT**ATG**TACTATTTAAAAAACACAAAC-3') contains a *SacI* site (in bold italics), a ribosome binding site (underlined), and a start codon (in bold and underlined). The 3' primer (5'-GT***CTCGAG***TTAAGCGACTTCATTCACCTG-3') contains an *XhoI* site (in bold italics) and stop codon (underlined). The amplified *lacY* fragment was inserted into the pGEM-T easy vector, yielding pT-LacY. The *lacY* fragment, obtained by *SacI* and *XhoI* digestions, was ligated to the *SacI* and *SalI* sites of pBQET to form pBQETY (Figure 3.2B).

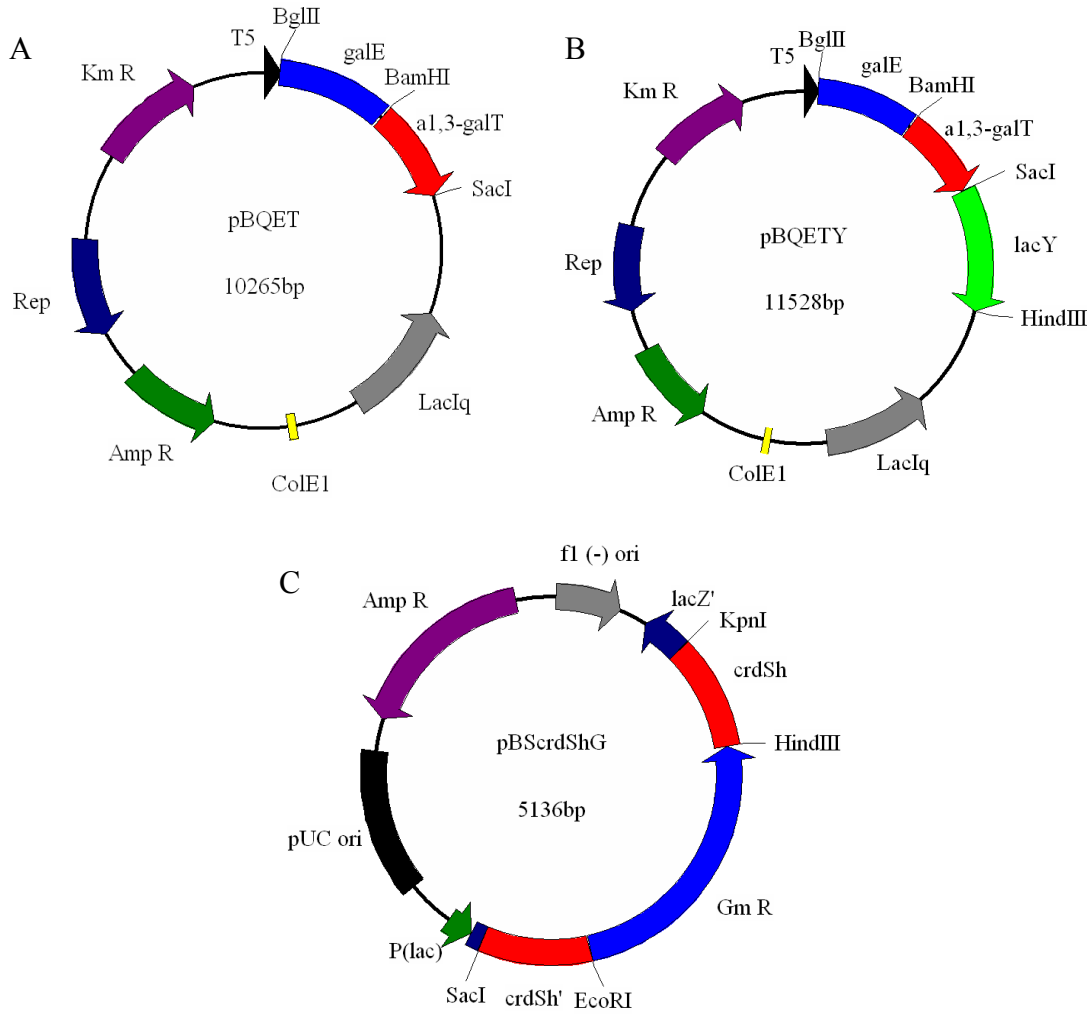


Figure 3.2 Plasmids constructed in this study. A: pBQET, containing the *galE:α1,3-galT* fusion enzyme for Gal-α1,3-Lac synthesis; B: pBQETY, containing the *galE:α1,3-galT* fusion enzyme and *lacY* for Gal-α1,3-Lac synthesis and lactose uptake; C: pBScrdShG, containing homologous regions of *crdS* interrupted by a gentamicin resistance cassette for *crdS* knockout.

3.5.4 Gene knockout of curdlan synthase (*crdS*) in ATCC 31749

The knockout plasmid for gene knockout of the curdlan synthase gene, *crdS*, was constructed using a gentamycin resistance cassette flanked on each side by 500 bp of homology to the curdlan synthase gene. The 5' primer for *crdS* amplification (5'-CG***GA GCTC***ATGTATTTCAGTGCTGAAGG-3') contains a *SacI* site (in bold italics) and the start codon for *crdS* (underlined). The 3' primer for *crdS* amplification (5'-CC***GGTACC***

TCACCCGAATGCCCGTGC-3') contains a *KpnI* site (in bold italics) and *crdS* stop codon (underlined). The amplified *crdS* gene was inserted into the pGEM-T easy vector to produce pT-crdS. Using pT-crdS as template, the pGEM-T easy vector along with 500 bp of homology at the 5' end of *crdS* and 500 bp of homology at the 3' end of *crdS* were amplified using phosphorylated primers. The 5' primer (5'-pGGCC***GAATTC***TCGCAA TAGGTTCTTACCTC-3') contains an *EcoRI* site (in bold italics) while the 3' primer (5'-pGGC***AAGCTT***TTCGTGACCCTGTCTTCGGC-3') contains a *HindIII* site. The amplified pGEM-T easy vector with homologous regions of *crdS* was self-ligated to form pT-crdSh. pT-crdSh was digested using SacI and KpnI, and the resulting fragment containing *crdS* homology (*crdSh*) was inserted into the corresponding restriction sites of pBluescriptII (KS^-), generating the plasmid pBScrdSh. The gentamicin resistance gene (Gm^R), along with its promoter and transcription terminator, was amplified from pYanni2 [28]. The 5' primer for Gm^R amplification (5'-CC***GAATTC***GTCTAGTGAGTAGTG GGTAC-3') contains an *EcoRI* site (in bold italics), and the 3' primer (5'-CG***AAGCTT*** GCTTGCAAACAAAAAAACCACC-3') contains a *HindIII* site. The amplified fragment containing Gm^R was inserted into the pGEM-T easy vector, forming pT-GmR. pT-GmR was digested using EcoRI and HindIII, and the fragment containing Gm^R was inserted into the corresponding restriction sites in pBScrdSh to form pBScrdShG (Figure 3.2C). The knockout plasmid, pBScrdShG, was transformed into ATCC 31749 using electroporation, followed by a 5 hour cultivation without antibiotics, and growth on LB/agar plates containing 50 µg/mL of gentamicin. The KO candidate colonies were grown four times in LB media at 30°C and 250 rpm to cure pBScrdShG.

3.5.5 Cell growth and GalE:α1,3GalT enzyme activity assay

ATCC 31749/pBQET, ATCC 31749/pBQETY, ATCC 31749Δ*crdS*/pBQET, and ATCC 31749Δ*crdS*/pBQETY inoculums were prepared at 30°C and 250 rpm in a culture tube with 4 mL of LB media containing 100 µg/mL of kanamycin and 100 µg/mL of ampicillin. The inoculums were diluted 500x and grown in 150 mL of freshly prepared LB media supplemented with antibiotics until an OD_{600} of 0.3 - 0.4 was reached, upon which IPTG was added to a final concentration of 1 mM. After IPTG addition, the cell culture was incubated at 30°C and 250 rpm for another 6 hours.

For the enzyme activity assay, the induced cells were collected by centrifugation at 5,000 x g and 4°C for 2 min. The cell pellet was resuspended in a buffer containing 25 mM Tris-HCl (pH 7.5), 10 mM $MnCl_2$, 100 mM NaCl, and 0.25% Triton X to a final cell concentration of 20% wet wt/v. Using a Branson Mode 250 Sonifer, the mixture was sonicated 6 times for 15 second intervals with 1 min rest on ice. The lysates were centrifuged for 2 min at 10,000 x g at 4°C, and 50 µL of supernatant was added to the substrate mixture for the enzyme activity assay. The substrate mixture consists of 20 mM lactose (10 mM final concentration) and 8 mM of UDP-glucose (4 mM final concentration) in the enzyme activity buffer described previously. The assay was conducted at 30°C for 40 min followed by 10 min at 97°C to stop the reaction. The deactivated enzyme mixture was diluted 50x, and the Gal-α1,3-Lac product was analyzed using the protocol detailed below (3.4.8.2 Carbohydrate analysis).

3.5.6 Cell viability

ATCC 31749 and ATCC 31749Δ*crdS* inoculums were prepared at 30°C and 250 rpm in a culture tube with 4 mL of LB media containing 10 µg/mL of gentamicin for the

crdS mutant. The inoculums were diluted 1000x and grown in 150 mL of freshly prepared LB media, supplemented with antibiotics for ATCC 31749Δ*crdS*. After 12 hours of growth in LB, the cells were collected by centrifugation at 3000 x g for 10 min at 4°C. The cell pellets were washed with 10% glycerol and then resuspended in nitrogen-free media. The nitrogen-free media consisted of 70 g/L sucrose, 1 g/L $K_2HPO_4 \cdot 3H_2O$, 5 g/L $MgSO_4 \cdot 7H_2O$, 5 g/L sodium citrate, 1 g/L $MnCl_2 \cdot 4H_2O$, and 50 mM Tris-HCl (pH 7.5). The cells were cultivated in nitrogen-free media at 30°C and 250 rpm in a biological shaker for 72 hours. Samples were taken at intervals of 12 hours and diluted using sterile LB media. The diluted cultures were spread on LB/agar plates, and colony forming units (cfu's) were counted as a measure of cell viability.

3.5.7 α-Gal epitope synthesis

The induced cells were prepared as described in 3.4.5 Cell growth and GalE:α1,3-GalT enzyme activity assay. The cells were then collected by centrifugation for 10 min at 3,000 x g and 4°C. The cell pellets were washed once with sterile 10% glycerol, and then resuspended in the reaction media. The reaction media contained 50 g/L sucrose, 25 g/L lactose, 1 g/L $K_2HPO_4 \cdot 3H_2O$, 5 g/L $MgSO_4 \cdot 7H_2O$, 5 g/L sodium citrate, 1 g/L $MnCl_2 \cdot 4H_2O$, and 50 mM Tris-HCl (pH 7.5). The final cell concentration was 10% wet wt/v, and the reaction vessel was a 50 mL Erlenmeyer flask. The reaction vessel was placed in a biological shaker at 30°C and 250 rpm. Samples were centrifuged for 3 min at 13,200 rpm. The supernatant was heated in boiling water for 10 min and then centrifuged again at 13,200 rpm for 3 min. The supernatant was appropriately diluted and analyzed as described in 3.4.8.2 Carbohydrate analysis.

3.5.8 Analytical techniques

3.5.8.1 SDS-PAGE

SDS-PAGE was used to confirm the successful expression of the GalE:α1,3-GalT fusion enzyme. JM109/pBQET was inoculated in 4 mL of LB media containing 100 µg/mL of kanamycin and 100 µg/mL of ampicillin. After overnight growth at 37°C with 250 rpm, 3 mL of the inoculum was transferred to 150 mL of LB media with antibiotics and grown at 37°C with 250 rpm. When an OD_{600} of 0.2 - 0.3 was reached, the cultures were induced with IPTG (final concentration 1 mM). Immediately after the addition of IPTG, the temperature was shifted to 25°C. After an induction period of 4 hours, the culture was centrifuged at 3,000 x g and 4°C for 10 min. The cell pellet was washed with 0.4 M NaCl and collected by centrifugation. A buffer containing 100 mM Tris-HCl and 100 mM NaCl (pH 8.0) with protease inhibitor cocktail (Sigma-Aldrich) was used to resuspend the cell pellet to a final concentration of 0.2 g/mL (wet weight). The cell suspension was sonicated 8 times for 10 second intervals with 1 min rest on ice. Cellular debris was removed by centrifugation at 3,000 x g and 4°C for 20 min. The GalE:α1,3-GalT fusion enzyme was purified from the supernatant using a HIS select HF nickel affinity gel (Sigma-Aldrich). A 12% Tris-HCl gel (Bio-Rad) was used for SDS-PAGE to analyze the His-tag purified and soluble protein fractions.

3.5.8.2 Carbohydrate analysis

Diluted samples were analyzed using a Dionex BioLC system with a CarboPac PA20 analytical column. The Dionex ED50 electrochemical detector measured carbohydrate concentrations through pulsed amperometry (waveform: t = 0.41 sec, p = -2.00 V; t = 0.42 sec, p = -2.00 V; t = 0.43 sec, p = 0.60 V; t = 0.44 sec, p = -0.10 V; t =

0.50 sec, p = -0.10 V). Sucrose, lactose, and Gal-α1,3-Lac concentrations were determined using calibration curves prepared from standards. The mobile phase consisted of degassed 200 mM sodium hydroxide (A) and 18 MΩ-cm water (B), pressurized with inert gas (He) at a flow rate of 0.5 mL/min. The following linear gradient was used for sucrose and lactose detection: t = 0 min, 5:95 (A:B); t = 5 min, 5:95; t = 10 min, 20:80; t = 25 min, 20:80; t = 25 min, 100:0; t = 40 min, 100:0; t = 40 min, 5:95; t = 55 min, 5:95. The linear gradient for Gal-α1,3-Lac detection entails the following steps: t = 0 min, 30:70 (A:B); t = 35 min, 30:70; t = 35 min, 100:0; t = 50 min, 100:0; t = 50 min, 30:70; t = 60 min, 30:70.

3.5.8.3 Intracellular lactose measurement

ATCC 31749/pBQET and ATCC 31749/pBQETY were prepared as described in 3.4.7 α-Gal epitope synthesis. Samples (1 mL) were taken from both reaction vessels at 34 and 45 hours after the start of the synthesis reaction and centrifuged for 5 min at 5,000 x g and 4°C. The supernatant was removed and used for analysis of extracellular lactose and Gal-α1,3-Lac. The cell pellet was washed with 50 mM of Tris-HCl buffer (pH 7.5) and centrifuged a total of three times to remove any residual extracellular lactose. After washing, the cell pellet was resuspended in 600 µL of 50 mM Tris-HCl buffer (pH 7.5). The cells were disrupted using sonication (6 x 10 sec, 1 min rest on ice). Cell debris was removed by centrifugation, and the supernatant containing the intracellular lactose was heated for 10 min in boiling water to denature any proteins. After centrifugation, the sample was diluted 50x and analyzed as described in 3.4.8.2 Carbohydrate analysis to determine the amount of lactose in mmol. The intracellular volume was determined by taking another 1 mL sample and collecting the cell pellet by centrifugation. The cell

pellet was washed with DI water and centrifuged three times to remove any residual media. The wet weight of the cell pellet was measured, and then, the cell pellet was dried in an oven at 80°C until a constant dry weight was obtained. The difference in weight between the wet and dry cell pellets was used to calculate the intracellular volume. The intracellular lactose concentration was then calculated using the amount of lactose determined through carbohydrate analysis and the intracellular volume.

3.5.8.4 Curdlan detection

The elimination of curdlan production in ATCC 31749Δ*crdS* was detected using an aniline blue staining method [22]. Candidate colonies were streaked on LB/agar plates containing 0.005% aniline blue. The plates were incubated for 5 days at 30°C, and the stained curdlan (blue) was detected by visual inspection.

3.6 References

1. Macher, B.A. and U. Galili, *The Galα1,3Galβ1,4GlcNAc-R (α-Gal) epitope: a carbohydrate of unique evolution and clinical relevance.* Biochimica et Biophysica Acta, 2008. **1780**(2): p. 75-88.

2. Galili, U., *The α-gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy.* Immunology and Cell Biology, 2005. **83**: p. 674-686.

3. Chen, X., P.R. Andreana, and P.G. Wang, *Carbohydrates in transplantation.* Current Opinion in Chemical Biology, 1999. **3**: p. 650-658.

4. Abdel-Motal, U.M., et al., *Immunogenicity of influenza virus vaccine is increased by anti-gal-mediated targeting to antigen-presenting cells.* Journal of Virology, 2007. **81**: p. 9131-9141.

5. Abdel-Motal, U.M., et al., *Increased immunogenicity of human immunodeficiency virus gp120 engineered to express Galα1-3Gal β1-4GlcNAc-R epitopes.* Journal of Virology, 2006. **80**: p. 6943-6951.

6. LaTemple, D.C., et al., *Increased immunogenicity of tumor vaccines complexted with anti-Gal: studies in knock out mice for α1,3galactosyltransferase.* Cancer Research, 1999. **59**: p. 3417-3423.

7. Galili, U., K. Wigglesworth, and U.M. Abdel-Motal, *Intratumoral injection of α-gal glycolipids induces xenograft-like destruction and conversion of lesions into endogenous vaccines.* Journal of Immunology, 2007. **178**: p. 4676-4687.

8. Chen, X., et al., *Production of α-galactosyl epitopes via combined use of two recombinant whole cells harboring UDP-galactose 4-epimerase and α-1,3-galactosyltransferase.* Biotechnology Progress, 2000. **16**: p. 595-599.

9. Fang, J., et al., *Synthesis of α-Gal epitope derivatives with a galactosyltransferase-epimerase fusion enzyme.* Carbohydrate Research, 2000. **329**: p. 873-878.

10. Hokke, C.H., et al., *One-pot enzymatic synthesis of the Galα1→3Galβ1→4GlcNAc sequence with in situ UDP-Gal regeneration.* Glycoconjugate Journal, 1996. **13**: p. 687-692.

11. Fang, J., et al., *Highly efficient chemoenzymatic synthesis of α-galactosyl epitopes with a recombinant α(1→3)-galactosyltransferase.* Journal of the American Chemical Society, 1998. **120**(27): p. 6635-6638.

12. Chen, X., et al., *Reassembled biosynthetic pathway for large-scale carbohydrate synthesis: α-Gal epitope producing "superbug".* ChemBioChem, 2002. **3**: p. 47-53.

13. Shao, J., T. Hayashi, and P.G. Wang, *Enhanced production of α-galactosyl epitopes by metabolically engineered Pichia pastoris.* Applied and Environmental Microbiology, 2003. **69**(9): p. 5238-5242.

14. Bettler, E., et al., *Production of recombinant xenotransplantation antigen in Escherichia coli.* Biochemical and Biophysical Research Communications, 2003. **302**: p. 620-624.

15. Ruffing, A., Z. Mao, and R.R. Chen, *Metabolic engineering of Agrobacterium sp. for UDP-galactose regeneration and oligosaccharide synthesis.* Metabolic Engineering, 2006. **8**: p. 465-473.

16. Gorke, B. and J. Stulke, *Carbon catabolite repression in bacteria: many ways to make the most out of nutrients.* Nature Reviews. Microbiology., 2008. **6**(8): p. 613-624.

17. Chen, X., et al., *Changing the donor cofactor of bovine α1,3-galactosyltransferase by fusion with UDP-galactose 4-epimerase.* The Journal of Biological Chemistry, 2000. **275**(41): p. 31594-31600.

18. Abramson, J., et al., *The lactose permease of Escherichia coli: overall structure, the sugar-binding site and the alternating access model for transport.* FEBS Letters, 2003. **555**(1): p. 96-101.

19. Phillips, K.R. and H.G. Lawford, *Curdlan: its properties and production in batch and continuous fermentations*, in *Progress in Industrial Microbiology*, D.E. Bushell, Editor. 1983, Elsevier Scientific Publishing Co.: Amsterdam.

20. Karnezis, T., et al., *Topological characterization of an inner membrane (1→3)-β-D-glucan (curdlan) synthase from Agrobacterium sp. strain ATCC 31749.* Glycobiology, 2003. **13**(10): p. 693-706.

21. Oberpichler, I., et al., *Phytochromes from Agrobacterium tumefaciens: Difference spectroscopy with extracts of wild type and knockout mutants.* FEBS Letters, 2006. **580**: p. 437-442.

22. Nakanishi, I., et al., *Demonstration of curdlan-type polysaccharide and some other β-1,3-glucan in microorganisms with aniline blue.* The Journal of General and Applied Microbiology, 1976. **22**: p. 1-11.

23. McIntosh, M., B.A. Stone, and V.A. Stanisich, *Curdlan and other bacterial (1→3)-β-D-glucans.* Applied Microbiology and Biotechnology, 2005. **68**: p. 163-173.

24. Lee, J.H. and I.Y. Lee, *Optimization of uracil addition for curdlan (β-1→3-glucan) production by Agrobacterium sp.* Biotechnology Letters, 2001. **23**: p. 1131-1134.

25. Gilewski, T., et al., *Immunization of metastatic breast cancer patients with a fully synthetic globo H conjugate: a phase I trial.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(6): p. 3270-3275.

26. Lee, I.Y., et al., *Influence of agitation speed on production of curdlan by Agrobacterium species.* Bioprocess Engineering, 1999. **20**: p. 283-287.

27. Lee, J.-H., et al., *Optimal pH control of batch processes for production of curdlan by Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1999. **23**: p. 143-148.

28. Graupner, S. and W. Wackernagel, *A broad-host-range expression vector series including a Ptac test plasmid and its application in the expression of the dod gene of Serratia marcescens (coding for ribulose-5-phosphate 3-epimerase) in Pseudomonas stutzeri.* Biomolecular Engineering, 2000. **17**: p. 11-16.

CHAPTER 4

CITRATE STIMULATION OF OLIGOSACCHARIDE SYNTHESIS IN METABOLICALLY ENGINEERED *AGROBACTERIUM* SP.

4.1 Abstract

Agrobacterium sp. ATCC 31749 was previously shown to be an advantageous host for oligosaccharide production. Unexpectedly, the addition of citrate to the oligosaccharide synthesis reaction resulted in up to a 7-fold improvement in the production *N*-aceytllactosamine, a disaccharide. The possible mechanisms for this citrate-induced stimulation of oligosaccharide production were investigated. The enhanced solubility of metal ions from citrate chelation was shown to have a negligible influence on oligosaccharide synthesis. Energy production from the metabolism of citrate proved to be a major contribution to the increase in oligosaccharide production, with other TCA cycle metabolites also having a stimulatory effect. ATCC 31749 was able to metabolize citrate along with sucrose, a preferred carbon source, and surprisingly, the consumption of sucrose was enhanced with the addition of citrate to the reaction. This phenomenon was not unique to sucrose, as glucose consumption also increased with the addition of citrate. The role of citrate as a buffer was also shown to be a major factor for enhanced oligosaccharide synthesis. Metabolic flux analysis revealed a dramatic increase in ATP for the citrate-containing reaction mixture, confirming the importance of energy for enhanced oligosaccharide production. Furthermore, the citrate stimulation of oligosaccharide synthesis was shown to be unique to *Agrobacterium* sp. ATCC 31749, as

a similarly engineered *E. coli* strain did not show significant improvement in oligosaccharide production with the addition of citrate.

4.2 Introduction

Oligosaccharides are important biomolecules, participating in various cellular processes, yet until recently, the significance of oligosaccharides in disease and other medical conditions has been largely overlooked. Currently, oligosaccharides are key components of potential vaccines and treatments for diseases including cancer, HIV, malaria, and anthrax [1-2]. While initial investigation of these treatments may show promise, they are often limited by the cost and difficulty of synthesizing the oligosaccharide component. Oftentimes, the oligosaccharides used in medical applications are naturally synthesized using enzymes, yet the sugar nucleotide cofactors required for enzymatic synthesis are also costly and challenging to produce. Alternatively, a whole cell synthesis strategy can produce the oligosaccharide from a cheap carbon source by generating the sugar nucleotide *in vivo*. Many efforts at whole cell oligosaccharide synthesis produced only small amounts of the desired oligosaccharide [3]. We have recently shown that using an unconventional bacterium, *Agrobacterium* sp., oligosaccharides can be synthesized at concentrations as high as 20 mM [4].

In our previous work, *Agrobacterium* sp. ATCC 31749 was engineered to produce *N*-acetyllactosamine (LacNAc), a disaccharide basal structure found in antigens overexpressed on the surface of cancerous cells. While LacNAc synthesis was successful in the engineered *Agrobacterium*, the amount of LacNAc produced was dramatically improved by the addition of sodium citrate to the reaction medium. This work

investigates the possible mechanisms for the observed citrate stimulation of oligosaccharide synthesis in the engineered *Agrobacterium*.

Agrobacterium sp. ATCC 31749 is primarily studied for its production of curdlan, a β1,3-glucose polymer. Initial investigation of ATCC 31749 indicates the presence of the Entner Doudoroff and pentose phosphate pathways as well as the TCA cycle, yet little else is known about the bacterium's metabolism [5]. ATCC 31749 has been shown to utilize citrate as a sole carbon source, presumably by the metabolism of citrate through the TCA cycle [6]. Since oligosaccharide synthesis requires a significant amount of energy, the energy generated from citrate consumption may explain the observed increase in oligosaccharide production. However, the oligosaccharide synthesis reaction includes sucrose, a preferred carbon source of ATCC 31749, which may prevent citrate uptake. Several other possible mechanisms may explain the citrate-induced stimulation of oligosaccharide production. Citrate is a known chelating agent, and as such, it may increase the solubility of metal ions in the reaction medium. Metal ions are important cofactors for many metabolic reactions. In particular, manganese is a cofactor for the β1,4-galactosyltransferase which synthesizes LacNAc [7]. An increase in manganese availability due to citrate chelation may explain the increase in oligosaccharide production. Citrate may also act as a buffer to control pH during oligosaccharide synthesis. This study investigates the possible mechanisms for the citrate stimulation of oligosaccharide production, utilizing metabolic flux analysis to gain insight about the effect of citrate on the entire metabolic network. Furthermore, we investigate whether the citrate-induced stimulation of oligosaccharide production is unique to *Agrobacterium* sp.

4.3 Results

4.3.1 Citrate stimulation of oligosaccharide synthesis

The addition of citrate to the reaction medium stimulated oligosaccharide synthesis in the engineered *Agrobacterium* strain, ATCC 31749/pKEL. To optimize the amount of citrate for oligosaccharide production, various concentrations of sodium citrate were added to the reaction buffer, ranging from 0 g/L to 10 g/L (Figure 4.1). The most dramatic improvement in oligosaccharide production was observed with the addition of 5 g/L sodium citrate. Increasing the concentration of sodium citrate from 2 g/L to 5 g/L, a 2.5-fold increase, led to nearly an equivalent improvement in oligosaccharide production, a 2.7-fold increase for LacNAc and a 2.4-fold increase for lactose. Doubling the concentration of sodium citrate to 10 g/L did not yield a similar increase in oligosaccharide production, indicating that an additional increase in citrate concentration will result in minimal improvement. The highest oligosaccharide concentrations were achieved with addition of 10 g/L sodium citrate, representing nearly a 7-fold improvement in LacNAc and over a 4-fold improvement in lactose production compared to the control with 0 g/L sodium citrate.

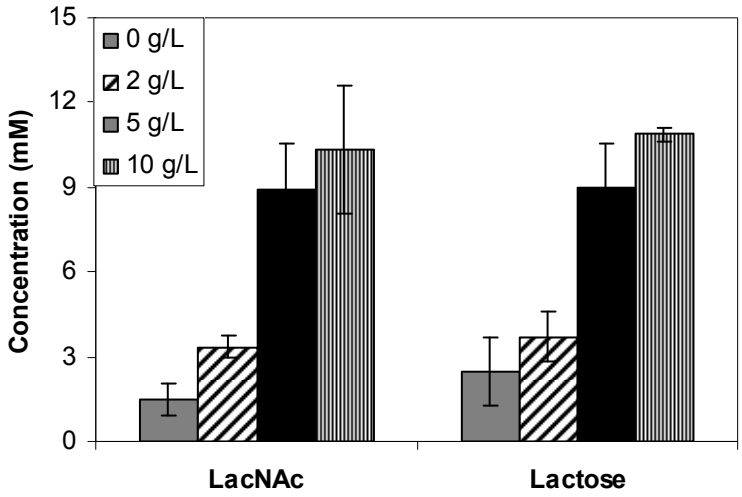


Figure 4.1 Oligosaccharide production in ATCC 31749/pKEL with varying concentrations of sodium citrate.

4.3.2 Mechanisms of citrate stimulation

From the known biological and chemical functions of citrate as a carbon and energy source, chelating agent, and buffer, there are several possible mechanisms for the observed stimulation of oligosaccharide production in ATCC 31749/pKEL. These potential mechanisms were investigated to gain insight into oligosaccharide synthesis in the engineered *Agrobacterium*.

4.3.2.1 Citrate as carbon source for energy production

Past reports have indicated that *Agrobacterium* sp. can grow using citrate as the sole carbon source [6]. As an intermediate of the TCA cycle, citrate can be directly consumed to provide both cellular energy and the carbon building blocks for other important metabolites. Citrate concentrations throughout the oligosaccharide synthesis reaction were measured to determine if citrate is consumed. Nearly 40% of the initial amount of citrate was consumed by ATCC 31749/pKEL during oligosaccharide synthesis, indicating that citrate is utilized as a carbon and energy source (Figure 4.2).

Surprisingly, the consumption of citrate as a carbon source did not prevent the uptake of sucrose for oligosaccharide production. In fact, sucrose consumption increases by 79% with the addition of citrate. Many microorganisms possess strict catabolite repression systems, preventing the simultaneous uptake of two carbon sources. However, it appears that ATCC 31749 does not employ this type of regulation with regard to citrate consumption, allowing simultaneous uptake of both citrate and sucrose.

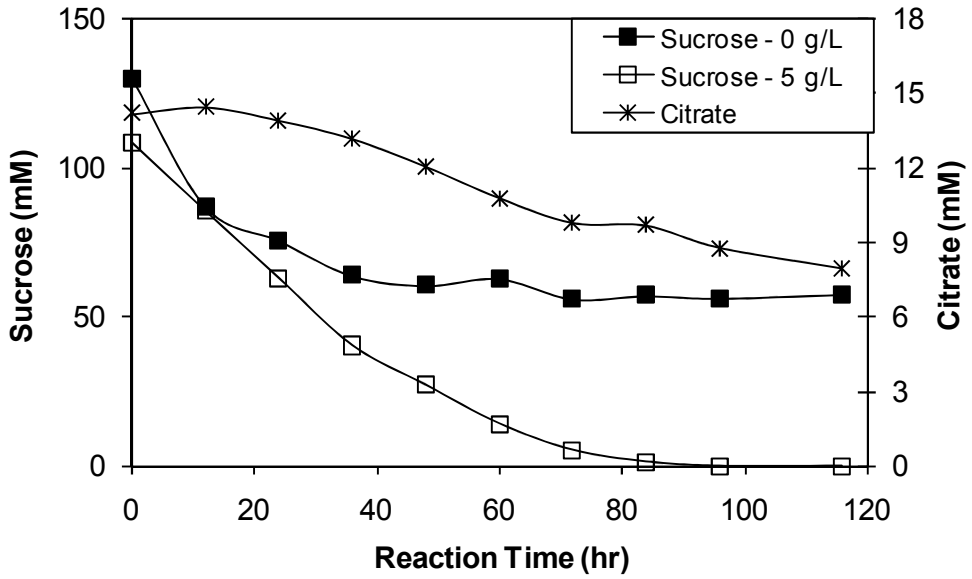


Figure 4.2 Sucrose and citrate concentrations during oligosaccharide synthesis. Sucrose concentration in citrate-free reaction (■). Sucrose concentration in reaction containing 5 g/L sodium citrate (□). Citrate concentration in reaction containing 5 g/L sodium citrate (*).

To confirm that the energy produced from citrate consumption leads to increased oligosaccharide synthesis, several other metabolites in the TCA cycle were investigated as potential energy sources. Both α-ketoglutarate and succinate stimulated oligosaccharide synthesis, with approximately a 3-fold increase in LacNAc (Figure 4.3) and 2-fold increase in lactose (data not shown). The effect of α-ketoglutarate and succinate on oligosaccharide synthesis is not as dramatic as the effect of citrate, yet this is

to be expected since the energy yield from α-ketoglutarate (9 ATP equivalents) and succinate (5 ATP equivalents) is less than that from citrate (12 ATP equivalents). However, both α-ketoglutarate and succinate, like citrate, can serve as chelating agents and buffers. These mechanisms must also be investigated.

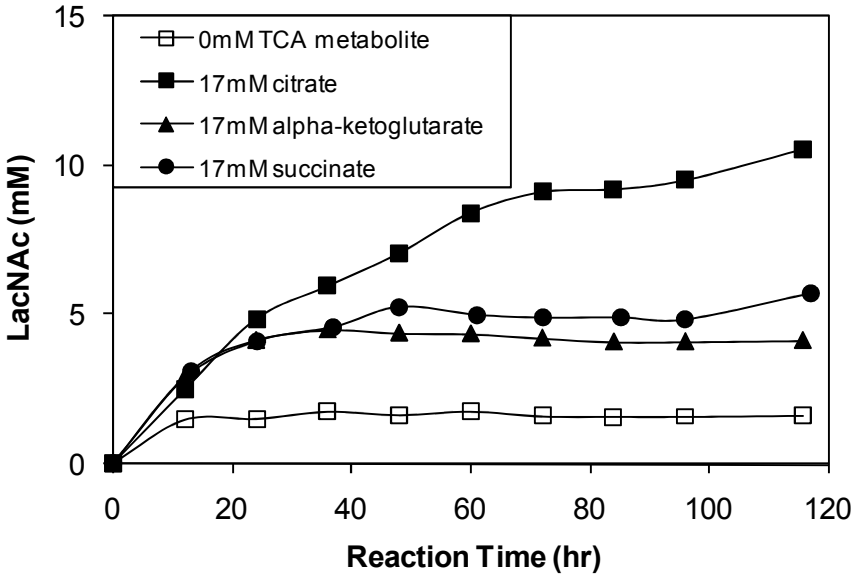


Figure 4.3 LacNAc synthesis in ATCC 31749/pKEL with addition of TCA cycle metabolites. Control with 0mM TCA cycle metabolites (□). Addition of 17mM citrate (■). Addition of 17mM α-ketoglutarate (▲). Addition of 17mM succinate (●).

