2015

Bioplastic Production in Cyanobacteria and Consensus Degenerate PCR Probe Design

Courtney Edward Lane
Louisiana State University and Agricultural and Mechanical College

BIOPLASTIC PRODUCTION IN CYANOBACTERIA AND CONSENSUS DEGENERATE PCR PROBE DESIGN

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

the Gordon A. and Mary Cain
Department of Chemical Engineering

by
Courtney Edward Lane
B.S., Louisiana State University, 2010
December 2015

ACKNOWLEDGEMENTS

This work embodies the culmination of my academic experience at Louisiana State University, forged through challenges, research, and guidance. I am excited to reach this milestone in my life and fortunately, I did not have to do it alone. There are a few very important people in my life to whom I am indebted for their contributions.

First and foremost, Dr. Michael G. Benton – my major professor. Through his knowledge, guidance, and criticism I was able to continually advance through obstacles and accomplish, or surpass, the goals of my projects. The patience and freedoms he awarded me presented the opportunity to instill confidence in myself and develop uniquely. These freedoms also allowed me to follow the work for which I am passionate and for that I am truly thankful. His approachable nature and open-door policy granted me motivation and determination in the face of failure and replaced fears of inadequacy with reaffirmed confidence. Dr. Benton, I am lucky to have had you as a mentor and will always carry with me the knowledge you conveyed. Thank you for giving me this great opportunity!

My committee members, Dr. Mary J. Wornat, Dr. Adam T. Melvin, and Dr. Seung-Jong Park all deserve my sincerest gratitude. Their criticism and insight helped to improve the quality of my work and my outlook as a scientific investigator. Thank you for your time and willingness to evaluate my work!

Looking back into my undergraduate and graduate studies, I would like to thank Dr. Kerry Dooley. As an instructor, you refused to "hold our hand" through the assignments. Out of necessity, we developed the key skills associated with true independent thinking. Dr. Dooley, you always said "One day you'll look back and thank me." Well, once again you are correct. You truly have the student's best interest at heart, thank you for preparing me for both graduate studies and industry. You played a major role in making me a better student.

I also wish to express my appreciation to LSU and the LSU Graduate School for not only allowing me to pursue this doctoral degree, but also for awarding me the Dissertation Year Fellowship. This award has helped me fulfill the requirements of my degree and has given me a sense of validation for my work. Thank you!

I owe a debt of gratitude to all of my classmates and colleagues, who provided me with advice, motivation, and support, but above all else someone with which to share the experience. Most notably, I had the good fortune of maturing academically alongside fellow research group member William Ainsworth, who never shied away from a good discussion. He was a great friend and asset during my time at LSU. Thank you Timothy Thibodeaux for being a great roommate and better friend. John Tate, Thomas Trieu, and Zenghui Zhang all also deserve recognition and my gratitude. I will always have an open door and an open ear for you all. Thank you for sharing in this experience with me.

Lastly, I must thank my family. Without the continued love, support, patience, and guidance of my family, I would have been unable to accomplish this goal. Thank you. To my parents Paige and Mike Lee, I could never thank you enough. You have given me every available opportunity in life and words cannot express how thankful I am. You always checked in on me, made time for me, helped me out, talked to me, and made me feel loved. Thank you. To thank my fiancée, Sarah Hains. You and little Vice were the last thing I expected to happen to me during this time, but I am so glad that you two came into my life. Sarah, without your support I could not have completed this goal. Even though I am certain you will not read this after hearing every detail for the past five years! Also, I am honored to have had the opportunity to watch Vice grow from an adorable four-month-old into the intelligent and kind little five-year-old he is today. I take comfort in knowing you two will continue to support me through my future academic and professional careers. Thank you. To my grandfather, William Darsey, whom is amongst the best of role-models. You showed me that a successful person is rich in family, friends, and intellect. The courage and confidence

you hold has shown me that that great risks can lead to great rewards, rewards which are only made sweeter through adversity and perseverance. Without these qualities I most certainly would not have even accepted this challenge. Thank you. To all of my family. I have truly been blessed to have such a wonderful family!

I would also like to show my appreciation to you, the reader. I have spent a significant amount of time organizing and materializing my thoughts to present. Without an audience, all of my work would be for naught. Thank you for your time and interest in my work and I hope you find it a valuable source of information.

TABLE OF CONTENTS

ABSTRACT

Cyanobacteria show much promise in reducing biodegradable thermoplastic production costs; however, most currently characterized strains are ill-equipped to do so. The result of Objective I produced a high-throughput assay designed to discover existing cyanobacterial strains and rapidly characterize them as PHA-producers or potential PHA-producers. This assay will play an instrumental role in the attainment of a novel cyanobacteria environmental isolate capable of accumulating high levels of PHA naturally. Objective II produced an open source computer program which dramatically speeds the design of similar assays for any arbitrary genetic screening purpose. The program is not limited to this implementation alone. In fact, there are as many uses for this program as there are consensus and/or degenerate oligonucleotide probe applications. The project was released as open source in order to provide a means of constant growth and development by those who need it most. The case studies investigated during the preliminary research of Objective III provided key insights into the complex mechanisms involved in in vitro PHA synthase polymerization kinetics. Additionally, multiple hypothetical physical phenomena are proposed, as inferred from data from literature, which are capable of explaining the kinetic model behavior. All difficulties encountered during the course of Objective III, namely the recombinant protein expression and purification failures, are detailed so that the methods used may be avoided in future experiments. Even though Objective III was completed using an impure PHA synthase sample, it was still found conclusively that the conserved cyanobacteria-specific insertion of the model cyanobacterium PHA synthase is required for proper functionality. This conclusion is significant because it is evidence that the PHA synthase of cyanobacteria may possess a unique catalytic mechanism or method of interaction for multimerization.

CHAPTER 1. INTRODUCTION

1.1 Chapter Preface

This chapter is intended to introduce the thesis of this work in the broadest of terms. This introduction presents the fundamental motivation behind the investigation of bioplastic production in cyanobacteria and the significance of the associated conclusions of this work. Specific examples and supporting information of the rationale will be detailed within the subsequent chapters.

1.2 Rationale

Plastics are a staple non-durable good material consumed in many end-use markets. One of the major benefits of plastic materials is that it can be used to reduce the weight of a component. According to the American Chemistry Council, many plastic components weigh half as much as their non-plastic counterparts (2013). The high rate of strength to density exhibited in plastic resins makes them an attractive class of materials to industries where the weight of the end product is a significant factor in the overall design efficiency. This phenomena is highlighted in the packaging and light-vehicle industries, where the use of plastic resins has consistently increased since their introduction to the market (ACC, 2013).

As it stands, the plastics market is dominated by petroleum-based synthetic polymers (ACC, 2013; Philip, Keshavarz, & Roy, 2007). The use of petroleum products in synthetic plastic production process has a significant impact on the petroleum consumption in the United States. The U.S. is currently a net importer of petroleum products and the amount used for plastic production is equivalent to eight percent of the annual net import value (US-EIA, 2014a, 2014c). An alternative source of plastic materials could significantly alleviate the dependence of the U.S. on foreign petroleum.

Additionally, synthetic plastics are an increasing environmental burden. This phenomena is primarily due to difficulties associated with synthetic plastic degradation. Some plastics are recoverable/recyclable, yet less than nine percent of the total generation of plastics is recovered in the U.S. annually (US-EPA, 2014). The amount of unrecovered plastics will only increase if this this issue is left unaddressed. The detrimental ecological impact of synthetic plastics will only continue to increase as the

annual generation rate is increasing over 7-times faster than the annual rate of recovery (confer Figure 1, page 10).

Furthermore, the number of landfills in the U.S. has plummeted by 70% over the last 25 years (US-EPA, 2014). The lack of available landfills means plastic wastes are typically either thermally degraded, or exported. The thermal degradation of synthetic plastics releases toxins into the atmosphere (Reddy, Ghai, Rashmi, & Kalia, 2003; Suriyamongkol, Weselake, Narine, Moloney, & Shah, 2007). And exportation of plastic wastes is a simply a method of reallocating the problem, and one which is becoming increasingly expensive (Miller, 2014). Both of these solutions address the symptoms of the problem, ignoring the cause. For this reason industry is being prompted to research and implement new environmentally favorable forms of plastics.

A biological alternative to petroleum-based polymers, polyhydroxyalkanoates (PHAs) encompass a diverse class of biodegradable polyesters capable of mitigating the ecological consequences in meeting the ever-growing demand for plastic commodities. PHAs are most commonly produced using heterotrophic microorganisms, which require expensive cultivation media (Choi & Lee, 1999; Ducat, Way, & Silver, 2011; Ienczak, Schmidell, & Aragão, 2013). The production costs associated with PHA production are the main obstacle to their application in industry (Bengtsson, Pisco, Johansson, Lemos, & Reis, 2010).

Photosynthetic prokaryotes capable of PHA accumulation, cyanobacteria possess the ability to harness light energy and fix atmospheric carbon dioxide in order to form these environmentally favorable bioplastics. The use of cyanobacteria to produce PHA would effectively eliminate the main contributor to PHA production costs – carbon feedstock (Choi & Lee, 1999). Unfortunately, these microalgae typically yield significantly less PHA than their heterotrophic counterparts.

A strain of cyanobacteria capable of high yields of PHA would be instrumental in the development of a more environmentally friendly, economically competitive, and sustainable form of plastic production.

1.3 Purpose

The purpose of this study is primarily to attain a more economically-viable strain of cyanobacteria for the design of a carbon-neutral bioplastic production process. A secondary goal, efficient consensus degenerate PCR primer design, presented itself during the course of this work and is also detailed herein.

1.4 Specific Objectives

Objective I. Design a PCR-based assay instrumental in the discovery of a cyanobacteria strain capable of natively accumulating high levels of PHA.

i) Devise robust quality assurance of cyanobacterial genomic DNA.
ii) Identify gene(s) necessary to polymerize PHA.
iii) Validate assay through comparison with traditional detection methods.

Objective II. Develop an efficient consensus degenerate PCR primer design computer application for amplification of arbitrary homologous DNA sequences.

i) Develop the application including methodology from previous work and DNA hybridization algorithms.
ii) Elucidate characteristics of a "good" primer set as described through hybridization algorithms.
iii) Test applicability of consensus primers.
iv) Allow selective addition of degeneracy.

Objective III. Investigate potential causes of low PHA accumulation in cyanobacteria via rational mutagenesis of PHA polymerization enzyme.

i) Analyze PHA synthase in cyanobacteria.
ii) Locate potential region(s) which may affect polymerization.
iii) Rationally mutate said region(s) using recombinant DNA methodologies.
iv) Study the impact of each mutation via *in vitro* enzymatic kinetics.

1.5 Significance of Study

Cyanobacteria show much promise in reducing biodegradable thermoplastic production costs; however, most currently characterized strains are ill-equipped to do so. The result of Objective I produced a high-throughput assay designed to discover existing cyanobacterial strains and rapidly characterize them

as PHA-producers or potential PHA-producers. This assay will play an instrumental role in the attainment of a novel cyanobacteria environmental isolate capable of accumulating high levels of PHA naturally.

Objective II produced an open source computer program which dramatically speeds the design of similar assays for any arbitrary genetic screening purpose. The program is not limited to this implementation alone. In fact, there are as many uses for this program as there are consensus and/or degenerate oligonucleotide probe applications. The project was released as open source in order to provide a means of constant growth and development by those who need it most.

The case studies investigated during the preliminary research of Objective III provided key insights into the complex mechanisms involved in *in vitro* PHA synthase polymerization kinetics. Additionally, multiple hypothetical physical phenomena are proposed, as inferred from data from literature, which are capable of explaining the kinetic model behavior. All difficulties encountered during the course of Objective III, namely the recombinant protein expression and purification failures, are detailed so that the methods used may be avoided in future experiments. Even though Objective III was completed using an impure PHA synthase sample, it was still found conclusively that the conserved cyanobacteria-specific insertion of the model cyanobacterium PHA synthase is required for proper functionality. This conclusion is significant because it is evidence that the PHA synthase of cyanobacteria may possess a unique catalytic mechanism or method of interaction for multimerization.

CHAPTER 2. REVIEW OF RELATED LITERATURE

2.1 Chapter Preface

The goal of this chapter is to provide the reader with a comprehensive collection of the history, phenomena, theories, and vocabulary associated with this work. This work encompasses the fundamental fields of polymers, cyanobacteria, and molecular biotechnology. Specifically, this work focuses on attaining a more economically-viable strain of cyanobacteria for the design of a carbon-neutral and biodegradable bioplastic production process. Therefore, this chapter will define and discuss terminology related to plastics, bioplastics, and biodegradation as well as, the significance of these articles in the economy and the environment. Cyanobacteria, and their bioplastic synthesis and accumulation mechanisms are also detailed herein. Finally, a synopsis of relevant molecular biotechnologies and theories thereof are provided.

2.2 Background

2.2.1 Polymers

The term *polymer* describes an entity of chain-like repeating unitary molecules (*monomer*) linked by *primary valence bonds* to form a macromolecule. A polymer's *primary structure* is defined by the chemical structure and the atomic composition of the monomer units. This term describes the nature, functionality, linking mode, and conferred chemical compositions of the monomers within the polymer (Ebewele, 2000).

The *secondary structure* describes the physiochemical properties of an isolated polymer molecule (i.e., polymer intra-molecular actions). The order in which the configuration units of a polymer are linked together is the *configuration*, which can only be altered through the breaking and reforming of valence bonds. In contrast, the *conformation* of the polymer is the rotation of the configuration units about the valence bonds. A polymer's conformation may be altered without the breaking of valence bonds (Ebewele, 2000). Polymers are either composed of a singular monomer type (*homopolymer*) or multiple monomer types (*copolymers*). Copolymer types include *random*, *alternating*, and *block* (Allcock, Lampe, & Mark,

2003). Polymers which are limited to one dimension are referred to as *linear* polymers, while *branched* polymers possess side chains of the same basic structure.

The *tertiary structure* of a polymer describes the physiochemical properties of the polymer molecules when in aggregate form (i.e., polymer inter-molecular actions). The aggregation of a polymer is the result of *secondary binding forces* and forms one of two essential states: an amorphous material or a crystalline material (Ebewele, 2000).

An *amorphous material* is composed of randomly coiled and intertwined polymers. Upon heating a solid state amorphous polymer, a transition occurs from a solid/brittle state to a molten/viscous state. The point at which this transition occurs is called the *glass transition temperature* (T_g), and any material exhibiting this property is considered a *glass*.

A *crystalline material* is composed of polymers, often folded upon themselves, ordered and packed as crystalline units. A polymer material rarely forms a perfect crystalline material due to the high levels of disorder inherent in molecules of this size and morphology (Ebewele, 2000). Therefore, the materials formed are actually *semi-crystalline* and the *degree of crystallinity* (or % crystallinity) is used to quantitate the similarity to the crystalline state. As a crystalline or semi-crystalline polymer is heated, it will undergo a transition to the amorphous state. The temperature at which this transition occurs is called the *melting temperature* (T_m).

An aggregated polymer may form *cross-links* or the linking of polymer chains via secondary valence bonds. Cross-links can be either intra- or inter-molecular relative to the initial molecular state. The presence of cross-links typically increases the degree of crystallinity of a polymer. Cross-linked polymers do not dissolve because the molecules cannot be separated, but they will swell in the presence of an unreactive solvent (Ebewele, 2000).

Polymer *degradation* is defined as the changes in the polymer's properties due to chemical, physical, or biological reactions resulting in bond scissions and subsequent chemical transformations. Polymer degradation can be classified as photo-oxidative, thermal, ozone-induced, mechanochemical, catalytic, or biodegradation, depending upon the degradation agent (Singh & Sharma, 2008).

2.2.2 Plastics

Plastics are organic polymeric materials. Two families of plastics, *thermosets* and *elastomers*, are typically composed of cross-linking branched polymers. The side chains allow for cross-links to form during vulcanization. For thermosets, this causes a high degree of crystallinity and rigidity. Elastomers, however, typically tend to have rather loose cross-link networks. The formation of these side-chain networks is irreversible and causes the local networking regions to lose their plastic properties.

On the other hand, *thermoplastics* are a family of resins composed of linear polymers or non-linking branched polymers. Amorphous thermoplastics do not have a distinct melting temperature. Therefore, the phase morphology demonstrated is liquid, gum, and rubber phase regimes above the glass transition temperature and only a glass phase below (Allcock et al., 2003). Semi-crystalline thermoplastics (micro-crystalline or crystalline thermoplastics) contain domains of highly ordered molecular arrays and domains of random dispersion. These polymers demonstrate a distinct melting temperature. Under the glass transition temperature crystalline and glassy domains form, above the T_m a molten phase is formed, and between these two temperatures resides the flexible thermoplastic phase regime. Thermoplastics are better known as the recyclable plastics. They can be melted and reformed for a given number of cycles before the polymer structure begin to degrade.

Due to the inherent stability of plastics, degradation is an energy intensive process. Since UV radiations possess energy sufficient to cleave C-C bonds, the principal degradation agent of plastics at ambient conditions is UV and visible light (Singh & Sharma, 2008). Thus, the lifetime of plastics for outdoor applications is assessed under light ranging from 290nm to 400nm (Singh & Sharma, 2008). Under normal conditions, photodegradation and thermal degradation are similar enough to both be classified under oxidative degradation. The main difference between the two degradation processes is that the thermal degradation mechanisms can occur in the bulk of the material, whereas photochemical degradation mechanisms are limited to the surface (Singh & Sharma, 2008). The presence of ozone in the atmosphere,

even trace amounts, causes strong degradation on plastic materials. Ozone-induced degradation can occur even under conditions where other oxidative degradation processes are slow (Singh & Sharma, 2008).

Biodegradation has many definitions, however, it essentially encompasses a process which utilizes the enzymatic action of microorganisms to break down materials into common metabolic products under conditions which accurately reflect the available disposal conditions.

2.2.3 Synthetic Plastics

2.2.3.1 Introduction

Plastics are an attractive material for a number of industrial processes because they can be tailor-made to suit a specific purpose. For example, if a more rigid material is required, then a polymer that incorporates a high degree of chain branching and cross-linking should be used. If transparency is desired, then a polymer which exhibits a glass transition temperature should be used. Plastics can also be blended to achieve desired material properties.

Plastics produced today are lightweight, durable, and inexpensive to produce. Polypropylene, for example, has an estimated production cost of \$0.185 kg^{-1} (Salehizadeh & Van Loosdrecht, 2004). In order for industrial processes to retain their low production costs, they require a readily available hydrocarbon source, most commonly fossil fuels.

The most prevalent industrial thermoplastics include polyethylene terephthalate (PET), low and high density polyethylene (LDPE and HDPE), polyvinyl chloride (PVC), polypropylene (PP), and polystyrene (PS) (ACC, 2007). Industrial thermoplastics exhibit high resistances to degradation at ambient conditions. This can be beneficial for uses such as packaging and storage, where an indefinite shelf-life may be required.

Polyolefins (C_nH_{2n} polymers), such as HDPE, LDPE, and PP, are prevalent in industry due to their low complexity and broad range of material properties (*see* Table 3). They are commonly manufactured as blown films and injection molds for wraps and storage containers.

2.2.3.2 Economic Influence in the United States

In the United States, the plastic resin manufacturing industry generates $87.1 billion in revenues and exports $30.5 billion of plastic resins annually while directly employing 50-60 thousand employees (ACC, 2013). The plastics and rubbers sector accounts for nearly nine percent of the non-durable goods manufacturing sector of the U.S. gross domestic product (US-BEA, 2014). Fossil fuels provide the vast majority of hydrocarbon raw materials for plastic resins in the U.S. (ACC, 2013). In 2010, the U.S. consumed 190 million barrels of liquid petroleum products for use as plastic resin feedstock – three percent of the total liquid petroleum product consumption for that year (US-EIA, 2014a). The United States is a net-importer of petroleum products at a rate of 6.2 million barrels per day (US-EIA, 2014c). In summary, the synthetic plastics industry is a crucial contributor to U.S. GDP and is heavily dependent on a hydrocarbon feedstock, which is unsustainable long-term.

2.2.3.3 Environmental Impact

32 million tons of plastic solid wastes were generated in the United States in 2012, yet only 2.8 million tons were recovered for recycling (US-EPA, 2014). High molecular weight polyolefins such as those used in plastic materials, are considered xenobiotic (Steinbüchel, 2005). Xenobiotics are compounds that do not occur within any known natural metabolic pathway; these compounds are predominantly non-biodegradable. The majority of these plastic wastes can persist in the environment for extended periods of time (Andrady & Neal, 2009). The U.S. is steadily increasing the generation of these plastic solid wastes (*see* Figure 1). The increase in the annual rate of generation from 1990-2012 is estimated to be 740(±60) thousand tons per year, while the increase in the annual rate of recovery is only estimated to be 100(±6) thousand tons per year (regressed from Figure 1). Compared to the municipal recovery of the same year (8.9%), industrial recovery of thermoplastics also leaves much room for improvement, as can be observed in Table 1.

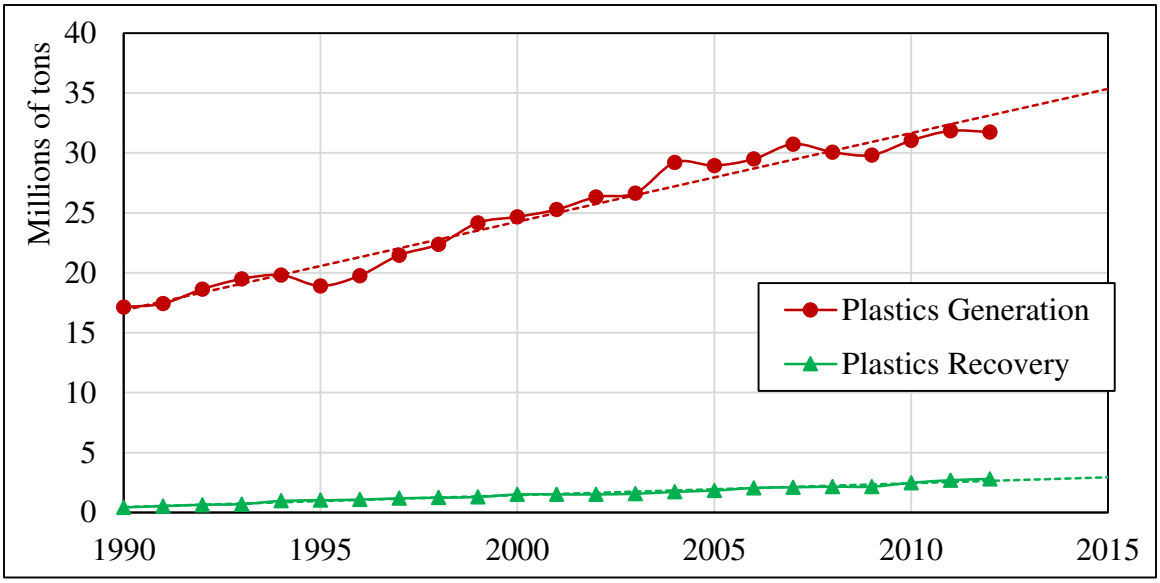


Figure 1: Total plastic materials generated in the municipal solid waste (MSW) and the total plastics recovery (recycling) from the plastics in MSW as adapted from the United States Environmental Protection Agency (US-EPA, 2014).

Table 1: Estimated recovery as percent generation of thermoplastic materials in the packaging industrial sector during 2012 (ACC, 2014). NA – Not Available.

Packaging Sector	HDPE	LDPE	PET	PP	PS
Beverage containers, bottles, jars	27.5%	NA	NA	8.33%	29.2%
Caps and closures	6.67%	NA	NA	NA	5.88%
Carrier bags/stretch and shrink	4.35%	17.6%	NA	NA	NA
Other rigid	19.3%	NA	16.4%	8.33%	5.88%
Other flexible	6.67%	NA	NA	1.8	5.88%

The number of U.S. landfills has also substantially decreased from 6,326 landfills in 1990 to 1,908 landfills in 2012 (US-EPA, 2014). The remainder of plastic wastes that do not enter the landfills undergo thermal degradation. The thermal degradation of halo-polymers, such as PVC, results in the formation of toxic halogenated hydrocarbons or inorganic chlorides (Bhaskar et al., 2003). Thermal degradation of non-chlorinated plastic polymers also results in the release of hazardous components in the form of airborne particulates (Michal, Mitera, & Tardon, 1976). Any plastic waste surplus undergoes exportation. Foreign government actions, such as China's "Green Fence" act of 2013, are increasing exportation costs and decreasing the speed at which plastic wastes can be exported (Miller, 2014).

Improper disposal of these non-biodegradable plastics leads to worldwide effects. The majority of the improperly disposed plastic eventually washes into the ocean. Unfortunately, the degradation of plastics in the marine environment is severely retarded and the accumulation of this anthropogenic debris has numerous detrimental effects on the ecosystem (Andrady, 1989). There are many documented cases of animals ingesting the materials or becoming entrapped in them, severely impeding proper development (Webb , Arnott, Crawford, & Ivanova, 2013). The debris also provides a means of safe transport for marine life over previously unreachable distances. The introduction of opportunistic travelers into a new ecosystem immediately increases the local available biodiversity and can have catastrophic effects on the indigenous biota (Barnes, 2002).

2.2.4 Bio-based Plastics

Viewed as a more sustainable approach to synthetic plastic production, *bio-based plastics* are generated using biological renewable resources, such as carbohydrate feedstock. While this method addresses the sustainability issues regarding fossil fuels, the environmental concerns associated with most of these plastics may remain.

However, one bio-based plastic has gained attention recently as a biodegradable alternative to polyolefins: poly(lactic acid) or PLA (Steinbüchel, 2005). PLA is typically produced using lactic acid generated via fermentation (Mecking, 2004). However, due to its simple primary structure (Figure 2), the material properties of the products are limited in comparison to existing synthetic plastics (*see* Table 3).

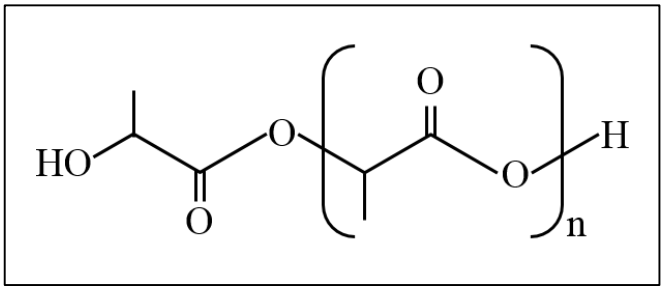


Figure 2: Chemical structure of poly(lactic acid) (PLA).

2.2.5 Biopolymers

Polymers produced by living organisms are known as *biopolymers*. These polymers host a plethora of functions, from catalytic activity to cell structure to energy and even information storage. The eight known classes of biopolymers are listed below in Table 2.

Table 2: Eight known classes of biopolymers as adapted from Lütke-Eversloh *et al.* (2001).

Class	Polymerase substrate(s)	Producers
Nucleic acids	Nucleoside triphosphates	Prokaryotes, Eukaryotes, Archaea
Proteins	Aminoacyl-transfer ribonucleic acids; amino acids	Prokaryotes, Eukaryotes, Archaea
Polysaccharides	Sugar-nucleoside diphosphates, sucrose	Prokaryotes, Eukaryotes, Archaea
Polyhydroxyalkanoates	Hydroxyacyl coenzyme A	Prokaryotes, Eukaryotes, Archaea
Polythioesters	Mercaptoacyl coenzyme A	Prokaryotes
Polyphosphates	Adenosine triphosphate	Prokaryotes, Archaea
Polyisoprenoids	Isopentenylpyrophosphate	Plants, Fungi
Lignin	Radical intermediates	Plants

Polymerases are the catalytically active proteins (*enzymes*) responsible for generating the primary covalent bonds of a biopolymer using a monomer or its precursor as substrate.