4.3.2.2 Citrate as chelating agent

Many cellular enzymes require metal ions as cofactors, including the recombinant β1,4-galactosyltransferase (LgtB) from *Neisseria meningitidis* which is used for production of the β1,4-galactose-containing disaccharides in ATCC 31749/pKEL. Manganese (Mn^{2+}) is an essential cofactor for LgtB activity, with magnesium (Mg^{2+}) and calcium (Ca^{2+}) substitutions yielding only half the activity observed with Mn^{2+} [8]. As a chelating agent, citrate may enhance the solubility of Mn^{2+} in the reaction medium, thereby increasing the intracellular concentration of Mn^{2+} available as cofactor for LgtB. Subsequently, the enhanced activity of LgtB may lead to higher production of the

disaccharides LacNAc and lactose. To test this theory, oligosaccharide production was measured using a manganese-free reaction buffer and a reaction buffer containing an alternative chelating agent, phytic acid (Figure 4.4).

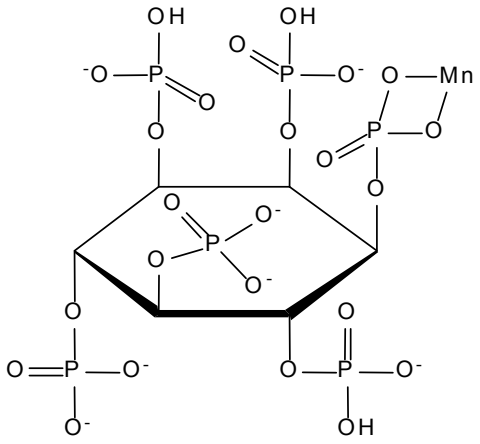


Figure 4.4 Chemical structure of phytic acid with manganese chelation.

When manganese, the preferred cofactor of LgtB, is eliminated from the reaction medium, the only available metal cofactor is magnesium. The magnesium ion will provide active LgtB, yet the activity will be reduced compared to that observed with Mn^{2+}. With the manganese-free reaction medium, LacNAc synthesis was reduced from 8.9 mM to 6.9 mM, and a similar decrease was observed for lactose. Despite this reduction, oligosaccharide synthesis is still more than 3-fold greater than that observed without citrate. This indicates that the citrate stimulation of oligosaccharide production is not solely due to increased manganese solubility. However, citrate may also influence magnesium solubility, requiring additional evidence to rule out chelation as a possible mechanism of the observed citrate stimulation.

Similar to citrate, phytic acid is a well-known chelating agent, yet unlike citrate, it is not a known metabolite of *Agrobacterium* sp. Furthermore, phytic acid has been used

in previous fermentations for oligosaccharide synthesis, presumably due to its chelating ability [9-10]. Substituting phytic acid for citrate in the reaction medium yields oligosaccharide levels comparable to the control reaction, containing neither citrate nor phytic acid (Figure 4.5). These results indicate that citrate chelation does not contribute significantly to the mechanism for citrate stimulation of oligosaccharide synthesis.

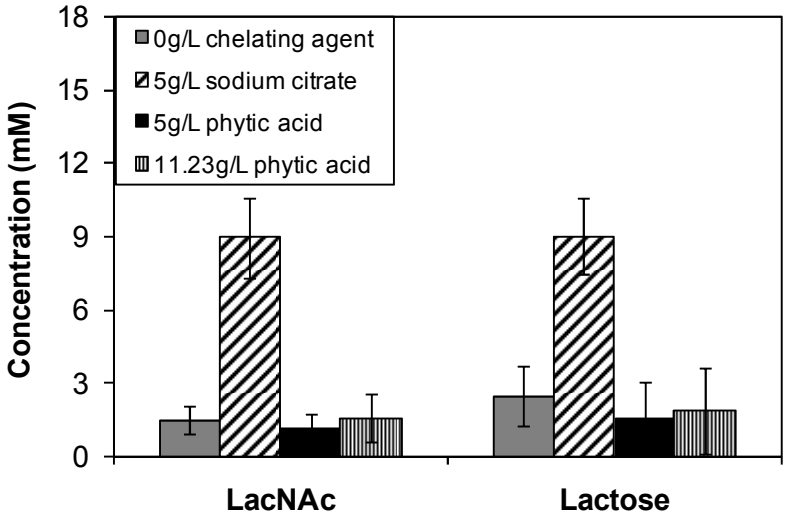


Figure 4.5 Oligosaccharide synthesis with the addition of a chelating agent.

4.3.2.3 Citrate as buffer

As a weak acid, citrate may act as a buffer to maintain the pH at an optimal level for oligosaccharide synthesis. In fact, citrate is a common buffer for enzymatic studies [11]. Citrate may either control the extracellular pH of the reaction, or if transported into the cell, citrate may regulate intracellular pH. During the course of oligosaccharide synthesis, the pH of the reaction becomes more acidic with time. In an effort to maintain constant pH, a Tris-HCl buffer (pH 7.5) was added to the reaction medium. With the presence of this buffer, it is unlikely that the addition of citrate had a significant buffering

effect on the reaction. Nevertheless, the influence of citrate's buffering capacity was investigated.

First, the optimum pH for LacNAc synthesis was determined. The optimal pH for polysaccharide production in ATCC 31749 is pH 5.5 [12], yet the optimal pH ranges for the recombinant enzymes GalE and LgtB are pH 7-8 and pH 6.5-7.0, respectively [7, 13]. Therefore, the oligosaccharide synthesis reaction was conducted at constant pH levels of 5.5, 6.0, 6.5, 7.0, and 7.5. At pH 6.0, ATCC 31749/pKEL synthesized the highest amount of LacNAc and lactose, indicating an optimum pH for oligosaccharide synthesis. Maintaining a constant pH of 6.0, ATCC 31749/pKEL was utilized for LacNAc synthesis both with and without the addition of citrate, and these results were compared reactions with no pH adjustment.

Surprisingly, the buffering capacity of citrate had a significant influence on oligosaccharide production (Figure 4.6). Over 60% of the increase in LacNAc production with the addition of citrate can be accounted for by the buffering mechanism. During oligosaccharide production, the pH of the reaction drops precipitously at the beginning of the reaction, reaching a minimum after 16 hours. Without the addition of citrate, the pH falls to 4.8, while with citrate, the pH of the reaction is 5.4. This small change in pH clearly has a dramatic effect on enzyme activity and oligosaccharide production. The increase in sucrose consumption with the addition of citrate can also be attributed to the buffering effect. With pH adjustment, the rate of sucrose consumption is similar for both reactions, regardless of the presence of citrate. While the buffering mechanism is a significant contribution to citrate stimulation of oligosaccharide production, a considerable increase in oligosaccharide production remains even with pH adjustment.

This additional improvement may be due solely to the use of citrate as a carbon and energy source as discussed above or due to some unidentified regulatory mechanism of citrate.

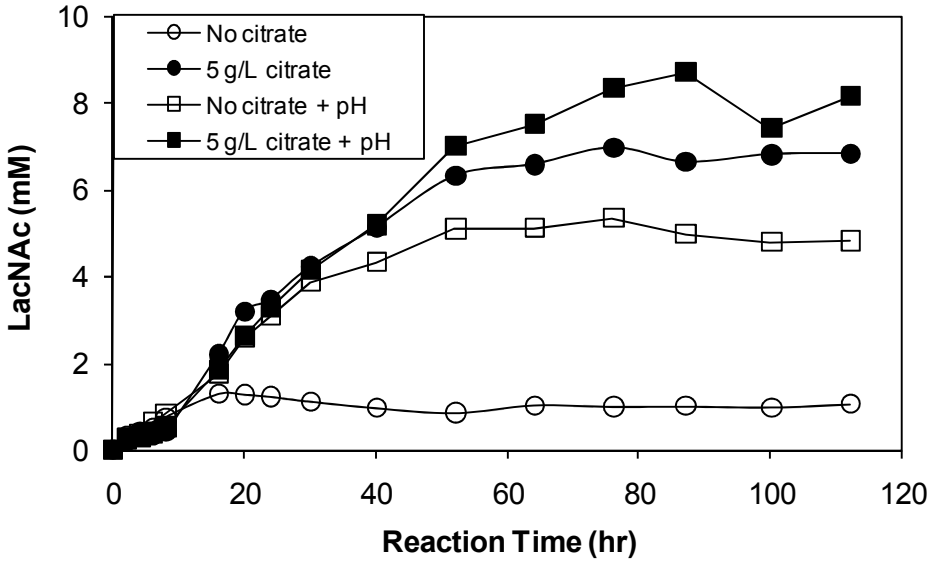


Figure 4.6 LacNAc synthesis by ATCC 31749/pKEL with and without citrate and pH adjustment. Conditions: 0 g/L sodium citrate, no pH adjustment (○); 5 g/L sodium citrate, no pH adjustment (●); 0 g/L sodium citrate with pH adjustment (□); 5 g/L sodium citrate with pH adjustment (■).

4.3.3 Simultaneous uptake of multiple carbon sources

With the ubiquitous evolution of catabolite repression mechanisms in bacteria, the simultaneous consumption of sucrose and citrate is an unexpected occurrence. To determine if this phenomenon is due to the specific combination of sucrose and citrate, glucose, an alternative carbon source, was tested with citrate for simultaneous consumption. ATCC 31749/pKEL consumed both glucose and citrate throughout the reaction, and similarly, oligosaccharide synthesis and glucose consumption increased with the addition of citrate (Table 4.1). However, the increase in oligosaccharide production is much higher when sucrose is used as the carbon source (4.5-fold increase)

than when glucose is employed (1.8-fold increase). This may be due to the fact that sucrose is a preferred carbon source of ATCC 31749. As evidence, the increase in sucrose consumption with the addition of citrate, a 79% increase, is greater than the increase in glucose consumption, a 55% increase. Overall, these results indicate that the simultaneous consumption of multiple carbon sources is not restricted to the use of sucrose, but instead, appears to be independent of the carbon source.

Table 4.1 Oligosaccharide synthesis with different carbon sources

Carbon Source[1]	Sodium citrate	LacNAc (mM)	Lactose (mM)	Carbon source consumed (mM)	Citrate consumed (mM)
Sucrose	0 g/L	1.48 ± 0.58	2.47 ± 1.22	80.1 ± 9.62	N/A
	5 g/L	8.93 ± 1.64	8.97 ± 1.56	143 ± 2.03	6.25 ± 2.04
Glucose	0 g/L	0.74 ± 0.18	5.30 ± 0.08	58.4 ± 19.6	N/A
	5 g/L	1.92 ± 0.33	8.93 ± 0.93	90.5 ± 18.1	4.64 ± 1.84

1. The concentration of carbon source is 50 g/L.

4.3.4 Metabolic flux analysis

From the results presented above, it is apparent that the addition of citrate influences multiple pathways including oligosaccharide synthesis, energy production, and carbon uptake. Metabolic flux analysis was performed to gain further insight into the effect of citrate on the entire metabolic network of the cell. Metabolic flux analysis is a method for calculating the intracellular mass fluxes using the stoichiometric reactions of the cell's metabolism and measured extracellular metabolites [14]. This method assumes a pseudo-steady state, where intracellular metabolites to not accumulate within the cell. The metabolic reaction network and equation used for the metabolic flux analysis are described in the materials and experimental methods section (5.4).

The first challenge in applying metabolic flux analysis is to determine which pathways are active during the oligosaccharide synthesis reaction. Previous experimental results indicate the Entner-Doudoroff (ED) pathway is utilized for sugar metabolism in *Agrobacterium* sp. [5]; therefore, this pathway was included in the metabolic network. The primary purpose of the pentose phosphate pathway is to produce reducing equivalents for anabolic reactions and to provide the cell with ribose-5-phosphate for synthesis of nucleotides. Under the conditions of the oligosaccharide synthesis reaction, both anabolic reactions and the need for nucleotides are minimal, allowing the pentose phosphate pathway to be excluded from the metabolic network. The oxygen level during oligosaccharide synthesis is a major influence on the metabolic network. The synthesis occurs under agitation, which provides a supply of oxygen; however, high cell concentration may lead to anaerobic conditions. To determine the oxygen status, the oligosaccharide synthesis reaction was conducted in a fermenter under two conditions: 1) dissolved oxygen controlled at 10% (aerobic) and 2) no air flow (anaerobic). Under aerobic conditions, the substrate, GlcNAc, was rapidly consumed as a nitrogen source for cell growth, resulting in no oligosaccharide synthesis. Under anaerobic conditions, very little carbon source was utilized, and oligosaccharide synthesis was minimal. These results suggest that oligosaccharide synthesis occurs under microaerobic conditions: oxygen is supplied to provide energy for oligosaccharide production, yet the oxygen level is too low to allow for the consumption of GlcNAc as a nitrogen source. To confirm that the anaerobic metabolism is active, the supernatant was analyzed for common anaerobic products, including lactate, acetate, ethanol, and succinate. Evidence of acetate production was found, and this anaerobic pathway was added to the metabolic network

along with the TCA cycle for aerobic respiration. To the best of our knowledge, this is the first report of anaerobic metabolism in ATCC 31749. Combining the oligosaccharide synthesis, ED, and acetate production pathways with the glyoxylate and TCA cycles provides the basic metabolic network for ATCC 31749 during oligosaccharide synthesis.

In metabolic flux analysis, the metabolic fluxes are assumed to be pseudo-steady state. In reality, however, fluxes are rarely pseudo-steady state, and metabolites frequently accumulate within the cell. Several metabolites were found to accumulate during oligosaccharide synthesis, namely glucose and fructose. Additional reactions were added to the metabolic network to account for the accumulation of these sugars. Furthermore, ATCC 31749 was found to consume the residual glycerol remaining from the wash step of cell preparation, and subsequently, the metabolic network was altered to reflect this additional carbon source. (See the Appendix A for a list of all reactions included in the network.)

With the metabolic network established, metabolic flux analysis was performed using fluxes measured from oligosaccharide synthesis reactions with and without the addition of citrate. (See Appendix A for calculations.) Figures 4.3 and 4.6 show three distinct phases within the LacNAc synthesis reaction. In the first phase (0-12 hours), both reactions, with and without citrate, synthesize LacNAc, and a negligible amount of citrate is consumed during this phase. In the second phase (12-60 hours), citrate is consumed, allowing the citrate-containing reaction mixture to continue LacNAc synthesis, while synthesis no longer occurs in the citrate-free reaction. In the third phase (60 – 120 hours), sucrose and citrate consumption slowly decrease for the citrate-containing reaction mixture, leading to a lower rate of oligosaccharide synthesis. Because the fluxes are very

similar during the first phase and the fluxes are not constant in the third phase, only the second phase will be compared using metabolic flux analysis. The results from this analysis are shown in Table 4.2, with the fluxes corresponding to the metabolic network in Figure 4.7. The citrate-containing reaction has higher rates of consumption for both sucrose and glycerol. In fact, no glycerol is consumed in the citrate-free reaction. The high rate of carbon consumption leads to a higher flux through the ED pathway, nearly a 400% increase compared to the citrate-free reaction, but surprisingly, the production of acetate only increased by 47%. Instead, the increased flux from the ED pathway was directed through the TCA cycle, leading to over a 6-fold increase in ATP production. This high rate of ATP generation may be responsible for the increase in oligosaccharide synthesis in the citrate-containing reaction. As expected, the flux through the glyoxylate cycle is negligible for both reactions since the glyoxylate cycle is primarily used for acetate and fatty acid metabolism [15].

Table 4.2 Metabolic flux analysis[2]

Flux	Enzyme(s)	Citrate-free	5 g/L sodium citrate
v_1*	sucrose hydrolase	19	100
v_2*	glucose accumulation	3.3	24
v_3*	fructose accumulation	4.3	8.8
v_4	glucokinase	16	73
v_5	fructokinase	15	91
v_6	phosphoglucose isomerase	15	90
v_7	phosphoglucomutase	-0.08	9.0
v_8	UTP glucose-1-phosphate uridylyltransferase	-0.08	9.0
v_9*	curdlan synthase	0.39	0.44
v_{10}	UDP-galactose 4'-epimerase	-0.47	8.6
v_{11}*	β1,4-galactosyltransferase (glucose as substrate)	-0.23	3.6
v_{12}*	β1,4-galactosyltransferase (GlcNAc as substrate)	-0.17	3.3
v_{13}*	β1,4-galactosyltransferase (mannose as substrate)	-0.07	1.6
v_{14}	glucose-6-phosphate dehydrogenase; 6-phospho-glucolactonase	31	153
v_{15}	6-phosphogluconate hydrolase; 2-keto-3-deoxy-gluconate aldolase	31	153
v_{16}	glycerol-3-phosphate dehydrogenase	-3.5	24
v_{17}*	glycerol kinase	-3.5	24
v_{18}	glyceraldehyde-3-phosphate dehydrogenase; phosphoglycerate kinase; phosphoglycerate mutase; pyruvate kinase	27	177
v_{19}	pyruvate dehydrogenase	58	330
v_{20}*	acetyl-CoA synthetase	17	25
v_{21}	citrate synthase	41	308
v_{22}*	citrate uptake	0	2.8
v_{23}	aconitase	41	310
v_{24}	isocitrate dehydrogenase	41	314
v_{25}	α-ketoglutarate dehydrogenase complex	41	314
v_{26}	succinyl-CoA synthetase	41	314
v_{27}	succinate dehydrogenase; fumarase	41	311
v_{28}	malate dehydrogenase	41	308
v_{29}	isocitrate lyase; malate synthase	0	-2.8
v_{30}	oxidative phosphorylation (NAD/NADH)	205	1466
v_{31}	oxidative phosphorylation ($FAD/FADH_2$)	41	311
v_{32}	maintenance energy (ATP consumption)	784	5516

2. Values normalized based on sucrose consumption in the 5 g/L sodium citrate reaction.
* Measured flux

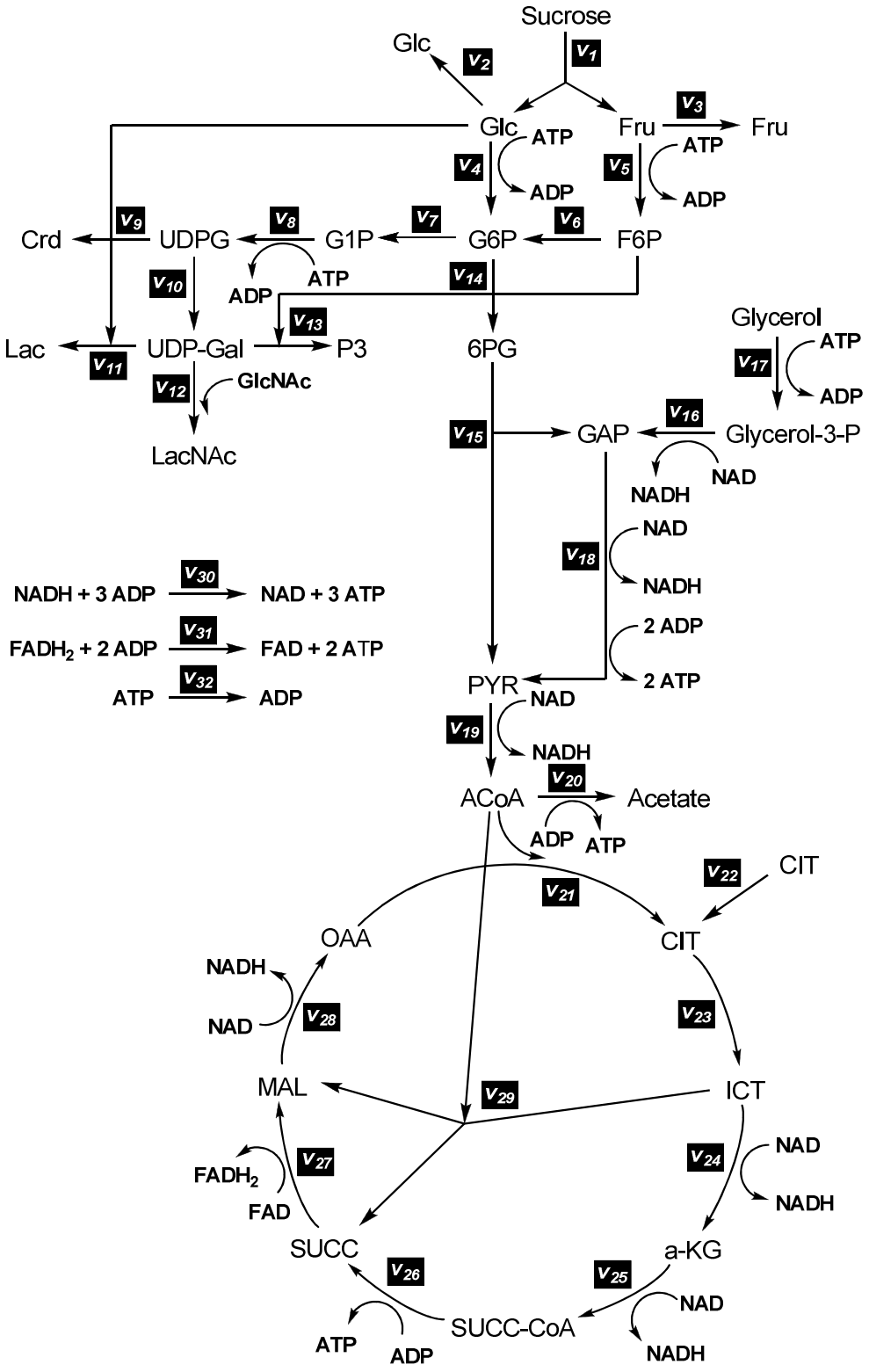


Figure 4.7 Metabolic network for ATCC 31749/pKEL. Metabolic fluxes are indicated by v_i with i = 1 to 32. Fluxes $v_{10} - v_{13}$ make up the engineered pathway for oligosaccharide synthesis, while all other fluxes are native to ATCC 31749.

4.3.5 Citrate effect unique to *Agrobacterium* sp.

To determine if the citrate stimulation of oligosaccharide synthesis is unique to *Agrobacterium* sp., an engineered *Escherichia coli* strain, AD202/pMUEL, was also tested. Oligosaccharide synthesis in this *E. coli* strain increased slightly with the addition of citrate (Table 4.3), yet this 30% increase is insignificant compared to the 350% improvement observed with the engineered *Agrobacterium* strain. Furthermore, citrate consumption by AD202/pMUEL is negligible throughout the synthesis reaction, and glucose consumption is similar for both the citrate-free and citrate-containing reactions. Based on these results, the citrate effect appears to be a unique property of *Agrobacterium* sp. ATCC 31749.

Table 4.3 Oligosaccharide synthesis and carbon consumption in AD202/pKEL with and without citrate addition

	citrate-free	5 g/L sodium citrate
LacNAc (mM)	0.30 ± 0.02	0.40 ± 0.01
Lactose (mM)	0.09 ± 0.05	0.12 ± 0.06
Glucose consumed (mM)	165 ± 15	180 ± 8.7
Citrate consumed (mM)	N/A	2.89 ± 0.33

4.4 Discussion and Conclusions

From the results presented in this study, it appears that the citrate-induced stimulation of oligosaccharide synthesis in ATCC 31749/pKEL is due to two primary mechanisms: 1) the consumption of citrate as a carbon and energy source and 2) pH maintenance due to the buffering capacity of citrate. While metal ions are important cofactors for oligosaccharide synthesis, citrate chelation does not appear to be significant.

A major finding of this work is the simultaneous uptake and utilization of multiple carbon sources by the engineered *Agrobacterium* sp. In many bacteria, catabolite repression prevents the simultaneous uptake of multiple carbon sources, making the co-metabolism of sucrose and citrate in ATCC 31749 a unique property. Only a few other bacterial species have been found to co-metabolize citrate and another carbon source. Microorganisms from the *Enterococcus*, *Leuconostoc*, *Lactobacillus*, and *Lactococcus* genera have been identified which can simultaneously utilize citrate and another carbon source such as glucose or lactose [16-21]. Most of these bacteria are grown under anaerobic or microaerophilic conditions, similar to the microaerobic conditions used for oligosaccharide synthesis with ATCC 31749/pKEL. Unlike ATCC 31749, these other citrate co-metabolizing organisms are all members of the Lactobacillales order. Several *Pseudomonas* species also co-metabolize citrate along with other carbon sources [22-23], yet to the best of our knowledge, this is the first report of multiple carbon source utilization by a member of the Rhizobiaceae family.

The synthesis of oligosaccharides in ATCC 31749/pKEL is an energy-intensive process, requiring two high-energy compounds, ATP and UTP, for each glycosidic linkage. The utilization of citrate by the TCA cycle provides additional energy for oligosaccharide production, with each citrate molecule generating 12 molecules of ATP for each turn through the TCA cycle. Assuming each citrate molecule goes through one complete turn of the TCA cycle, the consumption of 65 μmol of citrate would produce 780 μmol of ATP. The measured increase in oligosaccharide production due to citrate stimulation is 140 μmol, requiring 280 μmol of ATP. Thus, the energy required for

increased oligosaccharide production (280 μmol) can be completely accounted for by the ATP generated from citrate consumption (780 μmol).

Metabolic flux analysis revealed a large increase in the maintenance energy flux with the addition of citrate. The high maintenance energy flux, v_{32}, indicates that other ATP-consuming reactions may be active during oligosaccharide synthesis. For example, carbon uptake may occur via active transport. Close relatives of ATCC 31749, *A. radiobacter* and *A. tumefaciens*, are reported to use active transport for the uptake of glucose, galactose, fructose, xylose, and lactose [24-28]. It is likely that ATCC 31749 transports sucrose in a similar manner, requiring the use of ATP as an energy source. This method of sucrose transport may also explain the lack of carbon catabolite repression (CCR) in ATCC 31749 when both sucrose and citrate are present, for most bacterial CCR systems are associated with PTS-mediated sugar transport [29]. Assuming sucrose uptake does occur via active transport, this would only account for less than 3% of the citrate-induced increase in maintenance energy. Acidocalcisomes are another possible source of ATP consumption. Acidocalcisomes are intracellular organelles containing high amounts of inorganic polyphosphate (polyP), and these organelles are known to be produced in *A. tumefaciens*, a close relative of ATCC 31749 [30]. Furthermore, the genome sequence of ATCC 31749 contains two polyphosphate kinases for polyP synthesis, two exopolyphosphatases for polyP degradation, and a membrane-bound proton-translocating pyrophosphatase, enzymes shown to be associated with the acidocalcisomes of *A. tumefaciens.* Therefore, it is likely that ATCC 31749 also produces acidocalcisomes, requiring large quantities of ATP to produce polyP. The use of ABC

transporters for sucrose uptake and the storage of polyP in acidocalcisomes may explain the high maintenance energy flux for the citrate-containing reaction.

4.5 Materials and Experimental Methods

4.5.1 Materials

The chemicals used in this study were obtained from Sigma-Aldrich (lactose, LacNAc, GlcNAc, ATP, methyl blue), Mallinckrodt (glucose), Wako Chemicals USA (curdlan), and Fisher (all other chemicals).

4.5.2 Bacterial strains

The *Agrobacterium* sp. and *E. coli* strains used in this study, ATCC 31749/pKEL and AD202/pMUEL, were previously constructed [4, 31].

4.5.3 Oligosaccharide synthesis

The ATCC 31749/pKEL inoculums were prepared by overnight incubation at 30°C with agitation of 250 rpm in 3 mL of LB containing 100 μg/mL ampicillin and 100 μg/mL kanamycin. The cells were diluted 100 times and grown in freshly prepared LB medium supplemented with antibiotics until OD_{600} reached 0.3-0.4, upon which, IPTG was added to a final concentration of 0.75 mM, and cells were incubated for another 6 hours. The cells were harvested by centrifugation at 3,000 g and 4°C for 10 minutes. Cells were washed once with 10% glycerol prior to addition of the reaction buffer. The 10 mL reaction buffer contained 50 mM Tris-HCl (pH 7.5), 1 g/L $K_2HPO_4 \cdot 3H_2O$, 5 g/L $MgSO_4 \cdot 7H_2O$, 5 mM $MnCl_2$, 50 g/L sucrose (or glucose for *E. coli* strain), 10 g/L GlcNAc, and varying concentrations of sodium citrate. The cell concentration was 10% (wet weight), and the reaction vessel was a 50 mL flask. The reaction was carried out at

30ºC (or 25°C for *E. coli*) and 250 rpm in a biological shaker. For reactions with constant pH, HCl or NaOH were added every 2 hours to adjust the pH to the target level.

For reactions conducted in the Multifors fermenter system, two 150 mL inoculums were prepared as described above. These were concentrated to 50 mL by centrifugation at 3,000 g and 4°C for 10 minutes, followed by resuspension of the cell pellet. The 50 mL inoculums were added to 450 mL of sterilized LB media in the 500 mL fermenter vessels. The growth conditions include agitation at 600 rpm, an air flowrate of 0.5 vvm, temperature of 30°C, and pH 7.0. The pH was controlled using 1 M NaOH and 1 M HCl. Once the fermenter cultures reached an OD_{600} of 0.9, IPTG was added to a final concentration of 1 mM. After 10 hours of induction, the cells were collected by centrifugation. The cell pellets were resuspended in 300mL of sterile reaction buffer (described above). The reaction conditions for the fermenter are the same as the growth conditions with the following exceptions: pH was adjusted to 5.5, and there was no air flow for the anaerobic reaction.

4.5.4 Analytical methods

4.5.4.1 Carbohydrate analysis

Samples were heated in boiling water for 10 min. The supernatant was obtained by centrifugation and analyzed using a Dionex BioLC system with a CarboPac PA20 analytical column. The Dionex ED50 electrochemical detector measured carbohydrate concentrations through pulsed amperometry (waveform: t = 0.41 sec, p = -2.00 V; t = 0.42 sec, p = -2.00 V; t = 0.43 sec, p = 0.60 V; t = 0.44 sec, p = -0.10 V; t = 0.50 sec, p = -0.10 V). Sucrose, glucose, fructose, lactose and LacNAc concentrations were determined using calibration curves prepared from standards. The mobile phase consisted of

degassed 200 mM sodium hydroxide (A) and 18 MΩ-cm water (B), pressurized with inert gas (He) at a flow rate of 0.5 mL/min. The following linear gradient was used for LacNAc and lactose detection: t = 0 min, 5:95 (A:B); t = 5 min, 5:95; t = 10 min, 20:80; t = 25 min, 20:80; t = 25 min, 100:0; t = 40 min, 100:0; t = 40 min, 5:95; t = 55 min, 5:95. The following linear gradient was used for sucrose, glucose, and fructose detection: t = 0 min, 16:84 (A:B); t = 40 min, 16:84; t = 40 min, 100:0; t = 55 min, 100:0; t = 55 min, 16:84; t = 70 min, 16:84.

4.5.4.2 Citrate measurement

Samples (300 μL) were centrifuged at 13,200 rpm for 3 minutes. The supernatant was appropriately diluted, and citrate concentrations were measured using the pyridine-acetic anhydride method outlined by Marier and Boulet [32]. A Beckman Coulter DU 530 Life Sciences UV/Visible spectrophotometer was used for measurement.

4.5.4.3 Acetate and glycerol measurement

Samples (100 μL) were centrifuged at 13,200 rpm and 4°C for 10 min. A 10 % solution of trichloroacetic acid (100 μL) was added to 50 μL of the supernatant. The sample was placed on ice to allow for protein precipitation. After 10 min, the sample was centrifuged at 13,200 rpm and 4°C for 10 min. The supernatant was analyzed using an Agilent 1100 series HPLC with a 59346 Supelcogel H HPLC column. The mobile phase was a 0.1% (w/v) phosphoric acid solution. Analysis was performed using an isocratic program with a flowrate of 0.17 mL/min and a duration of 60 min.

4.5.4.4 Curdlan measurement

Samples for curdlan measurement were prepared using a standard method of curdlan isolation and dissolving the precipitated curdlan in 1 M NaOH [33]. Because

curdlan concentrations were below the level detectable by the common dry cell weight measurement, a more sensitive aniline blue assay for detecting β1,3-glucan was used [34-35]. Since aniline blue is very light sensitive, this assay was performed in the dark to prevent photobleaching. The microplate reader of a Molecular Devices SpectraMax M5 spectrophotometer was used for measurement, and for this instrument, the optimum excitation and emission wavelengths were determined to be 400 nm and 498 nm, respectively with a cutoff of 475 nm.

4.5.5 Metabolic flux analysis

The metabolic network for metabolic flux analysis was generated from the genome sequence of ATCC 31749 (to be described in a future publication) and from biochemical evidence found in the literature [36-38]. The specific pathways included in the metabolic network are discussed in the results. Overall, the network consists of thirty-two metabolic reactions and twenty-five metabolites. The metabolic reactions are listed in the appendix. Ten metabolic fluxes were measured (indicated in Appendix A), yielding an overdetermined system of equations. Using MATLAB, the system of equations was solved by applying the following equation:

$$\mathbf{A = -(inv(C*C')*C)*M'*V} \quad (1)$$

where **A** is a vector of calculated intracellular fluxes, **C** is the stoichiometric matrix of intracellular metabolic fluxes, **M** is the stoichiometric matrix of measured metabolic fluxes, and **V** is a vector of measured metabolic fluxes.