2.2.6 Bioplastics

Bioplastics are plastic materials derived from biopolymers. Naturally occurring bioplastics can be created from two types of biopolymers: polythioesters (PTEs) and polyhydroxyalkanoates (PHAs). It should be noted however, that PLA biosynthesis can be achieved in recombinant organisms (Yang et al., 2010). PTEs encompass an interesting class of biopolymers, most which are non-biodegradable (Steinbüchel, 2005). PTEs show promise as long-life bioplastics; however, large-scale degradation of this class of biopolymer still requires further investigation. In contrast, PHAs are a promising biodegradable alternative to polyolefins with a wide range of material properties (Rehm, 2007).

2.3 Polyhydroxyalkanoates

2.3.1 Introduction

Polyhydroxyalkanoates (PHAs) are a class of optically-active organic polyoxoesters which can be composed of over 150 known hydroxyalkanoic acid monomers (Anderson & Dawes, 1990; Rehm, 2007) (Figure 3). PHAs encompass a class of biopolymers which are both biocompatible and biodegradable (Chen & Wu, 2005; D. Jendrossek, Schirmer, & Schlegel, 1996). PHAs are categorized based on the size of their hydroxyl fatty acid monomers into short chain length (SCL) PHAs, medium chain length (MCL) PHAs, or long chain length (LCL) PHAs, each with monomers consisting of 3-5, 6-14, and >14 carbon atoms, respectively.

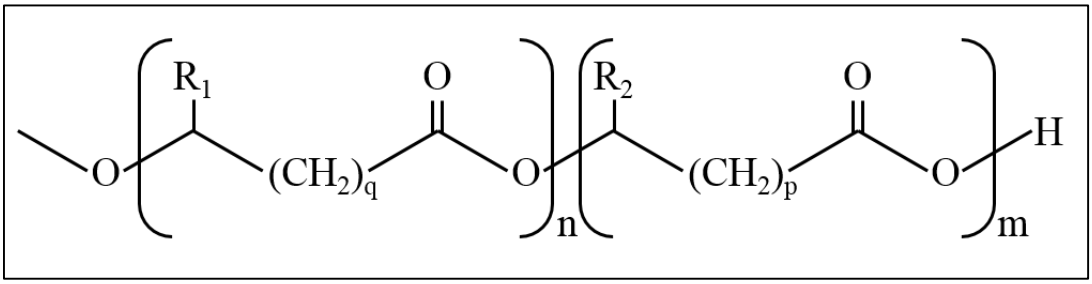


Figure 3: Chemical structure of polyhydroxyalkanoate copolymer. R1 and R2 represent alkyl groups from each respective monomer type. For short chain length, these groups will contain 0-2 carbons, medium chain length will contain 0-11 carbons in length, and large chain length will always contain a minimum of 8 carbons. The number of main chain methyl groups (p and q) can vary from 1 to 4. And the total number of monomers (N_{tot} = n + m) varies from 100 to 30,000 (Rai, Keshavarz, Roether, Boccaccini, & Roy, 2011).

The potential for customization is what makes PHAs such an attractive class of biomaterials because like synthetic plastics, they too can be tailor-made to accommodate specific utilization. Organisms have been found which can produce copolymers containing multiple chain length monomer units (Doi, Kitamura, & Abe, 1995; A. K. Singh & N. Mallick, 2009). Because PHAs can be composed of over 150 known monomers and form copolymers, this class of polymers can exhibit a diverse range of material properties rivaling those of synthetic plastics (*see* Table 3). PHAs are being applied in industry to produce biodegradable water-resistant surfaces, tissue scaffolding, medical devices, mulch films, controlled

pesticide delivery, nanocomposite materials, and various other plastic packaging and consumables (Philip et al., 2007).

Poly(hydroxybutyrate) (PHB) is the most commonly observed PHA. PHB is commonly compared to polypropylene because they possess similar melting temperatures, glass-transition temperatures, and degrees of crystallinity (Table 3). However, the chemical properties of the two differ significantly. PHB possesses less resistance to solvents than polypropylene, but exhibits a higher natural resistance to photodegradation (Holmes, 1985).

Table 3: Material properties of various synthetic plastics (PE, PET, PP, PS, PVC), bio-based plastics (PLA), and bioplastics (PHAs).

Polymer	T_m [°C]	T_g [°C]	E [GPa]	σ [MPa]	ε_b [%]
HDPE[1]	130 – 137		1.0 – 1.1	22 – 31	10 – 1200
LDPE[1]	98 – 115	(-25)	0.17 – 0.28	8.3 – 31	100 – 650
PET[1]	212 – 265	68 – 80	2.8 – 4.1	48 – 72	30 – 300
PP[1]	160 – 175	(-20)	1.1 – 1.6	31 – 41	100 – 600
PS[1]		74 – 105	2.3 – 3.3	36 – 52	1.2 – 2.5
PVC[1]		75 – 105	2.4 – 4.1	41 – 45	40 – 80
PLA[4]	100 – 180	55	3.0	50 – 70	4
P(3HP)[3]	78	(-18)	2.2 – 3.6	32 – 34	490 – 500
P(3HB)[2]	162 – 181	(-4.0) – 18	1.2 – 4.0	8 – 40	0.8 – 8.0
P(4HB)[3]	61	(-47)	0.12 – 0.24	12 – 15	650 – 740
P(3HB-*co*-3HV)[2]	64 – 172	(-13) - 13	0.082 – 8.7	1.8 – 50	0.17 – (>1200)
P(3HB-*co*-4HB)[2]	49 – 169	(-48) – (-2.0)	0.024 – 1.2	10 – 104	11 – 1300
P(3HB-*co*-3H4MV)[2]	126 – 162	(-2.0)		17 – 25	19 – 440
P(3HB-*co*-3HHx)[2]	96 – 142	(-2.0) – 0.0	0.14 – 0.99	4.5 – 26	3.0 – 850

[1](Harper & Baker, 2000), [2](Laycock, Halley, Pratt, Werker, & Lant, 2014), [3](Tripathi, Wu, Meng, Chen, & Chen, 2013), [4](Södergård & Stolt, 2002)
T_m – melting temperature, T_g – glass transition temperature, E – Young's modulus, σ – tensile strength, ε_b – elongation at break

2.3.2 Polyhydroxyalkanoate Metabolism

2.3.2.1 Introduction

PHAs are the only known polyesters existing in living organisms besides the water-soluble poly(malic acid) which occurs in lower level eukaryotes and the water-insoluble polyesters suberin and cutin which occur in plants (Steinbüchel & Hein, 2001). PHAs are produced as water-insoluble inclusions

known as *granules* in the cytoplasm of many microorganisms. These granules are produced as a form of carbon storage during times of carbon surplus and nutrient deficiency, although this is not a requirement in all PHA producers.

Aside from the source and availability of carbon, the degree of PHA accumulation can also be dependent on the relative levels of many elements in the producer's growth medium, such as oxygen, nitrogen, sulfur, phosphate, iron, magnesium, and potassium (Kessler & Witholt, 2001). Ionic strength can also impact PHA accumulation performance significantly; however, cultivation temperature has shown to have no notable effect on PHA accumulation (Grothe, Moo-Young, & Chisti, 1999).

2.3.2.2 Polyhydroxyalkanoate Biosynthesis

PHA biosynthesis can take many routes. For example, in polyhydroxybutyrate biosynthesis the upstream precursors to the hydroxybutyryl-CoA monomers are formed from the acetylation and reduction of acetyl-CoA by β-ketothiolase (PhaA) and acetoacetyl-CoA reductase (PhaB) respectively. An example metabolic pathway for the production of PHB is shown in Figure 4. Alternatively, MCL PHAs are typically formed utilizing fatty acid β-oxidation or *de novo* synthesis intermediates (Rehm, 2007). The single committed step in this pathway is the polymerization of the hydroxyacyl-CoA thioester monomers by PHA synthase (PhaC) (Rehm, 2007; Steinbüchel & Lütke-Eversloh, 2003). PHA production also requires non-catalytic proteins for stabilization of the intracellular granule (*phasins*). Excluding the phasin genes, the primary PHA biosynthesis genes are coded for within a single operon (Rehm, 2003).

2.3.2.2.1 PHA Synthase

PHA synthases have the most direct involvement in PHA production as they are solely responsible for PHA polymerization. Sequence alignments of the primary structures of PHA synthases show they are homologous to lipases and α/β hydrolase protein family members (Jia, Kappock, et al., 2000). PHA synthases are divided into four major categories (types I – IV) based on primary structure, subunit composition, and substrate specificity (Rehm, 2007).

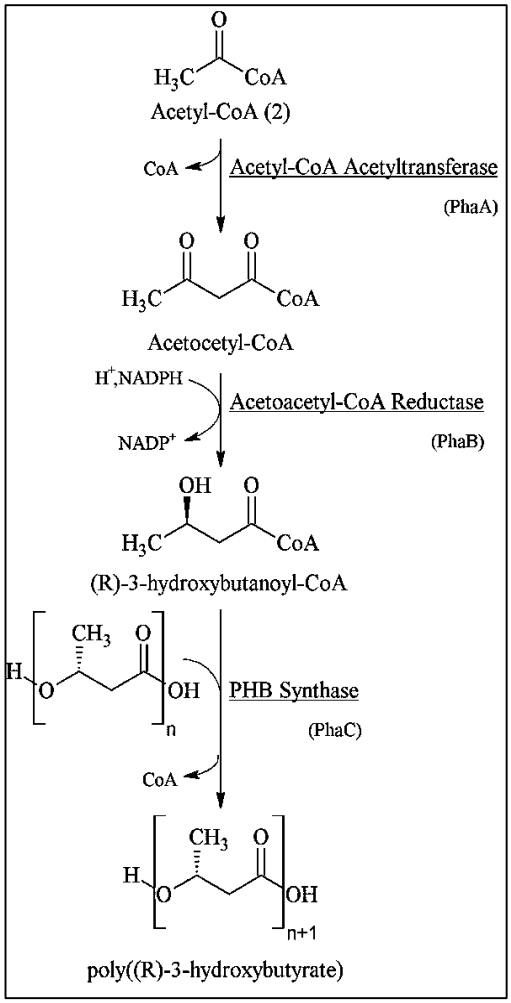


Figure 4: Poly(hydroxybutyrate) biosynthesis pathway.

Type I PHA synthases are comprised of a single enzymatic subunit (PhaC) which is typically 61-73 kDa in size (Rehm, 2007). All type I PHA synthases are homologous to the polyester synthase found in *Ralstonia eutropha*. These synthases catalyze the polymerization of hydroxy fatty acid monomers with 3-14 carbon atoms, thereby producing SCL to MCL polymer and copolymers (Nomura & Taguchi, 2007).

All type II PHA synthases are homologous to the synthase of *Pseudomonas aeruginosa*. Type II PHA synthases are comparable in size to type I PHA synthases. Type II synthases catalyze only coenzyme A thioester substrates with hydroxyl fatty acid components of 3-14 carbons atoms. While both SCL and MCL sized monomers can be incorporated, this type of enzyme has a strong selectivity to polymerize MCL

CoA thioesters (Nomura & Taguchi, 2007). Type II PHA synthases therefore generate predominantly MCL PHAs.

Type III PHA synthases are the first of the hetero-multimeric sub-classifications of polyester synthases. These type III synthases consist of two main subunits of relatively equal mass (~40 kDa ea.): the catalytic subunit 'PhaC' and the non-enzymatic subunit 'PhaE' (Rehm, 2007). The PhaC subunit shows 21-28% primary structure similarity to the type I and type II PHA synthases, while the PhaE subunit shows no similarity (Rehm, 2003). All type III PHA synthases are homologous to that found in *Allochromatium vinosum* (previously *Chromatium vinosum*). These synthases catalyze the polymerization of hydroxy fatty acid monomers with 3-5 carbon atoms, thereby producing SCL PHAs.

Type IV PHA synthases are the second of the hetero-multimeric sub-classifications of polyester synthases. Type IV synthases are composed of two subunits: the catalytic subunit 'PhaC' of similar size to type III PhaC subunits and 'PhaR' a 20 kDa non-enzymatic subunit (Rehm, 2003). All type IV synthases are homologous to the polyester synthase found in *Bacillus megaterium*.

Phylogenetically-related organisms commonly express the same type of PHA synthases. For example, *Pseudomonas* predominantly express type II and the *Bacillus* genus exclusively possesses type IV PHA synthases (Rehm, 2003).

It is believed that because the PHA synthase structure is so similar to that of lipases, it may also possess a similar catalytic mechanism. Lipases contain a catalytic triad of serine (`S`, Ser), histidine (`H`, His), and aspartate amino (`D`, Asp) acid residues. The catalytic serine is enclosed within a conserved region glycine (`G`, Gly) and alanine (`A`, Ala) residues. This conserved lipase box is approximately 6 amino acids in length, `GxSxG[G|A]` (Marchler-Bauer et al., 2015). In lipases, the histidine activates the serine for nucleophilic attack and subsequent covalent catalysis. PHA synthases contain a lipase-box variant, or "lipase-like box" which replaces the catalytic serine with cysteine (`C`, Cys). It is believed that for PHA synthases, the histidine forms a catalytic diad with cysteine, and activates it for a nucleophilic attack. While

it has proven necessary, the exact function of the aspartate residue is still open for debate (Jia, Kappock, et al., 2000). PHA polymerization requires dimers of PHA synthase (Jia, Kappock, et al., 2000). This mechanism is outlined in Figure 5.

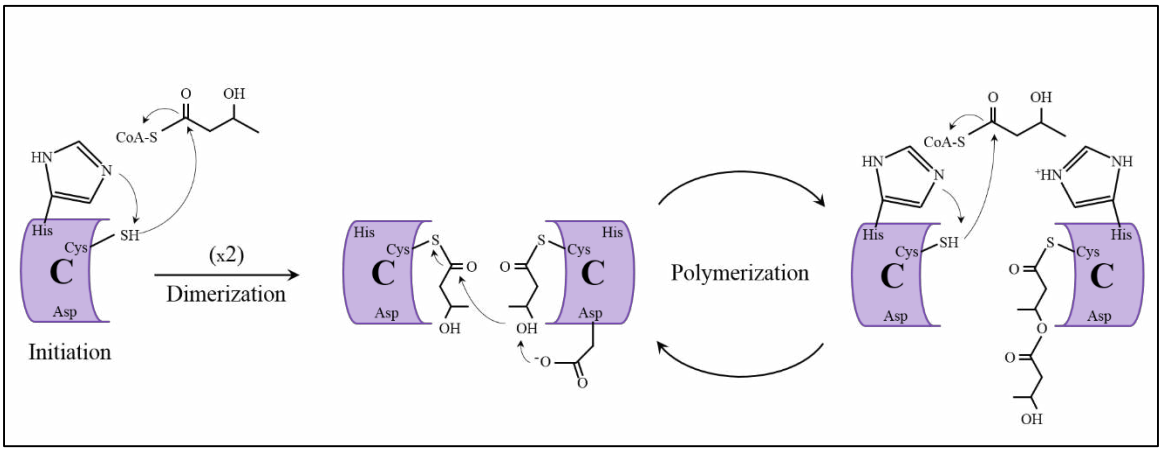


Figure 5: Proposed polymerization mechanism of PHA synthase ('PhaC') (Jia, Kappock, Frick, Sinskey, & Stubbe, 2000). Note that the PHA synthase 'C' structure may be bound to an additional subunits for types III and IV synthases.

2.3.2.3 Polyhydroxyalkanoate Biodegradation

Because PHAs function as energy storage within an organism, they must also be capable of degradation by said organism in times of energy deficiency. PHAs can be degraded through multiple enzymatic mechanisms. *PHA hydrolases* hydrolyze the primary bonds of PHA, degrading the polymer into shorter chain lengths. *PHA depolymerases* cleave the primary bonds of PHA and create hydroxyl fatty acid dimer products, which can subsequently be converted into the monomer substituents (Figure 6).

Both intracellular and extracellular PHA depolymerases have been observed. Extracellular PHA depolymerases allow for the metabolism of extracellular PHA from the surrounding lysed cells (Dieter Jendrossek, 2007). Organisms incapable of producing PHAs may still naturally express extracellular depolymerases (Dieter Jendrossek, 2007).

Once the hydroxyalkanoate monomers have been formed, they can enter either the pathway for ketone body synthesis/degradation or fatty acid degradation.

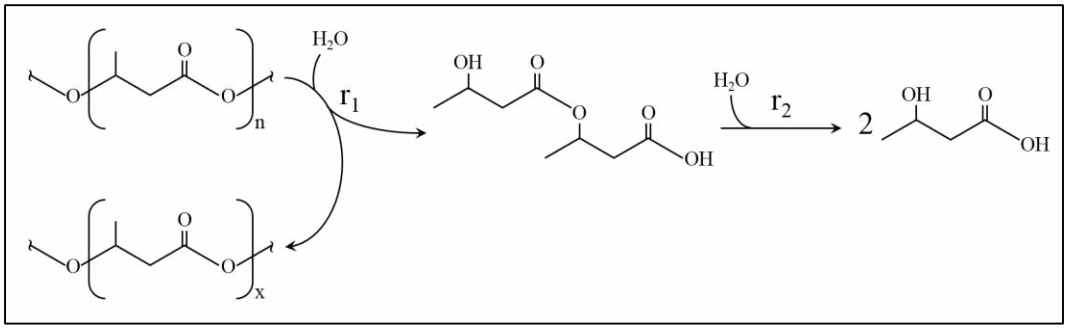


Figure 6: Poly(hydroxybutanoate) depolymerization. Reaction one, catalyzed by PHB depolymerase, depicts the hydrolysis of PHB to form 3-(3-Hydroxybutanoyloxy)butanoate. Reaction two, catalyzed by hydroxybutyrate-dimer hydrolase, shows the hydrolysis of the hydroxybutanoate dimer into its monomer constituents.

2.3.2.4 Polyhydroxyalkanoate Accumulation

PHAs are formed as water insoluble inclusions within a cell. The inclusion is stabilized by phasin proteins and phospholipids. PHA synthases, depolymerases, and regulator proteins are all located on the surface of the granule (Figure 7). There are two proposed mechanisms of granule formation: micelle formation and budding formation.

After initiation of the hydrophobic PHA polymers, the polymer remains covalently bound to the hydrophilic PHA synthase as depicted in Figure 5. This complex acts as a single amphipathic molecule and allows for traditional micelle formation to occur (Figure 8, top). This formation has been observed in reactions containing only PHA synthase and the hydroxyacyl-CoA monomers (Gerngross & Martin, 1995). Micelle formation represents the formation observed *in vitro* in the absence of phospholipids and may occur *in vivo* as well.

In the budding model (Figure 8, bottom), it is proposed that PHA synthase molecules are localized around/within the cytoplasmic membrane of the cell and as the polymer aggregate increases in size, the insoluble granule buds from the membrane. This model is supported by the work of Peters *et al.*, who observed that PHA synthase localizes at cell poles (2005).

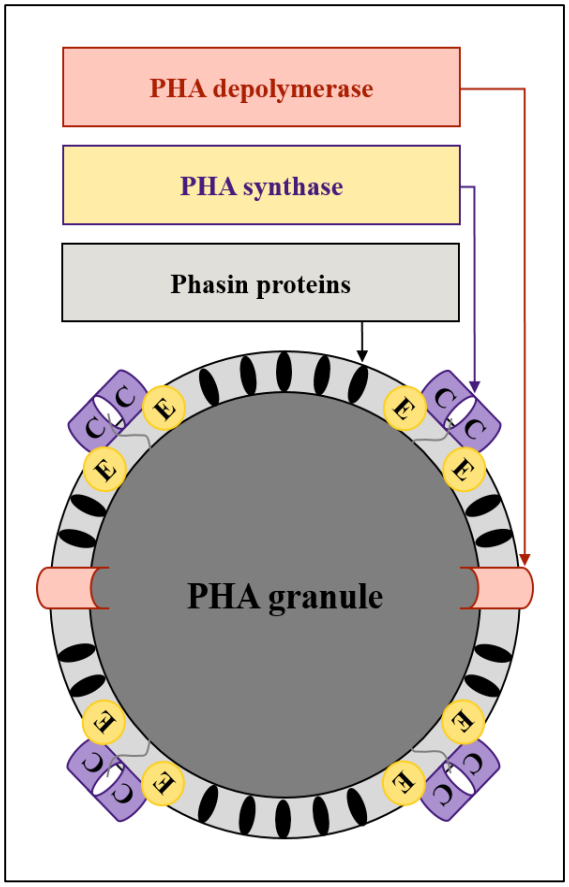


Figure 7: Intracellular PHA granule schematic with the type III PHA synthase depicted. 50-500nm diameter. Exterior area of granule is coated in phasin proteins and phospholipids. PHA synthase polymerizes PHA inwards, towards the bulk of the granule.

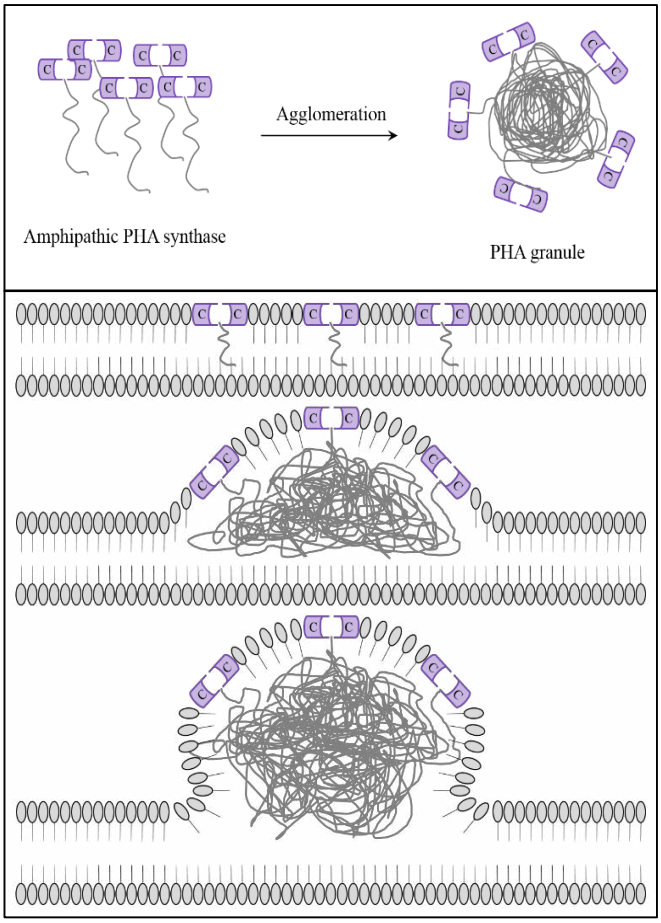


Figure 8: (TOP) Micelle formation of PHA granules in the absence of phospholipids. (BOTTOM) Budding formation of PHA granules from within the cytoplasmic membrane (Tian, Sinskey, & Stubbe, 2005).

In contrast, the observations of Tian and coworkers document that PHA granule formation only occur in the center of cells and away from the membrane (2005). They also record observations of dark structures near the PHA granule localization sites; these structures are believed to be mediation elements. The work of Tian *et al.* suggests that a third mechanism of PHA granule formation may be required to accurately describe this process.

2.3.3 Polyhydroxyalkanoate Production Limitations

PHA production research is predominantly focused on the use of recombinant or native heterotrophic microorganisms, such as *Bacillus megaterium*, *Cupriavidus necator*, *Escherichia coli*, *Pseudomonas*

aeruginosa, *Ralstonia eutropha*, or *Saccharomyces cerevisiae* (Agnew, Stevermer, Youngquist, & Pfleger, 2012; Breuer, Terentiev, Kunze, & Babel, 2002; Ienczak et al., 2013; Kahar, Tsuge, Taguchi, & Doi, 2004; RamKumar Pandian et al., 2010; A. Singh & N. Mallick, 2009). However these fermentation methods typically involve production costs substantial enough to limit their applicability in industry (Bengtsson et al., 2010). For example, industrial scale processes are currently producing PHAs at a cost of \$2.50 kg_{PHA}^{-1} ("DaniMer scaling up production of bio-based PHA resins," 2011).

The carbon source is a major contributor to the overall PHA production costs and many methods are being investigated to improve the cost-effectiveness of processes utilizing chemoheterotrophic organisms (Choi & Lee, 1999; Ducat et al., 2011; Ienczak et al., 2013). The use of agricultural/industrial wastes and other low-cost carbon sources improves carbohydrate feedstock costs, but unfortunately either compromises productivity or demands unit operation redesign (e.g., recycle), due to low specific yields (Ienczak et al., 2013).

To form a frame of reference for the impact of the carbon source costs, one can use the multi-stage bioreactor cascade described in Atlić *et al.*, which utilizes *Cupriavidus necator* to obtain a specific productivity comparable to what is required in industry (1.85 $g\,L^{-1}\,day^{-1}$) (2011). This process consumes glucose as its carbon source at a rate of 0.057 $kg\,h^{-1}$ and produces PHB at a rate of 0.019 $kg\,h^{-1}$, bringing the rate of glucose-to-PHB to 3.0 $kg_{glu}\,kg_{PHB}^{-1}$. Applying the industrial cost of glucose, \$0.493 kg_{glu}^{-1} (Salehizadeh & Van Loosdrecht, 2004), the contribution of the carbon source cost is \$1.5 kg_{PHB}^{-1}.

2.4 Cyanobacteria

2.4.1 Introduction

Over 2.8 billion years old, cyanobacteria are a diverse group of prokaryotic primary producers flourishing in both limnic (freshwater) and marine (saline) environments yet can be found in nearly every conceivable habitat on Earth (Abed, Dobretsov, & Sudesh, 2009; Buick, 2008). These robust organisms also exhibit high levels of biodiversity relative to the other Prokaryote sub-classifications (Garcia-Pichel,

Nübel, Muyzer, & Kühl, 1999). Although all cyanobacteria carry out oxygenic photosynthesis, some have proven capable of photomixotrophic and even heterotrophic growth (Rippka, Deruelles, Waterbury, Herdman, & Stainer, 1979; Stal & Moezelaar, 1997).

2.4.2 Morphology

The high level of biodiversity within Cyanobacteria is exemplified by the morphologies observed within this phylum. Cyanobacteria are generally considered to have a Gram-negative cell envelope, however many cyanobacteria exhibit various features of Gram-positive membranes (Hoiczyk & Hansel, 2000).

Cyanobacteria are currently categorized by their morphology, although genetic classification is becoming increasingly useful. They are categorized into five sections, or types, based on their cellularity and organization (Schirrmeister, Antonelli, & Bagheri, 2011). The simplest morphology is unicellular growth with random dispersion of cell aggregation, such as in the model cyanobacterium – *Synechocystis* sp. PCC6803. This growth more closely resembles typical bacterial growth. In some cases, as cyanobacteria cells divide, they remain enclosed in a gelatinous or mucilaginous sheath. This leads to the formation of filaments, which can be either linear or branched (Flores & Herrero, 2010). Furthermore, some of these filamentous cyanobacteria possess the capability to differentiate. For example, *Nostoc* possesses the ability to form nitrogen-fixing heterocysts (C.-C. Zhang, Laurent, Sakr, Peng, & Bédu, 2006). Heterocyst formation can be induced by the depravation of combined nitrogen. Additionally, akinetes or cells differentiated for storage and survival can also be formed (Flores & Herrero, 2010). Cyanobacterial colonies can exhibit morphologies ranging from standard unicellular growth to sets of highly organized multicellular filaments forming macro-level aggregates.