4.6 References

1. Dube, D.H. and C.R. Bertozzi, *Glycans in cancer and inflammation - potentials for therapeutics and diagnostics.* Nature Reviews: Drug Discovery, 2005. **4**: p. 477-488.

2. Seeberger, P.H. and D.B. Werz, *Synthesis and medical applications of oligosaccharides.* Nature, 2007. **446**: p. 1046-1051.

3. Ruffing, A. and R.R. Chen, *Metabolic engineering of microbes for oligosaccharide and polysaccharide synthesis.* Microbial Cell Factories, 2006. **5**: p. 25-33.

4. Ruffing, A., Z. Mao, and R.R. Chen, *Metabolic engineering of Agrobacterium sp. for UDP-galactose regeneration and oligosaccharide synthesis.* Metabolic Engineering, 2006. **8**: p. 465-473.

5. Kai, A., et al., *Analysis of the biosynthetic process of cellulose and curdlan using ^{13}C-labeled glucoses.* Carbohydrate Polymers, 1994. **23**: p. 235-239.

6. Phillips, K.R. and H.G. Lawford, *Curdlan: Its properties and production in batch and continuous fermentations*, in *Progress in Industrial Microbiology*, D.E. Bushell, Editor. 1983, Elsevier: Amsterdam. p. 201-209.

7. Park, J.E., et al., *Expression and characterization of beta-1,4-galactosyltransferase from Neisseria meningitidis and Neisseria gonorrhoeae.* Journal of Biochemistry and Molecular Biology, 2002. **35**(3): p. 330-336.

8. Park, J.E., et al., *Expression and characterization of β-1,4-galactosyltransferase from Neisseria meningitidis and Neisseria gonorrhoeae.* Journal of Biochemistry and Molecular Biology, 2002. **35**(3): p. 330-336.

9. Koizumi, S., et al., *Large-scale production of GDP-fucose and Lewis X by bacterial coupling.* Journal of Industrial Microbiology & Biotechnology, 2000. **25**: p. 213-217.

10. Koizumi, S., et al., *Large-scale production of UDP-galactose and globotriose by coupling metabolically engineered bacteria.* Nature Biotechnology, 1998. **16**: p. 847-850.

11. Gomori, G., *Preparation of buffers for use in enzyme studies*. Methods in Enzymology, ed. Colowick and Kaplan. Vol. 1. 1955, New York: Academic Press.

12. Lee, J.H., et al., *Optimal pH control of batch processes for production of curdlan by Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1999. **23**: p. 143-148.

13. Wilson, D.B. and D.S. Hogness, *The enzymes of the galactose operon in Escherichia coli. I. Purification and characterization of uridine diphosphogalactose 4-epimerase.* The Journal of Biological Chemistry, 1964. **239**(8): p. 2469-2481.

14. Stephanopoulos, G.N., A.A. Aristidou, and J. Nielsen, *Metabolic Engineering: Principles and Methodologies*. 1998, San Diego: Academic Press. 725.

15. Cozzone, A.J., *Regulation of acetate metabolism by protein phosphorylation in enteric bacteria.* Annual Review of Microbiology, 1998. **52**: p. 127-164.

16. Goupry, S., et al., *Metabolic flux in glucose/citrate co-fermentation by lactic acid bacteria as measured by isotopic ratio analysis.* FEMS Microbiology Letters, 2000. **182**: p. 207-211.

17. Jyoti, B.D., A.K. Suresh, and K.V. Venkatesh, *Diacetyl production and growth of Lactobacillus rhamnosus on multiple substrates.* World Journal of Microbiology & Biotechnology, 2003. **19**: p. 509-514.

18. Salou, P., P. Loubiere, and A. Pareilleux, *Growth and energetics of Leuconostoc oenos during cometabolism of glucose with citrate or fructose.* Applied and Environmental Microbiology, 1994. **60**: p. 1459-1466.

19. Sarantinopoulos, P., G. Kalantzopoulos, and E. Tsakalidou, *Citrate metabolism by Enterococcus faecalis FAIR-E 229.* Applied and Environmental Microbiology, 2001. **67**(12): p. 5482-5487.

20. Schmitt, P. and C. Divies, *Co-metabolism of citrate and lactose by Leuconostoc mesenteroides subsp. cremoris.* Journal of Fermentation and Bioengineering, 1991. **71**: p. 72-74.

21. Vaningelgem, F., et al., *Cometabolism of citrate and glucose by Enterococcus faecium FAIR-E 198 in the absence of cellular growth.* Applied and Environmental Microbiology, 2006. **72**(1): p. 319-326.

22. Molin, G., *Mixed carbon source utilization of meat-spoiling Pseudomonas fragi 72 in relation to oxygen limitation and carbon dioxide inhibition.* Applied and Environmental Microbiology, 1985. **49**: p. 1442-1447.

23. Ng, F.M.-W. and E.A. Dawes, *Chemostat studies on the regulation of glucose metabolism in Pseudomonas aeruginosa by citrate.* Biochemical Journal, 1973. **132**: p. 129-140.

24. Cornish, A., J.A. Greenwood, and C.W. Jones, *Binding-protein-dependent glucose transport by Agrobacterium radiobacter grown in continuous culture.* Journal of General Microbiology, 1988. **134**: p. 3099-3110.

25. Cornish, A., J.A. Greenwood, and C.W. Jones, *Binding-protein-dependent sugar transport by Agrobacterium radiobacter and A. tumefaciens grown in continuous culture.* Journal of General Microbiology, 1989. **135**: p. 3001-3013.

26. Greenwood, J.A., A. Cornish, and C.W. Jones, *Binding-protein-dependent lactose transport in Agrobacterium radiobacter.* Journal of Bacteriology, 1990. **172**(4): p. 1703-1710.

27. Williams, S.G., J.A. Greenwood, and C.W. Jones, *Molecular analysis of the lac operon encoding the binding-protein-dependent lactose transport system and beta-galactosidase in Agrobacterium radiobacter.* Molecular Microbiology, 1992. **6**(13): p. 1755-1768.

28. Williams, S.G., J.A. Greenwood, and C.W. Jones, *Agrobacterium radiobacter and related organisms take up fructose via a binding-protein-dependent active-transport system.* Microbiology, 1995. **141**: p. 2601-2610.

29. Deutscher, J., *The mechanisms of carbon catabolite repression in bacteria.* Current Opinion in Microbiology, 2008. **11**(2): p. 87-93.

30. Seufferheld, M., et al., *Identification of organelles in bacteria similar to acidocalcisomes of unicellular eukaryotes.* The Journal of Biological Chemistry, 2003. **278**(32): p. 29971-29978.

31. Mao, Z., H.-D. Shin, and R.R. Chen, *Engineering the E. coli UDP-glucose synthesis pathway for oligosaccharide synthesis.* Biotechnology Progress, 2006. **22**: p. 369-374.

32. Marier, J.R. and M. Boulet, *Direct determination of citric acid in milk with an improved pyridine-acetic anhydride method.* Journal of Dairy Science, 1958. **41**: p. 1683-1692.

33. Kim, M.K., et al., *Residual phosphate concentration under nitrogen-limiting conditions regulates curdlan production in Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 2000. **25**: p. 180-183.

34. Ko, Y.-T. and Y.-L. Lin, *1,3-β-Glucan quantification by a fluorescence microassay and analysis of its distribution in foods.* Journal of Agricultural and Food Chemistry, 2004. **52**: p. 3313-3318.

35. Shedletzky, E., C. Unger, and D.P. Delmer, *A microtiter-based fluorescence assay for (1,3)-β-glucan synthases.* Analytical Biochemistry, 1997. **249**: p. 88-93.

36. Arthur, L.O., et al., *Carbohydrate metabolism in Agrobacterium tumefaciens.* Journal of Bacteriology, 1973. **116**(1): p. 304-313.

37. Arthur, L.O., et al., *Carbohydrate catabolism of selected strains in the genus Agrobacterium.* Applied Microbiology, 1975. **30**(5): p. 731-737.

38. Fuhrer, T., E. Fischer, and U. Sauer, *Experimental Identification and quantification of glucose metabolism in seven bacterial species.* Journal of Bacteriology, 2005. **187**(5): p. 1581-1590.

CHAPTER 5

GENOME SEQUENCING OF *AGROBACTERIUM* SP. ATCC 31749

5.1 Abstract

Agrobacterium sp. ATCC 31749 is an important industrial microorganism for its production of the natural polysaccharide known as curdlan. Our previous work has also shown this organism to be a good host for the production of valuable oligosaccharides for medical applications. Despite the importance of ATCC 31749 for poly- and oligosaccharide production, very little is known regarding the genetic composition of this organism. In this study, the genome sequence of ATCC 31749 is obtained using the Genome Sequencer FLX system from 454 Life Sciences. Subsequent assembly, gene prediction, and annotation resulted in a genome size of 5.4 Mb with 5,585 predicted genes and metabolic pathways including the Entner-Doudoroff and pentose phosphate pathways as well as the TCA and glyoxylate cycles. Comparative genomics showed ATCC 31749 to be highly homologous to the plant pathogen *Agrobacterium tumefaciens*, with 87% of genes having homologs in *A. tumefaciens*. Lastly, genes were identified which may contribute to the regulation of curdlan synthesis, including those involved in exopolysaccharide, nitrogen-limited, oxygen-limited, and phosphate-limited regulation as well as genes important for energy production.

5.2 Introduction

The parent strain of *Agrobacterium* sp. ATCC 31749 was first isolated from a soil sample in 1964 by Harada and Yoshimura [1], and this microorganism was studied for its ability to produce two polysaccharides, succinoglycan and curdlan. After several

spontaneous mutations, the curdlan-producing strain, ATCC 31749, was isolated [2] (see Chapter 1, section 1.3.2 History of ATCC 31749, for details). Since its isolation in 1983, ATCC 31749 has primarily been used for the production of the curdlan polysaccharide. Curdlan itself is used in a wide variety of commercial applications, ranging from a thickener and gelling agent in the food industry to an additive in the production of superworkable concrete [3]. Curdlan and its derivatives have also been applied in the medical industry as a drug release vehicle, anticoagulant, HIV treatment, and anti-tumor agent [4, 5]. With the wide range of applications and extensive use of curdlan, Takeda Chemical Industries Ltd., a commercial provider of curdlan, produces 600 to 700 tons of curdlan annually [3]. The economical production of curdlan from ATCC 31749 and its close relatives enables the use of curdlan in applications ranging the relatively low-value products of the food industry to traditionally high-cost medical applications such as HIV treatments.

In more recent work, ATCC 31749 was employed for a process other than curdlan production. We have shown that ATCC 31749 can be successfully engineered to produce medically-relevant oligosaccharides, namely LacNAc and Gal-α1,3-Lac, and a polysaccharide, hyaluronan [6, 7] (Chapters 2 and 3). These oligo- and poly-saccharides are essential components of medical treatments including vaccines for the flu, HIV, and cancer, as well as treatment for inflammation and applications in surgical procedures. ATCC 31749 was selected as host for the whole-cell production of oligo- and poly-saccharides due to its high rate of curdlan production. Curdlan, a glucose polymer, is synthesized from UDP-glucose in ATCC 31749, and the high rate of curdlan production requires a high rate of UDP-glucose synthesis in this microorganism. The sugar

nucleotide UDP-glucose is also a precursor for UDP-galactose, the substrate for synthesizing the LacNAc and Gal-α1,3-Lac oligosaccharides, and UDP-glucuronic acid, a precursor for hyaluronan synthesis. Based on this reasoning, the high UDP-glucose production of ATCC 31749 can be exploited for oligosaccharide synthesis with the addition of only one or two enzymes. While the initial engineering of ATCC 31749 yielded modest amounts of the oligo- and polysaccharide products (7.5 g/L LacNAc, 1 g/L Gal-α1,3-Lac, and 0.3 g/L hyaluronan), a high production rate is required for industrial-scale synthesis. Further improvement of oligo- and poly-saccharide production in ATCC 31749 will require additional metabolic engineering.

The production of oligo- and polysaccharides in ATCC 31749 may be enhanced by modifying the genes involved in curdlan synthesis and regulation. Unfortunately, little is known regarding the genetics of ATCC 31749 and curdlan production. Initial investigation of ATCC 31749 identified 5 genes as important for curdlan production [8]. A phosphatidylserine synthase, which produces a component of the cell membrane, was shown to impact curdlan synthesis [9]. In addition, three genes of unidentified function, *crdA*, *crdC,* and *crdR*, were also necessary for producing curdlan. The fifth gene, curdlan synthase (*crdS*), was determined to be the enzyme catalyzing the transfer of the glucose residue from UDP-glucose to the growing polymer chain of curdlan [10]. Of these 5 genes, only one gene sequence is reported, that of the curdlan synthase. Based on the nucleotide sequence, Stanisich and coworkers suspect that *crdR* is somehow involved in regulating curdlan synthesis, but no additional information is known about the genes involved in the regulation of curdlan production [11]. These regulatory genes are primary targets for future metabolic engineering to improve oligo- and polysaccharide production

in ATCC 31749. The primary metabolic pathways of ATCC 31749 are also likely to impact curdlan synthesis. Biochemical studies have indicated that the Entner-Doudoroff pathway, TCA cycle, and pentose phosphate pathway are active in ATCC 31749 [12, 13]. However, the genetics of these pathways have not been confirmed and many of the organism's metabolic pathways remain unknown. Future improvement of oligo- and polysaccharide production by ATCC 31749 requires additional genetic information regarding the metabolism of ATCC 31749 as well as the identification of genes involved in regulating curdlan production.

5.3 Results

5.3.1 ATCC 31749 genome sequence and annotation

The genomic DNA of *Agrobacterium* sp. ATCC 31749 was isolated and sequenced according to the protocol described in the methods section (5.4). A total of 92,994,272 base pair (bp) were sequenced in 399,219 reads, with an average read length of 233 bp. The read length distribution, as illustrated in Figure 5.1, indicates good quality sequencing results, with the majority of reads falling between 220 to 270 bp. The GS De Novo Assembler from 454 Life Sciences assembled the reads into 176 contigs with an average contig size of 31,122 bp. The genome sequence assembly gives an approximate genome size of 5.5 Mbp with 59% GC content and approximately 16 x sequence coverage. The 176 contigs were arranged in random order with stop codons inserted between each contig to generate a pseudochromosome, and this pseudochromosome was used for gene prediction and annotation. The pseudochromosome of ATCC 31749 includes 5,585 predicted genes, of which 3,466 have assigned function, 1,635 are conserved hypothetical proteins, and the remaining 484 are hypothetical proteins.

Analysis of transfer RNA's identified tRNA genes for each of the 20 amino acids, often with multiple genes for the same amino acid (Table 5.1). The tRNA distribution is similar to other *Agrobacterium* species, yet as shown in Table 5.1, this distribution differs significantly from other organisms like *E. coli*. In general, *E. coli* K12 has more tRNA genes, which may explain why recombinant protein production is higher in *E. coli* compared to *Agrobacterium.* Using the ribosomal RNA sequences from other *Agrobacterium* as template, the 5S, 6S, 16S, and 23S rRNA genes were located in ATCC 31749.

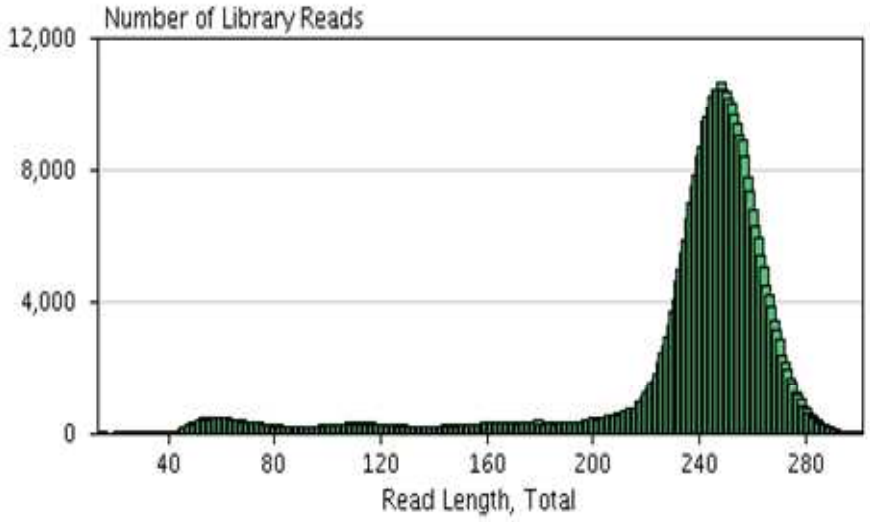


Figure 5.1 Read length distribution from genome sequencing

Table 5.1 tRNA gene distribution in ATCC 31749 genome sequence compared to an *E. coli* K12 strain.

tRNA	ATCC 31749	*E. coli* K12	tRNA	ATCC 31749	*E. coli* K12
Ala	2	5	Leu	5	8
Arg	3	7	Lys	2	6
Asn	1	4	Met	3	6
Asp	2	3	Phe	1	2
Cys	1	1	Pro	3	3
Gln	2	4	Ser	4	5
Glu	2	4	Thr	3	4
Gly	4	6	Trp	1	1
His	1	1	Tyr	1	3
Ile	1	5	Val	2	7

The genome sequence and annotation were uploaded into the Pathway Tools software for operon and metabolic pathway prediction. An overview of the predicted metabolic pathways is shown in Figure 5.2. This includes the Entner-Doudoroff (ED) pathway for glycolysis, the pentose phosphate pathway (PPP) for regeneration of reducing equivalents, and the tricarboxylic acid (TCA) and glyoxylate cycles for energy generation. Genetic evidence of these metabolic pathways confirms previous pathway predictions based on biochemical evidence [12, 13]. Unexpectedly, there are also genes present for a functional Embden-Meyerhof-Parnas (EMP) pathway for glycolysis. ATCC 31749 lacks a major enzyme of the EMP pathway: phosphofructokinase, the enzyme typically used to catalyze the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate with concomitant consumption of ATP. However, the genome does contain a pyrophosphate-fructose-6-phosphate 1-phosphotransferase gene which catalyzes the same conversion yet uses pyrophosphate (PPi) instead of ATP for the phosphorylation.

Despite possessing genes for a functional EMP pathway, experimental evidence shows that ATCC 31749 utilizes the lower energy-yielding ED pathway for glycolysis [12].

The ATCC 31749 genome also revealed genetic evidence to support other experimental observations. Sucrose has been shown to be a preferred carbon source of ATCC 31749, with higher levels of curdlan produced with sucrose as the carbon source compared to other sugars including glucose, fructose, galactose, lactose, and raffinose [14]. Several genes for sucrose degradation were identified in the genome: a sucrose hydrolase and three α-glucosidases. The genome sequence of *A. tumefaciens* includes a putative sucrose phosphorylase which converts sucrose and UDP directly to UDP-glucose and fructose. This enzyme saves two energy equivalents over the traditional pathway for UDP-glucose synthesis and could help explain why ATCC 31749 can produce more curdlan using sucrose compared to other sugars. Surprisingly, ATCC 31749 did not have a sucrose phosphorylase gene in the draft genome sequence, and the reason for sucrose as a preferred carbon source remains unknown. ATCC 31749 has also been shown to be capable of utilizing citrate as the sole carbon source [3]. In support of this observation, four iron (III) dicitrate ABC transporter genes were identified for citrate uptake as well as a citrate lyase gene for citrate metabolism. The ATCC 31749 genome provides crucial information for the future elucidation of carbon uptake and utilization in this microorganism.

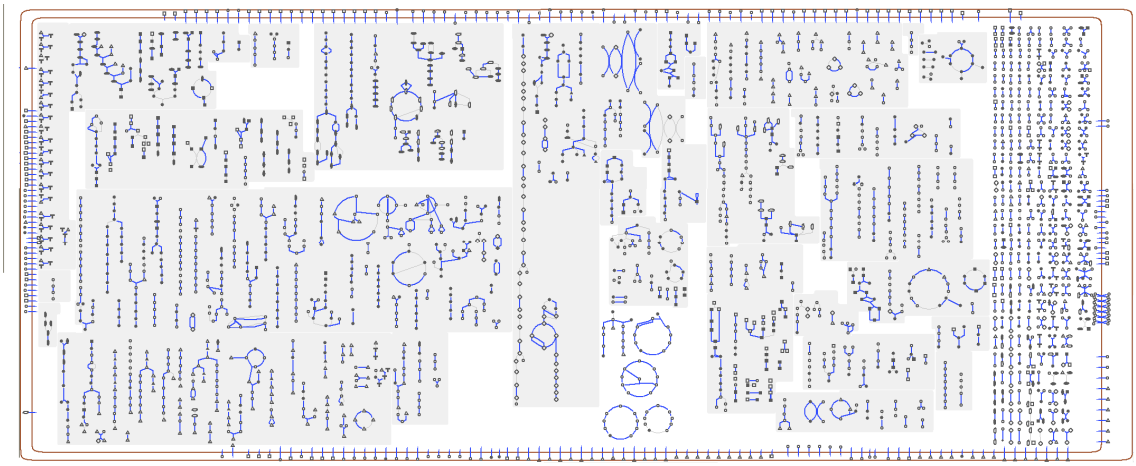


Figure 5.2 Overview of predicted metabolic pathways in ATCC 31749 using Pathway Tools. Each dot represents a metabolite, and each blue line represents an enzyme or protein. Transport reactions are shown on the red line representing the cell membrane.

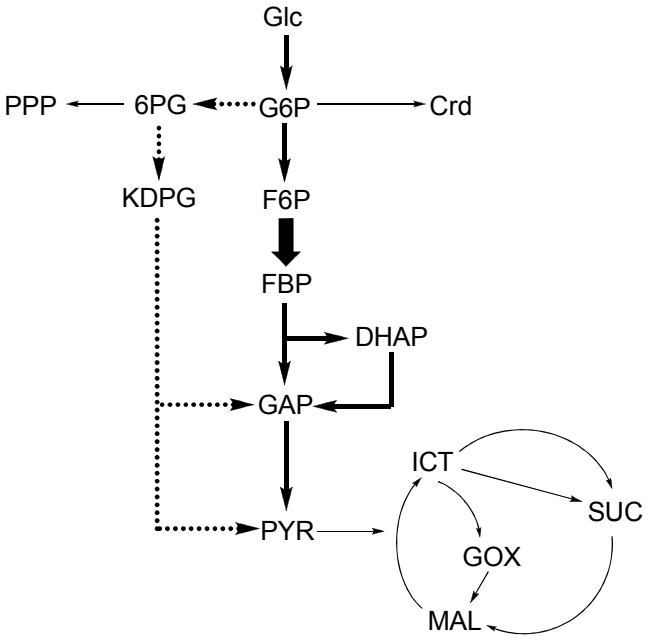


Figure 5.3 Major metabolic pathways predicted from the ATCC 31749 genome. Dashed arrows: ED pathway; bold arrows: EMP pathway; thick arrow: PPi-F6P 1-phosphotransferase.

As ATCC 31749 is primarily known for production of the curdlan polysaccharide, the genome sequence was also analyzed for genes involved in exopolysaccharide production. Of the five genes identified to be important for curdlan production, four genes were identified in the genome sequence: the three genes of the curdlan synthesis operon (*crdA, crdS, crdC*) and the phosphatidylserine synthase pss_{AG}. The fifth gene, *crdR*, is most likely present in the sequence, but since the nucleotide sequence for this gene has not been reported, it could not be identified at this time. Genes were identified for the synthesis of other polysaccharides, including succinoglycan, cyclic β1,3-glucan, a generic exopolysaccharide, and cellulose. It is of interest to note that both ATCC 31749 and *A. tumefaciens* C58 genomes contain the genes needed for both curdlan and cellulose synthesis, yet only ATCC 31749 is reported to produce curdlan and only *A. tumefaciens* is reported to produce cellulose. Cellulose, a β1,4-glucan polymer, and curdlan, a β1,3-glucan polymer, are very similar in chemical composition and structure (both are linear polymers). Thus, it would be interesting to investigate why each microorganism has a propensity for the production of just one particular glucan polysaccharide. As an initial investigation, the nucleotide sequences of the curdlan and cellulose synthesis operons for ATCC 31749 and *A. tumefaciens* C58 were compared to identify sequence changes that may explain this phenomenon. Comparison of the curdlan synthesis operons revealed only a few changes in the amino acid sequence of the *crdASC* genes and in the nucleotide sequence of the promoter and intergenic regions of the operon (Table 5.2). The only gene with known function and catalytic active sites is the curdlan synthase gene, *crdS*. Of the 5 amino acid mutations in *crdS* for *A. tumefaciens*, only one falls within the substrate binding site, but this residue is not known to directly participate in substrate binding [10].

As the mechanism and regulation of curdlan synthesis is largely unknown, the reason for the lack of curdlan synthesis in *A. tumefaciens* cannot be determined. Sequence alignment of the cellulose synthesis operons of ATCC 31749 and *A. tumefaciens* revealed a 852 bp insertion between the *celD* and *celE* genes in the ATCC 31749 *cel* operon. This insertion may prevent the transcription of *celE*, a gene which was shown to be necessary for cellulose synthesis in *A. tumefaciens* [15]. The absence of CelE in ATCC 31749 may explain why this microorganism does not produce the cellulose polysaccharide.

Table 5.2 Comparison of curdlan synthesis operons in ATCC 31749 and *A. tumefaciens*

Region	Total (bp)	Mismatches (bp)	Amino acid changes
Promoter	466	0	-
crdA	1458	13	2
crdA-crdS intergenic	131	2	-
crdS	1956	23	5
crdS-crdC intergenic	153	3	-
crdC	1166	20	6

5.3.2 Phylogeny and comparative genomics of ATCC 31749

ATCC 31749 is an α-proteobacterium of the Rhizobium family. Species in the Rhizobia family are typically plant-associated microorganisms. Many *Rhizobium* species are plant symbionts, forming root nodules on leguminous plants wherein the *Rhizobium* fixes nitrogen for the plant and the plant provides the bacterium with essential nutrients [16]. On the other hand, many *Agrobacterium* species are plant pathogens, causing gall (i.e. tumors) or a condition known as hairy root. It has been found that the pathogenicity of *Agrobacteria* is determined by a plasmid, either a Ti plasmid for tumor-inducing

species or a Ri plasmid for rhizogenic species [17]. As ATCC 31749 is not known to be either a plant symbiont or a phytopathogen, the relationship between ATCC 31749 and other members of the Rhizobia family was investigated.

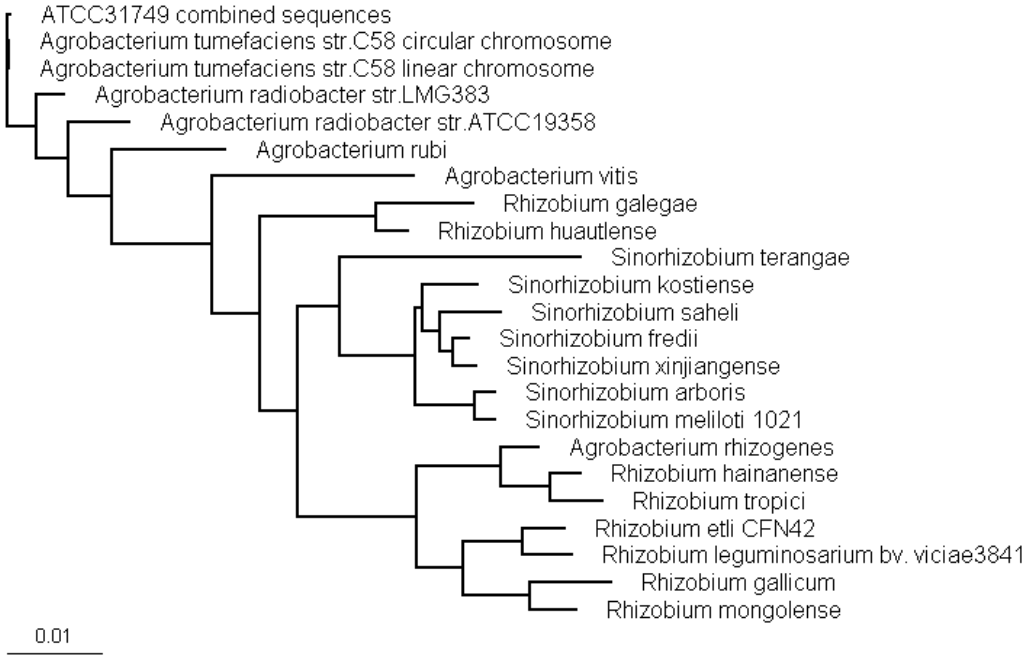


Figure 5.4 Phylogenetic tree constructed using 16S rRNA sequences of species in the Rhizobia family.

Prior to genome sequencing, the 16S rRNA gene of ATCC 31749 was identified and sequenced. Comparing the 16S rRNA sequence of ATCC 31749 to that of other Rhizobia revealed that ATCC 31749 is closely related to the phytopathogen, *A. tumefaciens* (Figure 5.4). After genome sequencing and annotation, the gene sequences of ATCC 31749 were compared to the protein-coding genes of the three other *Agrobacterium* species with sequenced genomes. Table 5.3 shows the number of genes in ATCC 31749 with predicted homologs in each respective genome. These results confirm that ATCC 31749 is closely related to *A. tumefaciens*, with more than 87% of genes having orthologs in *A. tumefaciens*. The protein-coding genes of the four sequenced

Agrobacteria (ATCC 31749, *A. tumefaciens* C58, *A. radiobacter* K84, and *A. vitis* S4) were compared to identify the core genome (homologous genes in all four genomes) and genes specific to ATCC 31749. Of the 599 genes specific to ATCC 31749, 74% are hypothetical proteins, 17% are conserved hypothetical proteins, and less than 9% have a predicted function. While all four *Agrobacteria* had homologs of curdlan synthase, only ATCC 31749 and *A. tumefaciens* C58 possessed all 3 genes of the curdlan synthesis operon. Similarly, all four strains had cellulose synthase homologs, but just three of the four strains (ATCC 31749, *A. tumefaciens* C58, and *A. radiobacter* K84) had homologs of the 6 genes required for cellulose synthesis. The 3,994 genes of the *Agrobacterium* core genome were also compared to other species of the Rhizobia family with available genome sequences (Table 5.4). Surprisingly, the six *Rhizobium* and two *Sinorhizobium* species have high homology with the *Agrobacterium* core genome, containing homologs for 95 to 98% of the genes of the core genome. This comparison suggests that the primary metabolic pathways are conserved throughout the Rhizobia family as well as many of the secondary pathways.

Table 5.3 Comparison of ATCC 31749 with three *Agrobacterium* species with sequenced genomes

Strain	# homologous genes	% of ATCC 31749 genes that are homologs
A. tumefaciens C58	4868	87.2 %
A. radiobacter K84	4311	77.2 %
A. vitis S4	4288	76.8 %

Table 5.4 Comparison of the *Agrobacterium* core genome with other Rhizobia species

Strain	# homologous genes	% of *Agrobacterium* core genes that are homologs
Rhizobium etli CFN 42	3884	97.2 %
Rhizobium etli CIAT 652	3852	96.4 %
R. leguminosarum bv. *trifolii* WSM 1325	3892	97.4 %
R. leguminosarum bv. *trifolii* WSM 2304	3888	97.3 %
R. leguminosarum bv. *viciae* 3841	3912	97.9 %
Rhizobium sp. NGR 234	3832	95.9 %
Sinorhizobium medicae WSM 419	3833	96.0 %
Sinorhizobium meliloti 1021	3829	95.9 %

5.3.3 Genes relevant to curdlan synthesis and regulation

The main motivation for sequencing the genome of ATCC 31749 is to discover genes involved in curdlan synthesis and regulation. From information available in literature, several gene categories were identified which may participate in curdlan synthesis and regulation. The ATCC 31749 genome was analyzed for genes associated with these categories, and the resulting list can be found in Table 5.5.