2.4.3 Photosynthesis

The primary enzyme responsible for all photosynthetic carbon-fixation, ribulose bisphosphate carboxylase-oxygenase (RuBisCo), possesses low selectivity for CO_2 versus O_2. It is believed this is due

to the lack of atmospheric O_2, and thereby evolutionary pressures, in the environment during the development stages of this metabolic pathway. Approximately 350 million years ago levels of atmospheric O_2 increased substantially, triggering the arise of CO_2 concentrating mechanisms (CCM) (Price, Sültemeyer, Klughammer, Ludwig, & Badger, 1998).

CCMs implement active transport to control the concentrations of inorganic carbon (mainly CO_2 and HCO_3^-) within the cell. For these mechanisms, HCO_3^- accumulates within the cell in concentrations in excess of 1000x equilibrium concentrations (Price et al., 1998). RuBisCo is encapsulated within a carboxysome allowing for increased local concentration of CO_2 surrounding the carboxylase which improves functionality. Both CCMs and RuBisCo are highly important factors in the overall inorganic carbon fixation rate of a photoautotroph.

Similar to plant photosynthesis, cyanobacterial photosynthesis utilizes water as the electron donor and chlorophyll *a* in combination with phycobilisomes to harvest light energy (Stal & Moezelaar, 1997). Phycobilisomes are phycobiliprotein complexes which act as light-harvesting antennae anchored to the thylakoid membrane within the chloroplast, directing energy to chlorophyll *a* (Robert MacColl, 1998). Cyanobacteria utilize two types of biliproteins: *phycocyanin* (blue pigment) and phycoerythrin (red pigment). Cyanobacteria have been deemed "blue-green algae" due to this abundance of phycocyanin. The electron flow generated from this light energy is used to drive inorganic carbon transport.

Naturally, Cyanobacteria exhibit the ability to regulate their local light conditions via movement, implementing both positive and negative phototaxis, typically by type IV pili (Bhaya, 2004). In order to do this, the cyanobacteria must possess a method of light-energy signal transduction.

Phytochromes are light harvesting complexes which function as light sensory complexes in plants. These complexes are also found in cyanobacteria (Hughes et al., 1997). They allow the cell to essentially sense where it is in relation to a light source. A remarkable characteristic of the phytochromes is their reversible photochromism. The inactive conformation (Pr) absorbs a photon at one periodicity ($\lambda = 665$

nm) which causes a shift to the active signaling conformation (Pfr). Pfr absorbs a photon at a different periodicity (λ = 730 nm) which results in the reversion back to the original conformation.

2.4.4 Cyanobacteria in Industry

The ability of cyanobacteria to fix atmospheric carbon through photosynthesis in order to produce a value-added products has been well documented and makes them attractive candidates in bioprocesses (Abed et al., 2009; Ducat et al., 2011; Simmons, Andrianasolo, McPhail, Flatt, & Gerwick, 2005; L. T. Tan, 2007). These organisms produce a variety of bioactive metabolites and commodity 'bio-products' such as isoprene, biofuels, and biopolymers at nearly carbon-neutral conditions (Ducat et al., 2011). These metabolites exhibit a variety of properties from anti(-viral, -fungal, -bacterial) to immunosuppressive and even anti-cancer properties (Abed et al., 2009; Simmons et al., 2005).

Cyanobacteria provide a promising platform for lessening PHA production costs due to expensive carbon sources because they are the sole prokaryotic native producers of PHA via oxygenic photosynthesis (Asada, Miyake, Miyake, Kurane, & Tokiwa, 1999; Sharma, Kumar Singh, Panda, & Mallick, 2007). A PHA production process utilizing cyanobacteria would essentially be creating biodegradable bioplastics while sequestering atmospheric carbon dioxide as a cost-free carbon source.

2.5 Polyhydroxyalkanoate Production in Cyanobacteria

The work of Hai *et al.* suggests that the PHA biosynthesis pathway within cyanobacteria occurs in a widespread and general fashion, and thus far, only type III PHA synthases have been observed in this phylum (2001).

Unfortunately, typical PHA accumulation observed in cyanobacteria under photoautotrophic growth is less than 10% dry cell weight (DCW), if detected at all (Bhati, Samantaray, Sharma, & Mallick, 2010; Ducat et al., 2011), an amount far lower than the yields observed in heterotrophic high-density processes, which can approach 87% DCW (Ienczak et al., 2013). However, Nishioka et al. demonstrated wild-type

cyanobacteria strains are capable of exhibiting significantly higher yields, 55% DCW PHA during photoautotrophic cultivation (2001).

Functional diversity, including this large deviation in PHA accumulation, is not a rarity within *Cyanobacteria* (Garcia-Pichel et al., 1999). With *Cyanobacteria*'s rich biodiversity, it is likely that there are undiscovered high-yielding cyanobacteria strains with PHA production capabilities rivaling those of the costly heterotrophs.

2.6 Synopsis of Relevant Molecular Biology

2.6.1 DNA

Deoxyribonucleic acid (DNA) is arguably the single most important biomolecule (Figure 9). It belongs to the nucleic acid class of biopolymers, and its unique structure allows for the storage and transfer of information. Single-stranded DNA (ssDNA) is a polymer of deoxyribonucleotides linked together by primary phosphodiester bonds. This structure acts as a backbone, where the nitrogenous base of each deoxyribonucleotide can vary. The nitrogenous bases can allow for antiparallel linking of ssDNA molecules via hydrogen bonding. However, the stability of this duplexed double-stranded DNA (dsDNA) is highly dependent upon the configuration of the polymers. The breaking of a dsDNA duplex at the hydrogen bonds to form two ssDNA molecules is known as *denaturation*. For dsDNA duplexes, the term *melting temperature* (T_m) describes the temperature at which half of the dsDNA denatures. Each nitrogenous base interacts with its neighboring base of the opposite strand, known as *base pairing*, in a way which can have either a stabilizing effect or a destabilizing effect. As expected, the more stable a duplex is, the more energy it will require to denature and the higher its T_m will be.

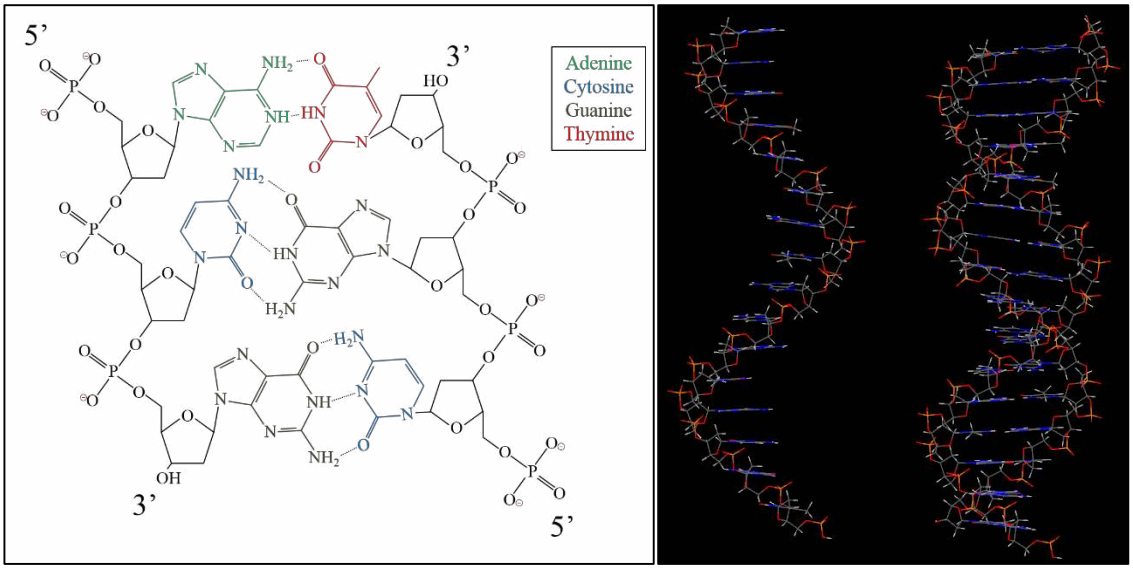


Figure 9: Chemical structure of Watson-Crick complementary DNA base pairs and ssDNA and B-form dsDNA duplex.

DNA containing information is primarily composed of four nitrogenous bases: adenine (A), cytosine (C), guanine (G), and thymine (T). The most stable base pairs of coding DNA are referred to as Watson-Crick base pairs and are A–T/T–A and C–G/G–C. These base pairs are observed so frequently in coding DNA that DNA is said to exhibit complementarity. For example, in complementary dsDNA if one position of strand one is A, then the respective nucleobase of strand two (the complement of A) would be T. In other words, A complements T and C complements G and vice versa. This means if the configuration/sequence of one strand of DNA in a complementary dsDNA duplex is known, then the opposite strand is also known by definition.

Any base pair variant other than Watson-Crick pairs (e.g., A–A or G–T) are known as *mismatches*. Mismatches generally contribute destabilizing effects, however a few have been observed which remain relatively stabilizing such as G–T (Allawi & SantaLucia, 1997; Hatim T. Allawi & John SantaLucia, 1998a, 1998c; H. T. Allawi & J. SantaLucia, 1998; Peyret, Seneviratne, Allawi, & SantaLucia, 1999).

DNA polymerases are key enzymes for the replication of DNA. These polymerases generate complementary dsDNA from a ssDNA template. The mechanism of this process is highly complex,

however for the sake of brevity, only essential concepts are discussed here. DNA polymerases require a short complementary single stranded nucleic acid sequence, known as a *primer*, bound to the template. The polymerase then consumes deoxynucleoside triphosphates to append new complementary DNA nucleotides to the primer beginning at the 3' ("three-prime") end. The product of this synthesis is a complementary dsDNA duplex with the primer incorporated into the helix. Application of this synthesis to both strands of a genome will result in two copies of said genome, allowing an organism to divide with both progeny containing genomic replicates.

2.6.2 Gene Expression

Living organisms utilize DNA as a central database, where discrete packets of information (*genes*) encoding proteins and other functional nucleic acids can be stored and passed along to progeny. The entire collection of all of the discrete packets of information is defined as an organism's *genome*. Gene expression is the process in which the information stored within a gene is used to synthesize the gene product.

2.6.2.1 Protein Expression and Degenerate Codons

In the case of proteins, first a complementary strand of ribonucleic acid (RNA) is synthesized using the gene as a template – the process known as *transcription*. RNA incorporates the modified nucleobase uracil (U) in lieu of thymine. At this point the RNA may undergo posttranscriptional modifications and when complete, this entity is referred to as mature messenger RNA (mRNA). Next, ribosomes utilize aminoacyl transcription RNA to parse the gene three nucleotides at a time – the process known as *translation*. With each parsing instance of three nucleotides, or *codon*, the ribosome appends a new amino acid onto the growing polypeptide until a termination codon is reached (Figure 10).

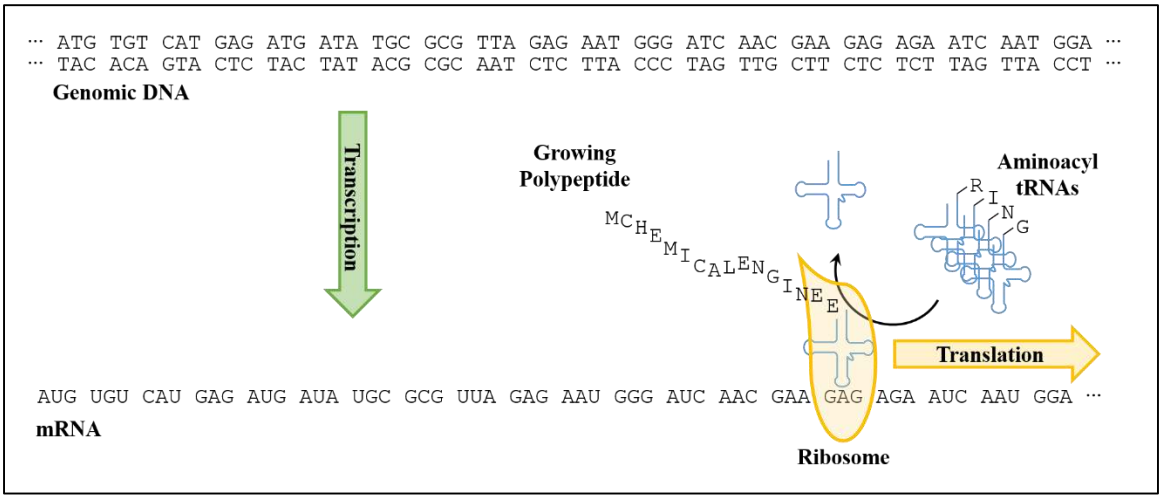


Figure 10: Simple schematic of gene expression using the standard genetic code.

Because there are three positions within a codon and four possible permutations of nucleobases, there are $4^3 = 64$ syntactic entities which need to be recognized by ribosomes. The relationship of codons and their amino acid translations are recorded as *genetic codes* or *translation tables*. And because there are only 20 naturally occurring amino acids coded by 64 codons, multiple codons may code for the same amino acid. Degenerate nucleotide nomenclature is a result of this phenomena to more easily express multiple permutations of a single DNA sequence. For example, under the standard genetic code (Appendix A.5) tyrosine (Tyr) can be coded for by `TAT` and `TAC` (`TA[T|C]`), which can be expressed more easily as or `TAY`.

Organisms parse genetic information differently. There have been more than 27 proposed translation tables since the discovery of the standard code (Sayers et al., 2009). These codes usually differ only slightly from the standard code.

2.7 Polymerase Chain Reaction

2.7.1 Introduction

The polymerase chain reaction (PCR) is a versatile utility which allows for the *in vitro* amplification of nucleic acids and underlies almost all of modern molecular biology (Sambrook & Russell, 2001). The speed, robustness, and flexibility of this molecular biotechnology are what rightfully grant its vast

popularity. There are numerous variations of PCR and a comprehensive list of these variants can be found in A.1. Some notable implementations include genetic detection, nucleic acid sequencing, whole genome assembly, and nucleic acid quantitation.

PCR requires seven essential components: (i) thermostable nucleic acid polymerase, (ii) pair of synthetic oligonucleotide primers, (iii) deoxynucleoside triphosphates, (iv) divalent cations, (v) monovalent cations, (vi) buffer to maintain pH, and (vii) a nucleic acid template. Once the seven essential components of PCR are combined, the reaction possesses everything required to synthesize new nucleic acids. The cations serve to promote dsDNA formation, divalent cations tend to dominate the contribution under standard conditions (Owczarzy, Moreira, You, Behlke, & Walder, 2008). The deoxynucleoside triphosphates provide substrate for polymerization beginning at the 3' end of each primer via the polymerase. The primers are typically used in excess to prevent this component from being the limiting reagent, however this is not always the case (confer A.1).

The central idea of PCR is the thermal cycle. First, the reaction temperature is increased to the point at which the template DNA is denatured. Next, the temperature is decreased to a point at which the primers are likely to anneal to the newly formed ssDNA template strands – known as the *annealing temperature* T_A. Finally, the temperature is raised once more to a degree at which the enzymatic activity of the polymerase is optimal to promote synthesis. Repetition of this thermal cycle will produce a short product spanning the region of the template bounded by the primers (Figure 11).

After initial short product formation, the short product will amplify exponentially and out compete other products. This results in essentially millions of copies of a single nucleic acid duplex.

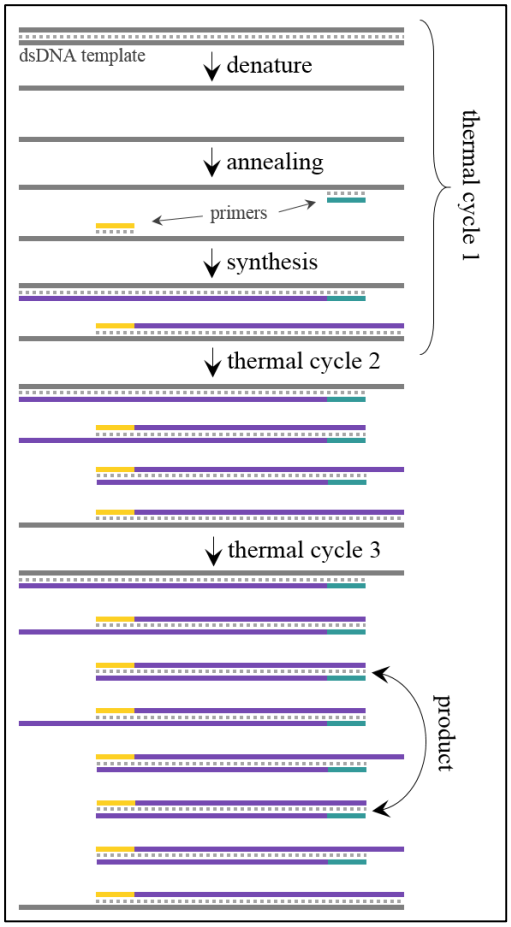


Figure 11: Graphic depiction of standard PCR initial product formation.

2.7.2 Basic Primer Design Considerations

There are many detailed reviews available on this matter, including Sambrook *et al.* (2001). However, some of the essential influencing factors will be covered here. Since PCR will amplify fragments of DNA exponentially, any undesired byproduct formation will greatly affect the product composition. Thus, specificity is key in the design of a PCR primer. In other words, the sequence which the primer complements should not occur anywhere else in the template DNA; otherwise primer extension will occur at multiple locations.

Thermodynamics also play a key role in primer design. In order for a primer to bind, it requires a certain degree of stability. This stability can be estimated by calculating the change in Gibbs free energy

(ΔG) of annealing for the primer-template duplex. This is typically performed using nearest-neighbor thermodynamics; for an in depth review on the subject read SantaLucia *et al.* (2004). This quantity is reflected in the melting temperature, length, and G | C content of the primer.

The melting temperatures are typically estimated using the nearest-neighbor thermodynamic parameters. A derivation of melting temperature with the relation to these parameters can be found in appendix section A.2. The melting temperature of the annealing region should typically be between 50°C and 70°C for a standard primer. Additionally, for a primer set, the two melting temperature values should not deviate by more than 5°C from one-another to prevent non-specific annealing of the more stable primer. While the length of a primer may vary depending on the application, it is typically held to 18-32bp. This range typically allows for proper and specific annealing within a template under standard PCR conditions.

G | C content is a quick approximation of the relative strength of the DNA duplex because as discussed earlier, the G-C base pair forms three hydrogen bonds while the A−T base pair only forms two. Thus, a DNA duplex of only G−C base pairs will require more energy to denature than an A−T DNA duplex of equivalent length. For this reason, the G | C content is typically held between 40-60%, and consecutive runs of G−C base pairs are typically held below five.

Finally, the primers should not be complementary to one another in any regions and should not be complementary to themselves. If this condition is not adhered to, the primers can anneal to one another and during synthesis the overhanging fragments will be complemented, forming short, non-functional primer-dimers. This can lead to numerous issues including primer consumption, weak amplification, and non-specific binding not previously possible.

Extra precaution should be taken in the design of the 3' end of the oligonucleotide primer. Since polymerase extension is facilitated on this end, it must bind completely and specifically. In other words, a long primer with a 5' end unbound to template DNA may still successfully synthesize a new strand if the

3’ end forms a stable hybridization. In fact, this phenomena is actually exploited in overlap extension PCR and polymerase chain assembly.

2.7.3 Optimization considerations

The ideal end product of a standard PCR is a desired number of replicates of a singular nucleic acid product. In practice, this can be a difficult result to obtain. As stated earlier, the key to successful PCR amplification is specificity. Some applications of PCR require the design of primers overlapping a specific portion of the template DNA. In these cases specificity may not be readily tailored through primer sequence alone.

Once primer design is complete, the specificity of a given reaction can be tuned through the annealing temperature. As the annealing temperature increases, the stability of a given primer-template hybridization decreases. If non-specific binding is occurring and forming a less stable hybrid than the desired primer-template duplex, then raising the annealing temperature may prevent the undesired annealing while still allowing desired annealing to occur. For this reason, the annealing temperature used in standard PCR is near the lowest melting temperature in the primer set, typically 5 °C less than $T_{m,\,low}$.

The secondary optimization parameter for standard PCR is the divalent cation concentration. Standard reactions utilize magnesium ions for this purpose. Cations tend to stabilize the dsDNA duplex formation by interacting with the negatively charged phosphate backbones of the nucleic acids. The specificity of a PCR amplification is inversely proportional to the magnesium concentration. This concentration is typically held between 1.0 mM and 3.0 mM.

The final standard PCR optimization parameter is the number of thermal cycles the reaction is subjected to. For standard PCR, increasing the number of thermal cycles increases the number of products according to: $X = (1+\eta)^n$, where X is the yield of PCR product copies, η is the efficiency of amplification per cycle, and n is the number of cycles (Booth et al., 2010). However, this parameter exhibits diminishing

returns typically after 35 cycles (Sambrook & Russell, 2001). After this point, the reaction becomes limited either due to polymerase thermal deactivation or primer availability.

2.8 Consensus Degenerate PCR Probe Design

In many applications it can be beneficial to design PCR primers capable of amplifying homologous sequences (Figure 12). The central idea in consensus primer design is that the primers can be designed within a common region shared by the target sequences so that the annealing ability of the primers remains relatively unchanged across multiple templates.

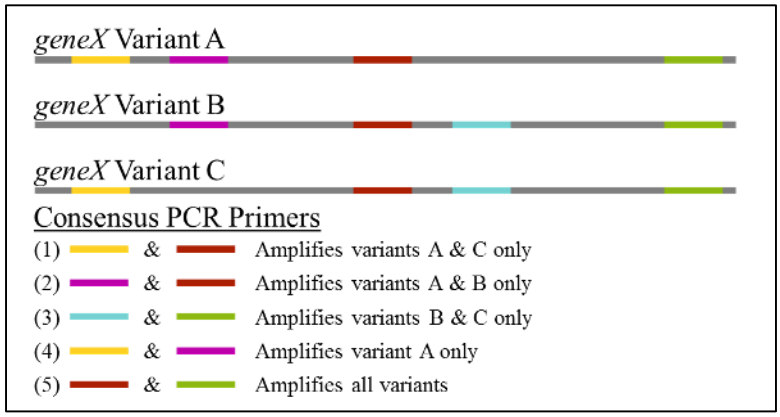


Figure 12: Example of consensus primer applications. Colored regions depicted are arbitrary consensus primer loci in the arbitrary gene *geneX*. Here the primer sets 1-4 could be used to discriminate groups of *geneX* variants. Primer set 5 could be used to detect the presence of any form of *geneX* and could potentially lead to isolation and characterization of the currently unknown *geneX* Variant D.

Multiple sequence alignment (MSA) algorithms analyze protein or nucleic acid homologous sequences and determine an optimal form of alignment of those sequences. MSAs allow for the visualization of regions conserved in a set of biological sequences. Alignments can grant insight on evolutionary relationships of the genes within the alignment and potential functional motifs of the sequence family. MSAs are an invaluable tool in the design of PCR primers for detection across homologous nucleic acid sequences.

In many cases a simple consensus primer set is not a viable option. Nucleotide variations are too numerous between the sequences for a consensus primer to effectively bind the templates in all sequences. For instance, if it is desired to design a four nucleotide primer across the following sequences: AATG, AAAG, AATG, and AATG. Then the consensus primer would be AATG, which may fail to amplify in the sequence AAAG. To account for this polymorphism one could use a small set of similar primers which represent some or all of the sequence permutations, known as a *degenerate primer*. The degenerate primer in the example above would be AA[A|T]G, also denoted as AAWG, and would possess a degeneracy of two (*see* Appendix A.6). Degenerate primers are a pragmatic approach to amplifying target sequences which may be unknown or are highly variable in nature. However, due to the inherent nature of incorporating multiple permutations of a sequence, there is a reduction in overall primer specificity as the degeneracy increases. The extreme case for the example above is NNNN, which has a degeneracy of 256 and is capable of annealing anywhere within any target sequence. Additionally, the initial concentration of primers in standard PCR is typically held constant. Because the overall concentration does not change, the initial concentration of the viable primer permutation is reduced under stringent conditions. In other words, if a primer has a degeneracy of two, the initial concentration of each permutation is reduced by half.

A consensus degenerate primer set allows for a single reaction condition across all template sequences and can greatly increase the likelihood of a successful amplification of an unknown targeted sequence. Currently, there are two primary methods of designing primers across multiple sequences.

2.8.1 Direct Alignment of Target DNA Sequences

The most direct approach is to simply subject the template sequences of DNA to multiple sequence alignment algorithms such as Clustal (Larkin et al., 2007). This is a readily available method and is best suited for cases when:

i) the target DNA is not a coding sequence
ii) there are many sequences available (large sample size, *n*)
iii) all of the sequence information for the experimental samples are known.

Directly aligning DNA sequences to produce an MSA is simple and rapid. Analysis of this MSA for primer design can yield low degeneracy primers based off of the consensus sequence. Unfortunately, many DNA MSA algorithms do not assume that the input DNA sequences are protein coding. Thus, application of these algorithms directly on the DNA sequences may result in alignments which lose nucleotide-grouping/codon information. Additionally, because DNA has a relatively low number of potential permutations (A, C, G, T, and deletion or Δ), a MSA of DNA may suggest conservation of a residue where it may not actually be present when provided with a larger sample size.

For example, under the equiprobability assumption for a random coordinate within an alignment of n sequences, the probability that the each residue element within that alignment column will match the residue of the first sequence due to random chance alone can be described by the equation: $P = 5^{(1-n)}$. Such an event would represent a sort of Type I error where the conservation would be present, but due only to random chance and not from evolutionary pressures. This sort of conservation is undesirable in probe design because it has no influencing factors to remain conserved. Furthermore, the use of primers based on these regions on an unknown DNA sequence may result in a loss of primer annealing ability and a false negative amplification result.

This method of consensus primer design has been implemented by applications such as PrimaClade and PriFi (Fredslund, Schauser, Madsen, Sandal, & Stougaard, 2005; Gadberry, Malcomber, Doust, & Kellogg, 2005). These applications are proficient in returning low-degeneracy consensus primers for nucleic acid MSAs; however, the methods used to align sequences risks losing vital codon grouping information since the alignments may not be designed with a focus on coding sequences. Because the codon information is lost for coding sequences in these methods, prediction of possible permutations of a given residue position is limited to the residues observed within the alignment column, unlike the known potential translations of a codon position. This can lead to a potential loss in robustness for a given primer set.

2.8.2 Protein-based Indirect Speculation of the Possible Nucleic Acid Sequences

The likelihood of a given alignment position appearing conserved due to random chance alone in a protein MSA is far less than that of a DNA MSA. Assuming equiprobability, the probability that the each residue element within that alignment column will match the residue of the first sequence due to random chance alone can be described: $P = 21^{(1-n)}$. Thus, one can be more confident that a conserved amino acid in a MSA is conserved due to evolutionary pressure than a conserved nucleic acid in a MSA.

This method of design requires the target DNA sequences be protein coding. The key theory behind this method is that one can analyze a MSA of protein sequences for possible regions of conservation then essentially guess what the nucleic acid sequence may be based off of a specific translation table.

For instance, a conserved protein region is found to be `CHES`. Using the standard translation table (Appendix A.5), one could obtain the possible codons for each amino acid. In this example, those would be `TGY`, `CAY`, `GAR`, and `WSN` for `C`, `H`, `E`, and `S` respectively. Thus, a degenerate primer for this example would be `TGYCAYGARWSN`, with a degeneracy of 128. Primers designed using only this method can result in highly degenerate primers. This method is highly robust and will likely amplify the intended target; however, due to the loss in specificity, it will also likely amplify undesired products. Therefore, amplification using primers designed through this method will likely require additional PCR product validation via DNA sequencing or nested PCR. Protein-level degenerate primer design is best suited when:

i) genes of interest are homologous globally
ii) genes of interest contain local homologous regions (conserved domains)
iii) there are few sequences available (small sample size, *n*)
iv) attempting to amplify homologues in a group organisms with high biodiversity

CODEHOP exemplifies the implementation of this methodology (Rose, Henikoff, & Henikoff, 2003). This application utilizes protein MSAs and codon frequency tables to generate moderate-to-high degeneracy primers. Because codon frequency tables are used in lieu of actual CDS information, it is quite possible that the consensus template sequence corresponding to a given primer was not represented as

accurately as possible. An additional downside to this method of design is the amount of noise produced in its output. This application typically results in numerous primer sets to sift through before the user obtains their desired primer set.