Table 5.5 ATCC 31749 genes with potential involvement in curdlan synthesis or regulation

Gene category	Genes in ATCC 31749 genome
EPS regulation (cyclic di-GMP)	GGDEF-containing proteins (31), EAL-containing proteins (12), 2',3'-cyclic nucleotide phosphodiesterase (1)
Nitrogen regulation	*rpoN*, *ntrB*, *ntrC*, *ntrX*, *ntrY*, *glnB*, *glnD*, *glnK*, *fixK*, *nifR*, *ptsN*
Oxygen regulation	*fnrN*, *nolR*
Phosphate regulation	*phoB*
pH regulation	*chvG*, *chvI*, *nolR*
Energy	PPi-F6P-phosphotransferase (*pfp*), exopoly-phosphatase (*ppx*), polyphosphate kinase (*ppk*), membrane-bound H^+-translocating pyrophosphatase (*rrpP*)

Several exopolysaccharides have been shown to be regulated by bis-(3',5')-cyclic-dimeric-guanosine monophosphate (c-di-GMP) [18-20]. This includes cellulose synthesis in *Gluconacetobacter xylinus* (formerly *Acetobacter xylinum*) and *A. tumefaciens*, alginate synthesis in *Pseudomonas aeruginosa*, and extracellular PEL polysaccharide synthesis also in *P. aeruginosa*. C-di-GMP is synthesized by diguanylate cyclases which contain a conserved GGDEF motif as the catalytic active site. On the other hand, c-di-GMP is degraded by phosphodiesterases containing EAL domains. The relative activities of GGDEF and EAL domain proteins regulate the level of the c-di-GMP signal. GGDEF and EAL domain proteins often contain N-terminal domains that respond to sensory input, allowing for activation or inactivation of the enzyme in response to environmental signals such as oxygen, light, starvation, redox conditions, and other signaling molecules [21]. This may explain how curdlan synthesis levels change in response to different environmental signals such as nitrogen or phosphate limitation,

oxygen, and pH. To regulate cellulose production, the c-di-GMP binds to the cellulose synthase to enhance the enzyme's activity. The c-di-GMP binding site on the cellulose synthase was identified as a PilZ domain. PilZ domains are also found on other enzymes regulated by c-di-GMP, including those necessary for alginate biosynthesis [22]. The amino acid sequence of the curdlan synthase in ATCC 31749 was analyzed using Pfam, yet the PilZ domain was not found. Despite the absence of the PilZ domain, curdlan synthesis may still be regulated by c-di-GMP as other proteins regulated by c-di-GMP also lack the PilZ domain. In addition to the direct binding of c-di-GMP to enhance enzyme activity, c-di-GMP has also been shown to act as a riboswitch, controlling the expression of genes involved in many cellular processes [23]. The genome sequence of ATCC 31749 was analyzed for GGDEF and EAL domain proteins via gene annotation and domain searches of all hypothetical proteins using Pfam. This search revealed 31 GGDEF domain proteins, 12 of which also contain EAL domains, and one 2',3'-cyclic nucleotide phosphodiesterase. Experimental evidence is required to determine what, if any, role c-di-GMP plays in regulating curdlan synthesis in ATCC 31749.

Previous investigations of ATCC 31749 and other closely-related strains identified several environmental factors that regulate curdlan synthesis. Nitrogen limitation is the main factor for curdlan synthesis; ATCC 31749 only produces curdlan after the nitrogen source has been exhausted [24]. Interestingly, nitrogen limitation is also an important factor for other Rhizobia. Under nitrogen-limited conditions, the *Rhizobium* species develop a symbiotic relationship with legumous plants by fixing nitrogen [25]. While ATCC 31749 cannot fix nitrogen due to the absence of a nitrogenase (*nifH*), it does possess many genes related to nitrogen metabolism and regulation (Table 5.5). In

addition to the sigma factor associated with nitrogen metabolism (*rpoN*), ATCC 31749 also contains genes involved in the nitrogen signal cascade (*ntrB*, *ntrC*, *ntrX*, *ntrY*, *glnB*, *glnD*, *glnK*) and in nitrogen fixation (*fixK*, *nifR*, *ptsN*). Oxygen is also an important environmental factor for curdlan synthesis, with higher oxygen levels corresponding to increased curdlan synthesis [26]. Thus, oxygen-responsive global regulators such as *fnrN* and *nolR* may play a role in curdlan synthesis regulation. Furthermore, phosphate concentration and pH also influence curdlan synthesis [27, 28], indicating that genes involved in phosphate (*phoB*) and pH (*chvG*, *chvI*) regulation may be relevant in regulating curdlan production. With the numerous environmental conditions effecting curdlan synthesis, there are many possible genes participating in the regulation of curdlan production in ATCC 31749.

Lastly, exopolysaccharide production is an energy-intensive process, and as such, energy-conserving and energy-generating enzymes may be required for curdlan production. Surprisingly, ATCC 31749 uses the lower energy producing ED pathway for glycolysis as opposed to the EMP pathway. However, ATCC 31749 does have a pyrophosphate-dependent F6P 1-phosphotransferase (*pfp*) which may allow for an active EMP pathway. Although the energy-conserving *pfp* may help to reserve energy for curdlan synthesis, there is likely another mechanism for generating the energy required for curdlan production. A close relative of ATCC 31749, *A. tumefaciens*, has been reported to generate acidocalcisomes containing polyphosphate (polyP) [29]. Genome sequencing shows that ATCC 31749 also has genes associated with acidocalcisomes and polyP (Table 5.5). PolyP can be broken down into pyrophosphate (PPi), a high energy compound, and thus, polyP may serve as an intracellular energy storage compound that is

utilized to provide energy for curdlan production. As energy is required for UDP-glucose synthesis, these energy-related genes are not only important for curdlan production but also for oligosaccharide production in the engineered ATCC 31749 strains developed in Chapters 2 and 3.

5.4 Discussion and Conclusions

Genome sequencing alone does not provide conclusive evidence of the mechanisms for curdlan synthesis and regulation, but the information garnered from this endeavor is essential in deciphering the metabolic pathways and genetic interactions contributing to curdlan production. The genome sequence of ATCC 31749 also contributes to the understanding of the *Agrobacterium* species and the Rhizobia family of microorganisms. Through comparative genomics, a core genome of the *Agrobacterium* species was identified and found to have significant homology with other sequenced *Rhizobium* and *Sinorhizobium* species. Despite the seemingly different phenotypes of symbiotic, pathogenic, and free-living species, microorganisms of the Rhizobia family share many common metabolic pathways. The genome sequence of ATCC 31749 contains many genes which may play a role in curdlan synthesis and regulation, including genes involved in exopolysaccharide, nitrogen, oxygen, phosphate, and pH regulation as well as genes for energy-conserving and energy-generating enzymes. Experimental investigation of these genes will reveal which are important for curdlan production.

5.5 Methods

5.5.1 16S rRNA sequencing and phylogenetic tree construction

The following universal primers were used for cloning the 16S rRNA gene from ATCC 31749: 5'- AGAGTTTGATCCTGGCTCAG – 3' (forward primer) and 5' –

AAGGAGGTGATCCAACCGCA – 3' (reverse primer). The amplified 16S rRNA gene was ligated into the pGEM T-easy vector (Promega) for sequencing. The 16S rRNA sequence of ATCC 31749 was aligned with 16S rRNA gene sequences of other Rhizobia using ClustalX. The Rhizobia gene sequences were obtained from the NCBI database. A phylogenetic tree of the aligned 16S rRNA gene sequences was constructed using TreeView.

5.5.2 Genomic DNA isolation, sequencing, and annotation

ATCC 31749 was cultivated overnight in a test tube containing 4 mL of LB media at 30°C with agitation at 250 rpm. The overnight culture was used for genomic DNA (gDNA) isolation with the GenElute Bacterial Genomic DNA Kit from Sigma-Aldrich. The genome of ATCC 31749 was sequenced at the BioMedical Genomics Center at the University of Minnesota using the Genome Sequencer FLX System from 454 Life Sciences. The reads obtained from gDNA sequencing were assembled into contigs using the GS De Novo Assembler from 454 Life Sciences. The contigs were arranged in random order with stop codons inserted between each contig to form a pseudochromosome. This pseudochromosome was used for gene prediction with two gene prediction programs: GeneMarkS and Glimmer v3.02. The predicted genes from GeneMarkS and Glimmer were combined, and any repeated open reading frame (orf) (i.e. predicted by both programs) was removed along with orfs with unusual start codons to obtain a list of unique predicted genes. The predicted genes were annotated by comparing the nucleotide sequences to those of the non-redundant (nr) database of NCBI using Netblast (blastcl3), an E-value cutoff of 10^{-4}, and XML format for the output. A perl script (blast_XML_extraction.pl) was written to extract the gene query, query length, hit

number, hit ID, hit length, hit definition, score, E-value, identities, and positives (see Appendix B for Perl script). The hit definition of the top hit for each gene was used to annotate the predicted gene list. All conserved hypothetical and hypothetical proteins were entered into Pfam to identify any conserved domains and assign a proposed function. Transfer RNA genes were identified using tRNAscan-SE 1.21, and ribosomal RNA genes were identified using blastn with the 5S, 6S, 16S, and 23S rRNA sequences from *A. tumefaciens* C58, *A. radiobacter* K84, and *A. vitis* S4 as query and the pseudochromosome of ATCC 31749 as database.

5.5.3 Operon and metabolic pathway prediction

The genome sequence and annotation was formatted using a Perl script (Pathologic_format.pl) (see Appendix B for Perl script) as specified by the PathoLogic component of the Pathway Tools software. With the formatted sequence and annotation file, PathoLogic was used to create a Pathway/Genome Database (PGDB) for ATCC 31749. Enzymes not automatically assigned by PathoLogic were manually annotated. PathoLogic functions were employed to predict transcription units (i.e. operons) and identify transporters. Lastly, the Pathway Hole Filler identified any genes missing from the pathways predicted for the genome.

5.4.4 *Crd* and *cel* operon analysis

The *crd* and *cel* operons of *A. tumefaciens* C58 were obtained from the genome sequence in the NCBI database. The *A. tumefaciens* C58 and ATCC 31749 operons were aligned using ClustalW2 to identify differences in the nucleotide sequences. The amino acid sequences of *crdS* from ATCC 31749 and *A. tumefaciens* C58 were aligned with

ClustalW2 to identify differences. The amino acid changes in the CrdS of *A. tumefaciens* C58 were compared to the catalytic active sites identified previously [10].

5.5.5 Comparative genomics

The genome sequences of the *Agrobacterium*, *Rhizobium*, and *Sinorhizobium* species listed in Tables 5.3 and 5.4 were obtained from the NCBI database. The predicted gene sequences of ATCC 31749 were compared to the protein-coding sequences of *A. tumefaciens* C58, *A. radiobacter* K84, and *A. vitis* S4 using blastx with an E-value cutoff of 10^{-3} and XML output format. The hits from each genome comparison were sorted using a Perl script (matchandsort.pl, Appendix B) to identify the genes of the *Agrobacterium* core genome. The nucleotide sequences associated with the genes of the core genome were compared to the protein-coding sequences of other Rhizobia using blastx with an E-value cutoff of 10^{-3} and XML output format.

5.6 References

1. Harada, T. and T. Yoshimura, *Production of a new acidic polysaccharide containing succinic acid by a soil bacterium.* Biochimica et biophysica acta, 1964. **83**: p. 374-376.

2. Phillips, K.R., et al., *Production of curdlan-type polysaccharide by Alcaligenes faecalis in batch and continuous culture.* Canadian Journal of Microbiology, 1983. **29**(10): p. 1331-1338.

3. Lee, I.-Y., *Curdlan*, in *Biotechnology of Biopolymers: From Synthesis to Patents*, A. Steinbuchel and Y. Doi, Editors. 2005, Wiley-VCH: Weinheim, Germany. p. 457-480.

4. Demleitner, S., J. Kraus, and G. Franz, *Synthesis and antitumour activity of sulfoalkyl derivatives of curdlan and lichenan.* Carbohydrate Research, 1992. **226**(2): p. 247-252.

5. Gordon, M., et al., *A phase I study of curdlan sulfate--an HIV inhibitor. Tolerance, pharmacokinetics and effects on coagulation and on CD4 lymphocytes.* Journal of Medicine, 1994. **25**(3-4): p. 163-180.

6. Mao, Z. and R.R. Chen, *Recombinant synthesis of hyaluronan by Agrobacterium sp.* Biotechnology Progress, 2007. **23**: p. 1038-1042.

7. Ruffing, A., Z. Mao, and R.R. Chen, *Metabolic engineering of Agrobacterium sp. for UDP-galactose regeneration and oligosaccharide synthesis.* Metabolic Engineering, 2006. **8**: p. 465-473.

8. Stasinopoulos, S.J., et al., *Detection of two loci involved in (1à3)-β-glucan (curdlan) biosynthesis by Agrobacterium sp. ATCC31749, and comparative sequence analysis of the putative curdlan synthase gene.* Glycobiology, 1999. **9**(1): p. 31-41.

9. Karnezis, T., et al., *Cloning and characterization of the phosphatidylserine synthase gene of Agrobacterium sp. strain ATCC 31749 and effect of its inactivation on production of high-molecular-mass (1à3)-β-D-glucan (curdlan).* Journal of Bacteriology, 2002. **184**(15): p. 4114 - 4123.

10. Karnezis, T., et al., *Topological characterization of an inner membrane (1à3)-β-D-glucan (curdlan) synthase from Agrobacterium sp. strain ATCC 31749.* Glycobiology, 2003. **13**(10): p. 693-706.

11. McIntosh, M., B.A. Stone, and V.A. Stanisich, *Curdlan and other bacterial (1à3)-β-D-glucans.* Applied Microbiology and Biotechnology, 2005. **68**: p. 163-173.

12. Kai, A., et al., *Analysis of the biosynthetic process of cellulose and curdlan using 13C-labeled glucoses.* Carbohydrate Polymers, 1994. **23**: p. 235-239.

13. Kai, A., et al., *Biosynthesis of curdlan from culture media containing 13C-labeled glucose as the carbon source.* Carbohydrate Research, 1993. **240**: p. 153-159.

14. Lee, I.-Y., et al., *Production of curdlan using sucrose or sugar cane molasses by two-step fed-batch cultivation of Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1997. **18**: p. 255-259.

15. Matthysse, A.G., S. White, and R. Lightfoot, *Genes required for cellulose synthesis in Agrobacterium tumefaciens.* Journal of Bacteriology, 1995. **177**(4): p. 1069-1075.

16. van Rhijn, P. and J. Vanderleyden, *The Rhizobium-plant symbiosis.* Microbiological Reviews, 1995. **59**(1): p. 124-142.

17. Zambryski, P., J. Tempe, and J. Schell, *Transfer and function of T-DNA genes from Agrobacterium Ti and Ri plasmids in plants.* Cell, 1989. **56**(2): p. 193-201.

18. Lee, V.T., et al., *A cyclic-di-GMP receptor required for bacterial exopolysaccharide production.* Molecular Microbiology, 2007. **65**(6): p. 1474-1484.

19. Romling, U., M. Gomelsky, and M.Y. Galperin, *C-di-GMP: the dawning of a novel bacterial signalling system.* Molecular Microbiology, 2005. **57**(3): p. 629-639.

20. Ross, P., R. Mayer, and M. Benziman, *Cellulose biosynthesis and function in bacteria.* Microbiological Reviews, 1991. **55**(1): p. 35-58.

21. Hengge, R., *Principles of c-di-GMP signalling in bacteria.* Nature Reviews Microbiology, 2009. **7**: p. 263-273.

22. Akmikam, D. and M.Y. Galperin, *PilZ domain is part of the bacterial c-di-GMP binding protein.* Bioinformatics, 2006. **22**(1): p. 3-6.

23. Sudarsan, N., et al., *Riboswitches in eubacteria sense the secondary messenger cyclic di-GMP.* Science, 2008. **321**: p. 411-413.

24. Kim, M.K., et al., *Higher intracellular levels of uridinemonophosphate under nitrogen-limited conditions enhance metabolic flux of curdlan synthesis in Agrobacterium species.* Biotechnology and Bioengineering, 1999. **62**(3): p. 317-323.

25. Fischer, H.-M., *Genetic regulation of nitrogen fixation in Rhizobia.* Microbiological Reviews, 1994. **58**(3): p. 352-386.

26. Lee, I.Y., et al., *Influence of agitation speed on production of curdlan by Agrobacterium species.* Bioprocess Engineering, 1999. **20**: p. 283-287.

27. Kim, M.K., et al., *Residual phosphate concentration under nitrogen-limiting conditions regulates curdlan production in Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 2000. **25**: p. 180-183.

28. Lee, J.-H., et al., *Optimal pH control of batch processes for production of curdlan by Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1999. **23**: p. 143-148.

29. Docampo, R., et al., *Acidocalcisomes - conserved from bacteria to man.* Nature Reviews Microbiology, 2005. **3**(251-261).

CHAPTER 6

TRANSCRIPTOME ANALYSIS OF *AGROBACTERIUM* SP. ATCC 31749 REVEALS GENES IMPORTANT FOR CURDLAN SYNTHESIS AND REGULATION

6.1 Abstract

To determine which genes are important for curdlan synthesis and regulation, the transcriptome of ATCC 31749 was analyzed under different environmental conditions. Two conditions were investigated using DNA microarrays: pH and nitrogen. The pH-dependent study compares gene expression levels during curdlan synthesis at pH 7 and pH 5.5, where increased curdlan production is observed at pH 5.5. The nitrogen-dependent study compares gene expression levels before curdlan synthesis (nitrogen-rich) and during curdlan synthesis (nitrogen-limited). Both transcriptome comparisons revealed genes significantly up- and down-regulated that may influence curdlan production. Genes associated with three functional categories were selected for gene knockout: 1) nitrogen-limited regulation, 2) energy-related, and 3) GTP-derived second messengers. Genes from all three categories were found to affect curdlan production. The *nifR/ntrB/ntrC* operon was shown to regulate curdlan production in an RpoN-independent manner. Polyphosphate was also determined to be an important factor for curdlan synthesis, but surprisingly, gene knockout of the two exopolyphosphatases in ATCC 31749 had opposing effects on curdlan production. Lastly, the GTP-derived second messengers, (p)ppGpp and c-di-GMP, were also shown to be important for curdlan

synthesis. Several mechanisms are discussed based on the curdlan produced by the gene knockout mutants and supporting information in the literature.

6.2 Introduction

Genome sequencing of ATCC 31749 identified many genes with possible influence over curdlan synthesis. In particular, regulatory genes associated with nitrogen, oxygen, and phosphate limitation were found, three environmental factors shown to influence curdlan production [1-3]. The genome also includes numerous genes participating in the synthesis and degradation of c-di-GMP, a second messenger shown to regulate production of polysaccharides such as cellulose, alginate, and PEL exopolysaccharide [4-6]. As curdlan synthesis requires high energy compounds (ATP and UTP), genes related to energy production and storage are also likely to be important for curdlan production. Lastly, conserved hypothetical proteins and hypothetical proteins may play a significant role in curdlan synthesis and regulation. With 5,585 predicted genes in the genome of ATCC 31749, how can those related to curdlan production be identified?

DNA microarrays are powerful tools to analyze global gene expression (i.e. the transcriptome) of a microorganism under various conditions. Transcriptome analysis via DNA microarray identifies genes with changes in transcript level, and as the level of gene expression generally correlates to the amount of functional protein present, this analysis can indicate enzymes, transporters, or transcriptional regulators that are active under specific environmental conditions. In this study, custom DNA microarrays were designed from the genome sequence of ATCC 31749. Two factors important for curdlan

production were investigated using transcriptome analysis: pH and nitrogen. The pH-dependent analysis evaluates changes in transcript level during curdlan synthesis at pH 7 and pH 5.5 while the nitrogen-dependent analysis compares gene expression levels before (nitrogen-rich) and after (nitrogen-limited) the start of curdlan synthesis. The pH and nitrogen-dependent transcriptome comparisons identified significantly up- and down-regulated genes associated with curdlan and enhanced curdlan production. From the list of up- and down-regulated genes, several genes were selected for gene knockout to investigate their potential influence on curdlan synthesis. The experimental results of these gene knockout studies are presented along with an analysis of their significance in context with other experimental results reported in the literature. Based on the data and subsequent analysis, several possible mechanisms are discussed for the regulation of curdlan synthesis in ATCC 31749.

6.3 Results

6.3.1 Microarray analysis of ATCC 31749 at pH 7 vs. pH 5.5

Many environmental conditions have been optimized to maximize curdlan production by *Agrobacterium* sp. ATCC 31749. One condition found to have a significant impact on curdlan synthesis is the pH of the cell culture. A two-phase reaction was found to be optimal for curdlan synthesis, where the growth phase is controlled at pH 7 and the nitrogen-limited phase is controlled at pH 5.5. By shifting the pH from 7 to 5.5 at the time of nitrogen depletion, there is a 50% increase in curdlan production compared to curdlan production at a constant pH of 7 [2]. In an attempt to determine which genes are responsible for the increased curdlan production under acidic conditions, gene

transcript levels were compared under the two different pH conditions (pH 7 and pH 5.5) using a DNA microarray. For this pH-dependent microarray analysis, samples were taken 70 hours after the point of nitrogen depletion, which corresponds to approximately 100 hours after the start of cultivation in the bioreactor. The concentrations of curdlan and dry cell weight under the two pH conditions are shown in Figure 6.1. Surprisingly, the dry cell weight measurement for the pH 7 reaction shows an increase after nitrogen depletion, which is not observed for the pH 5.5 culture. To determine if this increase is due to cell growth or a result of the dry cell weight measurement, the total protein content was measured for each sample using the Bradford assay. If the increase in the dry cell weight measurement is due to cell growth, there should also be a corresponding increase in the amount of total protein. For both the pH 7 and pH 5.5 reactions, the protein concentration increased during growth, reached a peak at the time of nitrogen depletion, and decreased slightly during nitrogen-limitation. This suggests that the measured increase in dry cell weight is artifact of the measurement technique. Based on the method of dry cell weight measurement, the production of an insoluble product or the accumulation of an intracellular product may account for the observed increase in dry cell weight for the pH 7 reaction. While this putative product is a notable difference between the pH 7 and pH 5.5 cultivations, it does not appear to have a major impact on curdlan production during this time.

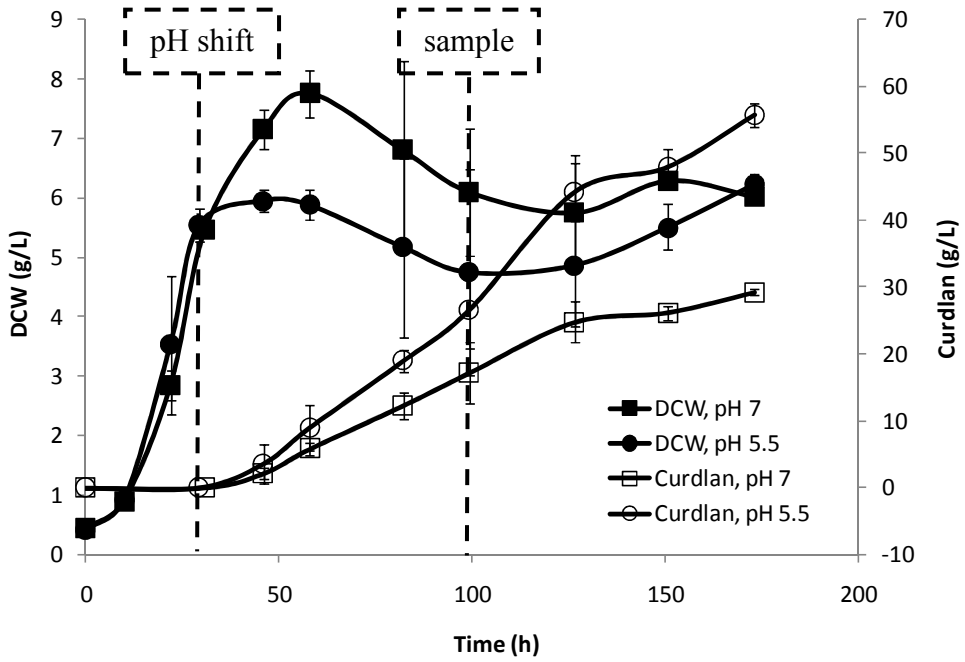


Figure 6.1 Dry cell weight (DCW) and curdlan concentrations for the pH-dependent cultivations. For the pH 7 reactor, the pH was held constant throughout the reaction. For the pH 5.5 reactor, the pH was set to 7 for the growth phase (0-30h) and shifted to pH 5.5 at the time of nitrogen depletion (~30h).

The microarray data for the pH comparison identified a total of 338 genes with more than 2-fold changes in gene expression level. The split between up and down-regulated genes was nearly equal, with 170 up-regulated genes and 168 down-regulated genes. A list of these genes and their corresponding fold-changes in expression will be made publicly available online after publication of this work. For the pH comparison, both samples are in the stationary phase, are subject to nitrogen-limitation, and are producing curdlan. The only observable differences between the cultures are the pH value, the amount of curdlan production, and the production of an insoluble or intracellular product as indicated by the dry cell weight measurement. The significantly up- and down-regulated genes were grouped according to function, and the distribution of differentially expressed gene functions is listed in Table 6.1. The majority of up-regulated

genes (61%) are hypothetical, suggesting that an unknown mechanism may be responsible for the increased curdlan production at pH 5.5. The substantial number of up-regulated regulatory genes indicates that transcriptional regulation may also be important. As expected, many of the down-regulated genes are transporters or metabolic genes, two processes known to have decreased activity under stress conditions like low pH.

Table 6.1 Distribution of up- and down-regulated genes for the pH-dependent microarray comparison

Category	Number of genes	Percentage (%)
Up-regulated		
cell replication and division	8	4.7
cell membrane and lipid synthesis and degradation	6	3.5
amino acid synthesis and metabolism	2	1.2
modification proteins	1	0.6
transporters	3	1.8
metabolism and cofactor synthesis	8	4.7
regulation	33	19.4
stress response and cell protection	4	2.4
symbiosis-related proteins	1	0.6
conserved hypothetical proteins	87	51.2
hypothetical proteins	17	10.0
Total	*170*	*100.0*
Down-regulated		
cell membrane and lipid synthesis and degradation	5	3.0
cell movement and extracellular appendages	6	3.6
nucleotide synthesis and degradation	3	1.8

Table 6.1 Continued

Category	Number of genes	Percentage (%)
Down-regulated (continued)		
amino acid synthesis and metabolism	5	3.0
protein folding, modification, secretion, and degradation	3	1.8
modification proteins	6	3.6
transporters	30	17.9
metabolism and cofactor synthesis	31	18.5
regulation	12	7.1
stress response and cell protection	6	3.6
polysaccharide synthesis and degradation	4	2.4
phage proteins	1	0.6
conserved hypothetical proteins	49	29.2
hypothetical proteins	7	4.2
Total	*168*	*100.0*

Genes with potential involvement or influence on curdlan production were identified from the list of differentially expressed genes for the pH comparison (Table 6.2). A decrease in production of other glucan-containing products would increase the availability of UDP-glucose for curdlan production and may explain the observed increase in curdlan synthesis at pH 5.5. Genes involved in the production of cyclic β1,2-glucan and an exopolysaccharide were found to be down-regulated at pH 5.5. In fact, a decrease in cyclic β1,2-glucan may explain the discrepancy in dry cell weight measurement after nitrogen exhaustion. Cyclic β1,2-glucan accumulates in the periplasm of the cell and therefore contributes to the dry cell weight measurement. So if cyclic β1,2-glucan is produced at pH 7 after nitrogen depletion, there would be an increase in the dry

cell weight measurement. On the other hand, there would be no increase in dry cell weight measurement at pH 5.5 if little cyclic β1,2-glucan is produced, as is suggested by the decreased expression level of *chvA* and *cgmA*. Since cyclic β1,2-glucan also utilizes UDP-glucose, a decrease in cyclic β1,2-glucan may contribute to the increase in curdlan production. However, this is not likely the sole contributing factor as the dry cell weight measurement only increases during the early stages of curdlan production (Figure 6.1). An increase in carbon metabolism may also enhance curdlan production by supplying additional carbon and energy. Enhanced expression of *aglE*, an α-glucoside transporter, may improve uptake of the carbon source, sucrose. In addition, reduced expression of *pfp* may help to direct the carbon flux towards curdlan synthesis rather than through glycolysis. Lastly, a regulatory mechanism may be responsible for the improved curdlan production at pH 5.5. The conserved hypothetical protein Atu1112 contains a GGDEF domain with potential for c-di-GMP synthesis. Increased levels of c-di-GMP have been shown to improve the production of exopolysaccharides in other microorganisms [4-6]; thus, increased levels of Atu1112 may regulate curdlan production. Similarly, a decrease in c-di-GMP degrading enzymes such as GTP cyclohydrolase may stimulate curdlan synthesis.

Table 6.2 Genes identified by the pH-dependent microarray analysis that may influence curdlan synthesis in ATCC 31749

Gene	Fold-change	Symbol	Definition
Glucan product synthesis			
r0991	-2.9	*chvA*	cyclic β1,2-glucan ABC transporter
r5576	-3.3	*cgmA*	cyclic glucan phosphoglycerol modification protein
r0171	-4.1		exopolysaccharide production protein

Table 6.2 Continued

Gene	Fold-change	Symbol	Definition
Metabolism			
r2734	2.2	*aglE*	ABC transporter, substrate binding protein (alpha-glucoside)
r0653	-2.0	*pfp*	pyrophosphate – fructose-6-phosphate 1-phosphotransferase
c-di-GMP synthesis and degradation			
r3955	2.4		hypothetical protein Atu1112
r3770	-2.3	*folE*	GTP cyclohydrolase I

6.3.2 Nitrogen-dependent microarray analysis of ATCC 31749

Curdlan synthesis in ATCC 31749 involves two phases. The first phase is a growth phase where a nitrogen source is present and no curdlan is produced. The second phase occurs when the nitrogen source is exhausted, causing the cell culture to enter the stationary phase and begin to produce curdlan. A nitrogen-dependent microarray analysis of gene expression, comparing the growth phase (nitrogen-rich) and curdlan-producing phase (nitrogen-limited), should reveal which genes are important for curdlan production and regulation. Since this comparison also includes a shift from growth to stationary phase, the microarray is expected to yield a large number of genes with significant changes in gene expression level. For the nitrogen-dependent microarray comparison, samples were taken during the exponential growth phase at 22 hours and when the rate of curdlan synthesis is high, approximately 70 hours after depletion of the nitrogen source (~100 hours after inoculation of the culture). Figure 6.2 shows the cell and curdlan concentrations at each sample time. The pH for the nitrogen-dependent cultivation was held constant at pH 7 throughout the reaction.

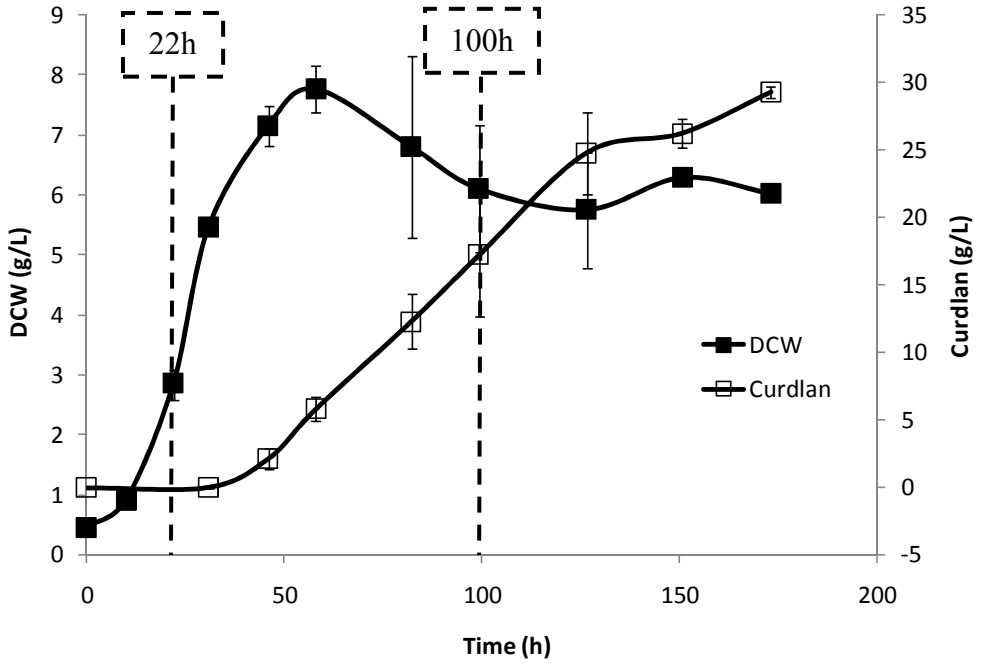


Figure 6.2 Dry cell weight (DCW) and curdlan concentrations for the nitrogen-dependent cultivation.

Analysis of the microarray data for the nitrogen-dependent comparison revealed a total of 2,456 genes with more than 2-fold changes in gene expression level. Of the 2,456 genes, 985 genes were up-regulated and 1,471 genes were down-regulated. A list of these genes and their corresponding fold-changes in expression will be made publicly available online after publication of this work. As stated previously, the large number of genes with changes in gene expression level is due to the fact that the initiation of curdlan synthesis also accompanies a shift from growth to stationary phase. The up- and down-regulated genes were grouped according to function, and Table 6.3 lists the distribution of genes across the functional categories. A majority of the up-regulated genes code for hypothetical proteins of unknown function. This indicates that the genes active during stationary phase for ATCC 31749 differ from that of more well-studied organisms such as *E. coli*, suggesting unique pathways to be investigated in the future. These hypothetical

proteins may also be important for curdlan synthesis and regulation. Many regulatory genes are up-regulated in the 100 hour sample and may be important for stationary phase survival or curdlan production. A significant number of stress response genes are also up-regulated during the stationary phase. The down-regulated genes include a large number of transporters and metabolic genes. This is expected since the cell no longer requires the high level of nutrient uptake and metabolism that is required during the growth phase.