2.9 Summary

To reiterate, this chapter was intended to provide the reader with an in depth overview of the history, phenomena, theories, and vocabulary associated with this work. The applications, relevance, and influence of plastics in large-scale social and economic settings were examined. Polyhydroxyalkanoates were defined as a biodegradable class of biopolymers and a potential alternative to polyolefins and covered from the initial scope of polymers, including the degradation thereof. Next, current PHA production methods and limitations were discussed and cyanobacteria were introduced as a potential method to reduce production costs. Mechanisms of cyanobacterial PHA biosynthesis were presented to provide a basis of the contributing factors influencing PHA accumulation in cyanobacteria. Finally, a review of the information supporting the tangent work involving consensus degenerate PCR primer design was given.

CHAPTER 3. CYANOBACTERIA METHODS

3.1 Chapter Preface

This chapter details the general techniques applied throughout Objective I (section 1.4). The work contained herein is only of secondary importance to the overall scope of this dissertation; however, the inferences derived within this chapter may prove valuable to any reader interested in studying cyanobacteria.

3.2 Materials and Methods

3.2.1 Strains and Cultivation Conditions

In order to characterize Cyanobacteria in any respect, a representative culture collection needed to be obtained. According to the National Center for Biotechnology Information Taxonomy database (accessed February, 2015), over 90% of the 13,000 classified entries in Cyanobacteria can be accounted for by only 2 classes: Nostocales and Oscillatoriophycideae (Benson, Karsch-Mizrachi, Lipman, Ostell, & Sayers, 2009; Sayers et al., 2009). Therefore, representative cyanobacteria were chosen from these main classes.

Plectonema sp. UTEX 1541, *Nostoc muscorum* UTEX1037, *Nostoc punctiforme* UTEXB1629, *Spirulina platensis* UTEX LB 2340 (alias *Arthrospira plastensis*), *Synechococcus leopoliensis* UTEX 2434, *Synechocystis* sp. PCC 6803, and *Synechocystis* sp. UTEX 2470 were chosen for study (Figure 13). These cyanobacteria were cultivated in a Forma Scientific Plant Tissue Culture Incubator Model 3750 at 29°C under fluorescent lighting at 60 μmol m^{-2} s^{-1} in BG-11 medium (Rippka et al., 1979) supplemented with 100 mmol TES buffer l^{-1} (pH 8.2). Cultures were agitated once daily. This collection allowed for a representative sample of the cyanobacteria phylum which could be tested and grown simultaneously at a single general mesophilic condition.

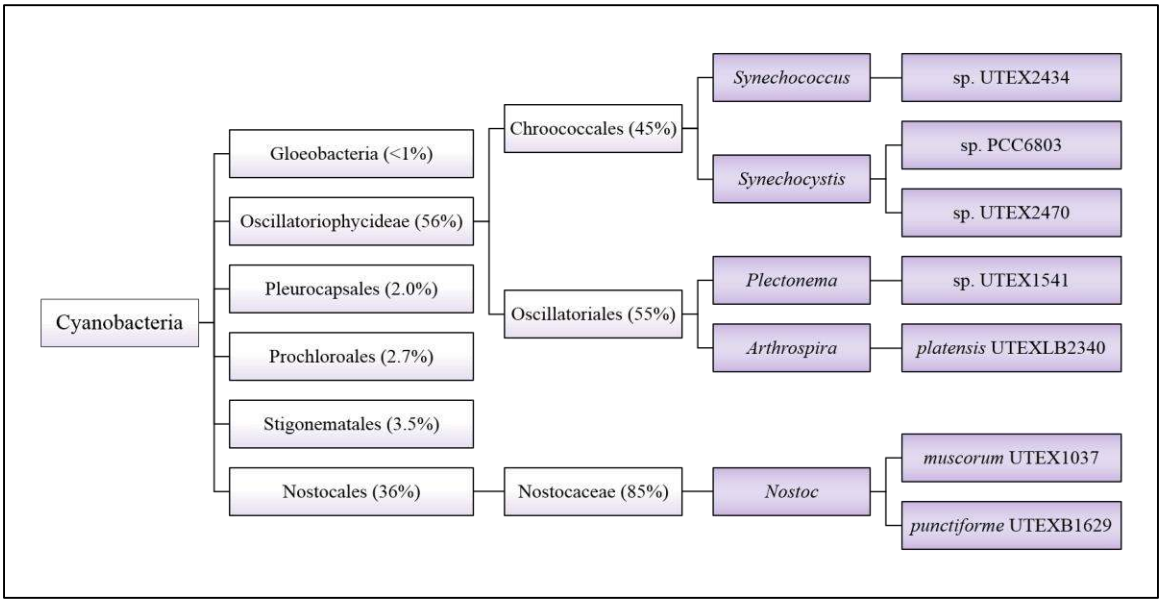


Figure 13: Abbreviated phylogenetic tree of organisms used in this study. Percentages depicted are of the relative number of taxonomic entries in the NCBI database compared to all classified entries of the parent node.

3.2.2 Cell Concentration Estimation

In order to approximate yields from DNA isolation, an estimate of the cell concentration is required. Since cyanobacteria possess such diverse morphology, high-throughput methods such as flow cytometry cannot be utilized. Instead, the traditional hand tallying via microscopy (ZWR Vistavision Inverted Microscope, 40x magnification) was performed. To accomplish this task, a Hausser Scientific hemocytometer with improved Neubauer ruling (0.1 mm chamber depth) and hand tally counter (Fisher Scientific) was implemented. The hemocytometer provides a ruled chamber of known volume from which the cells can be counted in order to estimate a cell density.

Equation 1: Regression equation for cell concentration as a function of optical density at 730nm (x).

$$C(x)\left[\frac{cell}{mL}\right] = \beta_1 x$$

These calculated concentrations were correlated to spectrophotometric absorbance targeting the far-red phytochrome (Pfr) at a 730 nm (OD_{730}) wavelength using a Beckman Coulter DU730 UV/Vis spectrophotometer. No-intercept linear regression (Equation 1) was performed using calculated cell concentrations as the dependent variable.

3.2.3 Growth Kinetic Modelling

The estimation of a bacterial growth phase is important in many microbiological applications. For example, when DNA is isolated during the exponential growth phase, many cells are undergoing DNA synthesis in preparation for cell division, which leads to higher mean DNA content per cell and subsequently higher yields (Figure 14).

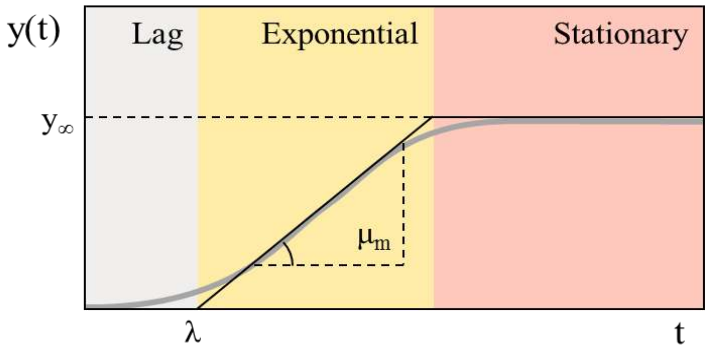


Figure 14: Example bacterial growth phase depicting lag phase, exponential phase, and stationary phase. The y-axis values are a measure of bacterial quantity and the x-axis values measure the fundamental dimension of time. The cell death phase following stationary phase is not depicted.

Growth data was recorded by measuring the OD_{730} at various time intervals for at least 3 biological replicates. These readings were normalized to their respective initial OD_{730}. The natural log of this normalized OD_{730} was then plotted against time for modeling. A simple logistic model with biologically relevant fitting parameters was obtained through an analogous derivation of Zwietering *et al.* (Equation 2) (1990).

Equation 2: Logistic model for cell growth (Zwietering et al., 1990).
y_∞ – stationary phase value ($t \to \infty$); μ_m – maximum specific growth rate; λ – lag time.

$$ln\left(\frac{OD_{730}|_t}{OD_{730}|_{t=0}}\right) = y(t) = \frac{y_\infty}{\left(1 + exp\left[\frac{4\mu_m}{y_\infty}(\lambda - t) + 2\right]\right)}$$

3.2.4 DNA Isolation

Due to the diverse morphologies of cyanobacteria, a generalized DNA isolation method is not always applicable. However, Neilan proposed a particularly practical general DNA isolation protocol

(2002). Six replicates of all cyanobacteria strains were subjected to this protocol with a minor modification (nucleic acid precipitation was performed in 0.2M NaCl final concentration). All DNA isolates were analyzed via spectrophotometry using the Beckman Coulter NanoVette (0.2-mm path length) at 4 fixed wavelengths: 260nm, 280nm, and 230nm, and a 320nm background correction.

3.3 Results

3.3.1 Cell Concentration Estimation

The hemocytometer measurements of the unicellular cyanobacteria were performed with no notable complications; however, the cell counts of the filamentous cyanobacteria (UTEX 1541, UTEX LB2340) proved more difficult. In order to estimate the number of counted cells, an average cell length was correlated via microscopy (Zeiss AxioObserver, 40x magnification) using the estimated length of each filament versus its respective number of cells (Figure 15). Then linear regression (no intercept) was performed in order to obtain an average cell length parameter. The resultant parameters for UTEX 1541 and UTEX LB2340 were 2.1(±0.45) and 6.0(±0.62) μm $cell^{-1}$, respectively (α=0.05). The regressor was then used to relate filament length measurements to cell count estimations. Error introduced from the cell length correlation was propagated into subsequent calculations.

UTEX 1037 proved even more difficult due to filament aggregation. Each individual cell in this filamentous cyanobacterium is readily observed; however, an effective method of disaggregating the filaments without causing significant cell lysis remains to be found. Repeated aspiration of sample using a 25 gauge syringe was used in this work. Due to the high difficulty of precisely quantifying the matting cyanobacteria, UTEX B1629 was assumed analogous to UTEX 1037 as these two strains have highly similar morphology and only order of magnitude estimates of cell concentration were necessary to perform DNA isolations.

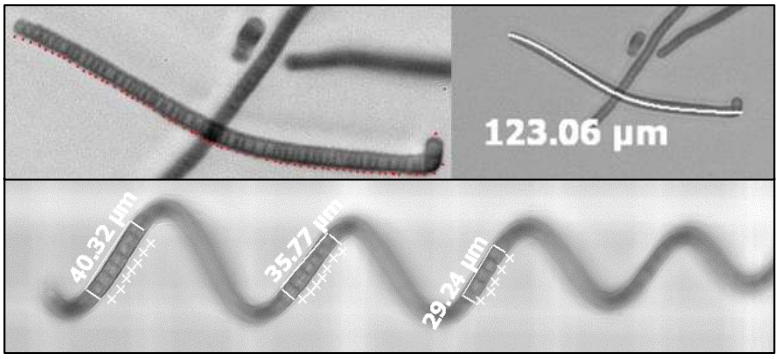


Figure 15: Example average cell length measurements of UTEX1541 (TOP) and UTEXLB2340 (BOTTOM) obtained using a Zeiss AxioObserver at 40x magnification.

Grouped measurements were first subjected to multiple comparisons testing, holding the family-wise error rate to 0.05, in order to ensure each representative data point was significantly different than the others (Initial Measurement Groups, Table 4). If these conditions were not met, measurements were pooled and analysis was repeated until all groups were significantly different from one-another (Significant Measurement Groups, Table 4). All error from filament length correlation was propagated in estimation of the regressor, when applicable (β_1, see Equation 1).

Table 4: Statistical summary of each independent cyanobacteria cell concentration regression from Equation 1. All error was propagated for confidence intervals (±(1-α)CI). α=0.05

Strain	β_1 x10^{-6}	Adj. R^2	$G_{Initial}$	$G_{Significant}$	$OD730_{MIN}$	$OD730_{MAX}$
PCC6803	85(±13)	0.66	5	4	0.065	0.82
UTEX1037	25(±9.6)	0.49	10	3	0.053	0.81
UTEX1541	10(±0.01)	0.74	5	5	0.026	0.65
UTEXLB2340	15(±7.0)	0.48	4	3	0.093	0.35
UTEX2434	180(±10)	0.83	7	7	0.067	0.66
UTEX2470	170(±7.6)	0.83	7	7	0.049	1.0

$G_{Initial}$ – Initial Measurement Groups; $G_{Significant}$ – Significant Measurement Groups.

3.3.2 Growth Kinetic Modelling

Nonlinear regression was performed minimizing a sum-of-squares objective function using the Gauss-Newton method, held to the constraints that all parameters must be positive or zero. If regression parameters were not statistically significant under 95% approximate confidence intervals, then regression was repeated on the nested model. In other words, the insignificant parameter was omitted from the full model and the regression was reiterated. These nested models were then compared using an F-test,

accounting for regression sum of square error (RSS) and regression degrees of freedom. A significance of 0.05 was used in hypothesis testing and the full model was used as the null hypothesis.

Table 5: Summary of the biologically relevant regressed parameters of the logistic model (Equation 2). Parameters and approximate confidence intervals given (±(1-α)ACI). NS: Not Significant. α=0.05.

Strain	y_∞	μ_m [day^{-1}]	λ [day]	Δt_{exp} [day]	AIC_{log}
PCC6803	2.06(±0.06)	0.21(±0.01)	NS	10	-638
UTEX1037	1.05(±0.23)	0.07(±0.02)	NS	15	-75.2
UTEX1541	2.57(±0.12)	0.16(±0.01)	NS	16	-240
UTEXLB2340	0.62(±0.04)	0.12(±0.03)	0.8(±0.7)	7	-123
UTEX2434	2.07(±0.06)	0.15(±0.01)	NS	14	-120
UTEX2470	1.83(±0.13)	0.13(±0.01)	NS	14	-138

y_∞ – stationary phase value ($t\rightarrow\infty$); μ_m – maximum specific growth rate; λ – lag time;
Δt_{exp} – duration of exponential growth phase.

It was observed that the growth kinetics were not well represented by the logistic model. Since lag-time had proven to be negligible in most cases, a simple saturation model was investigated (Equation 3). An analogous derivation of biologically relevant fitting parameters was performed as before (Appendix A.4). Akaike's Information Criterion (AIC) for sum-of-squares likelihood was used for non-nested model comparison. Based on information-theory, AIC is a measure of the relative quality of a statistical model (Burnham & Anderson, 2002). In this case, the AIC is proportional to the log of the RSS, so as RSS decreases (less residual error), AIC becomes more negative. AIC comparisons can only be performed between similar strains.

Equation 3: Saturation model for cell growth where lag time is insignificant ($\lambda = 0$).
y_∞ – stationary phase value ($t\rightarrow\infty$); μ_m – maximum specific growth rate;

$$ln\left(\frac{OD730|_t}{OD730|_{t=0}}\right) = y(t) = y_\infty\left(1 - exp\left[-\frac{\mu_m}{y_\infty}t\right]\right)$$

In each strain, the 2 parameter saturation model was a better fit than the 2 parameter logistic model (confer $\Delta AIC_{(sat\text{-}log)}$ Table 6). The 3 parameter model of 2340 was marginally better than the saturation model. However, UTEXLB2340 and UTEX 1037 showed markedly higher variation in optical density measurements than the other strains most likely due to aggregation (non-uniform cell suspensions). Additionally, in some cases (PCC 6803, UTEX 2434, and UTEX 2470) the biologically relevant parameters were statistically different than those estimated with the logistic model based off of the

approximate confidence intervals. A plot comparing the two models for the growth data in strain PCC 6803 is shown in Figure 16.

Table 6: Summary of the biologically relevant regressed parameters of the saturation model (Equation 2). Parameters and approximate confidence intervals given (±(1-α)ACI). NS: Not Significant. α=0.05.

Strain No.	y_∞	μ_m [day^{-1}]	Δt_{exp} [day]	AIC_{sat}	$\Delta AIC_{(sat-log)}$
PCC6803	2.33(±0.06)	0.32(±0.01)	7	-867	-229
UTEX1037	1.37(±0.69)	0.10(±0.04)	14	-75.4	-0.2
UTEX1541	3.14(±0.09)	0.22(±0.01)	14	-382	-142
UTEXLB2340	0.85(±0.22)	0.12(±0.03)	7	-122	1
UTEX2434	2.54(±0.13)	0.20(±0.01)	12	-171	-51
UTEX2470	2.13(±0.24)	0.19(±0.03)	11	-154	-16

y_∞ – stationary phase value ($t\rightarrow\infty$); μ_m – maximum specific growth rate;
Δt_{exp} – duration of exponential growth phase.

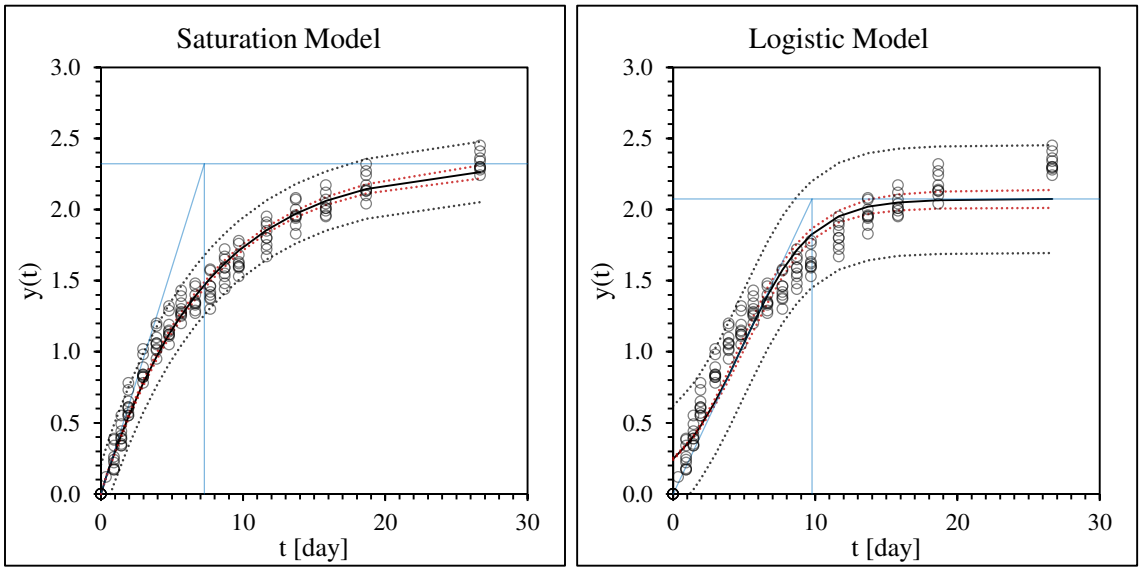


Figure 16: Plot of the growth kinetic models for the same *Synechocystis* sp. PCC 6803 data set. Comparison of logistic model on the left (Equation 2) and the saturation model on the right (Equation 3). Confidence and prediction intervals are depicted as red and black dotted lines respectively. Lines depicting relevant biological parameters are shown in blue (*see* Figure 14).

3.3.3 DNA Isolation

Spectrophotometric and agarose gel electrophoresis showed the DNA isolation protocol described in section 3.2.4 yields a sufficient quantity of DNA isolate for general molecular biological purposes. The DNA isolate yields can be observed in Figure 17.

However, the PCR genomic DNA quality assurance test, discussed in CHAPTER 4, indicated this method does not produce DNA isolate of high enough quality for general molecular biological purposes in

UTEX LB2340. A method optimized for *Arthrospira* (alias *Spirulina*) was required for this particular strain (Morin, Vallaeys, Hendrickx, Natalie, & Wilmotte, 2010).

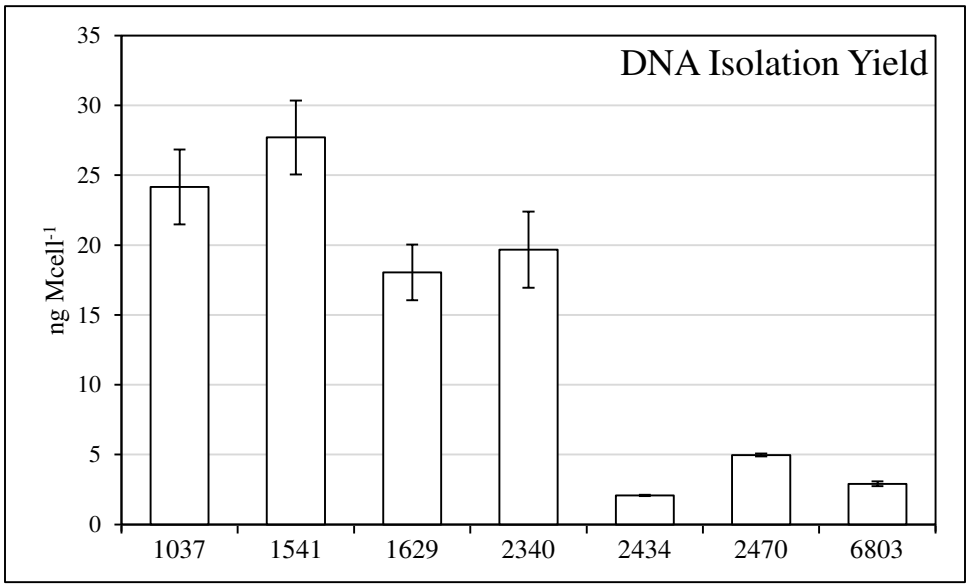


Figure 17: DNA isolation yields (μg $Mcell^{-1}$) of generalized protocol across applicable cyanobacteria strains. UTEX LB2340 is not depicted as it was determined to be of sufficiently low quality. Error from cell concentration regression was propagated (*see* section 3.3). Individual confidence intervals are depicted. ($n = 6$, $\alpha = 0.05$)

3.4 Summary

It can be concluded that the specific absorbance (i.e., the amount of light, absorbed by a fixed number of cyanobacteria cells) at a wavelength of 730 nm, will vary from strain to strain. Therefore, it is of the upmost importance to calibrate each strain individually when an experiments measurements are dependent on the cell count. It can also be concluded that under these growth conditions, the simple saturation model describes the observed growth kinetics of these cyanobacteria. Additionally, it can be concluded that most typical DNA isolation methods will work for a majority of cyanobacteria; however, more robust methods may be required for specific strains due to the cellular composition diversity observed in cyanobacteria.

CHAPTER 4. CYANOBACTERIA DNA CONTROL

4.1 Chapter Preface

This chapter details the development and testing of a cyanobacteria-specific PCR primer set for testing the quality of an arbitrary DNA isolation sample, a crucial element of Objective I (section 1.4). Much of this chapter has been reprinted from the original article in *Letters in Applied Microbiology* (C. E. Lane, Gutierrez-Wing, Rusch, & Benton, 2012).

4.2 Introduction

DNA-based techniques such as PCR are complicated when dealing with cyanobacteria, because cyanobacteria are morphologically diverse, and typical DNA isolation techniques are not always effective. Therefore, before PCR can be successfully used for cyanobacteria screening, verification that intact, genomic cyanobacterial DNA has been successfully isolated is a must.

Traditionally in PCR, check primers are used to verify successful DNA isolation. If a PCR primer binds successfully and amplification occurs, one can be confident that a valid template strand (often genomic DNA) is present. Often, there is a single housekeeping gene used to generate check primers, thereby serving as a positive control for multiple strains within a given organism class. For example, 18S rRNA genes are commonly used in yeast (Lantz, Stålhandske, Lundahl, & Rådström, 1999) and the glutamate decarboxylase and β-D-glucuronidase genes in *E. coli* (McDaniels et al., 1996). Generally, housekeeping genes targeted in PCR include those encoding ribosomal RNA, actin, tubulin, ubiquitin, and elongation factors (Filby & Tyler, 2007; Garg, Sahoo, Tyagi, & Jain, 2010; Jain, Nijhawan, Tyagi, & Khurana, 2006).

Check primers for the cyanobacteria phylum have been designed in one of two ways. The first method is to target a gene or operon which is uniquely specific to the phylum of interest. For example, in cyanobacteria the abundance of the phycobiliprotein phycocyanin (C-PC), one of two blue photosystem accessory pigments, aids chlorophyll *a* in energy harvesting in photosynthesis (Robert MacColl, 1998). In

the model cyanobacterium *Synechocystis sp*. PCC 6803, it is encoded by a five gene operon (Ughy & Ajlani, 2004). Two genes of importance (*cpcA* and *cpcB*) code for the phycocyanobilin-binding subunits (αPC and βPC respectively), while the other three genes code for rod linker polypeptides (Ughy & Ajlani, 2004). The amino acid which binds the chromophore via thioester linkage is very well conserved (Cys_{84} in αPC) (Robert MacColl, 1998). This operon is found primarily in cyanobacteria, but also in some cryptophyta and rhodophyta plastids (Eriksen, 2008; Robert MacColl et al., 1999). Neilan *et al.* have shown the α/β inter-genic spacer (IGS) is a novel region to investigate for phylogenetic classification of cyanobacteria due to its variability (1995). However, they report heterogeneity in amplification products across cyanobacterial strains (500-740 bp product) and no amplification product in *Nostoc punctiforme* PCC 73102 and *Nostoc commune* NIES 24 despite multiple reaction conditions (Neilan et al., 1995).

The second method targets a universal gene and achieves specificity through exploitation of cyanobacteria-specific consensus regions. An excellent example is the 16S ribosomal RNA primer sets developed by Nübel *et al*. (1997). The original intent of these sets, similar to those of Neilan and coworkers, was to show diversity within cyanobacterial populations using denaturing gradient gel electrophoresis (DGGE) of the PCR product fragments and for this purpose included up to 40-mer regions of GC-rich 5' tails in up to 62 bp primer length (Nübel et al., 1997). While the CYA359F/CYA781R(a and b mixture) set have been proven discriminant for cyanobacteria and plastids (Nübel et al., 1997), direct amplification using these primers have been shown to produce weak signal when investigated by Boutte *et al*. (2006). As a consequence of this weak signal, it has become typical when using these primers to perform semi-nested PCR with additional oligos in order to increase the product amplification (Boutte et al., 2006; Lymperopoulou, Kormas, Moustaka-Gouni, & Karagouni, 2011). It should be noted that a variation of the primer set developed by Nübel *et al*. has been proposed by McGregor and Rasmussen; however, to our knowledge, there has been no evidence of it recently being used as a cyanobacteria-specific PCR control (McGregor & Rasmussen, 2008). Also, an oxyphotobacteria-specific 16S ribosomal RNA primer set has been proposed by Rudi et al. (1997). While they observed positive amplification in cyanobacteria, including

Nostoc, they too observed best results when performing semi-nested PCR (Rudi et al., 1997). The two sets developed by Neilan *et al.* and Nübel *et al*. are among the most commonly used controls in cyanobacteria-related PCR (Saker, Welker, & Vasconcelos, 2007; Vaitomaa et al., 2003).

For high-throughput genetic screening and/or development of novel DNA isolation techniques for problematic strains, the primer sets described above have limitations when considered collectively. For example, some sets have shown weak signals, which can lead to false negatives or inconclusive results. Although this weakness is sometimes overcome with semi-nested PCR, this also is not ideal since semi-nested PCR requires two separate (non-tandem) reactions, a more expensive and time consuming process. Also, the sets have failed to detect certain species, limiting their utility as check primers for general cyanobacterial applications, especially with non-identified strains.