Table 6.3 Distribution of up- and down-regulated genes for the nitrogen-dependent microarray comparison

Category	Number of genes	Percentage (%)
Up-regulated		
cell replication and division	43	4.4
cell membrane and lipid synthesis and degradation	18	1.8
cell movement and extracellular appendages	9	0.9
amino acid synthesis and metabolism	9	0.9
protein folding, modification, secretion, and degradation	22	2.2
RNA-associated proteins	4	0.4
modification proteins	37	3.8
transporters	86	8.7
metabolism and cofactor synthesis	65	6.6
regulation	105	10.7
stress response and cell protection	27	2.7
polysaccharide synthesis and degradation	8	0.8
phage proteins	6	0.6
conserved hypothetical proteins	424	43.0
hypothetical proteins	122	12.4
Total	*985*	*100.0*

Table 6.3 Continued

Category	Number of genes	Percentage (%)
Down-regulated		
cell replication and division	82	5.6
cell membrane and lipid synthesis and degradation	69	4.7
cell movement and extracellular appendages	25	1.7
amino acid synthesis and metabolism	62	4.2
protein folding, modification, secretion, and degradation	42	2.9
RNA-associated proteins	74	4.9
modification proteins	66	4.5
transporters	210	14.3
metabolism and cofactor synthesis	290	19.9
regulation	101	6.9
stress response and cell protection	26	1.8
polysaccharide synthesis and degradation	24	1.6
phage proteins	2	0.1
invasion/virulence-associated proteins	3	0.2
conserved hypothetical proteins	368	25.0
hypothetical proteins	27	1.8
Total	*1471*	*100.0*

The list of up- and down-regulated genes identified by the nitrogen-dependent microarray comparison was analyzed for genes with functions potentially involved in curdlan synthesis and regulation. The few genes known to be involved in curdlan synthesis were significantly up-regulated. The gene for curdlan synthase (c*rdS*) was up-regulated more than 95-fold while the other genes in the curdlan synthesis operon, *crdA* and *crdC*, showed 99- and 66-fold increases respectively. Unfortunately, the nucleotide

sequence for the putative regulator *crdR* has not been reported so the change in expression level for *crdR* cannot be identified. Regardless, this gene expression data is the first reported evidence that the curdlan synthesis operon (c*rdASC*) is transcriptionally regulated with respect to nitrogen-limitation. As such, there must be some factor which controls transcription of the operon. Based on information available in the literature regarding curdlan synthesis and exopolysaccharide regulation, the genes involved in three processes were identified for further investigation: nitrogen-limited regulation, energy production, and GTP-derived second messengers (Table 6.4).

Table 6.4 Genes identified by the nitrogen-dependent microarray analysis that may influence curdlan synthesis in ATCC 31749

Gene	Fold-change	Symbol	Definition
Nitrogen-limited regulation			
r1024	5.2	*glnK*	nitrogen regulatory protein PII
r0418	-5.8	*glnD*	PII uridylyl-transferase
Energy-associated			
r2510	-6.9	*rrpP*	membrane-bound proton-translocating pyrophosphatase
r2544	-3.6	*ppx2*	exopolyphosphatase
r2335	-3.0		inorganic polyphosphate/ATP-NAD kinase
r3046	-2.4	*galU*	UTP:glucose-1-phosphate uridylyltransferase
GTP-derived second messengers			
r1444	6.7		hypothetical protein Atu0989 (GGDEF)
r1489	4.0	*spoT*	GTP pyrophosphohydrolase/synthetase
r3957	3.9		GGDEF/EAL family protein
r3770	3.8	*folE*	GTP cyclohydrolase I
r4126	-2.1		GGDEF/EAL family protein
r3899	-2.3		two component response regulator (GGDEF)
r5342	-2.5		hypothetical protein Atu0826 (GGDEF/EAL)

Table 6.4 Continued

Gene	Fold-change	Symbol	Definition
GTP-derived second messengers (continued)			
r5566	-2.8		GGDEF/EAL family protein
r5364	-2.9		hypothetical protein Atu4353 (GGDEF/EAL)
r1177	-3.3		GGDEF/EAL family protein
r2487	-3.4		GGDEF family protein
r2433	-3.6		GGDEF family protein
r2855	-3.7		GGDEF/EAL family protein
r0742	-3.8		GGDEF family protein
r4090	-7.0		GGDEF/EALfamily protein
r0078	-8.5		hypothetical protein Atu0784 (GGDEF/EAL)

6.3.3 Gene knockouts identify genes influencing curdlan production

6.3.3.1 Nitrogen-limited regulatory genes

The only factor that has been experimentally determined as necessary for curdlan production in ATCC 31749 is nitrogen limitation [7]. In other microorganisms, genes are transcriptionally regulated by nitrogen limitation via the nitrogen signaling cascade (Figure 6.3). Hence, the genes of the nitrogen signaling cascade may also regulate transcription of the curdlan synthesis operon. Briefly, the nitrogen signaling cascade includes a nitrogen sensor protein (GlnD), which under nitrogen-limited conditions, transfers UMP to a PII protein (either GlnB or GlnK). In its non-uridylylated form, the PII protein binds to NtrB, promoting dephosphorylating activity. The uridylylated form of the PII protein cannot bind NtrB, and subsequently, the free NtrB phosphorylates NtrC. The phosphorylated NtrC binds to the transcription factor RpoN. With NtrC•P, RpoN is active and initiates transcription of genes that contain an RpoN binding site in

their promoter region. The curdlan synthesis genes may be directly regulated by the nitrogen signaling cascade via activated RpoN, or the active RpoN may initiate transcription of another sigma factor or protein which regulates transcription of the curdlan synthesis operon (i.e. indirect regulation). From the nitrogen-dependent microarray analysis, two genes involved in the nitrogen signaling cascade were significantly up-regulated (*glnK* and *glnD*) (see Table 6.4), two proteins directly involved in sensing the nitrogen status of the cell. Other components of the nitrogen signaling cascade (*ntrB, ntrC*, and *rpoN*) are not transcriptionally regulated by nitrogen limitation [8], and therefore, these genes were not identified by the microarray analysis. In a review paper on curdlan, NtrB/NtrC were said to be required for curdlan synthesis in ATCC 31749, but no experimental evidence was presented [9]. The genome sequence of ATCC 31749 shows *ntrBC* to be in an operon with *nifR* (*nifR>ntrB>ntrC*). To determine the involvement of the nitrogen signaling cascade with respect to curdlan production, three genes (*glnK, nifR*, and *rpoN*) were targeted for gene knockout.

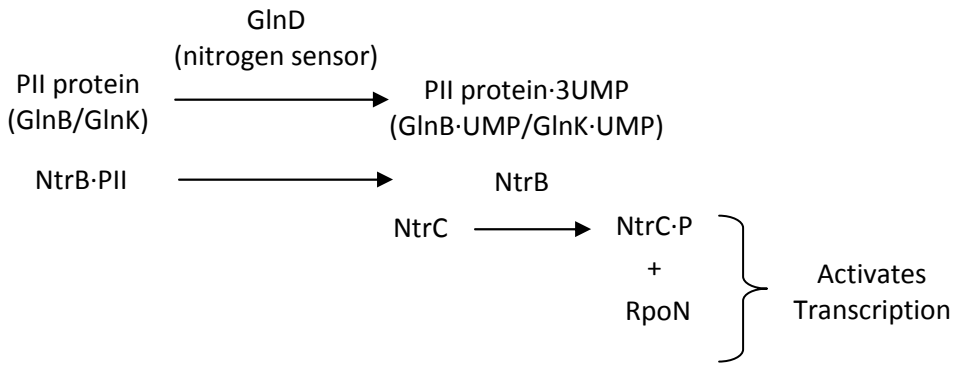


Figure 6.3 Nitrogen signaling cascade.

Successful gene knockout mutants (ATCC 31749Δ*rpoN* and ATCC 31749Δ*nifR*) were obtained using the method described in section 6.4. Despite numerous gene KO attempts targeting *glnK*, a successful KO mutant was not obtained for this target gene. This may indicate that *glnK* is essential for growth in ATCC 31749. The gene knockout technique used in this study may also affect transcription of any downstream genes contained in the same operon as the gene knockout target. Consequently, the results presented in this study may be attributed to other genes in the operon of the target gene, and any genes in the predicted operon which may be affected by the knockout will be included in the analysis.

RpoN is required for transcription of genes associated with nitrogen-limitation, and thus, it is expected that the *rpoN* knockout will eliminate transcription of the curdlan synthesis operon, thereby eliminating curdlan production. Surprisingly, the *rpoN* mutant not only synthesized curdlan but it produced more curdlan than the wild-type (Figure 6.4). This indicates that the nitrogen-limited regulation of curdlan production does not follow the traditional nitrogen signaling cascade. Furthermore, the results show that the nitrogen-limited regulation of curdlan production is RpoN-independent, and in fact, RpoN may inhibit curdlan synthesis. Combining this information with the literature claim that NtrB/C is required for curdlan synthesis suggests a novel mechanism for the regulation of curdlan production via nitrogen limitation.

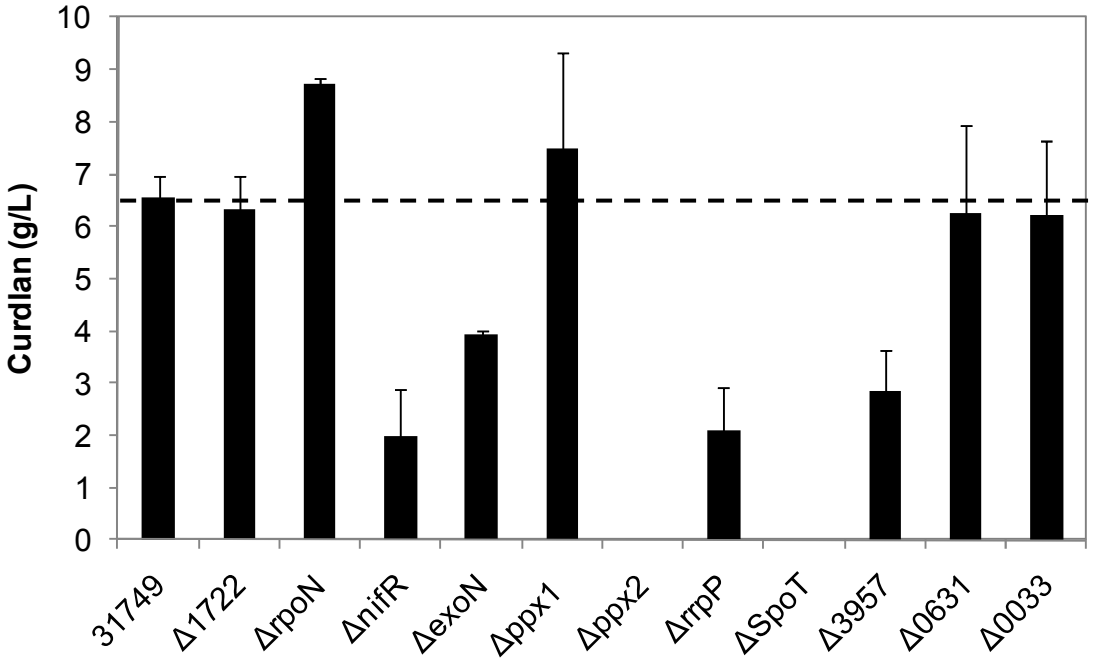


Figure 6.4 Curdlan synthesis in ATCC 31749 and gene KO mutants after 24 hours of cultivation in nitrogen-free media.

The role of *nifR* in nitrogen-limited regulation has not yet been determined in any microorganism; however, the *nifR/ntrB/ntrC* operon is conserved across several species. Knockout of *nifR* will not determine its function, but it should indicate if it is important for curdlan synthesis. As *nifR* is in an operon with other components of the nitrogen signaling cascade (*ntrB* and *ntrC*), *nifR* knockout is expected to reduce or eliminate curdlan production. Indeed, curdlan synthesis in the *nifR* mutant was significantly reduced (Figure 6.4). At this point, it is not known whether the interruption of *nifR* or possible polar effects on n*trBC* is responsible for the effect on curdlan production. However, this does suggest that at least some part of the nitrogen signaling cascade influences curdlan synthesis.

From the study of curdlan synthesis in the gene knockout mutants, it is apparent that RpoN does not regulate the *crd* operon despite the involvement of other elements of

the nitrogen signaling cascade. Therefore, other sigma factors were investigated as potential transcriptional regulators of curdlan synthesis. The promoter region of the *crd* operon was analyzed using Virtual Footprint, a program for predicting transcription factor binding sites. This analysis predicted binding sites for two known transcription factors: AlgU from *Pseudomonas aeruginosa* and ArcA from *E. coli*. Using these two proteins as query, a BLAST search of the ATCC 31749 genome revealed several potential homologs. Of these homologs, r1722, an ECF family RNA polymerase sigma factor, was shown to be up-regulated 47-fold in the nitrogen-dependent transcriptome comparison. Furthermore, r1722 was the RNA polymerase sigma factor with the most significant increase in expression. Based on this analysis, r1722 was targeted for gene KO as a possible transcriptional regulator of the *crd* operon. As shown in Figure 6.4, however, the r1722 mutant was able to produce curdlan at levels similar to the wild-type, indicating that it is not involved in regulating curdlan production.

6.3.3.2 Energy-related genes

Curdlan synthesis, similar to many other polysaccharides, is energy-intensive, requiring two high-energy molecules for each glucose monomer added to the polymer chain. With this energy requirement, it is surprising that curdlan is produced only under nitrogen-limitation, a condition of stress and typically associated with lower metabolic activity. How then can ATCC 31749 provide the energy required for curdlan synthesis?

The high-energy precursor for curdlan synthesis is the sugar nucleotide, UDP-glucose. The genome of ATCC 31749 includes two predicted genes for synthesis of UDP-glucose, *galU* (r3046) and *exoN* (r3316). These genes are predicted to be UTP:glucose-1-phosphate uridylyltransferases. *GalU* is down-regulated 2.4-fold during

curdlan synthesis, suggesting that it is not important for curdlan production. As *galU* is typically associated with cell growth, this result was anticipated. To determine if *exoN* is solely responsible for providing the UDP-glucose for curdlan synthesis, *exoN* was targeted for gene knockout.

Gene knockout of *exoN* was expected to have a deleterious effect on curdlan synthesis, as this gene participates in UDP-glucose synthesis. Hence, the main question is how significant will this impact be? Does *exoN* simply supply UDP-glucose for other polysaccharides produced by ATCC 31749, or is *exoN* the primary source for UDP-glucose in curdlan synthesis? As shown in Figure 6.4, *exoN* gene knockout does affect curdlan synthesis in ATCC 31749; however, it is clearly not the only means of UDP-glucose provision. Based on these experimental results, *exoN* supplies approximately 40% of the UDP-glucose used for curdlan production. Presumably, the remaining supply of UDP-glucose is derived from *galU*, the other UTP:glucose-1-phosphate uridylyltransferase in ATCC 31749. By expressing multiple pathways for UDP-glucose production, ATCC 31749 is able to ensure an adequate supply of sugar nucleotide for high levels of curdlan production. While this result reveals the genes directly involved in UDP-glucose synthesis, the question of how energy is supplied for this synthesis still remains to be answered.

As indicated in Table 6.4, several genes associated with the degradation of high-energy compounds are down-regulated during curdlan synthesis, namely *ppx2* and *rrpP*. Both of these genes are typically associated with acidocalcisomes. Acidocalcisomes are intracellular organelles that have been found in a limited number of organisms ranging from bacteria to man. The complete range of functions for acidocalcisomes remains to be

determined, yet they have been linked with nutrient storage (phosphate, Ca^{2+}, Zn, K^{+}, Na^{+}, Fe, Mg), pH homeostasis, and osmotic regulation [10]. Acidocalcisomes contain large amounts of polyphosphate (polyP), a compound containing high-energy phosphate bonds. Polyphosphate kinase (*ppk*) is responsible for polyP synthesis, but it can also catalyze the reverse reaction, synthesizing ATP or other nucleoside triphosphates from polyP [11-12]. Thus, polyP could serve as an energy source for curdlan production. To determine the influence of polyP and of the possible presence of acidocalcisomes, knockout mutants were constructed for *ppx2* (r2544), and *rrpP* (r2510). In addition to *ppx2*, the genome of ATCC 31749 contains another exopolyphosphatase (*ppx1*, r2763). As this gene may also degrade polyP, it too was targeted for gene knockout.

Acidocalcisomes, containing significant quantities of the high-energy compound polyP, may play an important role in curdlan synthesis. Three acidocalcisome-associated genes were targeted for gene knockout: *rrpP*, *ppx1*, and *ppx2*. The membrane-bound proton-translocating pyrophosphatase (*rrpP*) was down-regulated during curdlan production (Table 6.4). As RrpP consumes pyrophosphate, a high energy compound, it may compete with curdlan synthesis for the cell's energy resources, and down-regulation of this gene may enhance the availability of energy for curdlan production. Based on this reasoning, *rrpP* gene knockout is expected to improve curdlan production. As depicted in Figure 6.4, however, curdlan production was found to be reduced in the *rrpP* mutant. Several possible theories may explain these results. First, a functional *rrpP* may be required for polyP accumulation in the acidocalcisome. If polyP does not accumulate in the acidocalcisome, this eliminates a potential energy source for curdlan synthesis. Another possible explanation is that *rrpP* may help to maintain the intracellular pH at a

level optimal for curdlan synthesis. Extracellular pH regulation has been shown to have a significant impact on curdlan synthesis [2, 13], and *rrpP* may be responsible for intracellular pH regulation, one of the many functions ascribed to the acidocalcisome. The remaining two acidocalcisome-associated genes are both exopolyphosphatases, enzymes that degrade polyP into inorganic phosphate. Despite having the same function, the *ppx1* and *ppx2* knockout mutants had different effects on curdlan synthesis. The *ppx1* mutant produced curdlan at levels similar to the wild-type while curdlan production was eliminated in the *ppx2* mutant (Figure 6.4). This indicates that polyphosphate degradation via *ppx2* is important for curdlan synthesis. For the curdlan measurements shown in Figure 6.4, the cells were grown in rich media until stationary phase was reached and were then transferred to the nitrogen-free media. If the cells are transferred to the nitrogen-free media before reaching the stationary phase, curdlan synthesis is greatly reduced for the wild-type (Figure 6.5). Interestingly, the *ppx1* mutant is able to synthesize the same high level of curdlan as in the case of the stationary phase cells (Figure 6.5 and 6.4). Hence, the elimination of polyphosphate degradation during growth in the *ppx1* mutant allows for high levels of curdlan synthesis. This further emphasizes the different roles played by the two exopolyphosphatases, *ppx1* and *ppx2*. These seemingly contradictory results are discussed in section 6.2.4.

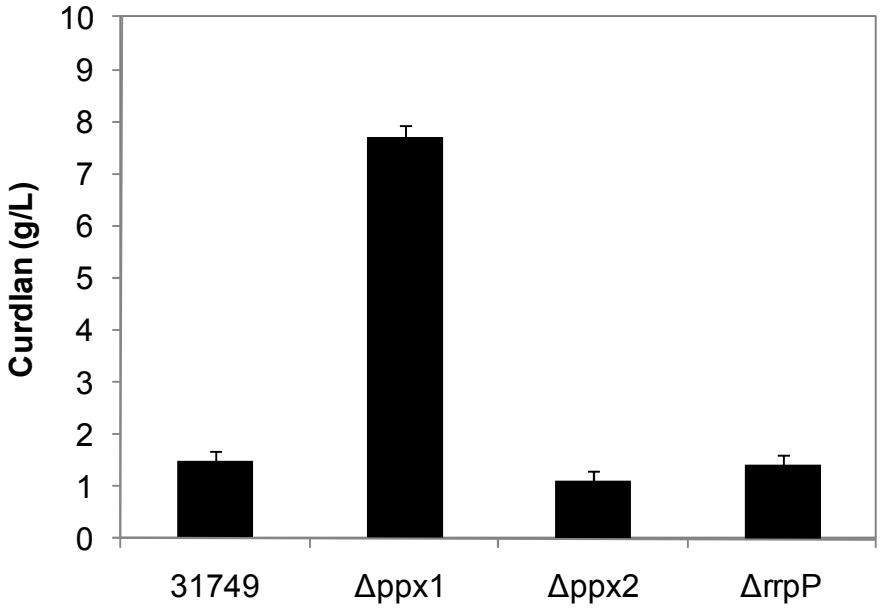


Figure 6.5 Curdlan synthesis in ATCC 31749 and acidocalcisome-associated gene KO mutants from cells transferred to nitrogen-free media prior to reaching stationary phase. These measurements were taken after 48 hours of cultivation in nitrogen-free media.

6.3.3.3 GTP-derived second messengers

Nucleotide-based second messengers are common regulatory molecules that are produced in response to environmental signals. They transmit the signal through interaction with cellular regulatory components to elicit the appropriate cellular response [14]. Two GTP-derived second messengers are of interest for their potential influence on curdlan production: (p)ppGpp and c-di-GMP.

Guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp) are collectively known as (p)ppGpp. These nucleotide-based second messengers were found to accumulate in response to multiple stress conditions, including nitrogen limitation. In turn, the elevated (p)ppGpp levels were shown to affect gene transcription and translation [15]. (p)ppGpp and its co-factor, DksA, can interact with RNA polymerase (RNAP) to directly promote or inhibit gene transcription, and (p)ppGpp/DksA can also indirectly

influence transcription by altering sigma factor utilization. Experimental evidence suggests that (p)ppGpp/DksA reduces transcription from RpoD (σ^{70}) promoters, increasing the availability of the core RNAP for alternative sigma factors such as RpoS (σ^{38}), RpoN (σ^{54}), and RpoE (σ^{24}) [16-17]. Hence, (p)ppGpp may play an important role in regulating transcription of the curdlan synthesis operon, *crdASC*. In *E. coli*, two proteins were shown to synthesize (p)ppGpp: RelA and SpoT [18]. A BLAST search of the ATCC 31749 genome revealed only one significant homolog of RelA (E-value = 7e-099) and SpoT (E-value = e-127): r1489, a GTP pyrophosphohydrolase/synthetase. Interestingly, this potential (p)ppGpp synthase is significantly up-regulated under curdlan-producing conditions (Table 6.4).

The SpoT homolog, r1489, was targeted for gene knockout, resulting in the mutant strain, ATCC 31749Δ*spoT*. This *spoT* mutant failed to synthesize the curdlan polysaccharide (Figure 6.4). The elimination of curdlan production, as opposed to simply a reduction in synthesis, suggests that (p)ppGpp may play an important role in regulating transcription of the *crdASC* operon in ATCC 31749. However, *spoT* may also affect the synthesis of GTP and other GTP-derived compounds, and these molecules may be responsible for the elimination of curdlan production in the *spoT* mutant.

Another GTP-derived second messenger, c-di-GMP, has been directly linked with the regulation of polysaccharide production in several organisms, including *Gluconacetobacter xylinus* (formerly *Acetobacter xylinum*)*, Agrobacterium tumefaciens,* and *Pseudomonas aeruginosa* [4-5, 19]. In *G. xylinus* and *A. tumefaciens*, c-di-GMP acts as an allosteric regulator of cellulose synthase, improving enzyme activity to promote cellulose synthesis [20]. C-di-GMP is also an allosteric regulator of alginate synthesis in

P. aeruginosa, yet instead of binding to the alginate synthase, c-di-GMP binds to Alg44, a protein involved in controlling either alginate polymerization or transport [6]. This second messenger has also been shown to effect gene transcription. In *P. aeruginosa*, c-di-GMP binds to FleQ, a transcriptional repressor of the *pel* operon involved in exopolysaccharide synthesis. While bound to c-di-GMP, FleQ can no longer bind to the promoter region of the *pel* operon, allowing for transcription of the *pel* genes and initiation of PEL exopolysaccharide biosynthesis [21]. A BLAST search of the ATCC 31749 genome did not reveal any homologs for Alg44 or FleQ, and the c-di-GMP binding domain of cellulose synthase (PilZ domain) is not present in curdlan synthase. However, c-di-GMP may play a role in regulating curdlan synthesis in ATCC 31749, as many c-di-GMP binding domains have not yet been identified. Two types of enzymes are involved in c-di-GMP synthesis and degradation: diguanylate cyclases, identified by GGDEF domains, and 5'-phosphoguanylyl-(3'-5')-guanosine phosphodiesterases, characterized by their EAL domains. In fact, it is common for a single protein to contain both GGDEF and EAL domains. From the nitrogen-dependent microarray comparison, only two GGDEF-domain genes were significantly up-regulated (r3957 and r1444), and twelve GGDEF-domain genes were significantly down-regulated, with eight of the twelve also having EAL domains (Table 6.4). While these genes may all influence c-di-GMP levels in ATCC 31749, only the two up-regulated GGDEF-domain genes (r1444 and r3957) were selected for gene knockout.

A successful gene KO mutant was obtained for r3957. Despite several attempts, gene KO of r1444 was unsuccessful, indicating this gene may be necessary for cell growth. The r3957 mutant showed a reduced level of curdlan synthesis with a 57%

decrease in curdlan synthesis (Figure 6.4). These results suggest that c-di-GMP may regulate curdlan production. To determine if the influence of c-di-GMP is specific to synthesis via r3957, two other GGDEF proteins (r0033 and r0631) were targeted for gene knockout. These two GGDEF proteins were expressed during curdlan synthesis but were not found to be significantly up- or down-regulated. The amount of curdlan produced by the r0033 and r0631 mutants is similar to that of the wild-type, confirming that the effect of c-di-GMP on curdlan production is specific to synthesis by the r3957 GGDEF domain-containing protein.

6.3.4 Analysis of possible mechanisms for the regulation of curdlan synthesis

From the curdlan synthesis experiments using the gene KO mutants, it is clear that multiple factors contribute to the regulation of curdlan synthesis in ATCC 31749. Figure 6.6 depicts the gene KO mutants with significant effects on curdlan synthesis and gives possible explanations for the role that each gene plays in regulating curdlan synthesis. The exact mechanisms of regulation remain to be determined, yet from the experimental results presented above and information available in the literature, one can speculate on the possible mechanisms of regulation. First, the nitrogen signaling cascade plays a role in regulating curdlan production. This is indicated by previous experiments which show that nitrogen-limitation and NtrB/C are required for curdlan synthesis [3, 9]. As the experimental results showing the requirement of NtrB/C have not been published, we confirmed this claim by showing that an intact *nifR/ntrB/ntrC* operon is important for curdlan synthesis. Despite the involvement of *ntrBC* in regulating curdlan production, the mechanism is RpoN-independent. This is reminiscent of *Rhodobacter capsulatus* in which NtrC has been shown to regulate the transcription of genes in an RpoN-

independent manner [22-24]. *R. capsulatus* is the only organism in which this RpoN-independent NtrC regulation has been shown to operate extensively. Interestingly, *R. capsulatus* is also the first organism in which *ntrBC* was shown to be transcribed in an operon with a *nifR*-like gene (*nifR3*) [25]. Hence, there may be some connection between the *nifR/ntrB/ntrC* operon and RpoN-independent nitrogen-limited regulation. Evidence presented in this work suggests that curdlan synthesis is regulated by this RpoN-independent mechanism. It still remains to be determined if this RpoN-independent activation is due to NtrC acting solely as a transcriptional activator or if NtrC interacts with another sigma factor to promote transcription. Not only is curdlan synthesis RpoN-independent, it also appears to be inhibited by RpoN. As no RpoN binding site is present in the promoter region of the *crd* operon, the effect of RpoN is likely indirect. Sigma factor competition is a plausible explanation. With a limited amount of RNA polymerase holoenzyme in the cell, sigma factors must compete for the available RNAP [26]. Therefore, through gene KO of *rpoN*, the amount of RNAP available for the sigma factor which promotes transcription of the *crd* operon is increased. This relief of sigma factor competition may explain the increase in curdlan synthesis observed with the *rpoN* mutant.

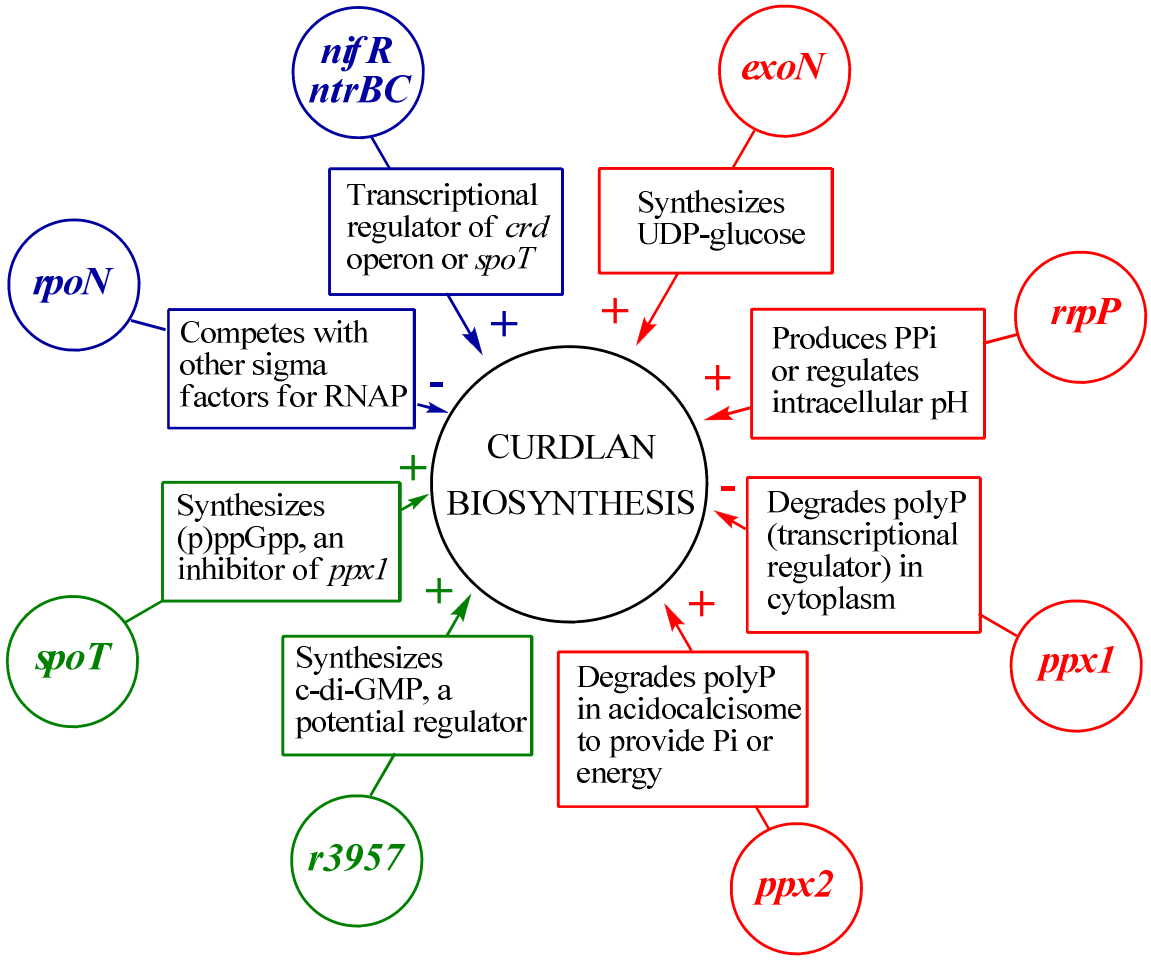


Figure 6.6 Schematic representation of the effects of gene KO on curdlan synthesis and possible explanations for the observed results. The colors indicate the functional category of the gene: nitrogen signaling cascade (blue), energy-related (red), and GTP-derived second messenger (green). The plus (+) or minus (-) signs indicate whether the effect on curdlan production is positive or negative.

In addition to regulation via the nitrogen signaling cascade, polyP also appears to be an important regulator of curdlan synthesis. Gene KO of the two exopolyphosphatases had opposing effects on curdlan production, with *ppx1* KO leading to increased curdlan synthesis (Figure 6.5) and *ppx2* KO resulting in the elimination of curdlan production (Figure 6.4). The exopolyphosphatase mutants also had different effects on cell growth. The *ppx1* mutant had a much slower growth rate than the wild-type while the *ppx2*

mutant grew at the same rate as the wild-type (Figure 6.7). Although the *ppx1* mutant displayed slower growth, the cell viability of ATCC 31749Δ*ppx1* during curdlan production remained similar to that of the wild-type. On the other hand, *ppx2* was found to be important for cell viability under nitrogen-limited conditions, with no viable cells remaining after 96 hours in the nitrogen-free media (Figure 6.8). While *ppx1* and *ppx2* both code for exopolyphosphatases, these genes clearly have different functions in ATCC 31749. These seemingly contradictory results may be understood by examining the subcellular locations of polyP accumulation. As discussed previously, polyP accumulates in intracellular organelles known as acidocalcisomes. Only a limited number of organisms have been shown to produce acidocalcisomes [10], but polyP is a biopolymer common to nearly all bacterial species [27-28]. PolyP accumulates in the cytoplasm of these non-acidocalcisome-producing organisms. In particular, polyP is found to accumulate during the stationary phase and under stress conditions. The mechanism which allows for polyP accumulation is the inhibition of the exopolyphosphatase by (p)ppGpp [29]. This connects polyP with the GTP-derived second messenger (p)ppGpp, another factor shown to be important for curdlan synthesis. It is likely that ATCC 31749 accumulates polyP both within the acidocalcisome organelle and in the cytoplasm.

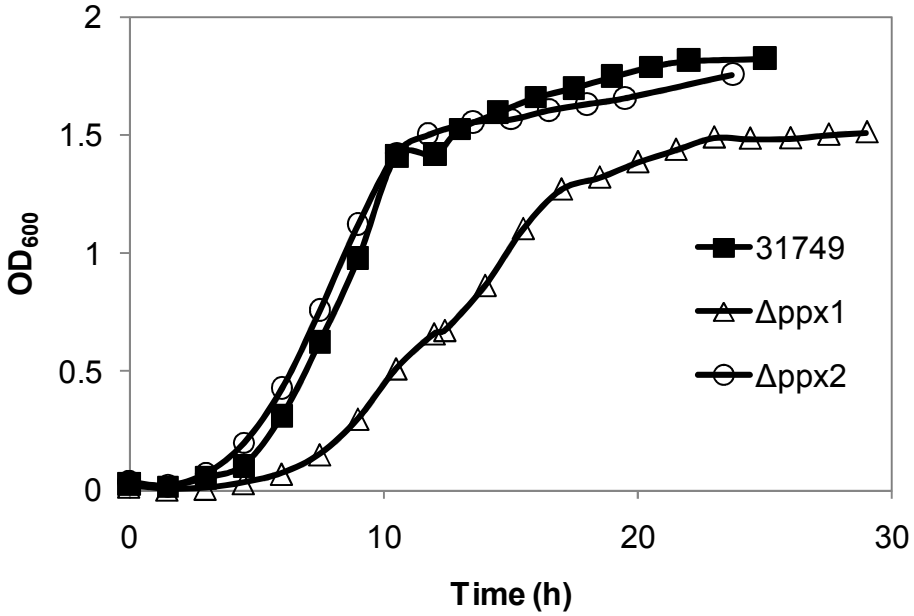


Figure 6.7 Cell growth profiles for wild-type ATCC 31749 (■) and exopolyphosphatase mutants, ATCC 31749Δ*ppx1* (Δ), and ATCC 31749Δ*ppx2* (○).

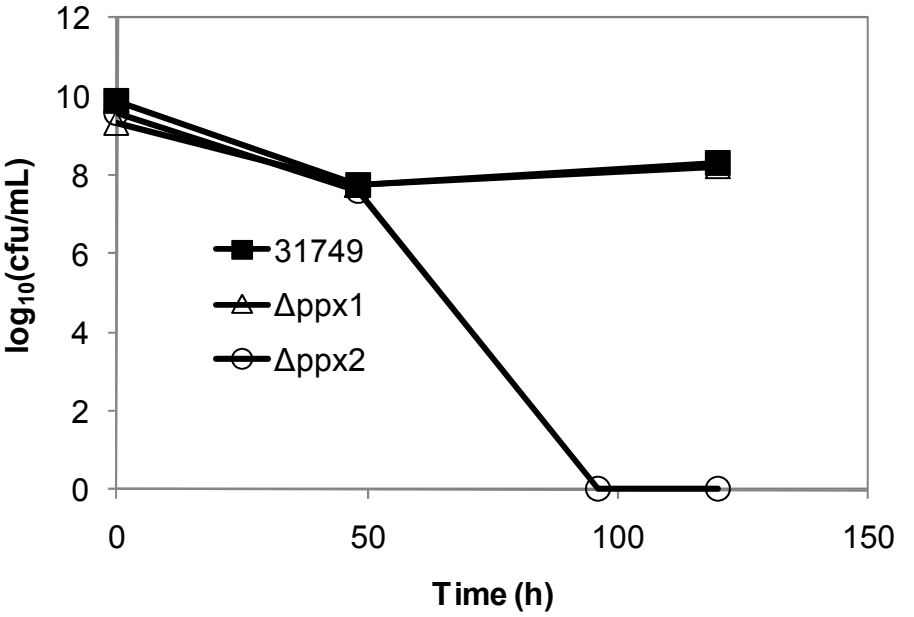


Figure 6.8 Cell viability for wild-type ATCC 31749 (■) and exopolyphosphatase mutants, ATCC 31749Δ*ppx1* (Δ), and ATCC 31749Δ*ppx2* (○).

To support this theory of two subcellular polyphosphate locations, the two exopolyphosphatase genes were analyzed with regard to genomic location and protein structure. In the literature, only two other bacterial species were shown to produce acidocalcisomes: *Agrobacterium tumefaciens* and *Rhodospirillum rubrum* [30-31]. Interestingly, the genomes of both *A. tumefaciens* and *R. rubrum* contain two exopolyphosphatase genes. In comparison, non-acidocalcisome-producing bacteria that have been shown to accumulate polyP in the cytoplasm, such as *E. coli* and *Pseudomonas aeruginosa*, only contain one exopolyphosphatase gene. The sole exopolyphosphatase gene in *E. coli* and *P. aeruginosa* is located adjacent to the polyphosphate kinase gene (*ppk*), with *ppx* and *ppk* in the same operon for *E. coli* (Figure 6.9). The acidocalcisome-producing *A. tumefaciens* and *R. rubrum* also contain an exopolyphosphatase gene in an operon with the polyphosphate kinase gene, similar to the *ppx2* gene of ATCC 31749. The second exopolyphosphatase gene of the acidocalcisome-producing microorganisms is located in an operon with *ftsJ*, coding for a cell division protein. This is analogous to *ppx1* in ATCC 31749. Gene knockout of *ppx1* in ATCC 31749 also likely prevents transcription of the cell division protein *ftsJ*, which may explain the slower growth rate of the *ppx1* mutant in Figure 6.7. Using protein BLAST, the protein domains of *ppx1* and *ppx2* were analyzed and compared to both acidocalcisome-producing and non-acidocalcisome-producing organisms (Figure 6.10). The Ppx-GppA catalytic domains of *ppx2* are located near the N-terminal for all five organisms. On the other hand, the Ppx-GppA domains of *ppx1* are located near the C-terminal, with a 50-200 amino acid sequence at the N-terminal with no predicted function. It is possible that this N-terminal sequence is used to direct the protein to the acidocalcisome organelle, similar to leader

sequences of periplasmic or extracellular proteins. Overall, the conservation of both genome organization and protein domain architecture of the two exopolyphosphatases suggests that *ppx1* is responsible for polyP degradation in the acidocalcisome while *ppx2* degrades polyP in the cytoplasm.

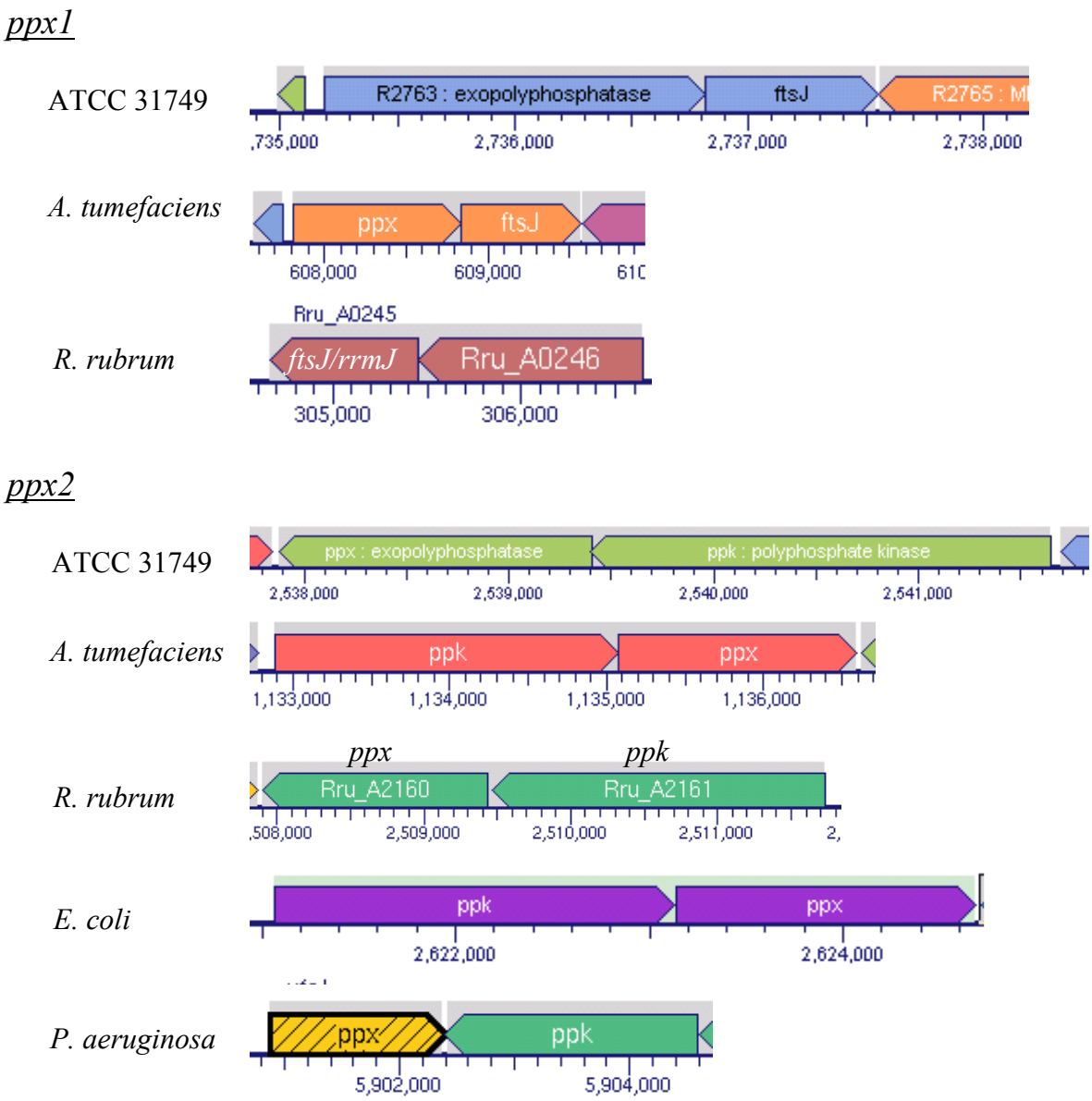
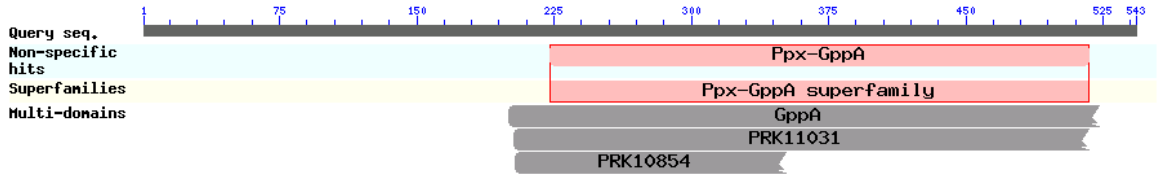


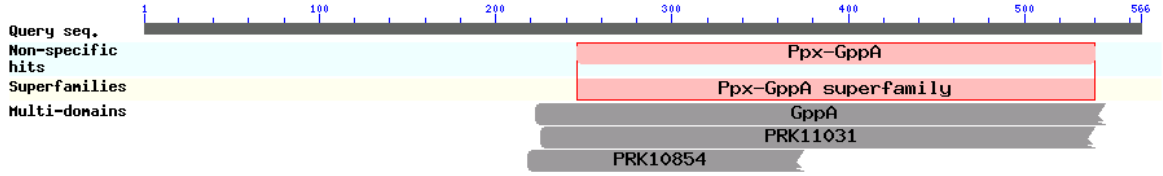
Figure 6.9 Genome organization of the exopolyphosphatase genes found in ATCC 31749 as well as acidocalcisome-producing (*A. tumefaciens* and *R. rubrum*) and non-acidocalcisome-producing (*E. coli* and *P. aeruginosa*) microorganisms.

Ppx1

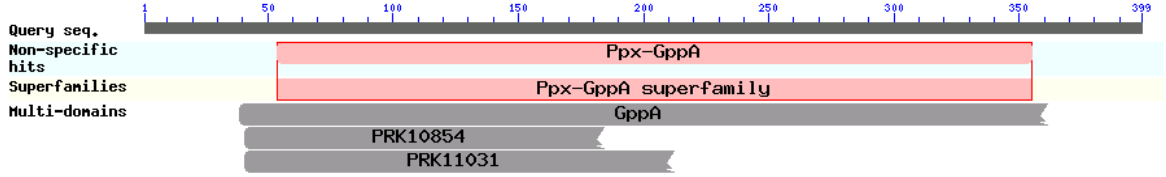
ATCC 31749



A. tumefaciens

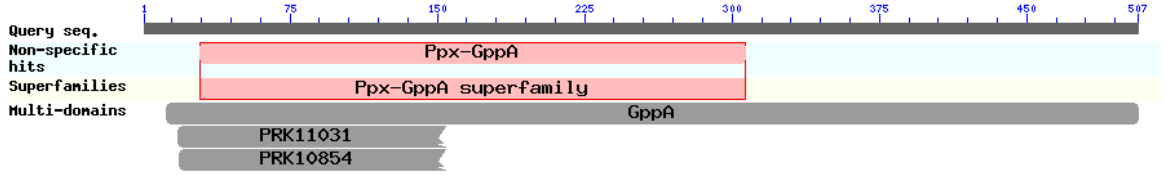


R. rubrum



Ppx2

ATCC 31749



A. tumefaciens

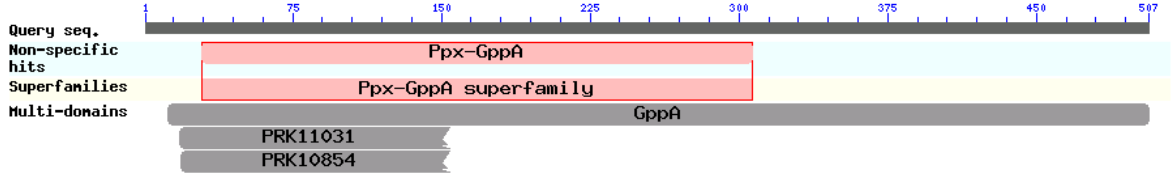
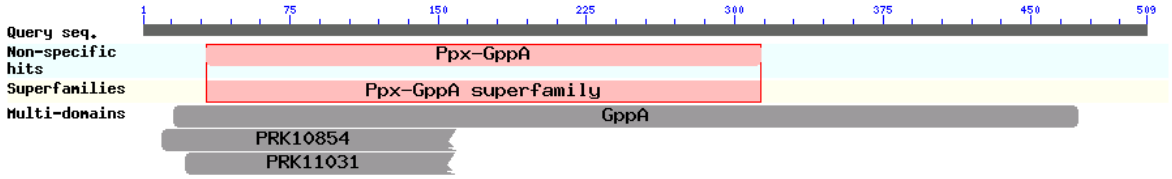


Figure 6.10 Predicted protein domains of the exopolyphosphatase genes found in ATCC 31749 as well as acidocalcisome-producing (*A. tumefaciens* and *R. rubrum*) and non-acidocalcisome-producing (*E. coli* and *P. aeruginosa*) microorganisms.

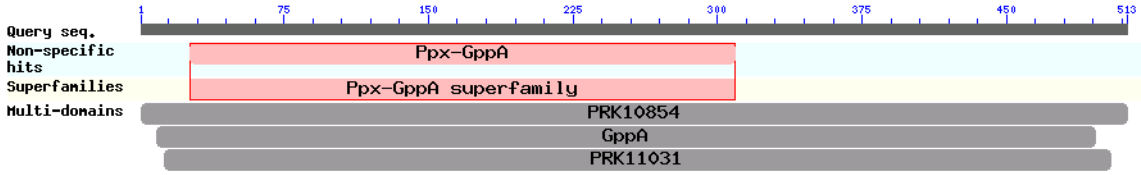
Figure 6.10 Continued.

Ppx2 (continued)

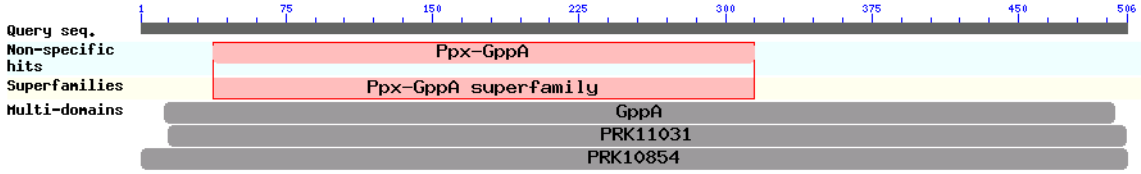
R. rubrum



E. coli



P. aeruginosa



The possibility of two different subcellular locations of polyP may explain the seemingly contradictory results of curdlan synthesis in the *ppx1* and *ppx2* mutants of ATCC 31749. The *ppx2* mutant was shown to have no curdlan synthesis (Figure 6.4). If *ppx2* is responsible for polyP degradation in the cytoplasm as predicted from the analysis of genome organization and protein structure, then the degradation of cytoplasmic polyP is important for curdlan synthesis. Perhaps the phosphate from polyP degradation is used for ATP generation, or cytoplasmic polyP may be a negative regulator of curdlan synthsis. On the other hand, *ppx1* is predicted to be associated with the acidocalcisome. Gene knockout of *ppx1* led to increased curdlan production from the pre-stationary phase cells (Figure 6.5). This suggests that the increased amount of acidocalcisome-associated

polyP in the *ppx1* mutant is important for curdlan synthesis. The acidocalcisome polyP may be used as an energy source for curdlan synthesis or it may act as a transcriptional activator.

While polyP and (p)ppGpp have not been shown to directly regulate exopolysaccharide production, there is experimental evidence suggesting a possible connection between exopolysaccharide synthesis and polyP, (p)ppGpp, GTP, and even NtrC. In 1998, a regulatory protein in *Pseudomonas aeruginosa* (AlgR2) was found to positively regulate production of the exopolysaccharide alginate along with GTP, ppGpp, and polyP [32]. During the same year, a study on polyP accumulation in *E. coli* revealed that NtrC is required for polyP accumulation upon nitrogen exhaustion [33]. Furthermore, this NtrC-mediated regulation of polyP accumulation was found to be RpoN-independent. It appears that the mechanisms regulating alginate synthesis and polyP accumulation in *P. aeruginosa* and *E. coli* are very similar to the regulation of curdlan synthesis in ATCC 31749. So while curdlan production may be unique to ATCC 31749, the mechanism of regulation may be a conserved stress response for many bacterial species. More than a decade has passed since the studies on alginate synthesis in *P. aeruginosa* and polyP accumulation in *E. coli*, yet there has been no additional investigation of the regulatory mechanisms involving ppGpp, polyP, and NtrC. With the addition of the results presented in this study, it appears that ppGpp and NtrC may be important regulators of polyP synthesis under nitrogen-limited conditions, and in turn, polyP may be essential for exopolysaccharide production under stress conditions.

6.4 Discussion and Conclusions

In this study, pH-dependent and nitrogen-dependent transcriptome comparisons were used to identify genes important for curdlan synthesis in ATCC 31749. The pH-dependent comparison revealed significant changes in gene expression for genes associated with glucan product synthesis, the primary metabolism, and c-di-GMP synthesis and degradation. These three gene categories may be responsible for the increase in curdlan production at pH 5.5. The down-regulation of biosynthetic genes associated with producing glucan products may increase the availability of carbon for curdlan synthesis. The up-regulation of a sucrose transporter and down-regulation of glycolytic genes may enhance the carbon flux for curdlan production. And if c-di-GMP is a positive regulator of curdlan production, up-regulation of c-di-GMP biosynthetic genes and down-regulation of c-di-GMP degrading genes would also promote curdlan synthesis. The nitrogen-dependent transcriptome comparison also showed significant changes in gene expression for several areas that may influence curdlan production: nitrogen-limited regulation, energy-related, and GTP-derived second messengers. Genes identified by the nitrogen-dependent transcriptome comparison were subjected to further experimental investigation.

Through gene KO of genes significantly up- or down-regulated under nitrogen-limited conditions and subsequent production of curdlan from these mutants, a possible mechanism for the regulation of curdlan synthesis was proposed. Three factors are proposed to be important for regulating curdlan synthesis: NtrC, (p)ppGpp, and polyP. NtrC is a transcriptional activator that may directly regulate the genes involved in curdlan biosynthesis, or alternatively, NtrC may regulate curdlan synthesis indirectly through

regulation of polyP accumulation. The GTP-derived second messenger, (p)ppGpp, is a known inhibitor of exopolyphosphatase [29] and may also indirectly regulate curdlan synthesis via polyP accumulation. PolyP is a known regulatory factor found to accumulate during the stationary phase and under stress conditions [27]. PolyP has been shown to be involved in transcriptional regulation [28] and may also serve as an energy source for curdlan production. Interestingly, gene KO of the two exopolyphosphatases resulted in opposite effects on curdlan production, suggesting different roles for cytoplasmic and acidocalcisome-associated polyP. The other GTP-derived second messenger investigated in this study, c-di-GMP, appears to also affect curdlan synthesis, but unlike (p)ppGpp, c-di-GMP is not required for curdlan production. Further investigation is necessary to determine how all the factors identified in this study function to regulate curdlan synthesis in ATCC 31749.

Interestingly, polyP may also be a contributing factor to the increased curdlan production at pH 5.5 compared to pH 7. Several investigations of polyP-accumulating microorganisms have shown pH 5.5 to be optimal for polyP accumulation [34-36]. Furthermore, PhoB, a transcriptional activator of the phosphate regulon, was shown to be required for polyP accumulation in *E. coli* [37], and likewise, ChvI, a phoB homolog, was shown to be up-regulated at pH 5.5 in *Agrobacterium tumefaciens* [38]. This suggests a possible mechanism for improved curdlan synthesis at pH 5.5. Increased transcription of the transcriptional activator *chvI* at pH 5.5 enhances polyP accumulation, which in turn regulates curdlan synthesis either though transcriptional regulation or increased energy supply.

While the details of the mechanism(s) regulating curdlan synthesis in ATCC 31749 must be determined, the main factors influencing curdlan synthesis have been identified. These factors will serve as targets for future metabolic engineering efforts to improve oligosaccharide synthesis in the *Agrobacterium* biocatalysts developed previously in this work (Chapters 2 and 3). Several recommendations for improving the oligosaccharide-producing biocatalysts are discussed in the next chapter.

6.5 Materials and Experimental Methods

6.5.1 Materials

The chemicals used in this study were obtained from Sigma-Aldrich ($FeSO_4 \cdot 7H_2O$ and $MnSO_4 \cdot H_2O$) and Fisher (all other chemicals). The restriction enzymes, DNA polymerase, and ligase were obtained from New England Biolabs. The Zyppy plasmid miniprep and Zymoclean gel DNA recovery kits from Zymo Research were used for plasmid isolation and DNA recovery. All kits used for RNA isolation and purification and microarray analysis are indicated in section 6.4.4.

6.5.2 Bacterial strains and plasmids

The bacterial strains used and generated in this study are listed in Table 6.5 along with the plasmids used for gene KO.

Table 6.5 Bacterial strains and plasmids used in this study

Strain or plasmid	Description	Source
ATCC 31749	Curdlan-producing *Agrobacterium* sp. (wild-type)	ATCC
ATCC 31749Δ*exoN*	ATCC 31749 mutant with gene KO of e*xoN*	This study
ATCC 31749Δ*nifR*	ATCC 31749 mutant with gene KO of *nifR*	This study
ATCC 31749Δ*ppx1*	ATCC 31749 mutant with gene KO of r2763	This study
ATCC 31749Δ*ppx2*	ATCC 31749 mutant with gene KO of *ppx*	This study

Table 6.5 Continued

Strain or plasmid	Description	Source
ATCC 31749Δ*rpoN*	ATCC 31749 mutant with gene KO of *rpoN*	This study
ATCC 31749Δ*rrpP*	ATCC 31749 mutant with gene KO of *rrpP*	This study
ATCC 31749Δ*spoT*	ATCC 31749 mutant with gene KO of r1489	This study
ATCC 31749Δ*0033*	ATCC 31749 mutant with gene KO of r0033	This study
ATCC 31749Δ*0631*	ATCC 31749 mutant with gene KO of r0631	This study
ATCC 31749Δ*1722*	ATCC 31749 mutant with gene KO of r1722	This study
ATCC 31749Δ*3957*	ATCC 31749 mutant with gene KO of r3957	This study
JM109	*E. coli* K12 strain used for cloning	Promega
DH5α/pUCP30T	*E. coli* strain containing suicide vector	H. Schweizer
pGEM-T easy	Vector used for cloning internal gene fragments	Promega
pUCP30T-*exoN*	Gene KO plasmid for interruption of *exoN*	This study
pUCP30T-*nifR*	Gene KO plasmid for interruption of *nifR*	This study
pUCP30T-*ppx1*	Gene KO plasmid for interruption of r2763	This study
pUCP30T-*ppx2*	Gene KO plasmid for interruption of *ppx*	This study
pUCP30T-*rpoN*	Gene KO plasmid for interruption of *rpoN*	This study
pUCP30T-*rrpP*	Gene KO plasmid for interruption of *rrpP*	This study
pUCP30T-*spoT*	Gene KO plasmid for interruption of r1489	This study
pUCP30T-*0033*	Gene KO plasmid for interruption of r0033	This study
pUCP30T-*0631*	Gene KO plasmid for interruption of r0631	This study
pUCP30T-*1722*	Gene KO plasmid for interruption of r1722	This study
pUCP30T-*3957*	Gene KO plasmid for interruption of r3957	This study

6.5.3 Fermentation for curdlan synthesis

The ATCC 31749 strains were grown from stock on solid LB media with appropriate antibiotics for plasmid-containing strains. After 1.5 days at 30°C, a single colony was inoculated into a test tube containing 4 mL of LB media (with antibiotics)

and grown overnight at 30°C with 250 rpm. The culture (100 µL) was used to inoculate a 500 mL Erlenmeyer flask containing 150 mL of LB media (with antibiotics). After 17 hours of growth, the cells were concentrated by centrifugation (3000 x g, 10 min, 4°C), reducing the liquid volume from 150 mL to 30 mL. The concentrated culture was added to 450 mL of the minimal fermentation media. The minimal fermentation media was described previously with 1.4 g/L of KH_2PO_4 [39]. NaOH and HCl (1 M) solutions were used for pH control. A Multifors fermenter system from Infors was employed for fermentation with four parallel 500 mL fermentation vessels. The agitation was set at 600 rpm with 1 vvm for the air florwate, and the pH was controlled at pH 7 for growth. The ammonium level was measured using the Berthelot reaction as described by Bailey and coworkers [40]. Upon ammonium depletion, the pH was either maintained at pH 7 or shifted to pH 5.5. After the start of curdlan synthesis (i.e. nitrogen exhaustion), 10 mL samples were taken every 24 hours for 5 to 7 days.

From each sample, 1 mL was used for measuring curdlan. An appropriate amount of 3 M NaOH (15-25 mL) was added to dissolve the water-insoluble curdlan polysaccharide. After waiting 30 min for dissolution, cells were removed from the solublized curdlan solution by centrifugation at 5,000 x g for 20 min. Curdlan was then precipitated from the solution by neutralization using 3 M HCl. The precipitated curdlan was collected by centrifugation (5000 x g, 20 min) and washed 3 times with DI water. The washed curdlan pellet was dried at 80°C until a constant weight was obtained.

To measure the dry cell weight, 5 mL of the culture sample was used. An appropriate amount of 0.5 M NaOH (15-40 mL) was added to dissolve the water-insoluble curdlan. The cells were collected by centrifugation at 5,000 x g and 4°C for 20

min. The cell pellet was washed twice with DI water, and the resulting cell pellet was dried at 80°C until a constant weight was obtained.

6.5.4 RNA isolation and DNA microarray processing

The samples for microarray analysis were collected at 22 h and 100 h of the curdlan synthesis fermentations. The RNA was immediately stabilized using Qiagen's RNAprotect Bacteria Reagent and stored at -80°C for no longer than 2 weeks prior to RNA isolation. RNA was isolated using Qiagen's RNeasy mini kit with on-column DNase digestion using the RNase-free DNase set. For efficient cell lysis, the concentration of the lysozyme solution was increased to 15 mg/mL, and the incubation time was increased to 15 min. For the 100 h samples containing high levels of curdlan, the volumes of the lysozyme solution, RLT buffer, and ethanol were doubled, and a centrifugation step (5,000 x g for 5 min at 10°C) was added after the addition of RLT buffer to remove curdlan before RNA precipitation. After RNA isolation, contaminating genomic DNA was removed using Ambion's DNA-free kit. RNA quality was determined using Agilent's 2100 bioanalyzer with the RNA 6000 Nano kit. RNA concentration was calculated from the OD_{260} readings measured using the SpectraMax M5 microplate reader with Corning's UV-transparent 96-well microplates. Purified RNA samples were stored at -80°C until analysis.

The additional RNA processing and the subsequent microarray processing described in this paragraph were performed by Gene Logic Inc. To prepare the RNA samples for microarray analysis, the RNA was converted to cDNA and labeled with either Cy-3 or Cy-5 using the Fairplay III Microarray Labeling kit from Agilent. The microarray is an 8 x 15k custom DNA microarray from Agilent with probes designed

using Agilent's eArray program. The eArray program was able to design probes for 5,580 of the 5,585 predicted genes in the ATCC 31749 genome. The 5 genes with no probes were all hypothetical and were less than 108 bp in length, with 4 of the 5 having less than 60 bp. Due to their small size and no matches to other known sequences, it is possible that these sequences are not real genes and were therefore excluded from the DNA microarray. For the 5,580 genes on the DNA microarray, there is one probe sequence with two technical replicates for each gene. Fifteen putative housekeeping genes had 10 replicate spots across the array for quality control analysis. (Putative housekeeping genes: *glk, rpoD, atpA, cscA, cisY, edd, pgl, zwf, rplB, rpsC, tkt, pykA, gapA, rpoB,* and *kdgA*). For each condition analyzed by the microarray, there are 4 biological replicates with dye swap so that two replicates were labeled with Cy-3 and two replicates with Cy-5. A microarray scanner model G2505B from Agilent Technologies was employed for imaging the microarray, and Agilent's Feature Extraction 10.5 Image Analysis Software was used to measure the signal intensities from the image of the microarray scan.

6.5.5 Microarray data analysis

Using the Agilent feature extraction software, the mean local background signal intensity was subtracted from the mean signal intensity for each individual spot on the microarray. The background-subtracted signals were then normalized using a lowess normalization to account for any dye-labeling bias. The lowess-normalized data was uploaded into GeneSpring GX 10 for additional processing. For the nitrogen-dependent comparison microarrays, a scaling normalization was applied to the lowess-normalized data, and for the pH comparison microarrays, a quantile normalization was used. The normalized data was then analyzed for statistically significant changes in gene expression

using a paired T-test with a p-value cutoff of 0.05. Plots of the microarray data before and after normalization and statistical analysis are shown in Figures 6.11 and 6.12. Statistically significant genes with more than a 2-fold change in gene expression were manually categorized according to function.

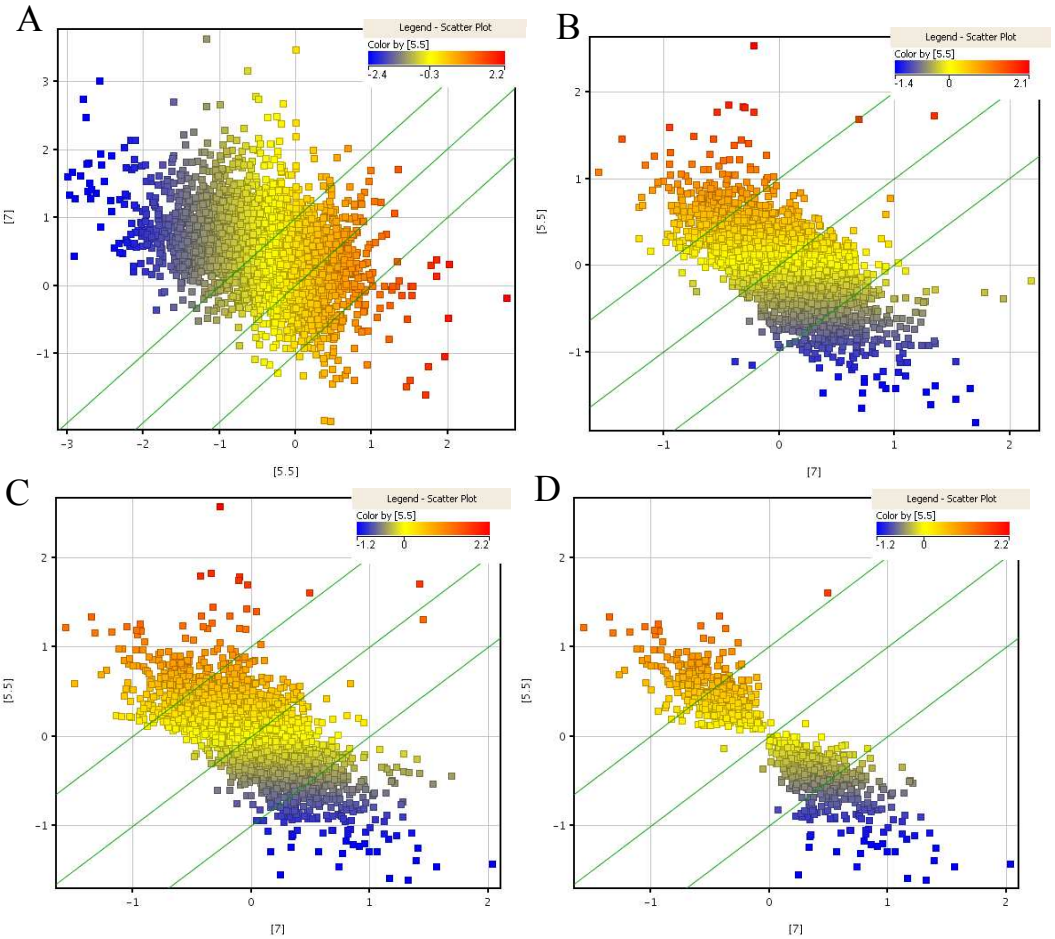


Figure 6.11 Microarray data for the pH-dependent DNA microarray with gene expression for pH 7 on the x-axis and gene expression for pH 5.5 on the y-axis. (A) Signal intensity after background subtraction and baseline transformation to the median; (B) Signal intensity after lowess normalization; (C) Signal intensity after quantile normalization; (D) Signal intensity for statistically significant changes in gene expression as predicted by the paired T-test. The green lines indicate genes with 2-fold changes in gene expression.