To address some of the limitations associated with the check primers above, a new primer set was developed capable of detecting high quality cyanobacterial DNA in a single-step reaction, even when little-to-no sequence information is available. The proposed set amplifies the majority of the αPC coding gene (*cpcA*).

4.3 Materials and Methods

4.3.1 Consensus PCR Primer Design

The primary structure for both αPC and *cpcA* from 22 various cyanobacteria were obtained from the NCBI RefSeq database (Sayers et al., 2009) and are listed in Appendix Table 2. A multiple alignment on the αPC amino acid sequences was performed using ClustalW2 (Larkin et al., 2007) with Gonnet weighting matrices (Gonnet, Cohen, & Benner, 1992) to determine areas containing highly conserved residues. Global alignment was chosen since the αPC sequences were highly related. The complete *cpcA* ORF sequences were then manually aligned in GeneDoc (Nicholas & Nicholas, 1997) to obtain a CEMA. For the initial primer design, Primer3 (Rozen & Skaletsky, 2000) was used on the consensus sequence to obtain primers in locations of both high amino acid residue conservation and codon bias. The forward

primer required redesign by inspection and included deoxyinosine modified bases at positions of high degeneracy (Ohtsuka, Matsuki, Ikehara, Takahashi, & Matsubara, 1985). Primer-BLAST was then performed with this detection set as an input and no specified template against the non-redundant cyanobacteria sequence library as a form of *in silico* PCR (Ye et al., 2012).

Next, 35 reference sequences of the 18S SSU gene of various fungi, green plants, and cryptomonads were obtained and are listed in Appendix Table 3. These nucleotide sequences were then aligned in ClustalW2. This is an example of consensus primer design via the method described in section 2.8.1. The consensus sequence was then subjected to Primer-BLAST with cyanobacteria mispriming library and again for *in silico* PCR against Chlorophyta (Ye et al., 2012).

4.3.2 PCR Conditions

To test the *cpcA* detection capabilities across all organisms in this study, 50 µl reactions consisting of 1.25 U *Taq* polymerase (Invitrogen), 20 mmol l^{-1} Tris-HCl (pH 8.4), 50 mmol l^{-1} KCl, 2.0 mmol l^{-1} $MgCl_2$, 0.2 mmol l^{-1} dNTPs (ea.), 0.5 µmol l^{-1} each primer, and 50 ng DNA template were prepared in 200 µl polypropylene tubes. These reactions underwent one cycle of [94°C for 3:00], 32 cycles of [94°C for 0:45, 53°C for 0:30, 72°C for 0:45], one cycle of [72°C for 10:00] and a final incubation at four centigrade in a Bio-Rad DNA Engine Peltier Thermal Cycler (model PTC0200).

As a positive control for the presence of PCR-quality DNA template in *E. coli*, the *gadA/B* detection primer set described by McDaniels (McDaniels et al., 1996) was used in four 50 µl reactions with 2.0 mmol l^{-1} $MgCl_2$. These reactions underwent one cycle of [94°C for 3:00], 30 cycles of [94°C for 0:45, 50°C for 0:30, 72°C for 0:45], one cycle of [72°C for 10:00] and a final incubation at four centigrade.

DNA extraction, amplification via PCR, and gel electrophoresis was performed at least twice for each species/primer combination tested.

4.3.3 Electrophoresis of PCR Products

PCR products were verified via agarose electrophoresis. Five microliters of each reaction product and one microliter of 6x Orange DNA Loading Dye (Fermentas #R0631) were loaded into a two percent low EEO agarose (US Biological #A1016) gel stained with 0.5 μg ml^{-1} ethidium bromide. An electric tension of 50 V was applied for 10 min to set samples in the gel and immediately followed by 100 V for 60 min in 0.5x TBE (Sambrook, 2001). Visualization was performed with a UVP Bioanalyzer in conjunction with UVP Visionworks LS (v6.5.2) Acquisition and Analysis for product molecular weight and concentration estimation.

4.4 Results

4.4.1 Alignment and Oligonucleotide Primer Results

The ClustalW2 multiple sequence alignment of the accessed sequences for αPC (Appendix Table 2) showed high conservation with limited gaps (Figure 19). Use of conserved areas allowed for the design of primers producing expected products of virtually equal length throughout these sequences (*G. violaceus* being the only exception, with a six bp gap) whose 5' termini locations are 0 and 423 with respect to $cpcA_{Syn6803}$. The amplified region includes the Cys_{84} phycocyanobilin-binding residue codon. Since single degeneracy was desired, some mismatching occurs within the annealing region. The mean number of mismatches over the 22 sequences was calculated (omitting modified bases) to be 1.45 for cpcA-F2, 2.50 for cpcA-R1, and 3.95 for the set. The multiple alignment and respective CEMA for the cpcA-F2 region for αPC/*cpcA* is shown in Figure 19.

Even under lax conditions in Primer-BLAST, the 18S detection set showed no potential mispriming of expected length in cyanobacteria. The overwhelming majority of expected *in silico* products were 550-560 bp in length. The forward and reverse primer 5' termini locations are 894 and 1450 with respect to the *S. cerevisiae* accession sequence (Z75578.1).

4.4.2 PCR Optimization

Under more specific reaction conditions during preliminary testing, *cpcA* was detected in all cyanobacteria strains with the exception of the currently un-sequenced *Plectonema sp*. Optimization was performed to maximize the amplification of *cpcA* product in this strain. The reaction products were subjected to typical gel quantitation with the results depicted in Figure 18. As expected, some mispriming did occur under low annealing temperature and high salt concentrations (indicated in Figure 18 by *).

Optimal conditions were used in all *cpcA* experiments and can be found in section 4.3.2 or abbreviated in Table 7.

Table 7: Sequences of the primers designed and abbreviated reaction conditions utilized.

Oligo		PCR Conditions		
Primer	Sequence	T_{Anneal} [°C]	Mg^{++} [mM]	N_{cycles}
cpcA-F2	ATGAAAACCCCICTIACIGAAG	53.0	2.00	32
cpcA-R1	ACCGTGGTTAGCTTTGATGT			
18SrDNA-F1	TGTCAGAGGTGAAATTCTTGGA	50.0	1.50	30
18SrDNA-R1	ACATCTAAGGGCATCACAGACC			

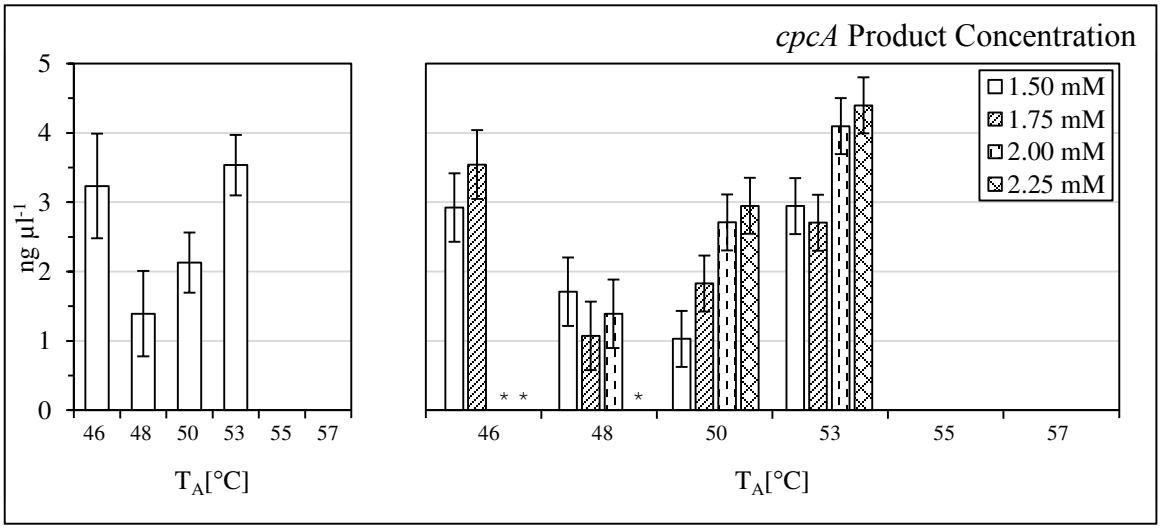


Figure 18: Resulting *cpcA* detection set PCR amplification products (8 µl from 50 µl reactions) from UTEX 1541 approximation of band mass from gel electrophoresis quantitation.
*Denotes non-negligible mispriming observed.
†55°C and 57°C were also tested under same Mg^{++} gradients and no products were observed (not depicted).

Figure 19: ClustalW2 multiple sequence alignment of αPC from the cyanobacteria sequences obtained through NCBI and listed in Appendix Table 2. A codon-equivalent alignment (CEMA) for the *cpcA* gene was produced in GeneDoc (Nicholas & Nicholas, 1997) to obtain regions of high codon bias. Primers were designed off of this CEMA in areas of high conservation. The cpcA-F2 (forward primer) region of the CEMA is shown (cpcA-F2 sequence is shown below) with respect to position of the αPC multiple alignment.

4.4.3 PCR Amplification Results

Electrophoresis results of the PCR products confirmed *cpcA* detection in cyanobacteria strains, as observed in (lanes 2-6, Figure 20, *cpcA*). No amplification was observed in any non-cyanobacteria, as observed in (lanes 8-10, Figure 20, *cpcA*). Under these conditions, we estimated the amplified *cpcA* PCR products final concentrations to be between nine and 20 ng μl^{-1} via gel quantitation. The PCR results for the 18S detection showed no amplification in cyanobacteria (lanes 2-6, Figure 20, 18S) and *E. coli* (lane 10, Figure 20, 18S) as was expected for prokaryotes. Positive detection was observed in *S. cerevisiae* (lane 9, Figure 20, 18S) and *Chlorella vulgaris* (lane 8, Figure 20, 18S) at the expected band lengths with final product concentrations of 10 and five ng μl^{-1} respectively. Additionally, no other products were observed for these two strains. The *E. coli gadA/B* control showed positive results for quadruplicate reactions (data not shown).

It should be noted that successful amplification of homologous products were also observed in cyanobacteria strains *Arthrospira maxima* UTEX LB2342 (CS-328), *Microcoleus vaginatus* PCC 9802 (FGP-2), *Microcystis aeruginosa* NIES-843, *Synechococcus leopoliensis* UTEX 2434, *Synechococcus* sp. PCC 9742, and *Synechocystis* sp. UTEX 2470 in subsequent experiments (CHAPTER 5).

4.5 Discussion

Cyanobacteria show much potential in driving down the production costs of many ecologically sound bioproducts. However, due to the low yields currently observed from products produced by these microorganisms improvements are needed before these processes can be considered economically sound. The difficulties associated with the high-throughput screening process make a robust check primer set prerequisite. Here, we have demonstrated the effectiveness of such a primer set.

Through the multiple alignment of a large population of cyanobacteria sequences, a detection primer set with single degeneracy was shown to amplify the *cpcA* target sequence specifically in 5 different cyanobacteria from 3 different classes. Three cyanobacteria (*Synechocystis sp*., *N. punctiforme*, and *S. platensis*) with known sequences for this gene were tested and successful detection was observed with as

high as 4 mismatches in the set. Notably, the primers were successful at *N. punctiforme* amplification, where other primers had failed on similar strains (Neilan et al., 1995). This set has the potential to be a valuable tool in the future high-throughput screening of cyanobacteria that the high-production bio-commodity field desperately requires. The *cpcA* detection primer set was demonstrated to be a robust tool for the detection of cyanobacterial DNA of adequate quality for routine molecular biology purposes. For axenic and unicyanobacterial cultures, the cpcA set alone is sufficient and recommended in regards to a DNA control. However the two sets proposed in this paper combined possess a wider variety of uses than just high throughput DNA control.

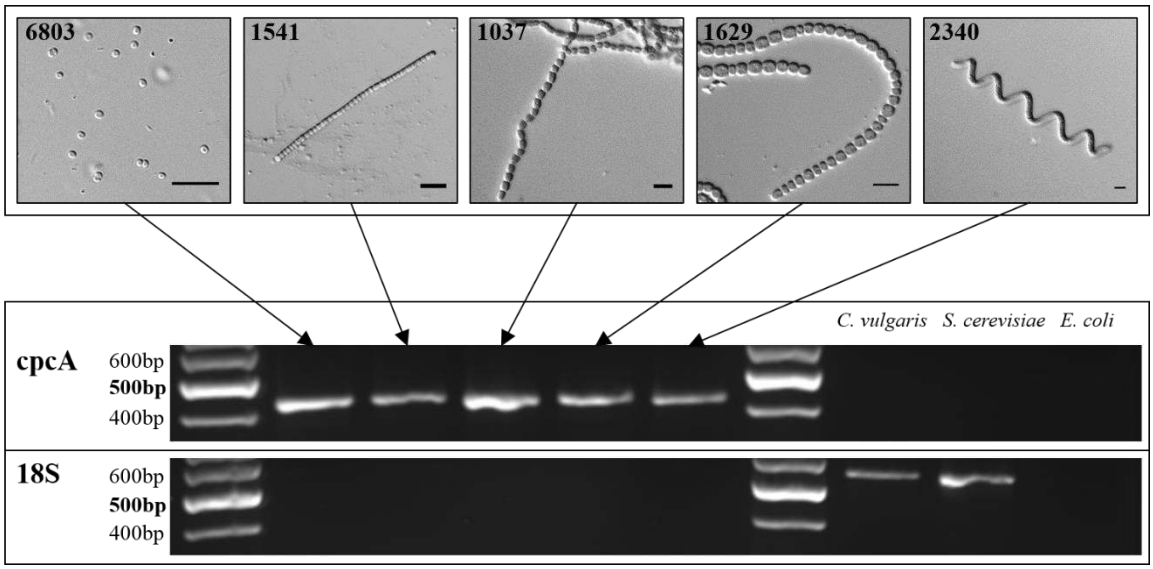


Figure 20: Brightfield microscopy images and agarose electrophoresis (2% standard agarose, 0.5 µg ml^{-1} ethidium bromide stain, 0.5xTBE) results of 5 µl of each PCR amplification products using the primer sets developed in this work. The first set (cpcA) detects the phycocyanin alpha-subunit coding gene and the second set (18S) 18S ribosomal RNA coding sequence (n = 2).
Lanes: (M) 5 µl 100bp O'GeneRuler DNA ladder, (1) *Synechocystis* sp. PCC 6803, (2) *Plectonema* sp. UTEX 1541, (3) *Nostoc muscorum* UTEX 1037, (4) *Nostoc punctiforme* UTEX B1629, (5) *Spirulina platensis* UTEX LB2340, (M), (6) Chlorella vulgaris, (7) Saccharomyces cerevisiae BY4741, (8) *Escherichia coli* DH5α.
**E. coli* positive control performed in quadruplicate using gadA/B amplification primers (McDaniels et al., 1996). (data not shown)

The result of $cpcA^{+}$/$18S^{+}$ could have multiple initial conditions since it only detects the presence of eukaryotic DNA, which could result in axenic algae cultures, mixed cultures of algae and cyanobacteria, and cyanobacteria with non-algae eukaryotic contamination (e.g., fungi). As a consequence, the primer sets proposed in this study could further be implemented in isolation of cyanobacteria from environmental

samples containing higher algae and other various eukaryotes such as the many common fungal contaminants. Implementation of this isolation procedure would involve a combination of streak plating and colony PCR in conjunction with traditional microscopy. This could ensure with higher degree of confidence than microscopy alone that a colony lacks any eukaryotic organisms. This pair of primer sets could be a valuable tool for any field of research involving environmental samples of cyanobacteria, such as the screening of toxic cyanobacteria in potable waters.

The cpcA controls could streamline many high-throughput methods involved in cyanobacteria research that will become more important in the near future as the interest in these unique organisms is consistently increasing. The current production capabilities of these organisms with respect to bioindustrial products of high value, has yet to approach their production potential.

CHAPTER 5. PCR-BASED PHA SYNTHASE DETECTION

5.1 Chapter Preface

This chapter presents the realization of Objective I (section 1.4) and includes the work originally published in *Molecular and Cellular Probes* (Courtney E. Lane & Benton, 2015). An additional section has been appended to expand upon the subject using the computer application described in CHAPTER 6.

5.2 Introduction

A biological alternative to petroleum-based polymers, polyhydroxyalkanoates (PHAs) encompass a diverse class of biodegradable polyesters capable of mitigating the ecological consequences in meeting the ever-growing demand for plastic commodities. This polymer class is well-studied and many informative reviews are available (Anderson & Dawes, 1990; Rehm, 2007; K. Sudesh, Abe, & Doi, 2000).

Currently, PHA production research is focused mainly on recombinant or native heterotrophic microorganism bioprocesses (Agnew et al., 2012; Atlić et al., 2011; Breuer et al., 2002; Kahar et al., 2004; RamKumar Pandian et al., 2010; A. Singh & N. Mallick, 2009). The industrial applicability of chemoheterotrophic processes is limited by the high production costs associated with the external carbon source (Bengtsson et al., 2010; Choi & Lee, 1999). The use of agricultural/industrial wastes and other low-cost carbon sources improves carbohydrate feedstock costs but unfortunately either compromises productivity or demands unit operation redesign (e.g., recycle) (Ienczak et al., 2013). A photoautotrophic PHA bioproduction process can exploit a phototroph's inherent photosynthetic carbon-fixation ability to consume atmospheric carbon dioxide, water, and light energy in order to produce PHAs, eliminating the requirement of a carbohydrate feedstock and the related costs.

Cyanobacteria are a diverse phylum of photoautotrophic prokaryotes inhabiting limnic and marine environments. These microalgae are the sole prokaryotic native producers of PHA via oxygenic photosynthesis (Asada et al., 1999; Sharma et al., 2007), along with a variety of other value-added products (Abed et al., 2009; Ducat et al., 2011; Simmons et al., 2005; L. T. Tan, 2007), making them attractive

candidates for phototrophic PHA bioproduction. Unfortunately, typical PHA accumulation observed in cyanobacteria under photoautotrophic growth is < 10% dry cell weight (DCW), if detected at all (Bhati et al., 2010; Ducat et al., 2011). This is far lower than the yields observed in heterotrophic high-density processes which can approach 87% DCW (Ienczak et al., 2013). However, Nishioka et al. demonstrated wild-type cyanobacteria strains are capable of exhibiting significantly higher yields – up to 55% DCW PHA during photoautotrophic cultivation (2001). Functional diversity, including this large deviation in PHA accumulation, is not a rarity within *Cyanobacteria* (Garcia-Pichel et al., 1999). With *Cyanobacteria*'s rich biodiversity, it is likely that there are undiscovered high-yielding cyanobacteria strains with PHA production capabilities rivaling those of the costly heterotrophs. Finding such species out of potentially millions of possibilities is certainly a logistical challenge, especially when accounting for the tedious culture optimization each strain requires to obtain the unique conditions for maximum PHA production. Since optimization is a laborious-but-necessary step, a high-throughput design implementing preemptive screening for strains ill-equipped for PHA production would greatly decrease avoidable costs.

Traditional qualitative screening of PHA accumulation in cyanobacteria involves the staining of the intracellular granules with basic oxazine/oxazone dyes, typically Nile blue/red, which bind to the PHA granule and allow for visualization through epifluorescence microscopy (Ostle & Holt, 1982). The fluorescent staining process first requires the growth of the isolate into the late-exponential phase (~10 days), followed by nutrient-limiting cultivation (~15 days), meaning it could take a month before the staining and visualization have been completed (Jau et al., 2005). It should be noted that the length of the nutrient deficient cultivation could potentially be shortened; however, cyanobacteria accumulate PHA at different rates (Bhati et al., 2010). Therefore, the minimum time required before PHA accumulation could be detected via staining would be strain dependent, and an assay implementing an insufficient nutrient deficient cultivation period would risk an increased rate of type II error. Since cyanobacteria have relatively slow growth rates ($t_d \geq 12$ hr), an assay with a lower detection limit, with respect to number of cells, would better conform to the demands of a high-throughput assay. PCR amplification is sensitive enough to detect

a single copy of a targeted DNA sequence, meaning the DNA of a single cell is sufficient in theory (Hahn et al., 2000). Using PCR to detect a gene necessary for PHA production would greatly reduce the number of generations an isolate requires before it can undergo downstream optimization. Such a process could efficiently decrease avoidable costs while advancing the attainment of a more environmentally friendly plastic production process.

PHA biosynthesis can take many routes. For example, in polyhydroxybutyrate biosynthesis the upstream precursors to the hydroxybutyryl-CoA monomers are formed from the acetylation and reduction of acetyl-CoA by β-ketothiolase (*phaA*) and acetoacetyl-CoA reductase (*phaB*) respectively. Alternatively, medium chain length PHAs are typically formed utilizing fatty acid β-oxidation or *de novo* synthesis intermediates (Rehm, 2007). The single committed step in this pathway is the polymerization of the hydroxyacyl-CoA thioester monomers by PHA synthase (Rehm, 2007; Steinbüchel & Lütke-Eversloh, 2003). It is also important to note that this enzyme is not the limiting factor in PHA accumulation in cyanobacteria. Numata et al. have demonstrated that the activity of the *Synechocystis* sp. PCC 6803 PHA synthase is comparable to that of a high-yielding heterotroph – *Ralstonia eutropha* (Numata et al., 2015). This suggests that in a hypothetical high-yielding cyanobacteria, the determining factor will not be the synthase itself, but will likely involve upstream metabolic elements. Some potential elements include the RNA Polymerase Sigma Factor SigE (Osanai et al., 2013) and nutrient dependent regulators (Hauf et al., 2013; Schlebusch & Forchhammer, 2010).

The PHA synthase protein family is divided into four major types (I – IV), under the basis of primary structure, subunit composition, and substrate specificity (Rehm, 2007). The work of Hai et al. suggests that the PHA biosynthesis pathway in *Cyanobacteria* occurs in a widespread and general fashion, and only type III PHA synthases have been observed (2001). Type III PHA synthases are a hetero-multimeric sub-classification of polyester synthases consisting of two main subunits with relatively equal mass (~40 kDa), PhaC and PhaE, typically coded in a single operon (Rehm, 2007). The type III PhaC subunit is capable of *in vitro* PHA polymerization in the absence of PhaE, albeit inefficiently (Müh et al.,

1999), and exhibits higher degrees of conservation than its counterpart, making it the ideal target for PCR-based PHA genetic characterization. If the gene encoding PhaC, *phaC*, is not present within an organism, then the organism is incapable of polymerizing PHA through any known mechanism.

In this work a simple and rapid PCR-based *phaC* detection assay was developed. Following an extensive sequence analysis of 29 cyanobacteria *phaC* sequences, a single low-degeneracy primer set was designed within regions of high conservation. This method of design was implemented in order to increase the probability of amplifying unknown sequences, while reducing the intrinsic probability of non-specific amplification associated with high-degeneracy primer sets. Such a design allows for the simultaneous testing of multiple cyanobacteria strains with minimal strain-dependent reaction alterations. In order to demonstrate how the presence of the *phaC* gene has the potential to be a good indicator for the ability to accumulate PHA, the PCR-based assay was then tested against five cyanobacteria strains alongside traditional staining methods. Next, a high-throughput screening assay was developed in the form of colony/quick prep PCR, which was then tested against nine diverse strains of cyanobacteria. This assay will help to rapidly categorize cyanobacteria as either potential producers or non-producers of PHA and is a major step in attaining a more economically-viable strain of cyanobacteria for the design of a carbon-neutral PHA production process.

5.3 Materials and Methods

5.3.1 Bacterial Strains and Media

Plectonema sp. UTEX 1541, *Spirulina platensis* UTEX LB2340, *Synechococcus leopoliensis* UTEX 2434, *Synechocystis* sp. PCC 6803 as donated by Terry Bricker, and *Synechocystis* sp. UTEX 2470 were cultivated in a Forma Scientific Plant Tissue Culture Incubator Model 3750 at 29°C under constant fluorescent lighting (60 μmol m^{-2} s^{-1}) in BG-11 medium (Rippka et al., 1979) supplemented with 100 mmol l^{-1} 2-[tris(hydroxymethyl)methylamino]-1-ethanesulfonic acid (TES) buffer (pH 8.2). Cultures were agitated once daily.

5.3.2 Nucleic Acid Isolation

Total DNA was isolated from PCC 6803, UTEX 1541, UTEX 2434, and UTEX 2470 using a slightly modified DNA isolation method originally designed for *Saccharomyces cerevisiae* through mechanical lysis (Harju, Fedosyuk, & Peterson, 2004). An additional chloroform/isoamyl alcohol (24:1) separation step was added to this protocol and performed on all of these isolates. An alternative method of DNA isolation, which was optimized for the *Arthrospira* (*Spirulina*) genus, was used for UTEX LB2340. This DNA isolation protocol included both mechanical and enzymatic lysis steps (Morin et al., 2010). All DNA template samples were analyzed through spectroscopy (Beckman Coulter, DU730; NanoVette, A44097).

5.3.3 Synthetic Oligonucleotide Primers

Primers were designed utilizing methodology from previous works (C. E. Lane et al., 2012) applied to the sequence accessions listed in the appendix (Appendix Table 4). The primer sequences used in this study were 5'-GGGATGTCTATTTGATTGAYTGG and 5'-GGTCGGGACTATCAAAAATCCA for the forward (phaC(3.1)-F) and reverse (phaC(3.1)-R) primers respectively. The reported melting temperatures were 52.5°C for phaC(3.1)-F and 54.7°C for phaC(3.1)-R (Integrated DNA Technologies). The forward and reverse primer annealing regions correspond to positions 344-366 and 805-826 of the *Synechocystis* sp. PCC 6803 accession (CP003265.1: (933155..933177, 933616..933637)). The standard free-energy deviations ($\Delta G^{\circ}_{37^{\circ}C,1M\ NaCl}$) of annealing were calculated using nearest neighbor thermodynamics for the longest consecutive complementary sequence of each primer-accession pair. The calculated values were then normalized against the sum of the complementary standard free-energy of annealing of the primers (perfect annealing, 100%) (SantaLucia, 1998). The normalized deviation in free-energy was used to represent the relative annealing ability of the primer set to each sequence in the alignment.

5.3.4 Standard PCR Amplification

Each 50 µl *phaC* detection reaction was prepared in a 200 µl polypropylene tube and consisted of 1.25 U Taq polymerase (Invitrogen, 10342), 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$,

0.2 mM dNTPs (ea.), 1.0 μM phaC(3.1)-F, 0.5 μM phaC(3.1)-R, and approximately 50 ng DNA template. Annealing temperature optimization was performed using PCC 6803, UTEX LB2340, and UTEX 1541 DNA templates. After the annealing temperature was determined, each reaction underwent an initial denaturation [94°C for 3:00], 32 thermal cycles [94°C for 0:45, 54°C for 0:30, 72°C for 0:45], a final elongation [72°C for 10:00], then maintained at 4°C in a Peltier thermal cycler (Bio Rad, PTC0200) until storage at 4°C. As a DNA quality control, all DNA isolates were checked for amplification using the phycocyanin α-subunit, *cpcA*, detection primer set previously described (C. E. Lane et al., 2012).

5.3.5 Electrophoresis of Standard PCR Product

Five microliters of each reaction product and one microliter of 6x Orange DNA Loading Dye (Fermentas, R0631) were loaded into a two percent standard agarose (US Biological, A1016) gel stained with 0.5 μg ethidium bromide ml^{-1}. Five microliters of 100 bp O'GeneRuler DNA ladder (Fermentas, SM1143) was used for each molecular weight marker. Visualization was performed with a UVP Bioanalyzer and UVP Visionworks LS Acquisition and Analysis (v6.5.2).

5.3.6 Standard PCR Product Sequence Confirmation

The respective biological replicates were combined for each strain that successfully amplified with the *phaC* detection set and the pooled reaction was then purified using a DNA cleanup micro kit (Thermo Scientific, K0831). Each of these pooled samples was then subjected to DNA sequencing using the Applied BioSystems BigDye Terminator (v3.1) on an ABI Prism 3130 using both primers independently. The consensus of the two sequencing runs, not including the primer regions, was then deposited into GenBank when it represented a unique variant of the *phaC* partial coding sequence (Benson et al., 2009).

5.3.7 Nutrient-Limited Growth

Three BG-11 media variants were used for nutrient-limited growth. For nitrogen-limited conditions, BG-11 was prepared with sodium nitrate omitted. For phosphate-limited conditions, BG-11 consisting of 5 mg dipotassium phosphate l^{-1} was used (Sharma et al., 2007). For both nitrogen-limited

and phosphate-limited conditions, media was prepared using both conditions above. All nutrient-limiting media were supplemented with 0.17% (w/v) sodium acetate as carbon source to accelerate accumulation (Sharma et al., 2007). Cyanobacteria strains were first washed and transferred from the normal medium to the 3 nutrient-limited media types in triplicate (n = 3) during exponential growth phase and cultured 5-10 days before sample collection and staining.

5.3.8 PHA Granule Visualization

Samples were heat-fixed and subjected to a Nile blue A fluorescent stain (Ostle & Holt, 1982) and visualized via epifluorescence microscopy (Zeiss AxioObserver.Z1) using filter-set 20 (excitation: BP 546/12, beam splitter: FT 560, emission: BP 575-640). Images were recorded using an Axiocam MRm.

5.3.9 Gas Chromatography-Mass Spectroscopy Confirmation

Cells grown under nutrient limiting conditions were acetone dried and subjected to methanolysis using a mixture of equal volumes of chloroform and acidified methanol (15% (v/v) sulfuric acid) and then heated to 100°C for 2 hours. Hydroxyalkanoate methylester content was assessed as described in Tan et al. (2014).

5.3.10 Colony/Quick Prep PCR

To demonstrate the use of PCR as a high-throughput screening method, whole cell PCR amplification was utilized. Additional genotyped strains were procured in order to further test the robustness of the primer sets while developing this assay. *Arthrospira maxima* UTEX LB2342 (CS-328), *Microcoleus vaginatus* PCC 9802 (FGP-2) as donated by Ferran Garcia-Pichel, *Microcystis aeruginosa* NIES-843 as donated by Christopher Gobler, and *Synechococcus elongatus* PCC 7942 as donated by Terry Bricker were all obtained for testing in addition to UTEX 1541, UTEX LB2340, UTEX 2434, UTEX 2470, and PCC 6803.

In addition to standard colony PCR, a method of quick template preparation was developed. First, samples were extracted (≤1 ml of OD_{730nm}~0.6) and cells were centrifuged at 10,000xg for five minutes to

obtain a pellet approximately one millimeter in diameter. Next, cells were washed in 200 µl TE buffer (pH 8.0) and pelleted again. Cells in the wet pellet were then mechanically lysed via micropipette in 10 µl dimethyl sulfoxide (DMSO). 90 µl of TE buffer (pH 8.0) was then added, well mixed via pipette, and the solution was centrifuged at 18,000xg (maximum) for five minutes to remove insoluble debris. The clarified lysate was then transferred to a fresh tube and used directly in amplification reactions as template. The clarified lysate was stored at −20°C.

PCR amplification was performed as described in the standard PCR methods section; however, the number of thermal cycles was increased from 32 to 35 for both *cpcA* and *phaC* reactions. The amplifications were carried out using whole cells, one, five, or 10 µl of the clarified cell lysate as template. 10 µl of the unpurified reactions were visualized via one percent agarose gel electrophoresis using five microliters of GeneRuler 1kb plus (Life Technologies, SM1333) as standard.

5.4 Results

5.4.1 PhaC Multiple Sequence Alignment

To investigate the level of conservation of the PHA synthase PhaC subunit within *Cyanobacteria*, the primary protein structures of 29 cyanobacteria (*see* Appendix Table 4) were globally aligned using Clustal Ω (Goujon et al., 2010; Sievers et al., 2011). The full alignment, with relevant annotations, is available in the appendix (Appendix Figure 1, A.7). 256 residue positions within the alignment showed conservation, with 155 of those positions showing identity-level conservation. As expected, the alignment showed little N-terminus conservation, high C-terminus conservation. The catalytic lipase-like box region (Rehm, 2007) (AGF51119.1:162..166) showed high conservation (`G[I|V]CQG`) throughout all sequences. The "Cyanobacterial box" described by Hai et al. (AGF51119.1:203..212) also exhibited conservation, although it was observed that both ends of this region contained positions with low levels of conservation (2001). In comparison with a previous alignment of 59 polyester synthase sequences from various organisms including one cyanobacterium (*Synechocystis* sp. PCC 6803) (Rehm, 2003), only amino

acids AGF51119.1: 205..209 bounded the cyanobacteria-specific insertion in the other type III PHA synthase sequences. This subregion remained well conserved (GC[S|T][L|I]G). This primary protein structure alignment, along with each respective genetic coding sequence, was used to generate a codon-equivalent multiple alignment.

5.4.2 *phaC* Codon-Equivalent Multiple Alignment and Primers

Using the coding sequences (Appendix Table 4) a codon-equivalent multiple alignment was generated in order to design a generalized set of primers. The full alignment, with relevant annotations, is available in the appendix (Appendix Figure 2, A.8). 415 positions within the alignment were completely conserved throughout all sequences. The 3' end of phaC(3.1)-R complemented a tryptophan codon which is believed to be involved in protein-protein interaction and was completely conserved throughout all 59 polyester synthase sequences of the Rehm alignment (Rehm, 2003). Within the alignment, the total number of mismatches ranged from 2-11 with a mean of 5. All of the mismatches occurred at least 6 nucleotides from the 3' end of phaC(3.1)-F and at least 5 nucleotides from the 3' end of phaC(3.1)-R. The average relative annealing ability was found to be 61% and ranged from 21-91%. The *Synechocystis* sp. PCC 6803 sequence (CP003265.1) possessed 5 total mismatches within the primer set and 60% of the relative annealing ability, placing it in the 31st percentile of the sequences (i.e., 69% of the sequences in the alignment had a higher relative annealing ability).

5.4.3 Standard PCR

Strains PCC 6803, UTEX 1541, and UTEX LB2340 were chosen for initial testing as these strains have known PHA production capabilities (Bhati et al., 2010; Hein, Tran, & Steinbüchel, 1998; Jau et al., 2005). In order to establish effective PCR amplification conditions, the cyanobacteria DNA isolates from UTEX 1541, UTEX LB2340, and PCC 6803 were used to optimize the annealing temperature. The product concentrations were estimated through gel quantitation. Undesired product formation occurred in the negative control (UTEX 1541) reactions only at annealing temperatures below 48°C; this artifact was approximately 360 bp in length and can be observed in Figure 21. Both PCC 6803 and UTEX LB2340

product formation followed comparable trends in desired product concentration at temperatures above 48°C, and the mean results are depicted in Figure 21. The experimental reaction annealing temperature selected was 54°C. The triplicate amplification results for UTEX 1541, UTEX LB2340, and PCC 6803 at these conditions can be observed in the right column of Figure 22 (a), (b), and (d) respectively. UTEX 1541 exhibited no amplification using the *phaC* detection primers, but showed positive amplification for the *cpcA* detection control. Both UTEX LB2340 and PCC 6803 exhibited positive amplification for both *phaC* and *cpcA* primer sets. The *phaC* product sequence of PCC 6803 confirmed the entirety of the expected accessed sequence. The *phaC* product sequence of UTEX LB2340, deposited under accession KR824842, was compared to the accession for strain NIES-39 and was confirmed with 11 single nucleotide discrepancies. Alignment of the translations from the experimental UTEX LB2340 and accessed NIES-39 sequences revealed only two of the single nucleotide polymorphisms would result in missense mutations (V200I and I202V). The characteristic lipase-like box region was confirmed for both *phaC* amplification product translations.

5.4.4 PHA Visualization

In order to verify PHA production capabilities, strains were subjected to nutrient-limited growth and fluorescent staining. A representative Nile blue A staining and visualization of UTEX 1541, UTEX LB2340, and PCC 6803 can be observed in the center column of Figure 22 (a), (b), and (d) respectively. PHA accumulation was not observed in UTEX 1541 under any of the conditions. PHA accumulation was exhibited in UTEX LB2340, and PCC 6803 under all three conditions. The presence of methyl 3-hydroxybutanoate in UTEX LB2340 and PCC 6803 was confirmed through gas chromatography-mass spectroscopy (Appendix Figure 3, A.9).

5.4.5 Screening of Unknowns

In the interest of testing the efficacy of a PCR-based PHA detection assay, two cyanobacteria in which PHA synthase presence, sequence information, and activity were unknown were assayed. PCR was performed on UTEX 2434 and UTEX 2470 at the conditions described in Materials and Methods. Product

amplification of the positive genomic control, *cpcA*, was observed in both strains. The type III PHA synthase *phaC* detection primer set failed to amplify in UTEX 2434, but successfully amplified in UTEX 2470 (Figure 22 (c) and (e), right column). The *phaC* product sequence of UTEX 2470, deposited as KR824841, was compared to the accessed sequence of PCC 6803 and 47 discrepancies occurred out of 438 residues. Alignment of the translations from the experimental UTEX 2470 and the accessed PCC 6803 sequences revealed five polymorphisms (N144T, I148V, D179E, G197S, and G255E). Upon comparison with the PhaC multiple sequence alignment, the N144T mutation of UTEX 2470 is the first non-asparagine residue observed in this position. The characteristic lipase-like box was confirmed within the detected UTEX 2470 PCR product. The PHA granule visualization experiment revealed no detectable PHA accumulation in UTEX 2434 under any conditions. PHA accumulation was observed in UTEX 2470 for all nutrient-limiting media types (Figure 22 (c) and (e), center column). The presence of methyl 3-hydroxybutanoate in UTEX 2470 was confirmed through gas chromatography-mass spectroscopy (Appendix Figure 3, A.9).

5.4.6 Colony/Quick Prep PCR

During the initial testing of the colony PCR protocol, it was determined that the number of thermal cycles needed to be increased by three in order to obtain yields comparable to those we found in amplification of high quality DNA isolates. Once the number of cycles had been determined, the colony PCR experiments were performed simultaneously for a given strain. The *cpcA* PCR amplification of whole cells unexpectedly failed in multiple cases (UTEX 1541, UTEX 2434, and UTEX LB2340), and there were similar unexpected failures for *phaC* whole cell amplification (UTEX LB2340, UTEX 2470, and PCC 6803). PCR amplification using 1 µl of the DMSO/TE clarified lysate exhibited more replicable results than the whole cells, but still exhibited a single unexpected failure (NIES-843).

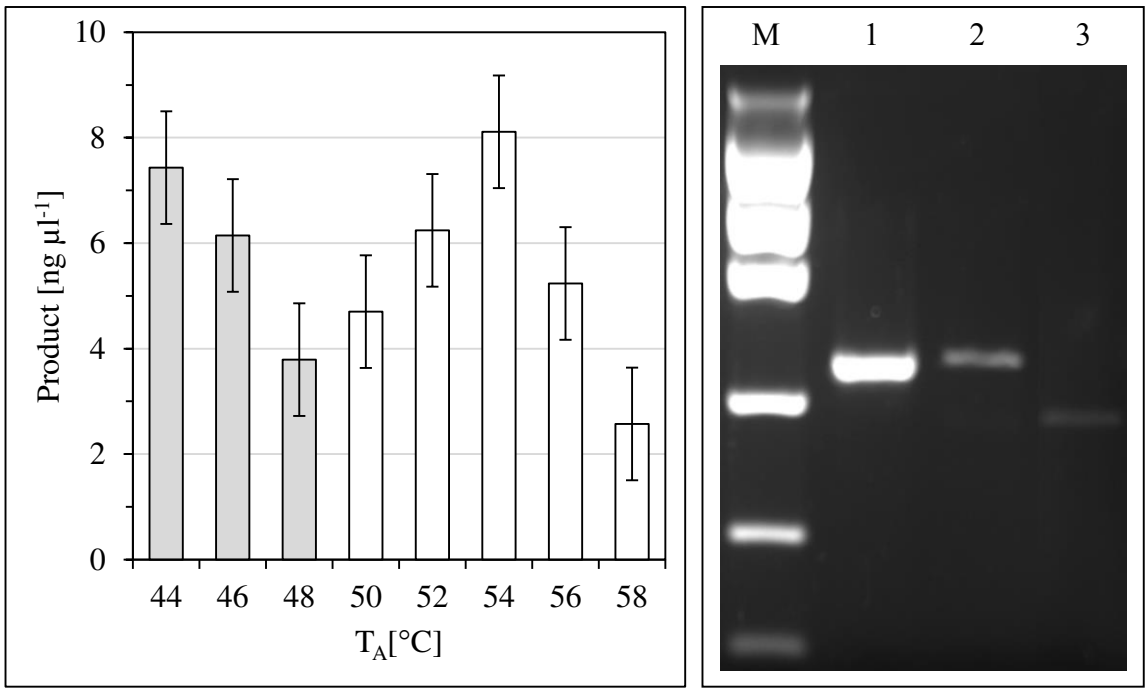


Figure 21: (LEFT) Mean phaC(3.1) amplification product concentrations from *Synechocystis* sp. PCC6803 and *Spirulina platensis* UTEX LB 2340. Shaded bars depict conditions which undesired byproduct formation was observed. Confidence intervals on pooled variance are depicted ($\alpha = 0.10$). (RIGHT) Agarose gel electrophoresis (two percent standard agarose, 0.5 µg ml^{-1} ethidium bromide stain, 0.5xTBE) of 6 µL of the raw PCR *phaC* detection primer set amplification product at an annealing temperature (T_A) of 44°C.

Lanes:

(M) Low DNA Mass Ladder (Invitrogen, 10068-013), (1) *Synechocystis* sp. PCC 6803, (2) *Spirulina platensis* UTEX LB2340, (3) *Plectonema* sp. UTEX 1541.

A significant improvement in the reliability of the assay was observed with an increase in the amount lysate template added to the reaction. PCR amplifications using both five and 10 µl of the DMSO/TE lysate showed no unexpected failures. The mean PCR product concentrations for these reactions can be observed in Figure 23 and electrophoresis images can be found in the appendix (Appendix Figure 4, A.10). There was no undesired byproduct formation, aside from primer-dimers, observed across all nine of the cyanobacterial genomes as can be observed in the appendix.

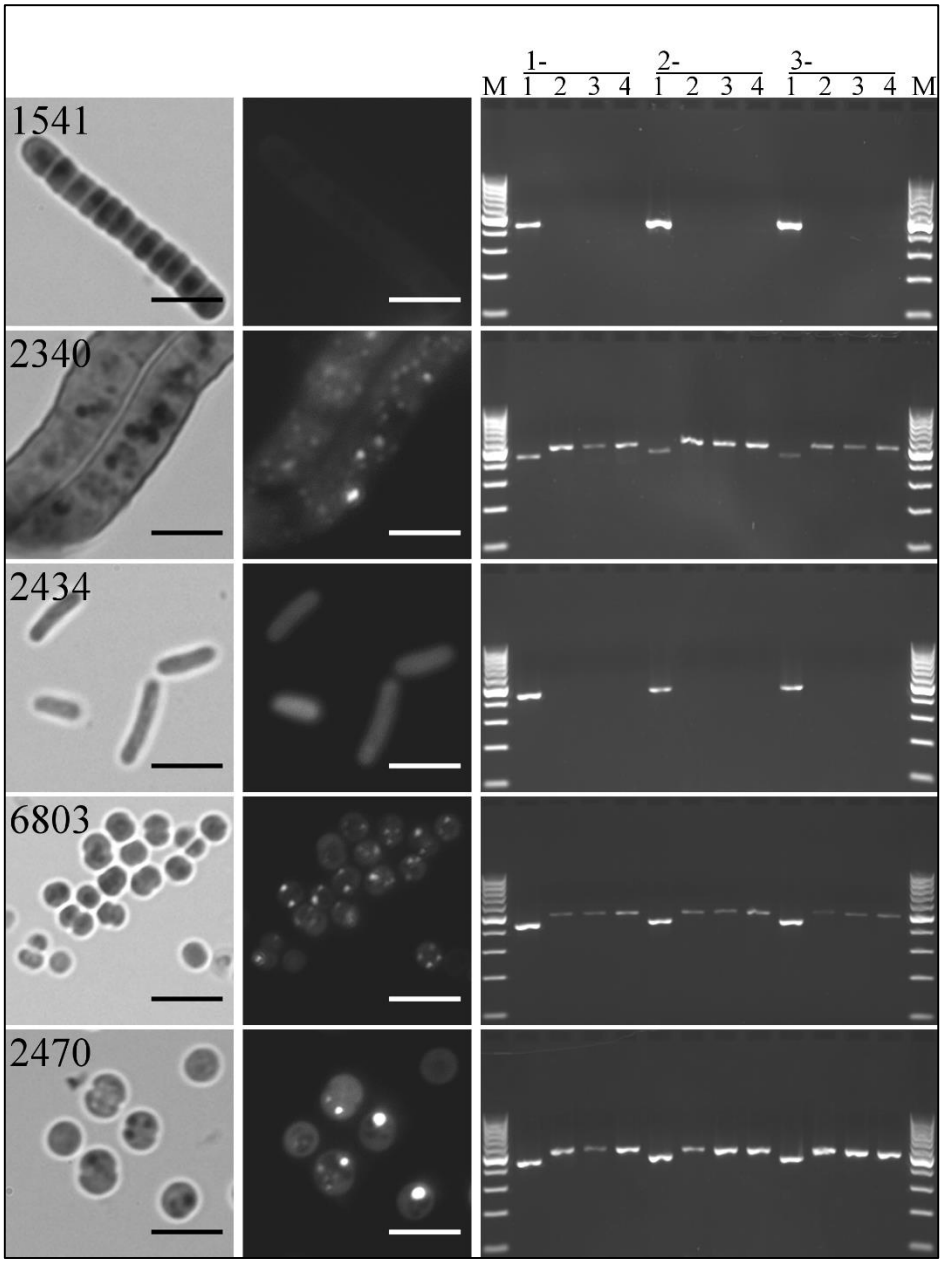


Figure 22: (LEFT and CENTER) Brightfield microscopy images (left) and epifluorescence microscopy images (center) of cyanobacteria stained with Nile blue A and grown in nutrient-limiting medium for PHA granule detection. Scale bar represents 5 µm.
(RIGHT) Agarose gel electrophoresis (2% standard agarose, 0.5 µg ml^{-1} ethidium bromide stain, 0.5xTBE) of 5 µL of each PCR amplification product using the primer set developed in this work for three samples (n = 3) (1-x, 2-x, 3-x) in triplicate for each respective cyanobacteria.

Lanes: (M) 5 µL of 100 bp O'GeneRuler DNA ladder, (y-1) *cpcA* control primer set, (y-2, y-3, y-4) *phaC* detection set developed in this work.

Rows: (1541) *Plectonema* sp. UTEX 1541 (2340) *Spirulina platensis* UTEX LB 2340, (2434) *Synechococcus leopoliensis* UTEX 2434, (6803) *Synechocystis* sp. PCC 6803,
(2470) *Synechocystis* sp. UTEX 2470.

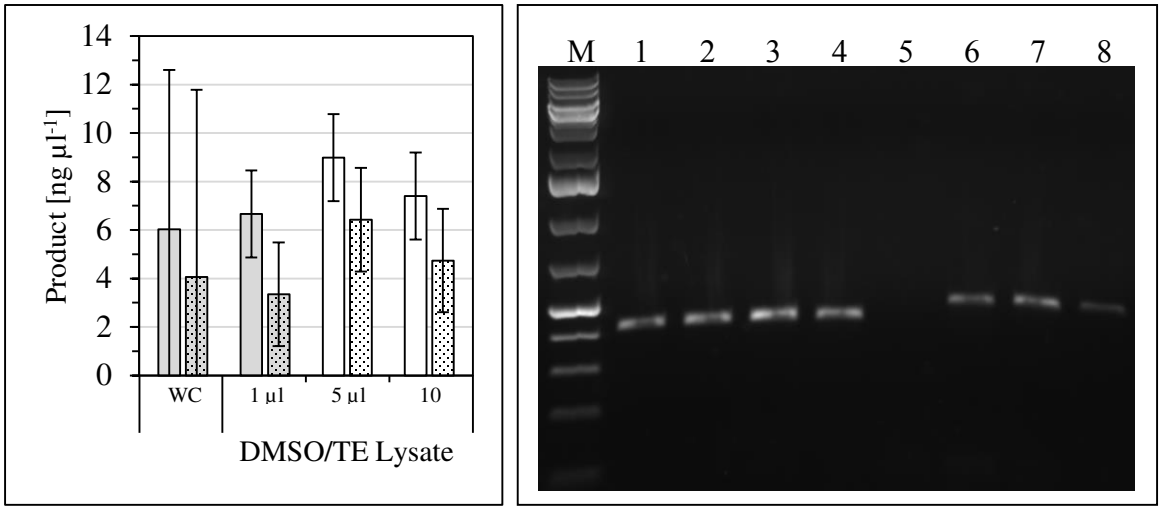


Figure 23: (LEFT) Mean PCR amplification product concentrations from the nine cyanobacteria strains tested using various template DNA sources from this study. Shaded bars depict conditions which unexpected failures were observed. Confidence intervals are depicted ($\alpha = 0.05$). (RIGHT) Agarose gel electrophoresis (one percent standard agarose, 0.5 µg ml^{-1} ethidium bromide stain, 0.5xTBE) of 10 µL of the raw colony PCR amplification products of *Synechocystis* sp. PCC 6803. Note that volumes associated with DMSO/TE describe the amount of lysate used as PCR template and do not denote volume loaded into gel.
Lanes:
(M) 5 µL GeneRuler 1kb plus DNA ladder, (1) WC *cpcA*, (2) 1 µL DMSO/TE *cpcA*, (3) 5 µL DMSO/TE *cpcA*, (4) 10 µL DMSO/TE *cpcA*, (5) WC *phaC*, (6) 1 µL DMSO/TE *phaC*, (7) 5 µL DMSO/TE *phaC*, (8) 10 µL DMSO/TE *phaC*.
WC – whole cells.

5.5 Discussion

The analysis of the type III PHA synthase PhaC subunit multiple sequence alignment showed high conservation in these enzymatic subunits throughout the characterized cyanobacteria. The conservation observed in the alignment is further support to the claims of Hai et al. that this enzyme class is widespread and general in PHA-accumulating cyanobacteria. The cyanobacteria-specific insertion, which was first described with four cyanobacteria sequences, was further analyzed with 29 sequences and proved to remain well conserved.

The results from the PCR-based assay were in agreement with the results observed in the traditional staining method, indicating that the presence of the *phaC* gene shows some merit in being a potential indicator of PHA accumulation abilities in cyanobacteria. The confirmed positive detection of *phaC* in PCC 6803 showed that this low degeneracy primer set is quite robust, since the majority (69%) of the

sequences within the alignment had a higher relative annealing ability. It should also be noted that the successful amplification of NIES-843 suggests this primer set should be viable for the 12 other *Microcystis* species in the *phaC* alignment because all of these accessed sequences possess identical primer-template annealing regions to NIES-843. Together, these conclusions suggest that this rapid PCR assay would make a worthwhile addition to any cyanobacteria PHA accumulation capability assay in order to greatly decrease avoidable costs in downstream optimization.

The DMSO/TE method of quick preparation for colony PCR proved far more reliable than whole cell colony PCR amplification, especially when using five or 10 µl for template. Additionally, the DMSO/TE method of preparation provides approximately 20x 50 µl-reactions-worth of testable DNA template, meaning this assay could be expanded with more primer sets to simultaneously test for other key PHA metabolism associated genes/regulatory sequences of interest as they are discovered.

Algae strain isolation can be a difficult and time consuming task which is typically accomplished by serial dilution and plating (Yeesang & Cheirsilp, 2011). This PCR assay could be implemented on small polyalgal colonies to detect if any of the algae possess the *phaC* gene and warrant further isolation steps. Since this PCR assay is inherently sensitive, it could be a valuable tool in the very early in the stages of strain isolation.

Following the discovery of desirable PHA producing cyanobacteria, typical culture condition optimization could improve PHA yields. Genetic modification could also be employed since some cyanobacteria have proven readily, if not spontaneously, transformable (Kufryk, Sachet, Schmetterer, & Vermaas, 2002; Vermaas, 1996). Antibiotic-free genetic modifications are already being implemented in typical PHA producers (Akiyama et al., 2011). These steps could ultimately lead to the discovery of an environmentally friendly and economically viable plastic production process. A carbon-neutral bioplastic production process would not only lessen dependence on petrochemicals, but also play a key role in slowing the accumulation of non-biodegradable solid wastes

5.6 (*ex post facto*) *In Silico* Analysis

Since the work performed in the previous sections of this chapter was completed before the existence of the computer application developed in CHAPTER 6, it was decided to try to gain more insight on the robustness of the phaC(3.1) primer set using the computer application and newly available *phaC* sequences (if extant).

Four additional cyanobacteria PhaC/*phaC* sequences were obtained. Two reference accessions: *Microcystis aeruginosa* NIES-44 (GAL94730.1) and *Microcystis aeruginosa* SPC777 (EPF22606.1), and two hypothetical proteins: *Nostoc punctiforme* PCC73102 (ACC80869.1) and *Xenococcus* sp. PCC7305 (ELS02364.1). These new sequences were included with the sequences used to construct the previous PhaC MSA and aligned using Clustal Ω (Sievers et al., 2011). The hybridization algorithm was implemented across all templates using the actual PCR conditions described in section 5.3. Statistical analysis was applied using R and Statistical Analysis System (SAS v9.4).

In order to gain further insight on the phaC(3.1) primer set *ex post facto*, the Gibbs free energies of annealing were estimated using the computer application designed in CHAPTER 6.

5.6.1 Alignment Results

The protein MSA including the newly accessed sequences can be found in the appendix section A.7. Interestingly, the two hypothetical proteins (*Nostoc punctiforme* and *Xenococcus* sp. accessions) show very little conservation in comparison to all other cyanobacterial PhaC sequences. These hypothetical proteins even lack the "Cyanobacterial-box", which could reshape what is known about PhaC in cyanobacteria. After similarity searches, it was found that these PhaC sequences more closely resemble those found in proteobacteria. These two accessions were omitted from further calculations since the existence of these protein sequences has yet to be validated.

5.6.2 Hybridization Results

It was observed that estimation of ΔG at actual reaction conditions, in lieu of the conditions of section 5.3.3 (37°C, 1.0 M NaCl), had little impact on the overall results. The new hybridization method estimated the phaC(3.1)-PCC 6803 hybrid to be within the 32nd percentile, a decrease from the 31st percentile obtained in previous calculations. In other words, the primers anneal to 68% of the accessed sequences in a more stable fashion than PCC6803, which was proven to successfully amplify. The most stable primer-template hybridization ($\Delta G_{i,\ MIN}$) of all permutations of the primers ranged between $-13 \leq \Delta G_{i,\ MIN} \leq -2.0$ kcal mol^{-1} for phaC(3.1)-F, $-15 \leq \Delta G_{i,\ MIN} \leq -1.0$ kcal mol^{-1} for phaC(3.1)-R, and $-27 \leq \Delta G_{i,\ MIN} \leq -4.6$ kcal mol^{-1} for the set.

5.6.3 Proposed Alternative Primer Set

It was observed that many primer-template hybrids would flag as "unlikely to amplify" under the CEMAsuite default conditions ($\Delta G_{F|R} > -6$ kcal mol^{-1}, $\Delta G_{F+R} > -16$ kcal mol^{-1}, confer section 6.6), including the PCC 6803 accession, which successfully amplified. However, as an exercise of the application, an alternative primer set was designed and proposed with slightly increased degeneracy to potentially improve the stability of the primer set across all accessed templates.