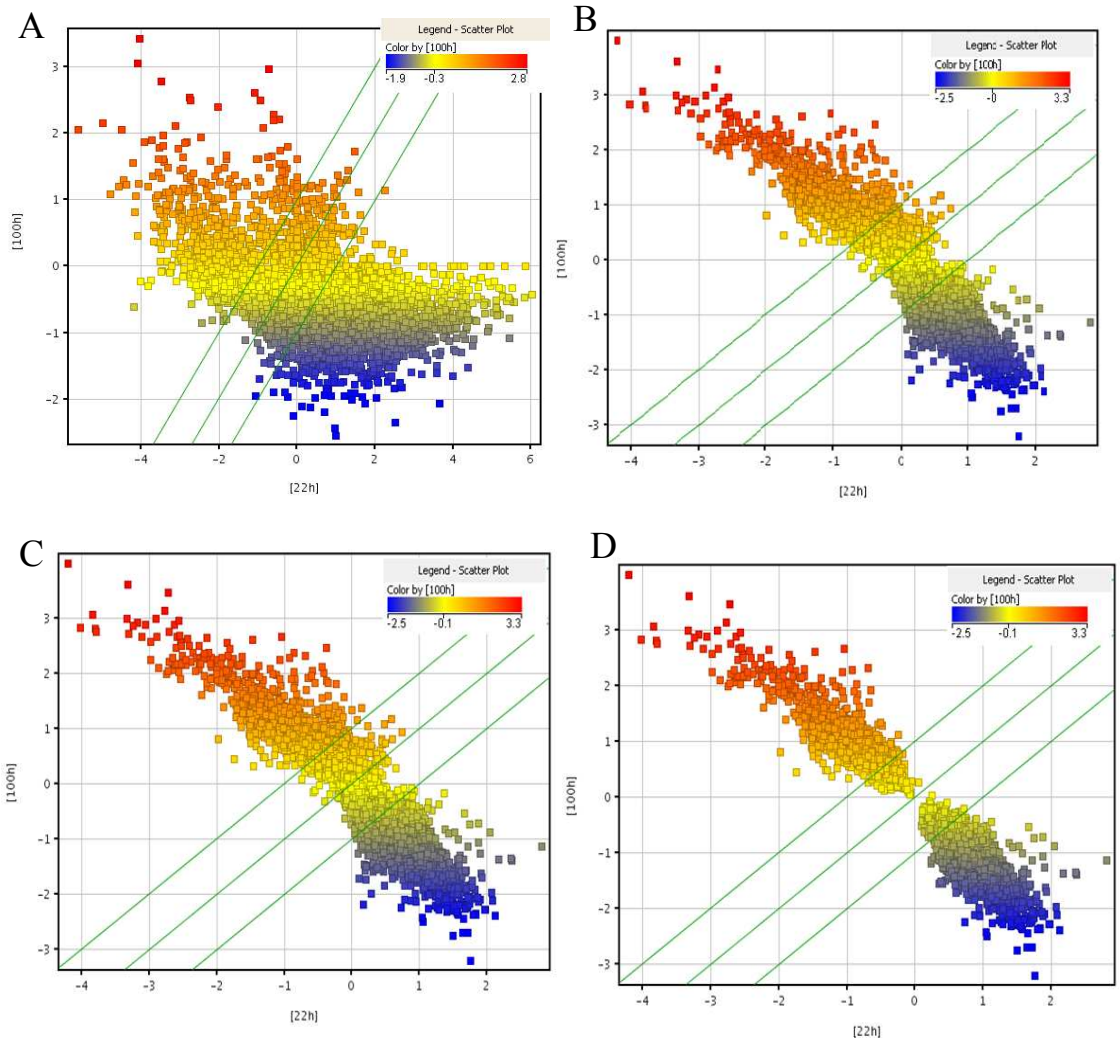


Figure 6.12 Microarray data for the nitrogen-dependent DNA microarray with gene expression for the 22h sample on the x-axis and gene expression for the 100h sample on the y-axis. (A) Signal intensity after background subtraction and baseline transformation to the median; (B) Signal intensity after lowess normalization; (C) Signal intensity after 75th percentile normalization; (D) Signal intensity for statistically significant changes in gene expression as predicted by the paired T-test. The green lines indicate genes with 2-fold changes in gene expression.

6.5.6 Gene knockouts

A gene KO method used previously with *A. tumefaciens* was followed for gene knockout in ATCC 31749 [38]. A small, internal fragment of the target gene (200-500bp) was cloned and inserted into the *SacI* and *HindII* sites of the suicide vector, pUCP30T. The primers used for cloning the internal fragments are listed in Table 6.6 with the

restriction enzyme sites underlined. The resulting KO plasmids (1 µL) were transformed into electrocompetent ATCC 31749 (20 µL) via electroporation using the Biorad micropulser Agr program with an ice-chilled 0.1 cm cuvette. Following electroporation, 1 mL of sterile LB media was added, and the culture was transferred to an eppendorf tube (1.5 mL). The culture was then incubated for 4 hours at 30°C with 250 rpm. After incubation, the culture was concentrated using centrifugation for 5 min at 5,000 rpm and resuspended in 100 µL of supernant. The concentrated culture was spread on LB agar plates containing 50 µg/mL of gentamicin for selection of the gene KO candidates. Successful gene KO was confirmed with PCR using primers that span the junction between the chromosome and inserted KO plasmid. For each gene, the primer homologous to the gentamicin resistance gene (GM^R) was used along with a primer homologous to the sequence upstream of the target gene. The primer sequences are listed in Table 6.6.

Table 6.6 Primers used for gene KO plasmid construction and gene KO confirmation

Target Gene	Forward/ Reverse	Primer Sequence
Internal gene fragment		
exoN	F	5'- CTAAGCTTCGGCTGCAGGATATAACGGC-3'
	R	5'- GAGAGCTCAAGACCGCCATCGAGGACT-3'
nifR	F	5'- GCAAGCTTCCCAATCCGCCCGTCCCTC-3'
	R	5'- CAGAGCTCGAAATGGTCGCCAGCCGC-3'
r2763 (*ppx1*)	F	5'- CTAAGCTTGCAGCGAGGTCCAGTGGGTG-3'
	R	5'- CAGAGCTCCAGTTCCGGGTCGTCGATG-3'
ppx (*ppx2*)	F	5'- CTAAGCTTGCTCCAATCCGCCAGTTCGC-3'
	R	5'- CAGAGCTCGGTCGGTGGTACATGGCGAA-3'

Table 6.6 Continued

Target Gene	Forward/ Reverse	Primer Sequence
rpoN	F	5'-GCAAGCTTCATCGAATCGTAACCCGCG-3'
	R	5'- GCGAGCTCAGCGACCACCTCAATCAG -3'
rrpP	F	5'- GTAAGCTTGATCAGCGCCGTCACGACAAG-3'
	R	5'- CAGAGCTCCCGCGCAACCCTGCTACCAT-3'
r1489 (*spoT*)	F	5'- GTAAGCTTGTATGCTCCAGGAACTCTTC-3'
	R	5'- GAGAGCTCCGTCAAGGGACGTCAGAAAA-3'
r0033	F	5'- CGTGACAAGCTTTTCATAGATGCAGGCAACGTC-3'
	R	5'- GAACGTGAGCTCCTGGTGGGCGGATGATTGT-3'
r0631	F	5'- CCTAGAAAGCTTCACAATAGTCGCTGCCAAGT-3'
	R	5'- CCTAGAGAGCTCGCTTGTCTCGTGACGCTCATT-3'
r1722	F	5'- CGTGACAAGCTTCGGGATGGC AATGGTCTCGT-3'
	R	5'- CTTGTAGAGCTCCTGGCGTTGATACCCATGCT-3'
r3957	F	5'- GAGTGTAAGCTTTCCAGCACATAATAGGGCGA-3'
	R	5'- CTTGGAGAGCTCGACACTTGCGGTTGTCATCG-3'
Gene KO confirmation		
GM^R	F	5'- GATGCCCATACTTGAGCCACCTAAC -3'
exoN	R	5'- GAACTGTCAGAAAAGCCGTCATTCC -3'
nifR	R	5'-AAGATCATCATTTGCCTCTTCCGGAG -3'
r2763 (*ppx1*)	R	5'- GCGAAGCGATCCCGTCGTGGCAAGA -3'
ppx (*ppx2*)	R	5'- GACTCGATCAGAAGCACAGGGGC -3'
rpoN	R	5'- ATGACGCATTTCGAGCTGACGCAGTT-3'
rrpP	R	5'- ATGCGGTGTCCTGTCCGTGGTTTA -3'
r1489 (*spoT*)	R	5'- CGCTTTCTTTTTGTGGAATATCGCA -3'
r0033	R	5'- GTCGTGTTTCAAGTTTCAGGGTCCG -3'
r0631	R	5'- GATGTTGGGTCTTTGGAAAAAACCG -3'
r1722	R	5'- ATGAATGGTGAAAAGCGGGGGAGC -3'
r3957	R	5'-GGTGAATCGCAGATGATGGAATTGAT -3'

6.5.6.1 ATCC 31749 electrocompetent cell preparation

For electrocompetent cell preparation, all glassware used for cell culture was rinsed repeatedly with purified water (resistivity = 18 MΩ/cm), and the same high purity water was used for all media and solution preparation. ATCC 31749 was inoculated in a test tube containing 4 mL of LB media and grown overnight at 30°C with 250 rpm. This culture was diluted 1000 times into 150 mL of LB media and grown at 30°C with 250 rpm until an OD_{600} of 0.5 was reached. After a 30 min rest on ice, the cells were transferred to chilled 50 mL centrifuge tubes and centrifuged for 10 min at 2,000 x g and 4°C. The cell pellet was washed repeatedly with 10 mL of sterile 10% glycerol. Following 3 wash cycles, the cell pellet was resuspended in 0.5 mL of sterile 10% glycerol, separated into 50 µL aliquots, and stored at -80°C until use.

6.5.7 Growth and cell viability studies of gene KO mutants

To determine if the gene KO has any effect on cell growth, the growth profiles of each gene KO mutant were obtained. For each mutant, a single colony was inoculated into 4 mL of LB media and grown overnight at 30°C with 250 rpm. The seed culture was diluted 100 times into 150 mL of LB media. Growth was monitored measuring the optical density of the cell culture using a Beckman Coulter DU530 spectrophotometer. Samples were taken every 1.5 hr.

A reduction or elimination of curdlan synthesis in a mutant strain may be due to a direct effect of the target gene on curdlan production or due to cell death of the mutant strain. To determine if the curdlan-deficiency is merely a result of cell death, the cell viability of curdlan-deficient mutants was assayed. At time periods of 0 hr, 48 hr, and 120 hr during the curdlan synthesis reaction, the cell culture was diluted and spread on LB

agar plates. After incubation at 30°C, the number of colony forming units was counted for each mutant to serve as a measure of cell viability.

6.5.8 Shake-flask curdlan synthesis for analysis of KO mutants

Due to the large number of gene KO mutants generated in this study, shake-flask cultivation was employed to analyze curdlan synthesis in the mutant strains. For each mutant, a single colony was inoculated into 4 mL of LB media and incubated overnight at 30°C and 250 rpm. The seed culture was diluted 100 times into two flasks containing 150 mL LB media. The cultures were grown at 30°C and 250 rpm until stationary phase was reached (approximately 12 hours for mutants with growth similar to the wild-type) or prior to stationary phase (OD_{600} = 1.4). The cells were collected by centrifugation at 3,000 x g and 4°C for 10 min. After one wash step with 10% glycerol, the cell pellets were resuspended in nitrogen-free minimal media. The nitrogen-free minimal media contains 40 g/L sucrose, 0.5 g/L KH_2PO_4, 0.5 g/L $MgSO_4 \cdot 7H_2O$, and 10 mL/L of trace element solution, with the pH adjusted to 7 prior to autoclaving using 6N NaOH. Concentrated solutions of sucrose, $MgSO_4 \cdot 7H_2O$, and trace element solution were autoclaved separately and added to the nitrogen-free media prior to the addition of cells. The trace element solution is comprised of 5 g/L $FeSO_4 \cdot 7H_2O$, 2 g/L $MnSO_4 \cdot H_2O$, 1 g/L $CoCl_2 \cdot 6H_2O$, 1 g/L $ZnCl_2$ in 0.1 M HCl. For curdlan synthesis, the cells, resuspended in the nitrogen-free media, were incubated at 30°C and 250 rpm. Samples were taken at 24 hr intervals to analyze the curdlan concentration. The procedure for measuring curdlan is the same as that described in section 6.4.3, yet the sample size is 5 mL for the shake-flask reactions.

6.6 References

1. Lee, I.Y., et al., *Influence of agitation speed on production of curdlan by Agrobacterium species.* Bioprocess Engineering, 1999. **20**: p. 283-287.

2. Lee, J.-H., et al., *Optimal pH control of batch processes for production of curdlan by Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1999. **23**: p. 143-148.

3. Phillips, K.R., et al., *Production of curdlan-type polysaccharide by Alcaligenes faecalis in batch and continuous culture.* Canadian Journal of Microbiology, 1983. **29**(10): p. 1331-1338.

4. Amikam, D. and M. Benziman, *Cyclic diguanylic acid and cellulose synthesis in Agrobacterium tumefaciens.* Journal of Bacteriology, 1989. **171**(12): p. 6649-6655.

5. Lee, V.T., et al., *A cyclic-di-GMP receptor required for bacterial exopolysaccharide production.* Molecular Microbiology, 2007. **65**(6): p. 1474-1484.

6. Merighi, M., et al., *The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in Pseudomonas aeruginosa.* Molecular Microbiology, 2007. **65**(4): p. 876-895.

7. Kim, M.K., et al., *Higher intracellular levels of uridinemonophosphate under nitrogen-limited conditions enhance metabolic flux of curdlan synthesis in Agrobacterium species.* Biotechnology and Bioengineering, 1999. **62**(3): p. 317-323.

8. Merrick, M.J. and R.A. Edwards, *Nitrogen control in bacteria.* Microbiological Reviews, 1995. **59**(4): p. 604-622.

9. McIntosh, M., B.A. Stone, and V.A. Stanisich, *Curdlan and other bacterial (1à3)-β-D-glucans.* Applied Microbiology and Biotechnology, 2005. **68**: p. 163-173.

10. Docampo, R., et al., *Acidocalcisomes - conserved from bacteria to man.* Nature Reviews Microbiology, 2005. **3**(251-261).

11. Ishige, K., H. Zhang, and A. Kornberg, *Polyphosphate kinase (PPK2), a potent, polyphosphate-driven generator of GTP.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(26): p. 16684-16688.

12. Shiba, T., et al., *Inorganic polyphosphate and polyphosphate kinase: Their novel biological functions and applications.* Biochemistry (Moscow), 2000. **65**(3): p. 375-384.

13. Jin, L.H., et al., *Proteomic analysis of curdlan-producing Agrobacterium sp. in response to pH downshift.* Journal of Biotechnology, 2008. **138**(3-4): p. 87-87.

14. Pesavento, C. and R. Hengge, *Bacterial nucleotide-based second messengers.* Current Opinion in Microbiology, 2009. **12**: p. 170-176.

15. Srivatsan, A. and J.D. Wang, *Control of bacterial transcription, translation and replication by (p)ppGpp.* Current Opinion in Microbiology, 2008. **11**: p. 100-105.

16. Costanzo, A. and S.E. Ades, *Growth phase-dependent regulation of the extracytoplasmic stress factor, σ^E, by guanosine 3',5'-bispyrophosphate (ppGpp).* Journal of Bacteriology, 2006. **188**(13): p. 4627-4634.

17. Szalewska-Palasz, A., et al., *Properties of RNA polymerase bypass mutants.* Journal of Biological Chemistry, 2007. **282**(25): p. 18046-18056.

18. Wu, J. and J. Xie, *Magic Spot: (p) ppGpp.* Journal of Cellular Physiology, 2009. **220**: p. 297-302.

19. Weinhouse, H., et al., *c-di-GMP-binding protein, a new factor regulating cellulose synthesis in Acetobacter xylinum.* FEBS Letters, 1997. **416**: p. 207-211.

20. Ross, P., R. Mayer, and M. Benziman, *Cellulose biosynthesis and function in bacteria.* Microbiological Reviews, 1991. **55**(1): p. 35-58.

21. Hickman, J.W. and C.S. Harwood, *Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor.* Molecular Microbiology, 2008. **69**(2): p. 376-389.

22. Cullen, P.J., et al., *Translational activation by an NtrC enhancer-binding protein.* Journal of Molecular Biology, 1998. **278**: p. 903-914.

23. Foster-Hartnett, D., et al., *A new type of NtrC transcriptional activator.* Journal of Bacteriology, 1994. **176**(20): p. 6175-6187.

24. Masepohl, B., et al., *Urea utilization in the phototrophic bacterium Rhodobacter capsulatus is regulated by the transcriptional activator NtrC.* Journal of Bacteriology, 2001. **183**(2): p. 637-643.

25. Foster-Hartnett, D., et al., *Sequence, genetic, and lacZ fusion analysis of a nifR3-ntrB-ntrC operon in Rhodobacter capsulatus.* Molecular Microbiology, 1993. **8**(5): p. 903-914.

26. Farewell, A., K. Kvint, and T. Nystrom, *Negative regulation by RpoS: a case of sigma factor competition.* Molecular Microbiology, 1998. **29**(4): p. 1039-1051.

27. Kulaev, I., V. Vagabov, and T. Kulakovskaya, *New aspects of inorganic polyphosphate metabolism and function.* Journal of Bioscience and Bioengineering, 1999. **88**(2): p. 111-129.

28. Seufferheld, M.J., H.M. Alvarez, and M.E. Farias, *Role of polyphosphates in microbial adaptation to extreme environments.* Applied and Environmental Microbiology, 2008. **74**(19): p. 5867-5874.

29. Kuroda, A., et al., *Guanosine tetra- and pentaphosphate promote accumulation of inorganic polyphopshate in Escherichia coli.* The Journal of Biological Chemistry, 1997. **272**(34): p. 21240-21243.

30. Seufferheld, M., et al., *Identification of organelles in bacteria similar to acidocalcisomes of unicellular eukaryotes.* The Journal of Biological Chemistry, 2003. **278**(32): p. 29971-29978.

31. Seufferheld, M., et al., *The H^+-pyrophosphatase of Rhodospirillum rubrum is predominantly located in polyphosphate-rich acidocalcisomes.* The Journal of Biological Chemistry, 2004. **279**(49): p. 51193-51202.

32. Kim, H.-Y., et al., *Alginate, inorganic polyphosphate, GTP and ppGpp synthesis co-regulated in Pseudomonas aeruginosa: implications for stationary phase survival and synthesis of RNA/DNA precursors.* Molecular Microbiology, 1998. **27**(4): p. 717-725.

33. Ault-Riche, D., et al., *Novel assay reveals multiple pathways regulating stress-induced accumulations of inorganic polyphosphate in Escherichia coli.* Journal of Bacteriology, 1998. **180**(7): p. 1841-1847.

34. McGrath, J.W., et al., *Acid-stimulated phosphate uptake by activated sludge microorganisms under aerobic laboratory conditions.* Water Research, 2001. **35**(18): p. 4317-4322.

35. McGrath, J.W. and J.P. Quinn, *Intracellular accumulation of polyphosphate by the yeast Candida humicola G-1 in response to acid pH.* Applied and Environmental Microbiology, 2000. **66**(9): p. 4068-4073.

36. Moriarty, T.F., et al., *Effect of reduced pH on inorganic polyphosphate accumulation by Burkholderia cepacia complex isolates.* Letters in Applied Microbiology, 2006. **42**: p. 617-623.

37. Rao, N.N., S. Liu, and A. Kornberg, *Inorganic polyphosphate in Escherichia coli: the phosphate regulon and the stringent response.* Jounral of Bacteriology, 1998. **180**(8): p. 2186-2193.

38. Yuan, Z.-C., et al., *Transcriptome profiling and functional analysis of Agrobacterium tumefaciens reveals a general conserved response to acidic conditions (pH 5.5) and a complex acid-mediated signaling involved in Agrobacterium-plant interactions.* Jounral of Bacteriology, 2008. **190**(2): p. 494-507.

39. Kim, M.K., et al., *Residual phosphate concentration under nitrogen-limiting conditions regulates curdlan production in Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 2000. **25**: p. 180-183.

40. Srienc, F., B. Arnold, and J.E. Bailey, *Characterization of intracellular accumulation of poly-β-hydroxybutyrate (PHB) in individual cells of Alcaligenes eutrophus H16 by flow cytometry.* Biotechnology and Bioengineering, 1984. **26**: p. 982-987.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

7.1 Conclusions

The experimental results and analysis presented in this dissertation accomplish the two main objectives outlined in the introduction (Chapter 1): 1) biocatalysts were developed for oligosaccharide synthesis utilizing ATCC 31749 as host and 2) the factors influencing curdlan production in ATCC 31749 were determined.

7.1.1 Biocatalyst construction for oligosaccharide production

Two biocatalysts were successfully constructed for the production of medically-relevant oligosaccharides through the metabolic engineering of *Agrobacterium* sp. ATCC 31749. The engineered *Agrobacterium* strain, ATCC 31749/pKEL, synthesized over 7 g/L of *N*-acetyllactosamine (LacNAc), an oligosaccharide moiety involved in the process of inflammation and a potential target for cancer treatments. The second biocatalyst, ATCC 31749/pBQET, produced an α-Gal epitope, Gal-α1,3-Lac, an essential component of vaccines under development for the treatment of flu, HIV, and cancer.

Additional metabolic engineering efforts improved the production capacity of the α-Gal-producing biocatalyst. Heterologous expression of a lactose transporter, *lacY*, increased uptake of the acceptor sugar, lactose, resulting in increased production of the α-Gal epitope. α-Gal epitope synthesis was further improved by increasing the supply of the sugar nucleotide precursor, UDP-glucose. Gene knockout of the curdlan synthase gene, *crdS*, eliminated the competitive consumption of UDP-glucose for curdlan synthesis. The

crdS mutant strain, ATCC 31749Δc*rdS*/pBQETY, synthesized nearly 1 g/L of Gal-α1,3-Lac. These results illustrate the benefit of applying metabolic engineering for biocatalyst improvement.

The synthesis of gram-scale quantities of oligosaccharides using the engineered *Agrobacterium* biocatalysts demonstrates the potential application for large-scale synthesis. Additionally, these biocatalysts offer advantages over other available methods for oligosaccharide synthesis. Unlike chemical synthesis and the bacterial coupling technique which require multiple steps for oligosaccharide synthesis [1-2], the engineered *Agrobacterium* synthesizes the glycosidic bond in only one step. This is ideal for industrial production where multiple reactors or fermenters can drive up both capital and production expenses. Furthermore, the *Agrobacterium* biocatalysts require limited and fairly inexpensive nutrients (sucrose, acceptor sugar, phosphate, buffer, and trace amounts of co-factors Mg^{2+} and Mn^{2+}), again reducing production cost in relation to other methods such as enzymatic and bacterial coupling which often require expensive, high energy compounds and co-factors [1, 3-5]. Other one-step, whole-cell biocatalysts for oligosaccharide synthesis employ traditional microbial hosts such as *Escherichia coli*. In this work, we show *Agrobacterium* sp. ATCC 31749 to be a superior host, synthesizing up to 11-fold more LacNAc compared to a similarly engineered *E. coli* strain. While the development of two biocatalysts for the production of LacNAc and Gal-α1,3-Lac is significant in and of itself, this dissertation also demonstrates the potential for ATCC 31749 to serve as host for the production of many other UDP-glucose derived oligosaccharides.

7.1.2 Multiple factors influence curdlan and oligosaccharide synthesis in ATCC 31749

Agrobacterium sp. ATCC 31749 is a beneficial host for oligosaccharide synthesis due to its natural production of curdlan polysaccharide. The large amount of curdlan synthesized by ATCC 31749 implies efficient production of the sugar nucleotide precursor, UDP-glucose. Through metabolic engineering efforts, the efficient UDP-glucose production pathway of ATCC 31749 was hijacked for production of the oligosaccharides LacNAc and Gal-α1,3-Lac. Thus, oligosaccharide production in the biocatalysts is dependent upon the same metabolic pathways for curdlan synthesis. At the start of this project, however, only limited information was available regarding curdlan synthesis. Previous work had identified several extracellular factors important for high curdlan production: nitrogen-limitation, a high level of dissolved oxygen, a pH shift from 7.0 to 5.5 after nitrogen exhaustion, and a low phosphate concentration [6-9]. In addition, four genes were identified as necessary for curdlan production: *crdA, crdS, crdC,* and *crdR*, with only *crdS* having a determined function and reported nucleotide sequence [10-11]. With this limited amount of information on curdlan synthesis, further improvement of the *Agrobacterium* biocatalysts was not feasible. To identify metabolic engineering targets for biocatalyst development, additional extracellular, intracellular, and regulatory factors affecting oligosaccharide and curdlan production in ATCC 31749 were investigated.

Several extracellular factors were found to enhance oligosaccharide synthesis in the *Agrobacterium* biocatalysts. Various carbon sources were investigated to optimize LacNAc synthesis in ATCC 31749/pKEL, including glucose, sucrose, fructose, glycerol,

and GlcNAc. Sucrose was found to be the preferred carbon source for efficient oligosaccharide synthesis. Presumably, ATCC 31749 can transport and metabolize sucrose more efficiently than other carbon sources, as sucrose was also shown to lead to higher amounts of curdlan production [12]. Another extracellular factor found to influence oligosaccharide synthesis was the addition of citrate to the synthesis medium, which resulted in up to a 10-fold improvement in LacNAc production. Further investigation revealed that ATCC 31749 can simultaneously metabolize both sucrose and citrate as sources of carbon and energy. Citrate, a weak acid, also acted as a buffer to help maintain the pH at a level optimal for oligosaccharide synthesis. Both the addition of citrate and the identification of a preferred carbon source led to improved oligosaccharide production in the *Agrobacterium* biocatalysts.

With the successful determination of extracellular factors influencing oligosaccharide and curdlan synthesis, intracellular and regulatory factors became the focus of further investigation. With very little genetic information available for ATCC 31749, the first step in identifying intracellular factors for curdlan synthesis was to sequence the genome of this microorganism. Through genome sequencing, the metabolic pathways relevant to curdlan production were determined, including the pathway for UDP-glucose synthesis, the Entner-Doudoroff pathway for glycolysis, and the TCA cycle for energy production. The genome sequence also revealed regulatory processes and genes with possible influence over curdlan synthesis. Regulatory genes involved in nitrogen, oxygen, phosphate, pH, and exopolysaccharide regulation in other organisms were identified in ATCC 31749 and may be responsible for regulating curdlan synthesis in response to various environmental signals. While genome sequencing revealed the

genes present in ATCC 31749 and also provided the nucleotide sequences needed for metabolic engineering, additional information was required to pinpoint the genes relevant to curdlan production. For this purpose, the transcriptome profile of ATCC 31749 was studied before and after the start of curdlan production, and genes with significant changes in gene expression level were selected for further experimental investigation.

Using gene knockout, several intracellular and regulatory factors were found to influence curdlan production. As expected, the nitrogen signaling cascade was found to be important for curdlan synthesis. However, this nitrogen-limited regulation was shown to be RpoN-independent, indicating a unique mechanism of NtrC-mediated regulation. Another important factor for curdlan synthesis is polyP. Gene knockout of two exopolyphosphatases was found to have opposing effects on curdlan production as well as cell growth and cell viability. The exopolyphosphatase mutant Δ*ppx1* was found to have slow cell growth but was able to produce curdlan at levels similar to the wild-type. Additionally, Δ*ppx1* was shown to produce more than 5 times the amount of curdlan compared to the wild-type when transferred to nitrogen-free media before reaching the stationary phase. On the other hand, the *ppx2* mutant showed normal growth but failed to produce curdlan and had reduced cell viability in the nitrogen-free media. I propose that these results indicate two subcellular locations of polyP accumulation and degradation via exopolyphosphatase: the acidocalcisome and the cytoplasm. PolyP has been shown to accumulate in acidocalcisomes, possibly serving as energy and phosphate storage [13-15]. PolyP has also been shown to accumulate during the stationary phase and regulate gene transcription [16]. Thus, polyP may serve as a source of energy or phosphate during curdlan synthesis, or polyP may act as a transcriptional regulator. The GTP-derived

second messengers, (p)ppGpp and c-di-GMP, were also found to affect curdlan synthesis. Gene knockout of a *spoT* homolog eliminated curdlan synthesis in ATCC 31749. SpoT synthesizes the GTP-derived second messenger, (p)ppGpp. In many microorganisms, (p)ppGpp is found to accumulate at the start of the stationary phase and has been shown to be involved in regulating transcription, translation, and replication [17]. It is therefore plausible that (p)ppGpp is required for transcription or translation of the curdlan synthesis operon *crdASC*. Evidence suggests that c-di-GMP, another GTP-derived second messenger, also influences curdlan production. C-di-GMP may be an allosteric regulator of the curdlan synthesis proteins.

Interestingly, the two intracellular factors, (p)ppGpp and polyP, may not act independently to influence curdlan synthesis. Previous data has shown that (p)ppGpp is required for polyP accumulation. (p)ppGpp inhibits the exopolyphosphatase to allow for cytoplasmic polyP accumulation in the stationary phase [18]. Therefore, (p)ppGpp may regulate curdlan production via polyP accumulation. Additional evidence implies a possible connection between polyP and nitrogen regulation; a study on polyP accumulation in *E. coli* showed that polyP failed to accumulate in *ntrC*$^-$ mutants while polyP accumulation occurred in *rpoN*$^-$ mutants [19]. This suggests a similar mechanism of regulation for polyP and curdlan synthesis. Moreover, NtrC may regulate curdlan production indirectly by influencing polyP accumulation. While the exact mechanism for the regulation of curdlan synthesis remains to elucidated, this work identified four important factors for curdlan production in ATCC 31749: the *nifR/ntrB/ntrC* operon, polyP, (p)ppGpp, and c-di-GMP.

7.2 Significanct Contributions

The research presented in this dissertation makes significant contributions to the fields of oligosaccharide synthesis and microbial exopolysaccharide synthesis and regulation. The *Agrobacterium* whole-cell biocatalysts developed in this dissertation (Chapters 2 and 3) are the first examples of engineering a polysaccharide-producing microorganism for oligosaccharide synthesis. This metabolic engineering strategy takes advantage of the natural capability of the microorganism to produce the sugar nucleotide precursor for oligosaccharide synthesis. The expression of *lacY*, a lactose permease, constitutes a new strategy for increased acceptor uptake in whole-cell oligosaccharide synthesis: selective uptake of the acceptor sugar via transport protein. This strategy was shown to be successful for enhancing oligosaccharide synthesis in the engineered *Agrobacterium* and can be applied to other hosts for improved oligosaccharide production. These contributions are important advances in the field of oligosaccharide synthesis.

The knowledge and understanding of microbial ecopolysaccharide synthesis and regulation are also advanced through the results presented in this dissertation. The transcriptome analysis reveals the first evidence that the curdlan synthesis operon is transcriptionally regulated. This information contributes to the understanding of the regulation of curdlan production in ATCC 31749 and will help direct future efforts to elucidate the regulatory mechanism. The general understanding of microbial exopolysaccharide synthesis and regulation is advanced by the identification of important regulatory factors for curdlan synthesis: polyP, (p)ppGpp, and c-di-GMP. C-di-GMP was previously identified to influence synthesis of several exopolysaccharides [20-23], but

polyP and (p)ppGpp are factors first identified through this dissertation. These factors are known to accumulate during stress conditions in many microorganisms and may be components of a conserved regulatory mechanism for exopolysaccharide synthesis. Furthermore, this is the first evidence linking curdlan synthesis to the stress response of the cell, providing additional insight into the understanding of why this biopolymer is produced. This dissertation presents evidence to suggest two subcellular locations for polyP accumulation: the acidocalcisome and the cytoplasm. While this must still be confirmed experimentally, it may be a key component in understanding the regulation of curdlan synthesis. Lastly, the results of this research indicate a unique mechanism for nitrogen-limited regulation of curdlan synthesis. Some components of the nitrogen signaling cascade are shown to be important for curdlan synthesis, but this regulation is shown to be RpoN-independent. This is likely to be a key factor in elucidating the mechanism for transcriptional regulation of curdlan synthesis. The results of this dissertation provide important contributions to the understanding of microbial exopolysaccharide synthesis and will likely guide future research endeavors in the field.

7.3 Recommendations for Future Work

The two main objectives of this dissertation were achieved: biocatalysts were successfully developed to produce medically-relevant oligosaccharides in just one step and requiring inexpensive co-factors and substrates, and several extracellular and intracellular factors influencing curdlan synthesis were determined. With the identification of genes affecting curdlan synthesis and regulation, future research efforts will focus on two areas: 1) additional metabolic engineering to improve oligosaccharide

synthesis in the *Agrobacterium* biocatalysts and 2) investigation of the mechanism(s) regulating curdlan synthesis in ATCC 31749.

7.3.1 Biocatalyst development for improved oligosaccharide production

7.3.1.1 Large-scale oligosaccharide synthesis

The *Agrobacterium* biocatalysts developed for oligosaccharide synthesis offer several advantages over other synthesis techniques, namely inexpensive co-factor and substrate requirements and a one-step synthesis process. These features are important for large-scale oligosaccharide production. Despite these beneficial attributes, initial attempts at large-scale oligosaccharide synthesis were unsuccessful using the engineered *Agrobacterium* biocatalysts.