The proposed forward primer, phaC(3.1)-F*, sequence is 5'-GGCTTAGATATCTAYTTYATTGAYTGG and the proposed reverse primer, phaC(3.1)-R*, sequence is 5'-GCATAAACTAAAGGTTCCNCCYTGRCA. The forward and reverse primer annealing regions correspond to positions 340-366 and 490-516 of the *Synechocystis* sp. PCC 6803 accession (CP003265.1: (933151..933177, 933301.. 933327)). The 3' end of phaC(3.1)-R* falls exactly on the CQG residues within the lipase-like box. Hybridization was also performed on this set at identical conditions as phaC(3.1). The results from each set are displayed in Figure 24. The $\Delta G_{i,\ MIN}$ value for phaC(3.1)* ranged between $-12 \leq \Delta G_{i,\ MIN} \leq -2.5$ kcal mol^{-1} for phaC(3.1)-F*, $-20 \leq \Delta G_{i,\ MIN} \leq -10$ kcal mol^{-1} for phaC(3.1)-R*, and $-32 \leq \Delta G_{i,\ MIN} \leq -14$ kcal mol^{-1} for the set. The proposed set dramatically increased the minimum

observed ΔG_{F+R} estimate from -4.6 to -14 kcal mol^{-1}. The proposed set reduced the number of individual primers warnings in CEMAsuite from 21 to 9 and the primer set warnings from 5 to 4.

The mean $\Delta G_{i, MIN}$ value for each respective primer category was estimated to be more stable in phaC(3.1)* when compared to phaC(3.1) ($p_F = 0.0013$, $p_R = 0.0018$, $p_{F+R} = 0.0002$).

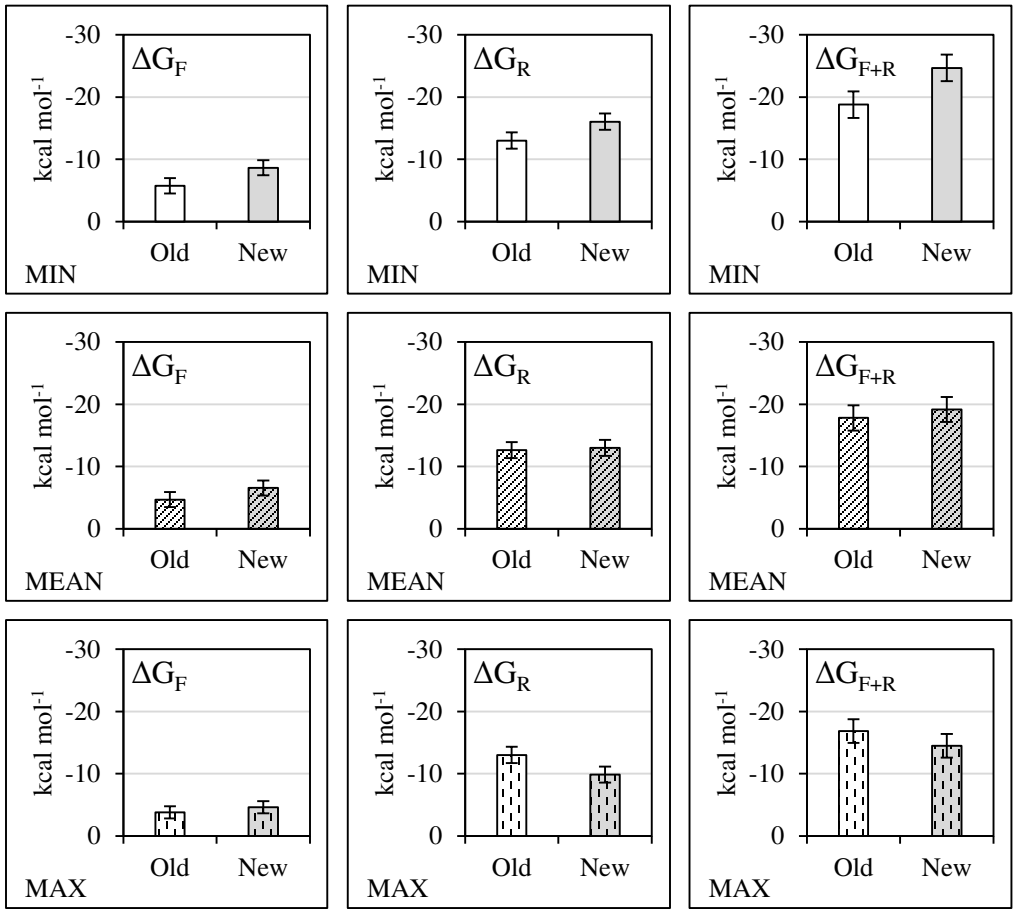


Figure 24: Gibbs free energies of annealing estimated via CEMAsuite hybridization algorithm on 33 cyanobacteria *phaC* sequences. Confidence intervals depicted as error bars ($\alpha = 0.05$). Confer Figure 30 for ΔG estimates.

The mean minimum predicted stability of the new primer set is significantly more stable than the original low degeneracy set for any case. The minimum predicted $\Delta G_{i, MIN, F+R}$ for the set was reduced from -1.0 to -14 kcal mol^{-1}, which should result in a dramatic increase in the stability of the primer-template interaction. Because of the loss of specificity due to the increase in degeneracy for the new primer set, these results do not suggest the new primer set is better suited for a high-throughput assay. Before this conclusion can be tested, the effects of the added degeneracy should be studied experimentally.

5.7 Summary

The PCR primers designed within this section proved to be a robust primer set capable of detecting cyanobacteria which have the potential to produce PHA. The PCR-based assay was hastened though the implementation of the colony PCR method developed in this work. *Ex post facto* analysis revealed the PCR primers designed within this work, while proven experimentally, where not optimal in regards to the default CEMAsuite hybridization criterion and a second, potentially more robust, primer set was proposed. The proposed set was found to be more stable across the cyanobacteria *phaC* sequences than the original set, but the specificity of the proposed set remains to be tested.

CHAPTER 6. CEMASUITE

6.1 Chapter Preface

This chapter discusses the end product of Objective II (section 1.4) – CEMAsuite. CEMAsuite is the consensus degenerate primer design computer application first described in Oxford Journals' *Bioinformatics* (Courtney E. Lane, Hulgan, O'Quinn, & Benton, 2015). Due to the brief nature of said publication, this chapter expands upon the key dynamics and implementations of CEMAsuite. For an overview of existing consensus degenerate primer design software confer section 2.8.

6.2 Introduction

It was observed that the CEMA-based primer design methodology implemented throughout Objective I was applicable to any arbitrary set of homologous genes. So, in order to reduce the time and effort required for similar primer design, a computer application was developed. This application, named CEMAsuite, was written in the Java™ coding language (Java 7 SDK) and developed in the NetBeans integrated development editor (v. 8.0).

CEMAsuite was developed in an attempt to find a compromise between the two consensus primer design methods described in section 2.8. Its intent is to aid in the design of a sort of minimum-degeneracy primer set which is robust enough for the assay at hand, while retaining as much specificity and product homology (*i.e.*, little variance in product length) as possible. It was desired to grant the user total control over the degenerate primer construction process, which in turn allows the balancing of the specificity and sensitivity of the consensus degenerate PCR primers.

CEMAsuite possesses the following capabilities:

(i) Construction of a CEMA from protein MSA file
(ii) Generation and scoring of each position in the consensus DNA sequence using multiple algorithms
(iii) Design of single-degeneracy primer backbones using Primer3 (Untergasser et al., 2012)
(iv) Estimation of degenerate primer stability on each of the coding sequences within the CEMA
(v) Manual editing of primer residue degeneracy
(vi) Intuitive presentation to allow for rapid analysis of alignments and speculation of primer degeneracy incorporation

6.3 CEMA Construction

Under the equiprobability assumption, the probability that all residues of an arbitrary column in an arbitrary MSA will match the first residue in that column due to random chance alone can be described by Equation 4, where R is the number of observed residue states (+1 for deletion), and n is the number of sequences in the MSA. This type of conservation can be considered a sort of Type I error where the conservation is present, but due to random chance alone, and not necessarily from evolutionary pressures. This conservation is undesirable in probe design since it has no influencing factors to remain conserved, meaning the use of primers based on these regions on a homologue with unknown sequence may result in a loss of primer annealing ability. For this reason, CEMA-suite begins the conservational analysis for primer design on the protein level, where $R = 20$, as opposed to coding DNA, where $R = 4$.

Equation 4: Probability that all residues of an arbitrary column in an arbitrary MSA will match the first residue in that column due to random chance alone under the equiprobability assumption.

$$\mathrm{P} = (R + 1)^{(1-n)}$$

A CEMA is generated by obtaining the primary coding sequences of proteins in an extant MSA and expanding each position within the protein MSA to the observed codon representing each amino acid. For example, a protein sequence `CLANE` is encoded by `TGC CTA GCA AAC GAA`. In a protein MSA including `CLANE`, the following gaps are included `C-LANE-`. The resulting analogous CEMA sequence

generated would be TGC---CTAGCAAACGAA---. This process is repeated for all sequences within the input alignment to generate a CEMA.

A consensus sequence is then generated from the CEMA. This consensus sequence will not allow residue values for alignment positions containing gaps. In other words, if a gap occurs anywhere within a given column of the CEMA, the consensus residue of that column will be a gap value.

6.4 CEMA Positional Scoring

Since many primer design applications can account for positional quality in their objective functions, CEMAsuite can score each position within a CEMA's consensus sequence by one of four algorithms. The *Percent Identity* algorithm scores each position based on the normalized frequency of the consensus nucleotide. The *Identity Runs* algorithm scores positions on identity and then adjusts the value based on the number of consecutive completely conserved positions within the location. The *Potential Degeneracy* algorithm scores positions on identity and then adjusts the value based on the potential degeneracy of the consensus codon positions according to one of 18 translation tables. The final algorithm, *Runs & Degeneracy*, scores values sequentially using each of the three methods described above.

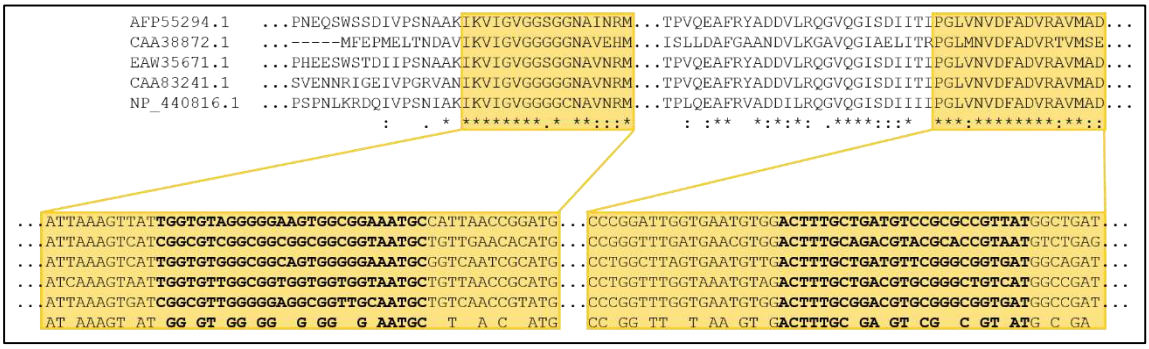


Figure 25: Region of example protein MSA (TOP) and respective CEMA (BOTTOM) covered within the following score examples. Regions depicted in score examples are highlighted in yellow above.

6.4.1 Percent Identity

In this method, CEMA positions are scored simply on the normalized frequency of the consensus nucleotide throughout the sequences (Figure 26). This method is most useful when many sequences are

available/within the alignment OR when the amplification of unknown sequences will likely not be attempted. This is the default scoring method.

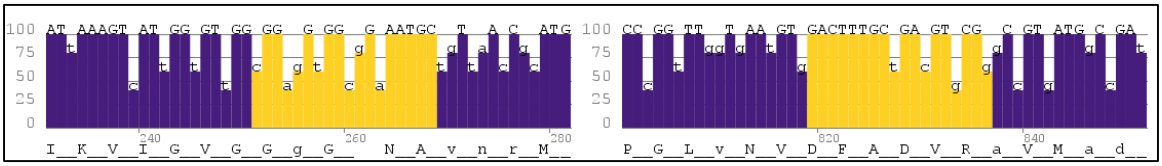


Figure 26: Plot of quality from highlighted region of Figure 25 scored using the Percent Identity method. The CDS residue for each position is printed on top of the bar, while the conserved amino acid is printed below the bars.

6.4.2 Identity Runs

In this method, CEMA positions are scored on Percent Identity and then the scores are adjusted based on the number of *consecutive completely conserved positions* within the location (Figure 27). The score adjustment value is specified by the user. The run weight should always be a positive integer less than or equal to 100.

To illustrate how this algorithm adjusts a score, a simple example will be discussed. Using a block weight of BW and an isolated region of n completely conserved nucleotides, the value of n×BW will be added to the score of each nucleotide within that locus. No value is added to any region which is not completely conserved. Once all conserved regions have been adjusted throughout the sequence, all position scores within the sequence are then normalized to the new maximum position score. For example, note the scores of `GACTTTGC` and `AATGC` in Figure 27, which were calculated with a block weight of 10. Here `GACTTTGC` (n = 8) is the largest region of the most consecutive completely conserved residue positions within the entire CEMA. The region's score was calculated via 100 × ((100 + 8×10) ÷ (100 + 8×10)) = 100. Meanwhile, `AATGC` (n = 5) possesses less weighting even though it is a completely conserved region. This regions score was calculated via 100 × ((100 + 5×10) ÷ (100 + 8×10)) = 83⅓.

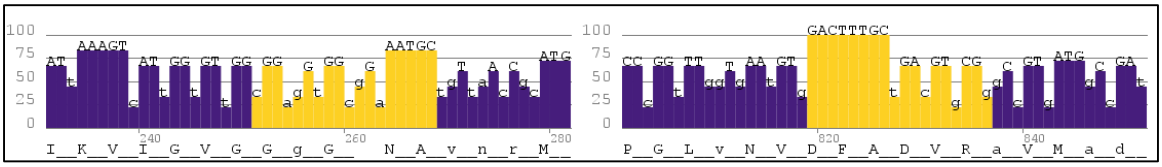


Figure 27: Plot of quality from highlighted region of Figure 25 scored using the Identity Runs method and a run weight of 10. The CDS residue for each position is printed on top of the bar, while the conserved amino acid is printed below the bars. Note that positions of 100% conservation (denoted by capital letters) possess quality scores of less than 100.

This method is most useful when many sequences are available and/or implemented within the alignment. This scoring method can be used to filter out the regions where runs of perfect matches will not occur. The regions of high quality are key regions to investigate for the 3' end of the primer.

6.4.3 Potential Degeneracy

In this method, CEMA positions are scored on Percent Identity and then then the scores are adjusted based on the potential degeneracy of the consensus codon positions according to 18 translation tables (Figure 28). In other words, if the consensus codon within the alignment is `CGT` and it is desired to adjust scores using the standard translation table (`CGT` codes for Serine which can be coded by [`CGN`, `AGR`] ~ `MGN`), then the quality of `C`'s position will be divided by 2, the quality of `G`'s position will be divided by 1, and the quality of `T`'s position will be divided by 4.

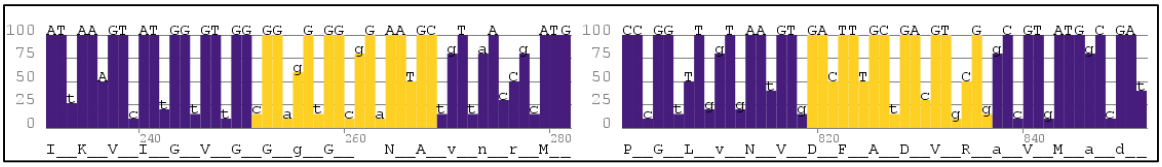


Figure 28: Plot of quality from highlighted region of Figure 25 scored using the Potential Degeneracy method and the standard translation table. The CDS residue for each position is printed on top of the bar, while the conserved amino acid is printed below the bars. Note that positions of 100% conservation (denoted by capital letters) possess quality scores of less than 100.

This method is most useful when there are *few* sequences available/within the alignment as it attempts to filter out regions of low conservation and high potential degeneracy. This is a method which can be useful for the cases where the primers will be used to try to amplify on organisms with unknown target sequences.

6.4.4 Identity Runs & Potential Degeneracy

In this method, CEMA positions are scored using each of the 3 scoring methods described above (Figure 29). This method can help to discover regions of high conservation (from Identity Runs) with low potential degeneracy (from Potential Degeneracy).

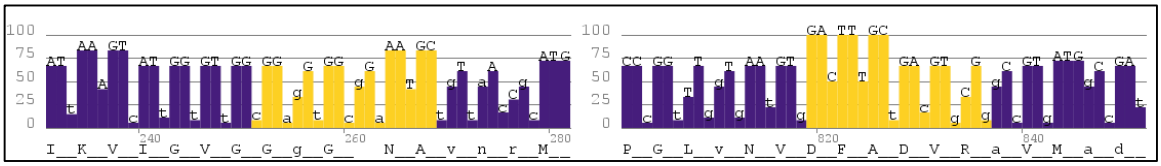


Figure 29: Plot of quality from highlighted region of Figure 25 scored using the Runs & Degeneracy method with a block weight of 10 and the standard translation table. The CDS residue for each position is printed on top of the bar, while the conserved amino acid is printed below the bars. Note that positions of 100% conservation (denoted by capital letters) possess quality scores of less than 100.

6.5 Consensus Primer Design

CEMAsuite incorporates a compiled Primer3 (v 2.3.6) executable and offers a streamlined user interface (Untergasser et al., 2012). Briefly, Primer3 is an open source primer design project which efficiently produces a list of potential primer sets for a given DNA template. Primer3 allows for the input of a DNA template sequence and the positional quality of that sequence which is typically implemented for DNA sequencing trace files. CEMAsuite utilizes this capability for a different purpose. CEMAsuite inputs the consensus sequence generated from the CEMA and inputs the scoring algorithm results for each positional quality input.

For a complete list of inputs, please reference the Primer3 documentation. One notable input is the MIN_END_QUALITY. Briefly, the MIN_END_QUALITY input will set a threshold on the minimum quality score allowed in the 3' end of the primer. This will only allow the output of primer sets which possess a 3' end with qualities greater than the specified value.

Implementation of this process should result in a single degeneracy primer of high quality for the consensus CEMA sequence. However, it should be noted that the score plot (Figure 26 - Figure 29) can greatly speed up the process of primer design by inspection if desired.

6.6 Hybridization Stability Estimation

One of the key elements of CEMAsuite is the ability to anneal the primers to each template and output an estimated Gibbs free energy for the designated conditions. This allows the user to pinpoint cases where the primer set is likely to fail and improve the primers as they see fit. It is recommended that the conditions set are the actual PCR reaction conditions and the annealing temperature of the thermal cycles.

The algorithm for the thermodynamic parameter estimation first locates the primer-template region (columns) with the least mismatches throughout ALL sequences for each primer. Next, it simulates annealing for each primer-template pair *in this region* (*i.e.*, iterates down through the columns for new templates) utilizing the nearest-neighbor parameter estimation methods described by Allawi, Bommarito, Peyret, and SantaLucia and their respective coworkers (Allawi & SantaLucia, 1997; Hatim T. Allawi & John SantaLucia, 1998a, 1998c; H. T. Allawi & J. SantaLucia, 1998; Bommarito, Peyret, & Jr, 2000; Peyret et al., 1999; SantaLucia, 1998; SantaLucia & Hicks, 2004; SantaLucia J Jr & N., 2001). In order to account for the entropic dependence of hybridization on the cationic concentration, two methods of adjustment have been implemented (Owczarzy et al., 2008; SantaLucia & Hicks, 2004).

If a primer is degenerate, *each permutation of that degenerate primer is simulated individually*, and the most stable conformation is used to populate the mean/min/max Gibbs free energy values. The mean value is the average of the most stable conformation of all permutations of a primer annealing. The minimum value is the most stable conformation of the *most* stable permutation of a degenerate primer. The maximum value is the most stable conformation of the *least* stable permutation of a degenerate primer. This part of the algorithm is outlined in Figure 30.

In order to obtain some insight on exactly what constituted a “good” primer set based on our hybridization algorithms, 94 data points were obtained through literature and subjected to the stability analysis at the specified conditions (de Roda Husman, Walboomers, van den Brule, Meijer, & Snijders, 1995; Ishii & Fukui, 2001; Snijders et al., 1990; Yamamoto & Harayama, 1995). For these calculations,

the annealing temperature (T_A) of the thermal cycles was used as the input temperature and positive detection was taken as it was cited within the literature (as were failures). Overall, there were 29 observations of failures and 65 observations of strong amplification.

The stability of the individual primers was analyzed first, these were sorted based on the relative stabilities of the oligos within the set (*i.e.*, one deemed "more stable" and one deemed "less stable"). The resulting $\Delta G_i(T_A)$ values of the individual primers were binned into 1.0 kcal mol^{-1} bins and plotted on a histogram.

To see the effects of the overall binding ability of the primer set, the sum of the two binding energies ($\Delta G(T_A) = \Delta G_F(T_A) + \Delta G_R(T_A)$) was estimated and plotted on a 2.5 kcal mol^{-1} binned histogram.

In order to discern potentially "good" primers from potentially "bad" primers, the hybridization algorithm in CEMAsuite was implemented on samples from literature. The stability estimations were calculated using both entropic adjustment methods (Owczarzy et al., 2008; SantaLucia & Hicks, 2004).

To attempt to describe failed PCR amplifications due to the annealing ability of a single primer, the estimated Gibbs free energy of annealing was calculated for each primer in the set at the conditions stated in literature. These values were then categorized by their predicted stability for each set. The results of this analysis are depicted in Figure 31.

It was observed that the Owczarzy *et al.* adjustment method generally predicts more stable hybridizations than that of SantaLucia and Hicks. This is believed to be due to the ability of the Owczarzy correction to account for stabilization effects of the divalent cations in solution. The Owczarzy *et al.* adjustment was chosen as the default correction since there was less observed overlap between the less stable primers from the failed and successful amplifications as observed by the red bars in Figure 31.

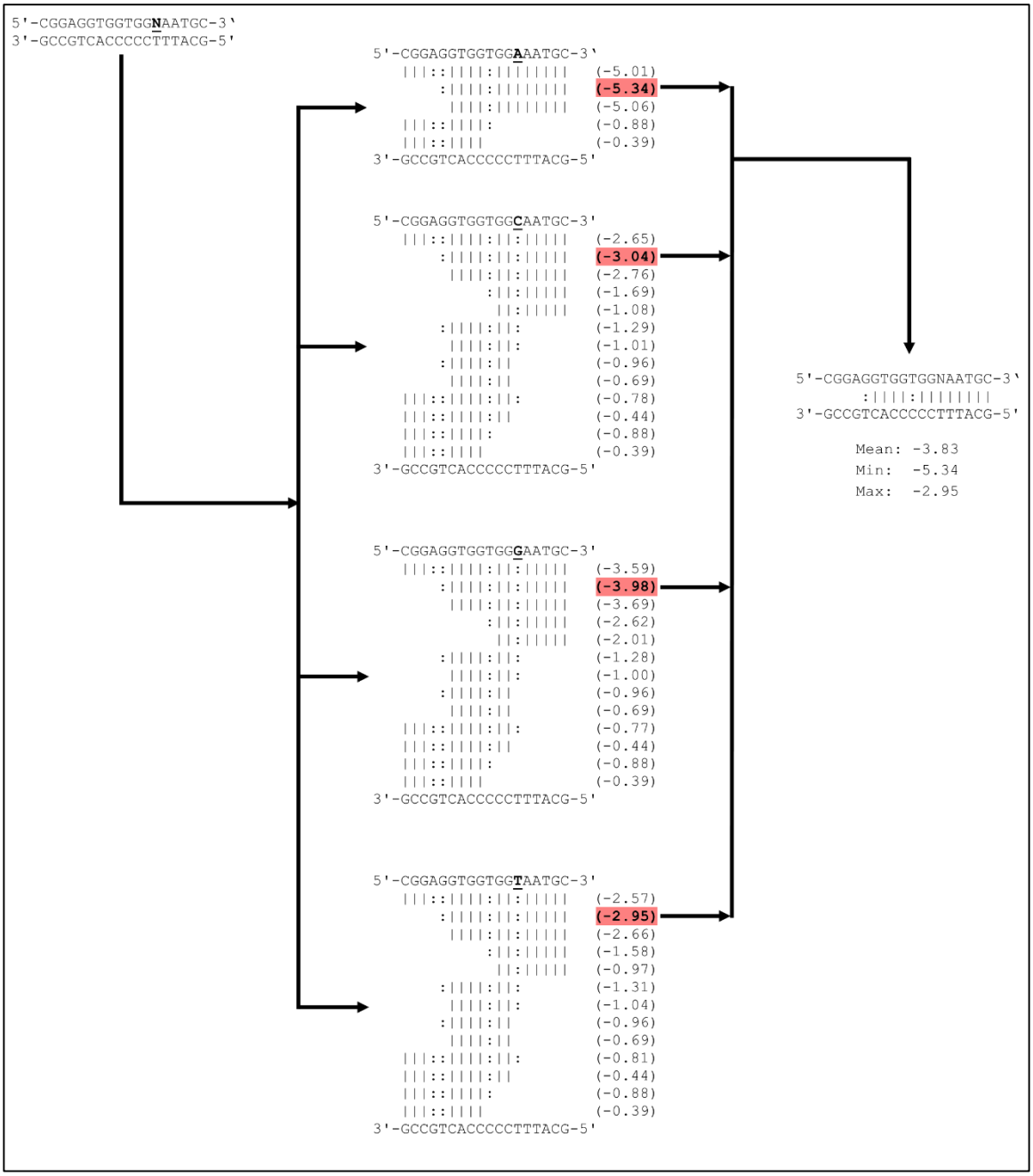


Figure 30: Primer-template annealing algorithm. First, all possible single-degeneracy permutations of a degenerate primer are created. Next, each of these single-degeneracy primers is used to simulate annealing to the template region. The stability of each potential hybridization conformation is estimated for each single-degeneracy primer via nearest neighbor thermodynamics (see text for details). Each possible primer-template hybridization confirmation hydrogen bonding is depicted using ' | ' and ' **:** '. The result is a Gibbs free energy value (kcal mol^{-1}) at the specified conditions as indicated by the numbers in parentheses next to each possible conformation. Statistical analysis is then performed on the most stable conformation of each single-degeneracy primer-template hybridization and returned as the output values associated with that particular degenerate primer-template hybridization.

' | ' – Watson-Crick base pairing, ' **:** ' – interacting mismatch, ' ' – no interaction.

It was also observed that strong amplification begins to fail when the weaker of the two oligos had a $\Delta G_i(T_A)$ value approximately −6.0 kcal mol^{-1} when applying the default correction. This value was then used as the default *individual primer warning threshold* in CEMAsuite. In other words, if a predicted primer-template hybridization $\Delta G_i(T_A)$ value is less stable (more positive) than the *individual primer warning threshold* value, then that primer set will be flagged. The mean $\Delta G_i(T_A)$ value of the weaker oligos for successful amplifications was −9.5 (±0.7) kcal mol^{-1} and −5.9 (±0.7) kcal mol^{-1} when applying the Owczarzy *et al.* and SantaLucia and Hicks corrections, respectively.