The main factor preventing large-scale synthesis is insufficient recombinant protein expression. In small-scale synthesis, this problem is overcome by using a two-step process. In the first step, the engineered *Agrobacterium* is grown in rich media, and recombinant protein expression is induced. After successful protein expression, the cells are collected, washed, and transferred to the minimal, nitrogen-free media for oligosaccharide synthesis. As this two-step process is undesirable for large-scale synthesis, the conditions established for large-scale curdlan production were employed for one-step oligosaccharide synthesis. Unfortunately, the minimal media used for large-scale synthesis did not support recombinant protein expression, as indicated by the negligible activity measured after induction. To overcome this problem, I suggest that the fusion enzyme (*galE:lgtB* or *galE:α1,3galT*) be placed under the control of the promoter for the curdlan synthesis operon, *crdASC*. The genes for curdlan synthesis are adequately expressed in the minimal media following nitrogen-limitation, with a 75-fold increase in

gene expression detected for *crdS* in the nitrogen-dependent microarray comparison. The nucleotide sequence for cloning the promoter of the *crdASC* operon is available from the genome sequence. If gene expression of the fusion enzyme remains limited under the *crd* promoter, the nucleotide sequence of the amino acid codons for the fusion enzyme should be optimized for expression in ATCC 31749. These two steps, *crd* promoter expression of the fusion enzyme and codon optimization, should allow for a one-step process for large-scale oligosaccharide synthesis using the *Agrobacterium* biocatalysts. Moreover, oligosaccharide synthesis in the bioreactor should improve oligosaccharide yields, as the conditions promoting high curdlan production such as high dissolved oxygen concentration and pH maintenance are well controlled in the bioreactor.

7.3.1.2 Targets for metabolic engineering

In this dissertation, several intracellular factors were found to be important for curdlan synthesis, including (p)ppGpp and polyP. Gene knockout of the GTP pyrophosphohydrolase /synthetase (*spoT*) and the exopolyphosphatase (*ppx2*) eliminated curdlan production. Presumably, the reverse may also be true: overexpression of *spoT* and *ppx2* may enhance curdlan production. The two genes should be cloned and overexpressed in ATCC 31749 to determine their effect on curdlan synthesis. If this does in fact improve curdlan synthesis, it may also benefit oligosaccharide synthesis in the *Agrobacterium* biocatalysts. The fusion enzyme for oligosaccharide synthesis and the two genes (*spoT* and *ppx2*) can be simultaneously expressed using the pBQ expression vector to determine the influence of (p)ppGpp and polyP on oligosaccharide production.

Gene knockout of the other exopolyphosphatase (*ppx1*) led to a significant improvement in curdlan production when cells in the growth phase were transferred to

the nitrogen-free media. The same mutant may also be able to produce higher levels of oligosaccharide when engineered with a plasmid containing the requisite fusion enzyme. The plasmids required for LacNAc- and Gal-α1,3-Lac production can be directly transformed into ATCC 31749Δ*ppx1* to test this hypothesis. If successful, a double knockout mutant with gene deletions in both *ppx1* and *crdS* can be constructed to determine if the elimination of curdlan production along with increased polyP will further improve oligosaccharide synthesis.

The development of biocatalysts with both *spoT* and *ppx2* overexpression and *ppx1* gene deletion may further improve oligosaccharide synthesis in the engineered *Agrobacterium* sp. ATCC 31749. These biocatalysts will also determine if (p)ppGpp and polyP regulate only the curdlan synthesis operon or if they also affect UDP-glucose synthesis. If (p)ppGpp and polyP only regulate the genes for curdlan synthesis, the biocatalysts with *spoT* and *ppx2* overexpression and the *ppx1* gene deletion will not show improvement in oligosaccharide synthesis. So in addition to the potential for improving oligosaccharide production, these results will also provide insight into the genes and pathways affected by (p)ppGpp and polyP.

7.3.2 Determination of the mechanisms regulating curdlan synthesis in ATCC 31749

The gene knockout experiments detailed in this dissertation identified several important factors influencing curdlan synthesis in ATCC 31749; however, additional work is required to elucidate the details of the mechanism(s) involved in regulating curdlan production. The gene knockout results indicate that polyP, (p)ppGpp, and c-di-GMP effect curdlan production in ATCC 31749. To confirm this hypothesis, the levels of polyP, (p)ppGpp, and c-di-GMP during both growth and curdlan synthesis should be

measured for the wild-type strain and the Δ*nifR*, Δ*ppx1*, Δ*ppx2,* Δ*spoT*, and Δ*3957* mutants. (p)ppGpp and polyP levels in the Δ*nifR* mutant should help to determine if the nitrogen signaling cascade directly regulates curdlan production or if it acts through the regulation of either (p)ppGpp or polyP. PolyP levels in the Δ*ppx1* mutant should be enhanced compared to the wild-type while (p)ppGpp and polyP levels should be negligible in the Δ*spoT* mutant. Furthermore, a decrease in polyP during the course of curdlan synthesis may indicate that polyP is utilized as an energy source for curdlan synthesis as opposed to simply acting as a transcriptional or translational regulator. Alternatively, if polyP increases along with curdlan production, it will suggest a regulatory role for polyP.

Determining the role of polyP in the regulation curdlan production is an interesting challenge, as the two exopolyphosphatase knockouts showed very different effects on growth and curdlan production. To test the hypothesis that acidocalcisome-associated polyP and cytoplasmic polyP play different roles in regulating growth and curdlan synthesis, a reporter protein such as green fluorescent protein (*gfp*) should be linked to the two exopolyphosphatases and the two polyphosphate kinases. Visualization of cells expressing the *gfp*-modified genes will determine if the polyP synthesizing and degrading proteins are localized in distinct subcellular regions. This may be the first step in unraveling the mechanism of polyP regulation with regard to both curdlan synthesis and other stress-induced processes influenced by polyP.

To determine if the four factors identified through gene knockout, the *nifR/ntrB/ntrC* operon, polyP, (p)ppGpp, and c-di-GMP, effect curdlan production through transcriptional regulation, gene expression levels of the *crd* operon should be

measured in the wild-type and the Δ*nifR*, Δ*ppx1*, Δ*ppx2,* Δ*spoT*, and Δ*3957* mutants. The operon expression level can be measured by fusing the *crd* promoter region to a reporter gene, either *gfp* or *lacZ* (β-galactosidase). The estimated gene expression from measurement of the reporter gene in the background of the wild-type and mutant strains should reveal if transcription of the *crd* operon is influenced by any of the gene knockouts.

The knowledge gained through investigation of the mechanisms regulating curdlan synthesis may suggest additional genes that influence curdlan production. These genes can serve as targets for another round of metabolic engineering to further improve the *Agrobacterium* biocatalysts for oligosaccharide production. The work described in this dissertation establishes the foundation for using engineered *Agrobacterium* biocatalysts for oligosaccharide synthesis and also identifies extracellular and intracellular factors influencing the natural production of curdlan in ATCC 31749. By building upon this work with the recommend experiments detailed in this section, the resulting biocatalysts should enable large-scale and economical production of oligosaccharides for medical applications.

7.4 References

1. Endo, T., et al., *Large-scale production of N-acetyllactosamine through bacterial coupling.* Carbohydrate Research, 1999. **316**: p. 179-183.

2. Flowers, H.M., *Chemical synthesis of oligosaccharides.* Methods in Enzymology, 1978. **50**: p. 93-121.

3. Endo, T., et al., *Large-scale production of CMP-NeuAc and sialylated oligosaccharides through bacterial coupling.* Applied Microbiology & Biotechnology, 2000. **53**: p. 257-261.

4. Koizumi, S., et al., *Large-scale production of GDP-fucose and Lewis X by bacterial coupling.* Journal of Industrial Microbiology & Biotechnology, 2000. **25**: p. 213-217.

5. Koizumi, S., et al., *Large-scale production of UDP-galactose and globotriose by coupling metabolically engineered bacteria.* Nature Biotechnology, 1998. **16**: p. 847-850.

6. Kim, M.K., et al., *Higher intracellular levels of uridinemonophosphate under nitrogen-limited conditions enhance metabolic flux of curdlan synthesis in Agrobacterium species.* Biotechnology and Bioengineering, 1999. **62**(3): p. 317-323.

7. Kim, M.K., et al., *Residual phosphate concentration under nitrogen-limiting conditions regulates curdlan production in Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 2000. **25**: p. 180-183.

8. Lee, I.Y., et al., *Influence of agitation speed on production of curdlan by Agrobacterium species.* Bioprocess Engineering, 1999. **20**: p. 283-287.

9. Lee, J.-H., et al., *Optimal pH control of batch processes for production of curdlan by Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1999. **23**: p. 143-148.

10. Karnezis, T., et al., *Topological characterization of an inner membrane (1à3)-β-D-glucan (curdlan) synthase from Agrobacterium sp. strain ATCC 31749.* Glycobiology, 2003. **13**(10): p. 693-706.

11. Stasinopoulos, S.J., et al., *Detection of two loci involved in (1à3)-β-glucan (curdlan) biosynthesis by Agrobacterium sp. ATCC31749, and comparative sequence analysis of the putative curdlan synthase gene.* Glycobiology, 1999. **9**(1): p. 31-41.

12. Lee, I.-Y., et al., *Production of curdlan using sucrose or sugar cane molasses by two-step fed-batch cultivation of Agrobacterium species.* Journal of Industrial Microbiology & Biotechnology, 1997. **18**: p. 255-259.

13. Docampo, R., et al., *Acidocalcisomes - conserved from bacteria to man.* Nature Reviews Microbiology, 2005. **3**(251-261).

14. Ishige, K., H. Zhang, and A. Kornberg, *Polyphosphate kinase (PPK2), a potent, polyphosphate-driven generator of GTP.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(26): p. 16684-16688.

15. Shiba, T., et al., *Inorganic polyphosphate and polyphosphate kinase: Their novel biological functions and applications.* Biochemistry (Moscow), 2000. **65**(3): p. 375-384.

16. Kulaev, I., V. Vagabov, and T. Kulakovskaya, *New aspects of inorganic polyphosphate metabolism and function.* Journal of Bioscience and Bioengineering, 1999. **88**(2): p. 111-129.

17. Srivatsan, A. and J.D. Wang, *Control of bacterial transcription, translation and replication by (p)ppGpp.* Current Opinion in Microbiology, 2008. **11**: p. 100-105.

18. Kuroda, A., et al., *Guanosine tetra- and pentaphosphate promote accumulation of inorganic polyphopshate in Escherichia coli.* The Journal of Biological Chemistry, 1997. **272**(34): p. 21240-21243.

19. Ault-Riche, D., et al., *Novel assay reveals multiple pathways regulating stress-induced accumulations of inorganic polyphosphate in Escherichia coli.* Journal of Bacteriology, 1998. **180**(7): p. 1841-1847.

20. Amikam, D. and M. Benziman, *Cyclic diguanylic acid and cellulose synthesis in Agrobacterium tumefaciens.* Journal of Bacteriology, 1989. **171**(12): p. 6649-6655.

21. Hickman, J.W. and C.S. Harwood, *Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor.* Molecular Microbiology, 2008. **69**(2): p. 376-389.

22. Lee, V.T., et al., *A cyclic-di-GMP receptor required for bacterial exopolysaccharide production.* Molecular Microbiology, 2007. **65**(6): p. 1474-1484.

23. Merighi, M., et al., *The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in Pseudomonas aeruginosa.* Molecular Microbiology, 2007. **65**(4): p. 876-895.

APPENDIX A

METABOLIC FLUX ANALYSIS

A.1 Reactions included in the metabolic network of ATCC 31749/pKEL

(1) Sucrose → Glc + Fru

(2) Glc → Glc (extracellular)

(3) Fru → Fru (extracellular)

(4) Glc + ATP → G6P + ADP

(5) Fru + ATP → F6P + ADP

(6) F6P → G6P

(7) G6P → G1P

(8) G1P + ATP → UDP-Glc + ADP

(9) UDP-Glc → Crd

(10) UDP-Glc → UDP-Gal

(11) UDP-Gal + Glc → Lactose

(12) UDP-Gal + GlcNAc → LacNAc

(13) UDP-Gal + F6P → P3

(14) G6P → 6PG

(15) 6PG → GAP + PYR

(16) Glycerol-3-P + NAD → GAP + NADH

(17) Glycerol + ATP → Glycerol-3-P + ADP

(18) GAP + 2 ADP + NAD → PYR + 2 ATP + NADH

(19) PYR + NAD → ACoA + NADH

(20) ACoA + ADP → Acetate + ATP

(21) ACoA + OAA → CIT

(22) CIT (extracellular) → CIT

(23) CIT → ICT

(24) ICT + NAD → aKG + NADH

(25) aKG + NAD → SUCC-CoA + NADH

(26) SUCC-CoA + ADP → SUCC + ATP

(27) SUCC + FAD → MAL + $FADH_2$

(28) MAL + NAD → OAA + NADH

(29) ICT + ACoA → SUCC + MAL

(30) NADH + 3 ADP → NAD + 3 ATP

(31) $FADH_2$ + 2 ADP → FAD + 2 ATP

(32) ATP → ADP

A.2 MATLAB code for metabolic flux analysis

```
% This m-file contains the stoichiometric matrix C for ATCC31749/pKEL.
% 0 = C(transpose)v
% There are 32 total fluxes and 25 metabolites, giving 7 degrees of
freedom.

% Since most values in the matrix are zero, start by identifying a
% matrix of zeros with rows = number of unmeasured fluxes, and columns
% = number of metabolites.

% With: ED, TCA, Glyoxylate, curdlan, ATP, NAD/NADH, FAD/FADH
% No: PPP, EMP, PEP carboxylase, PYR carboxylase

% Sucrose is used as the carbon source.
C = zeros(22,25);

% Now specify stoichiometric coefficients for each flux. Reactants are
% negative and products are positive.

%(1) Glycerol-3-P + NAD -> GAP + NADH
C(1,1)= -1; C(1,2)= -1; C(1,3)= 1; C(1,4)= 1;

%(2) Fru + ATP -> F6P + ADP
C(2,5)= -1; C(2,6)= -1; C(2,7)= 1; C(2,8)= 1;
```

```
%(3) Glc + ATP -> G6P + ADP
C(3,9)= -1; C(3,6)= -1; C(3,10)= 1; C(3,8)= 1;

%(4) F6P -> G6P
C(4,7)= -1; C(4,10)= 1;

%(5) GAP + 2 ADP + NAD -> PYR + 2 ATP + NADH
C(5,3)= -1; C(5,8)= -2; C(5,2)= -1; C(5,11)= 1; C(5,6)= 2; C(5,4)= 1;

%(6) PYR + NAD -> ACoA + NADH
C(6,11)= -1; C(6,2)= -1; C(6,12)= 1; C(6,4)= 1;

%(7) ACoA + OAA -> CIT
C(7,12)= -1; C(7,13)= -1; C(7,14)= 1;

%(8) CIT -> ICT
C(8,14)= -1; C(8,15)= 1;

%(9) ICT + NAD -> aKG + NADH
C(9,15)= -1; C(9,2)= -1; C(9,16)= 1; C(9,4)= 1;

%(10) aKG + NAD -> SUCC-CoA + NADH
C(10,16)= -1; C(10,2)= -1; C(10,17)= 1; C(10,4)= 1;

%(11) SUCC-CoA + ADP -> SUCC + ATP
C(11,17)= -1; C(11,8)= -1; C(11,18)= 1; C(11,6)= 1;

%(12) SUCC + FAD -> MAL + FADH2
C(12,18)= -1; C(12,19)= -1; C(12,20)= 1; C(12,21)= 1;

%(13) ICT + ACoA -> SUCC + MAL
C(13,15)= -1; C(13,12)= -1; C(13,18)= 1; C(13,20)= 1;

%(14) MAL + NAD -> OAA + NADH
C(14,20)= -1; C(14,2)= -1; C(14,13)= 1; C(14,4)= 1;

%(15) G6P -> 6PG
C(15,10)= -1; C(15,22)= 1;

%(16) 6PG -> GAP + PYR
C(16,22)= -1; C(16,3)= 1; C(16,11)= 1;

%(17) G6P -> G1P
C(17,10)= -1; C(17,23)= 1;

%(18) G1P + ATP -> UDP-Glc + ADP
C(18,23)= -1; C(18,6)= -1; C(18,24)= 1; C(18,8)= 1;

%(19) UDP-Glc -> UDP-Gal
C(19,24)= -1; C(19,25)= 1;

%(20) NADH + 3 ADP -> NAD + 3 ATP
C(20,4)= -1; C(20,8)= -3; C(20,2)= 1; C(20,6)= 3;

%(21) FADH2 + 2 ADP -> FAD + 2 ATP
C(21,21)= -1; C(21,8)= -2; C(21,19)= 1; C(21,6)= 2;
```

```
%(22) ATP -> ADP
C(22,6)= -1; C(22,8)= 1;

% This m-file contains the stoichiometric matrix M for ATCC31749/pKEL.
% There are 32 fluxes and 25 metabolites, giving 7 degrees of freedom.

% Since most values in the matrix are zero, start by identifying a
% matrix of zeros with rows = number of measured fluxes and columns =
% number of metabolites.

% no sucrose phosphorylase, overdetermined

% Sucrose is used as the carbon source.
M = zeros(10,25);

% Now specify stoichiometric coefficients for each flux. Reactants are
% negative and products are positive.

%(1) Sucrose -> Glc + Fru
M(1,9)= 1; M(1,5)= 1;

%(2) Glc -> Glc(ex)
M(2,9)= -1;

%(3) Fru -> Fru(ex)
M(3,5)= -1;

%(4) Glycerol + ATP -> Glycerol-3-P + ADP
M(4,6)= -1; M(4,1)= 1; M(4,8)= 1;

%(5) ACoA + ADP -> Acetate + ATP
M(5,12)= -1; M(5,8)= -1; M(5,6)= 1;

%(6) CIT(ex) -> CIT
M(6,14)= 1;

%(7) UDP-Glc -> crd + UDP
M(7,24)= -1;

%(8) UDP-Gal + Glc -> Lactose + UDP
M(8,25)= -1; M(8,9)= -1;

%(9) UDP-Gal + GlcNAc -> LacNAc + UDP
M(9,25)= -1;

%(10) UDP-Gal + F6P -> LacN + UDP
M(10,25)= -1; M(10,7)= -1;


% Measured values for 5g/L citrate (mM/h)

% Sucrose consumption
V(1,1)= 2.737826;

% Glucose accumulation
V(2,1)= 0.65301;
```

```
% Fructose accumulation
V(3,1)= 0.239839;

% Glycerol consumption
V(4,1) = 0.649883;

% Acetate production
V(5,1)= 0.679199;

% Citrate consumption
V(6,1)= 0.077918;

% Curdlan formation
V(7,1)= 0.012129;

% Lactose formation
V(8,1)= 0.099341;

% LacNAc formation
V(9,1)= 0.089897;

% Gal/Man formation
V(10,1)= 0.044844;


% Measured values for 0g/L citrate (mM/h)

% Sucrose consumption
V(1,1)= 0.520147;

% Glucose accumulation
V(2,1)= 0.091229;

% Fructose accumulation
V(3,1)= 0.11749;

% Glycerol consumption
V(4,1)= -0.09471;

% Acetate production
V(5,1)= 0.460595;

% Citrate consumption
V(6,1)= 0;

% Curdlan formation
V(7,1)= 0.010739;

% Lactose formation
V(8,1)= -0.00619;

% LacNAc formation
V(9,1)= -0.00473;

% Gal/Man formation
V(10,1)= -0.00189;
```

APPENDIX B

PERL SCRIPTS

B.1 blast_XML_extraction.pl

```
# C:\Perl\blast_XML_extraction2.pl
# Purpose: To extract blast information from the generated XML file

# set warnings
use strict;
use warnings;

# Declare and initialize variables
my $file_blast = 'Pangenome_S_meliloti1021.out';
my @XML = ( );
my @query = ( );
my @qlength = ( );
my @number = ( );
my @id = ( );
my @definition = ( );
my @hlength = ( );
my @score = ( );
my @evalue = ( );
my @identity = ( );
my @positive = ( );
my $count = -1;
my $hit_query = '';
my $hit_qlen = '';
my $hit_num = '';
my $hit_id = '';
my $hit_def = '';
my $hit_hlen = '';
my $hit_score = '';
my $hit_evalue = '';
my $hit_identity = '';
my $hit_pos = '';

# Open blast XML file and store as array
open(XML_BLAST, $file_blast);
@XML = <XML_BLAST>;
close XML_BLAST;
```

```
foreach my$line (@XML) {

        # Find query and store in query array
        if ($line =~ /^ {6}<Iteration_query-def>/) {

                # Add one to count
                $count++;

                # Remove XML notation from line
                $line =~ s/^ {6}<Iteration_query-def>//;
                $line =~ s/<\/Iteration_query-def>//;

                # Store in query array
                $hit_query = $line;
                $query[$count] = $hit_query;

        # Find query length and store in qlength array
        }elsif ($line =~ /^ {6}<Iteration_query-len>/) {

                # Remove XML notation from line
                $line =~ s/^ {6}<Iteration_query-len>//;
                $line =~ s/<\/Iteration_query-len>//;

                # Store in qlength array
                $hit_qlen = $line;
                $qlength[$count] = $hit_qlen;

        # If there are no hits, store no hits message in definition array
        }elsif ($line =~ /^ {6}<Iteration_message>/) {

                # Remove XML notation from line
                $line =~ s/^ {6}<Iteration_message>//;
                $line =~ s/<\/Iteration_message>//;

                # Store line in definition array
                $hit_def = $line;
                $definition[$count] = $hit_def;

                # All other arrays should add a newline \n, with number set to 1.
                $number[$count] = "1\n";
                $id[$count] = "\n";
                $hlength[$count] = "\n";
                $score[$count] = "\n";
                $evalue[$count] = "\n";
                $identity[$count] = "\n";
                $positive[$count] = "\n";
```

```
	# Find hit number and store in number array
	}elsif ($line =~ /^ {10}<Hit_num>/) {

		# Remove XML notation from line
		$line =~ s/^ {10}<Hit_num>//;
		$line =~ s/<\/Hit_num>//;

		# Store in number array
		$hit_num = $line;

		# if hit number is greater than 1, add one to count and add another line to
query
		# and query length
		if ($hit_num > 1) {
			$count++;
			$query[$count] = $hit_query;
			$qlength[$count] = $hit_qlen;
		}

		$number[$count] = $hit_num;

	# Find hit id and store in id array
	}elsif ($line =~ /^ {10}<Hit_id>/) {

		# Remove XML notation from line
		$line =~ s/^ {10}<Hit_id>//;
		$line =~ s/<\/Hit_id>//;

		# Store in id array
		$hit_id = $line;
		$id[$count] = $hit_id;

	# Find hit definition and store in definition array
	}elsif ($line =~ /^ {10}<Hit_def>/) {

		# Remove XML notation from line
		$line =~ s/^ {10}<Hit_def>//;
		$line =~ s/<\/Hit_def>//;

		# Store in definition array
		$hit_def = $line;
		$definition[$count] = $hit_def;

	# Find length and store in length array
```

```
}elsif ($line =~ /^ {10}<Hit_len>/) {

        # Remove XML notation from line
        $line =~ s/^ {10}<Hit_len>//;
        $line =~ s/<\/Hit_len>//;

        # Store in length array
        $hit_hlen = $line;
        $hlength[$count] = $hit_hlen;

# Find score and store in score array
}elsif ($line =~ /^ {14}<Hsp_score>/) {

        # Remove XML notation from line
        $line =~ s/^ {14}<Hsp_score>//;
        $line =~ s/<\/Hsp_score>//;

        # Store in score array
        $hit_score = $line;
        $score[$count] = $hit_score;

# Find e-value and store in evalue array
}elsif ($line =~ /^ {14}<Hsp_evalue>/) {

        # Remove XML notation from line
        $line =~ s/^ {14}<Hsp_evalue>//;
        $line =~ s/<\/Hsp_evalue>//;

        # Store in evalue array
        $hit_evalue = $line;
        $evalue[$count] = $hit_evalue;

# Find identities and store in idenitity array
}elsif ($line =~ /^ {14}<Hsp_identity>/) {

        # Remove XML notation from line
        $line =~ s/^ {14}<Hsp_identity>//;
        $line =~ s/<\/Hsp_identity>//;

        # Store in identity array
        $hit_identity = $line;
        $identity[$count] = $hit_identity;

# Find positives and store in positive array
}elsif ($line =~ /^ {14}<Hsp_positive>/) {
```

```
                # Remove XML notation from line
                $line =~ s/^ {14}<Hsp_positive>//;
                $line =~ s/<\/Hsp_positive>//;

                # Store in positive array
                $hit_pos = $line;
                $positive[$count] = $hit_pos;
        }
}

# Print arrays to files

my$outputfile1 = "S_meliloti1021_query.txt";
unless(open(blast_nr_query, ">>$outputfile1")) {
        print "Cannot open file \"$outputfile1\" to write to!!\n\n";
        exit;
}
print blast_nr_query @query;
close (blast_nr_query);

my$outputfile5 = "S_meliloti1021_qlength.txt";
unless(open(blast_nr_qlength, ">>$outputfile5")) {
        print "Cannot open file \"$outputfile5\" to write to!!\n\n";
        exit;
}
print blast_nr_qlength @qlength;
close (blast_nr_qlength);


my$outputfile2 = "S_meliloti1021_number.txt";
unless(open(blast_nr_number, ">>$outputfile2")) {
        print "Cannot open file \"$outputfile2\" to write to!!\n\n";
        exit;
}
print blast_nr_number @number;
close (blast_nr_number);

my$outputfile3 = "S_meliloti1021_id.txt";
unless(open(blast_nr_id, ">>$outputfile3")) {
        print "Cannot open file \"$outputfile3\" to write to!!\n\n";
        exit;
}
print blast_nr_id @id;
close (blast_nr_id);

my$outputfile4 = "S_meliloti1021_definition.txt";
```

```
unless(open(blast_nr_definition, ">>$outputfile4")) {
        print "Cannot open file \"$outputfile4\" to write to!!\n\n";
        exit;
}
print blast_nr_definition @definition;
close (blast_nr_definition);

my$outputfile6 = "S_meliloti1021_hlength.txt";
unless(open(blast_nr_hlength, ">>$outputfile6")) {
        print "Cannot open file \"$outputfile6\" to write to!!\n\n";
        exit;
}
print blast_nr_hlength @hlength;
close (blast_nr_hlength);

my$outputfile7 = "S_meliloti1021_score.txt";
unless(open(blast_nr_score, ">>$outputfile7")) {
        print "Cannot open file \"$outputfile7\" to write to!!\n\n";
        exit;
}
print blast_nr_score @score;
close (blast_nr_score);

my$outputfile8 = "S_meliloti1021_evalue.txt";
unless(open(blast_nr_evalue, ">>$outputfile8")) {
        print "Cannot open file \"$outputfile8\" to write to!!\n\n";
        exit;
}
print blast_nr_evalue @evalue;
close (blast_nr_evalue);

my$outputfile9 = "S_meliloti1021_identity.txt";
unless(open(blast_nr_identity, ">>$outputfile9")) {
        print "Cannot open file \"$outputfile9\" to write to!!\n\n";
        exit;
}
print blast_nr_identity @identity;
close (blast_nr_identity);

my$outputfile10 = "S_meliloti1021_positive.txt";
unless(open(blast_nr_positive, ">>$outputfile10")) {
        print "Cannot open file \"$outputfile10\" to write to!!\n\n";
        exit;
}
print blast_nr_positive @positive;
close (blast_nr_positive);
```

exit;

B.2 Pathologic_format.pl

```
# C:\Perl\Pathologic_format.pl
# Purpose: To format the ATCC 31749 annotation into PathoLogic format for upload into Pathway Tools

# set warnings
use strict;
use warnings;

# Declare and initialize variables
my $file_ORF = 'Pathologic_ORFs.txt';
my $file_gene = 'Pathologic_gene.txt';
my $file_startb = 'Pathologic_startbase.txt';
my $file_endb = 'Pathologic_endbase.txt';
my $file_function = 'Pathologic_function.txt';
my $file_comment = 'Pathologic_comment.txt';
my @orf = ( );
my @gene = ( );
my @startbase = ( );
my @endbase = ( );
my @function = ( );
my @genecomment = ( );
my $count = 0;
my $id_line = '';
my $name_line = '';
my $startbase_line = '';
my $endbase_line = '';
my $function_line = '';
my $genecomment_line = '';
my $id = "ID\t";
my $name = "NAME\t";
my $sbase = "STARTBASE\t";
my $ebase = "ENDBASE\t";
my $func = "FUNCTION\t";
my $dblink_line = "DBLINK PID:\t\n";
my $producttype_line = "PRODUCT-TYPE\tP\n";
my $comment = "GENE-COMMENT\t";
my $divider = "//\n";
my $gene1 = '';
my $startbase1 = '';
my $endbase1 = '';
my $function1 = '';
my $comment1 = '';
```

```
# Open input files and store as arrays
open(ORF_file, $file_ORF);
@orf = <ORF_file>;
close ORF_file;

open(gene_file, $file_gene);
@gene = <gene_file>;
close gene_file;

open(STARTB, $file_startb);
@startbase = <STARTB>;
close STARTB;

open(ENDB, $file_endb);
@endbase = <ENDB>;
close ENDB;

open(FUNC, $file_function);
@function = <FUNC>;
close FUNC;

open(COMMENT, $file_comment);
@genecomment = <COMMENT>;
close COMMENT;

# Remove tabs from comment lines
foreach my$line1 (@genecomment) {

        $line1 =~ s/\t//g;
}

# Remove spaces from function file
foreach my$line2 (@function) {
        $line2 =~ s/^  //;
}

# Format each ORF and write to Pathologic file
foreach my$line (@orf) {

        $gene1 = $gene[$count];
        $startbase1 = $startbase[$count];
        $endbase1 = $endbase[$count];
        $function1 = $function[$count];
        $comment1 = $genecomment[$count];

        $id_line = "$id$line";
```

```
        $name_line = "$name$gene1";
        $startbase_line = "$sbase$startbase1";
        $endbase_line = "$ebase$endbase1";
        $function_line = "$func$function1";
        $genecomment_line = "$comment$comment1";

        # Open output file and print lines
        my$outputfile1 = "Pathologic_ATCC31749.txt";
        unless (open(pathologic_ATCC31749, ">>$outputfile1")) {
                print "Cannot open file \"$outputfile1\" to write to!!\n\n";
                exit;
        }
        print pathologic_ATCC31749 "$id_line";
        print pathologic_ATCC31749 "$name_line";
        print pathologic_ATCC31749 "$startbase_line";
        print pathologic_ATCC31749 "$endbase_line";
        print pathologic_ATCC31749 "$function_line";
        print pathologic_ATCC31749 "$dblink_line";
        print pathologic_ATCC31749 "$producttype_line";
        print pathologic_ATCC31749 "$genecomment_line";
        print pathologic_ATCC31749 "$divider";
        close(pathologic_ATCC31749);

        # add one to count
        $count++;

}
exit;
```

B.3 matchandsort.pl

```
# C:\Perl\matchandsort.pl
# Purpose: To compare hits from comparative genomics and form a core genome and
differences list of genes.

# set warnings
use strict;
use warnings;

# Declare and initialize variables
my $file_vitis = '31749_vitis_AA_match.txt';
my $file_orf = '31749_ORF_query.txt';
my $file_radio = '31749_radiobacter_AA_match.txt';
my $file_tume = '31749_tumefaciens_AA_match.txt';
my @que = ( );
my @vgene = ( );
```

```
my @rgene = ( );
my @tgene = ( );
my @corgene = ( );
my @vdiff = ( );
my @rdiff = ( );
my @tdiff = ( );
my @qdiff = ( );
my $count1 = 0;
my $count2 = 0;
my $count3 = 0;
my $count4 = 0;
my $que1 = '';
my $rgene1 = '';
my $vgene1 = '';
my $tgene1 = '';

# open each file and store as array

open(NR_VITIS, $file_vitis);
@vgene = <NR_VITIS>;
close NR_VITIS;

open(NR_ORF, $file_orf);
@que = <NR_ORF>;
close NR_ORF;

open(NR_RAD, $file_radio);
@rgene = <NR_RAD>;
close NR_RAD;

open(NR_TUM, $file_tume);
@tgene = <NR_TUM>;
close NR_TUM;

# for each line of gene array, if it equals the query, store in newque array.
# if not, add line to newque array and compare query to next gene.

foreach my$line (@que) {

        $rgene1 = $rgene[$count1];
        $tgene1 = $tgene[$count1];
        $vgene1 = $vgene[$count1];

        $count4 = $count2;

        if ($line eq $rgene1) {
```

```
                if ($line eq $tgene1) {
                        if ($line eq $vgene1) {
                                $corgene[$count2] = $line;
                                $count2 ++;
                        }
                }
        }
        if ($count2 == $count4) {
                $qdiff[$count3] = $line;
                $rdiff[$count3] = $rgene1;
                $tdiff[$count3] = $tgene1;
                $vdiff[$count3] = $vgene1;
                $count3 ++;
        }
        $count1 ++;
}

# Print array to new files

my$outputfile1 = "Coregenome.txt";
open(COR_GEN, ">>$outputfile1");
print COR_GEN"@corgene";
close (COR_GEN);

my$outputfile2 = "31749_differences.txt";
open(QUE_DIF, ">>$outputfile2");
print QUE_DIF"@qdiff";
close (QUE_DIF);

my$outputfile3 = "radiobacter_differences.txt";
open(RAD_DIF, ">>$outputfile3");
print RAD_DIF"@rdiff";
close (RAD_DIF);

my$outputfile4 = "tumefaciens_differences.txt";
open(TUM_DIF, ">>$outputfile4");
print TUM_DIF"@tdiff";
close (TUM_DIF);

my$outputfile5 = "vitis_differences.txt";
open(VIT_DIF, ">>$outputfile5");
print VIT_DIF"@vdiff";
close (VIT_DIF);

exit;
```