To investigate the effects of the overall binding ability of the primer set, the sum of the two binding energies ($\Delta G(T_A) = \Delta G_F(T_A) + \Delta G_R(T_A)$) was estimated and analyzed. It was observed that this value approached approximately −16 kcal mol^{-1} before failures became prevalent when applying the default correction. This value was used as the default *primer set warning threshold*. The mean sum binding energy for successful amplification was −24 (±1.2) kcal mol^{-1} and −16 (±1.2) kcal mol^{-1} when applying the Owczarzy *et al.* and SantaLucia and Hicks corrections respectively. The results of this analysis are depicted in Figure 32.

6.7 Usage

A Java swing-based interface is deployed containing tabs that allow the user to visualize the stepwise primer design process. Before using CEMAsuite, the user must first possess a clustal format protein MSA file of the homologue-of-interest. This file can be imported and displayed under the *Protein MSA* tab.

Next, the user may choose to either upload the coding sequences from a local file, or attempt to retrieve them from the National Center for Biotechnological Information (NCBI) database using the Entrez programming Efetch utility (Geer et al., 2010). Once the sequences have been successfully uploaded, they will be displayed under the *CDS* tab.

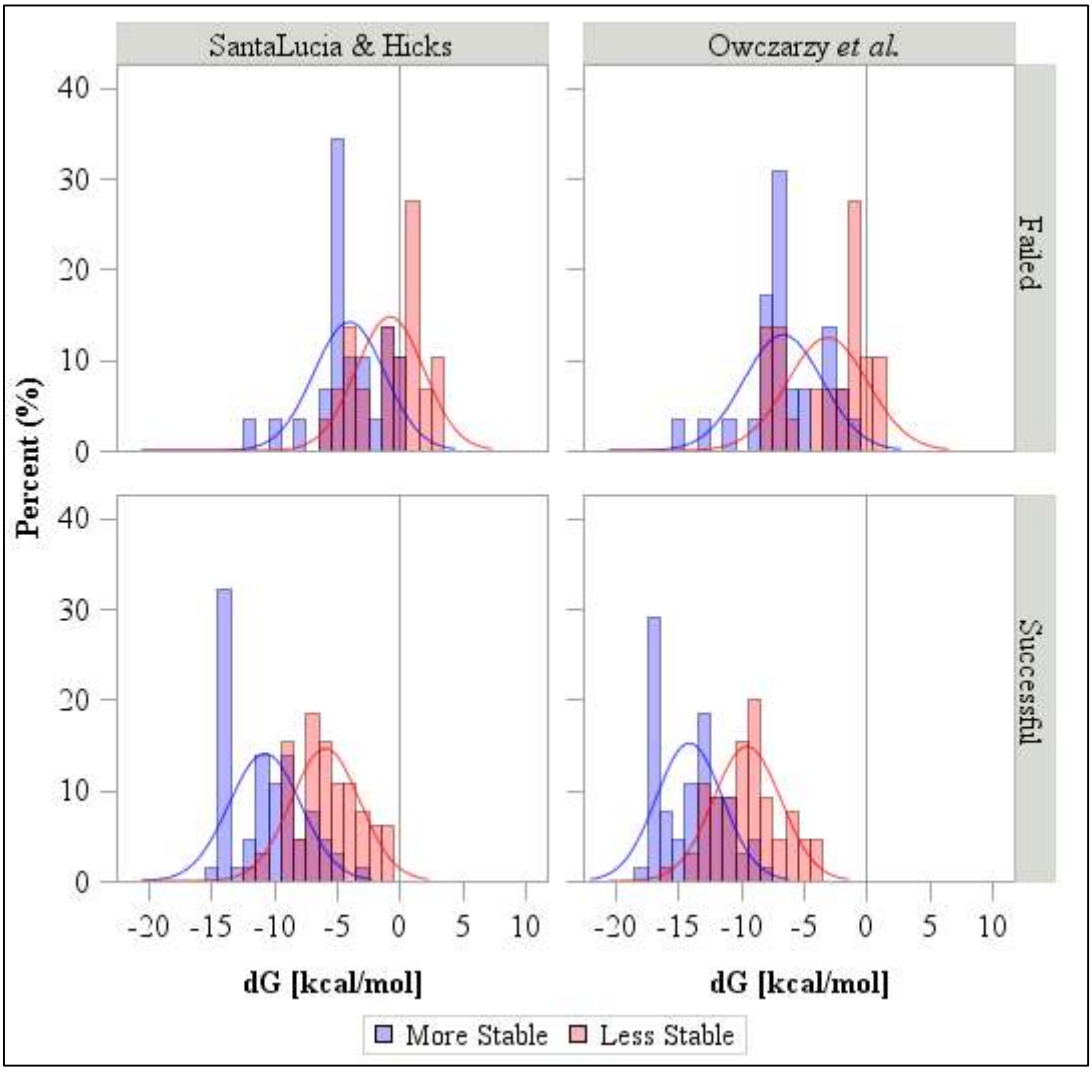


Figure 31: Histograms of the estimated $\Delta G_i(T_A)$ values from data taken from literature and calculated using the hybridization algorithm within CEMAsuite. Detection value was based off of considerations listed within literature. The colored lines on the plot represent the normal density of each sample population.

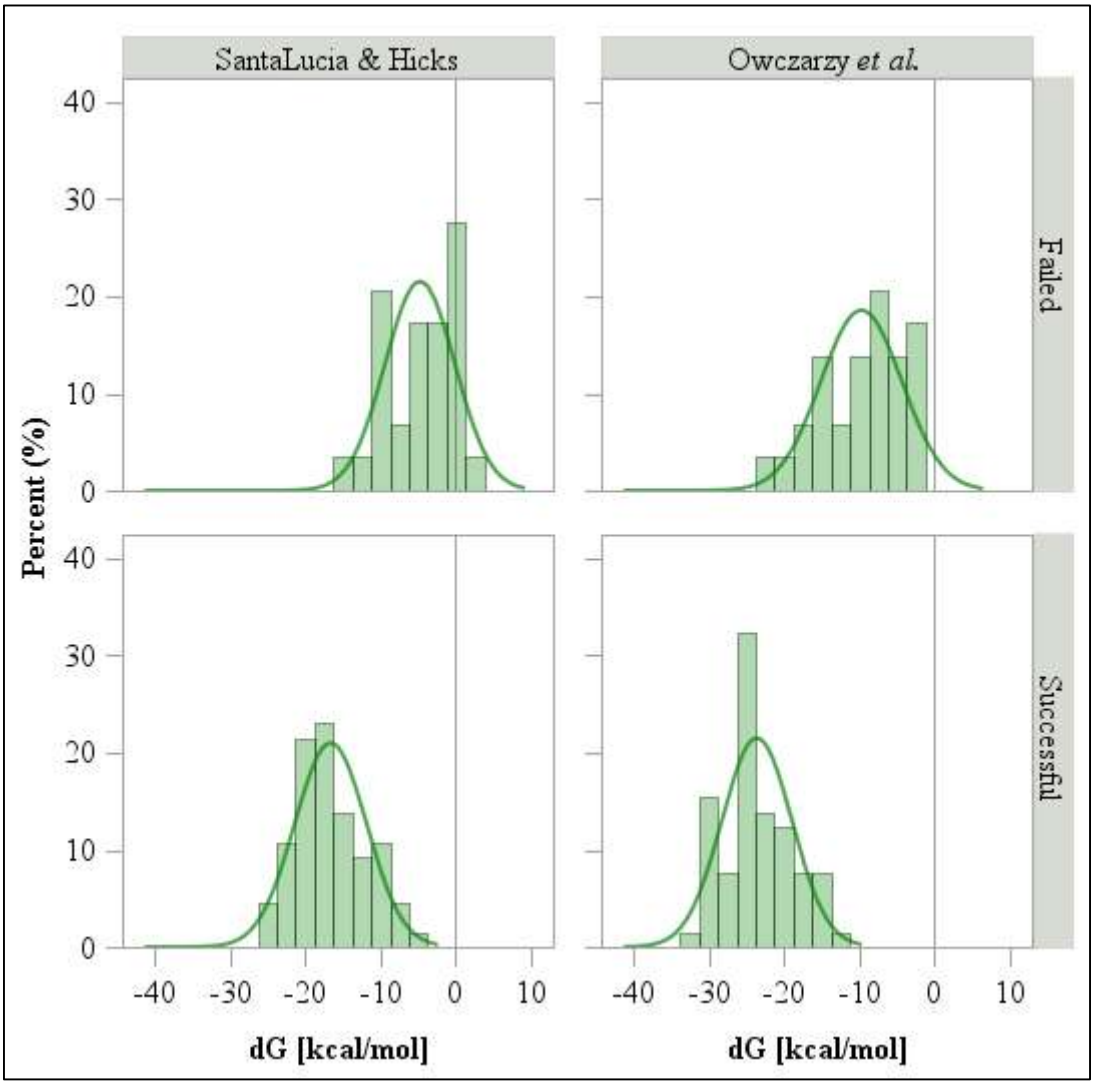


Figure 32: Histograms of the sum of the forward and reverse $\Delta G(T_A)$ values from data taken from literature and calculated using the hybridization algorithm within CEMAsuite. Detection value was based off of considerations listed within literature. The colored lines on the plot represent the normal density of each sample population.

The CEMA can be generated and displayed under the *CEMA* tab (Figure 33, left). The user may select one of the four consensus scoring methods described in section 6.4. At this point, the scored consensus sequence will be visible as a bar plot underneath the tabs (Figure 33, left).

Once a scored consensus sequence has been obtained, CEMAsuite can access a compiled Primer3 executable which can return potential primer sets for the consensus sequence displayed under the *Primer Design* tab. This functionality is limited to Windows systems, however the scored consensus information can be readily exported for input into alternative design applications if required.

After the primers have been designed, each primer-template hybridization combination can be calculated and displayed under the *Hybridization* tab (Figure 33, right). The program will highlight primer-template pairs which are unlikely to successfully amplify as shown in Figure 33. 94 PCR experimental results were collected from literature and subjected to the CEMAsuite hybridization algorithm to obtain the default values of the warning thresholds; however, these values can be adjusted readily. Selective degeneracy can be added by the user to increase the stability of the primer-template pairs. CEMAsuite simplifies this task by also highlighting the regions of the plot and CEMA which correspond to the primer locations. This allows the user to quickly account for the overall amino acid conservation and/or nucleic acid conservation for each degeneracy position to add.

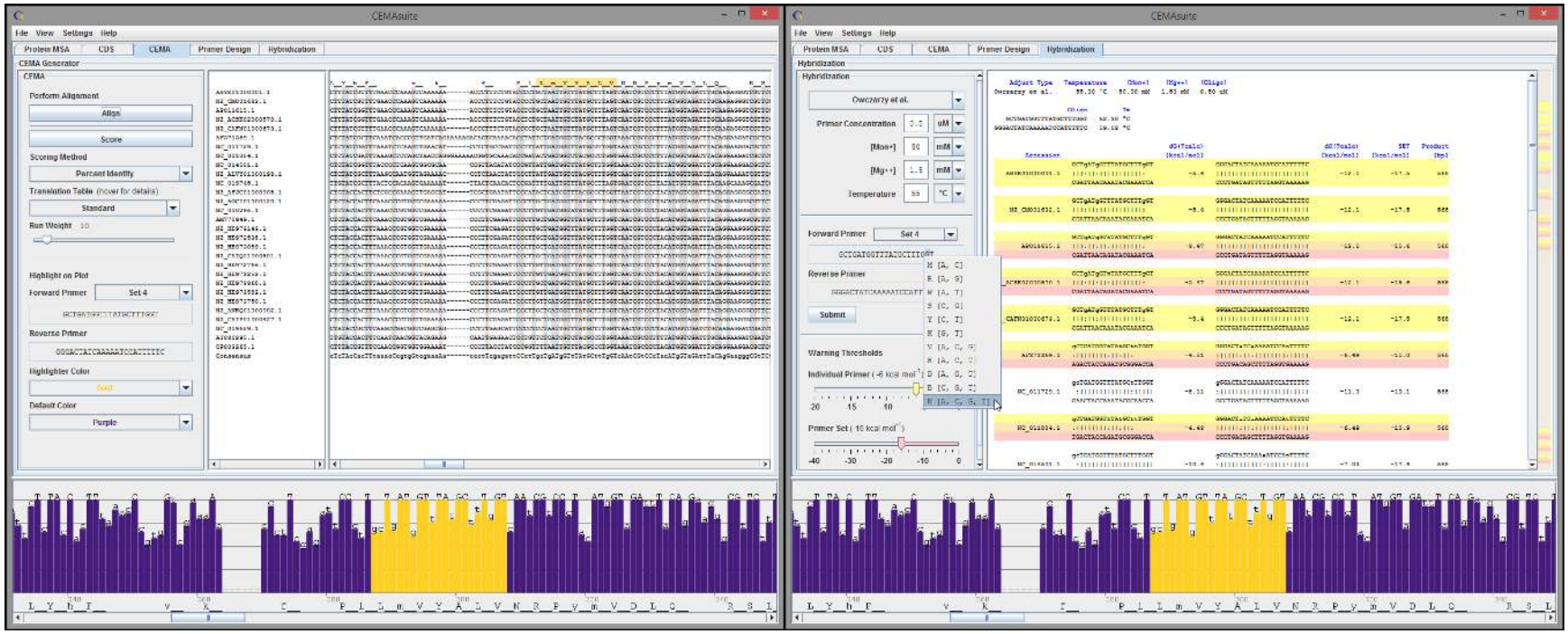


Figure 33: Screenshots of the CEMAsuite application for a sample project.

(LEFT) CEMAsuite CEMA tab with a successfully generated CEMA, scored using the Percent Identity algorithm shown in the bar plot. The forward primer region is highlighted (gold) in the plot and the first line of the CEMA alignment text.

(RIGHT) CEMAsuite Hybridization tab with each calculated Gibbs free energy value at the specified conditions for each primer-template pair. Primer-template pairs which the user would like to flag as "likely to fail amplification" are highlighted if they possess an individual primer less stable than the user input Individual Primer Warning threshold (yellow) and/or the sum of the stability values of both primers is less stable than the Primer Set Warning Threshold (red). Selective addition of degeneracy via the popup menu is shown in the forward primer input region.

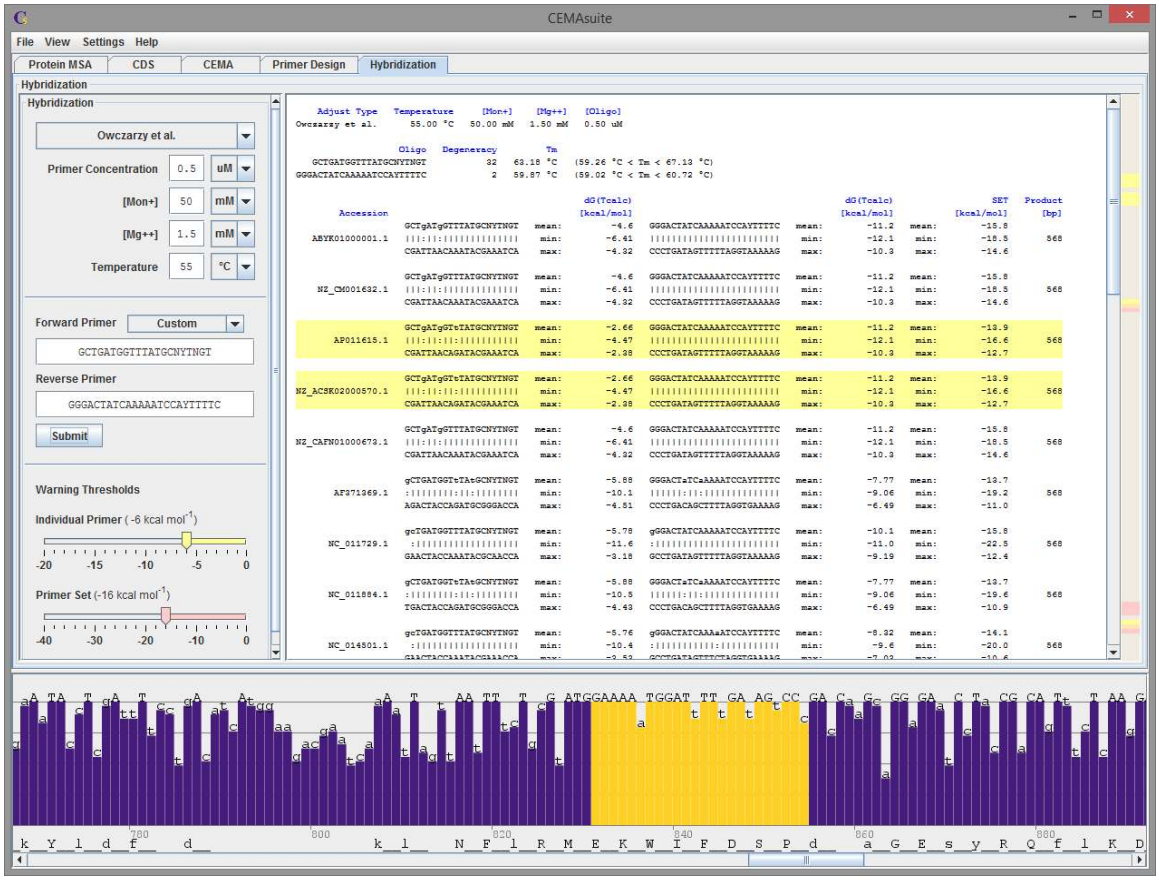


Figure 34: CEMAsuite example output after some degenerate nucleotide positions have been incorporated into the active primer set.

6.8 Investigation of Primer3 for Consensus Primer Design

6.8.1 Introduction

In order to test the efficacy of primers developed in Primer3 (Untergasser et al., 2012) from a CEMA consensus sequence, an arbitrary homologous gene was selected for primer design.

The targeted gene, named *ftsZ*, encodes a tubulin-like cell division protein (FtsZ). FtsZ is the major cytoskeletal protein involved in prokaryotic cytokinesis, meaning its functionality is essential for cell propagation. Since the targeted gene was chosen arbitrarily, the functionality will not be discussed in detail. For an informative review of FtsZ functionality in cytokinesis, please confer Erickson *et al.* (2010). 67 FtsZ and *ftsZ* sequences were obtained through the NCBI database and are listed in Appendix Table 5 (Sayers et al., 2009).

6.8.2 Materials and Methods

6.8.2.1 Experimental Design

As stated in section 2.8, the likelihood of MSA conservation type I error is inversely proportional to the alignment sequence sample size. So, to investigate the effects of sample size on consensus PCR primer design, three alignments of 5, 20, and 67 sequences were generated using the obtained accessions. CEMAs were generated in a preliminary version of CEMAsuite which lacked hybridization functionality. The consensus sequence of each alignment was subjected to primer design via Primer3 (Untergasser et al., 2012). Two primer sets were designed from each alignment, one implementing the Primer3 quality score capabilities, and one without (Table 9).

Table 8: Primer set nomenclature used throughout this work.

Number of Sequences	Positional Scoring −	Positional Scoring +
5	FtsZ5.0	FtsZ5.1
20	FtsZ20.0	FtsZ20.1
67	FtsZ67.0	FtsZ67.1

Next, various bacterial DNA isolates from strains *Escherichia coli* DH5α, *Nostoc muscorum* UTEX1037, *Plectonema* sp. UTEX1541, *Synechococcus* sp. UTEX2434, *Synechocystis* sp. UTEX2470, and *Synechocystis* sp. PCC6803 were subjected to PCR using the primers. Each primer-template combination was performed in triplicate. The eukaryotic genomic DNA of *Saccharomyces cerevisiae* BY4741 was used as a negative control in conjunction with a reaction containing no template DNA.

6.8.2.2 PCR Conditions

To test the *ftsZ* amplification capabilities of each primer set, PCR was performed using 50 µl reactions consisting of 1.25 U *Taq* polymerase (Invitrogen), 20 mmol l^{-1} Tris-HCl (pH 8.4), 50 mmol l^{-1} KCl, 2.0 mmol l^{-1} $MgCl_2$, 0.2 mmol l^{-1} dNTPs (ea.), 0.5 µmol l^{-1} each primer, and 50 ng DNA template prepared in 200 µl polypropylene tubes. These reactions underwent a standardized method of touchdown PCR which consisted of one cycle of [94°C for 3:00], 10 cycles of [94°C for 0:45, T_A for 0:30(-1°C $cycle^{-1}$) , 72°C for 0:45] where $T_A = T_{m,low(1M\ NaCl)} + (5°C - N_{cycles} \times 1°C\ cycle^{-1})$, followed by 32 cycles of [94°C

for 0:45, $T_{A,final}$ for 0:30, 72°C for 0:45], one cycle of [72°C for 10:00] and a final incubation at four centigrade in a Bio-Rad DNA Engine Peltier Thermal Cycler (model PTC0200).

6.8.2.3 Electrophoresis of PCR Products

Five microliters of each reaction product and one microliter of 6x Orange DNA Loading Dye (Fermentas #R0631) were loaded into a 1.5 percent standard agarose (US Biological #A1016) gel stained with 0.5 μg ethidium bromide ml^{-1}. Five microliters of 100 bp O'GeneRuler DNA ladder (Fermentas # SM1143) was used for each molecular weight marker. Visualization and analysis was performed with a UVP Bioanalyzer and UVP Visionworks LS Acquisition and Analysis (v6.5.2).

6.8.2.4 In Silico Analysis

In order to obtain further insight on the influence of various factors on consensus primer design using Primer3, primers were designed using alignments of the FtsZ5, FtsZ20, and FtsZ67 sequence groups and subjected to hybridization analysis via the CEMAsuite algorithm. The hybridization acted as a method to qualitatively describe the ability of primers to anneal for comparisons.

First, the influence of the method of nucleic acid MSA generation was investigated. Using traditional direct nucleic acid multiple sequence alignment, as described in section 2.8, each set of *ftsZ* sequences was aligned using Clustal Ω (Sievers et al., 2011). The CEMAs used in section 6.8.1 were recycled for this work.

In order to test the consensus sequence generation method, two consensus sequences were generated for each alignment. The first consensus sequence was generated by appending the mode residue for each column within the alignment, treating a gap as a residue. The second consensus sequence was generated using the CEMAsuite consensus sequence generation method, where only fully populated positions/columns of an alignment will append residue values and all others will be treated as gaps. Three primer sets were designed for each consensus sequence using Primer3.

Next, in order to test the impact of score weighting on Primer3 output, each scoring algorithm available in CEMAsuite was used to score each CEMA gap excluded consensus sequence, as would be the output of a typical CEMAsuite implementation. Only the score weighting inputs were varied and three primers were designed using the Percent Identity scoring method. Then three primers were designed using Identity Runs with the default run weight of 10. Finally, three primers were designed using Potential Degeneracy.

Once primers were designed using all methods outlined in Figure 35, they were subjected to hybridization analysis using the CEMAsuite algorithm and all *ftsZ* sequences collected. The hybridization was used to simulate the effects on primer design when limited sequence information is available. In other words, if one designs a primer set using only a five sequence MSA, such as in FtsZ5, how efficiently can those primers anneal unknown sequences? The remaining 62 sequences in the FtsZ collection act as unknown cases in this investigation.

The hybridization conditions implemented for each set were standard PCR conditions: 70 mM monovalent cations and 1.5 mM magnesium ions. In order to prevent minor variations in primer length and G | C content from influencing the results, the annealing temperature for each primer set hybridization was standardized. The T_A used for each set was equal to the estimated $T_{m,\,low}$ output by Primer3 minus five centigrade, which is a standard initial annealing temperature. Next, the $\Delta G_{i,j}(T_A)$ values of the least stable primer within a set were then normalized against the Gibbs free energy value representing perfect annealing of that primer ($\Delta G_i(T_A)^*$). This value was used to describe the minimum relative annealing ability of a single primer within a primer set (Equation 5). In other words, a value representing the likelihood of a PCR amplification failing due to a single primer only. The same calculations were performed on the sum of the primers to determine a relative annealing ability of the primer set (Equation 6).

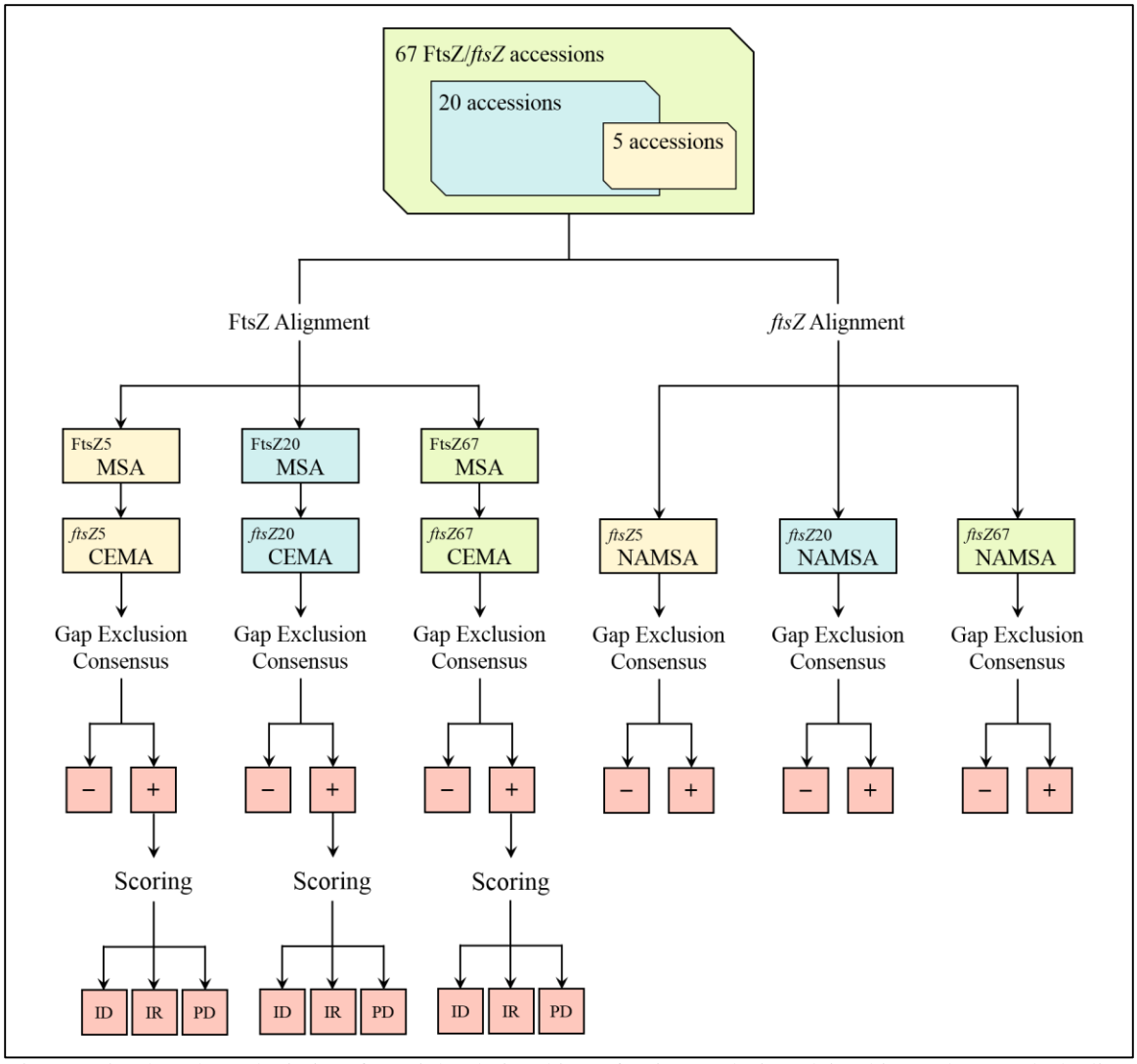


Figure 35: Flowchart depicting factorial relationships of primers designed in this study. Red shading indicates a primer design stage.

Equation 5: Minimum relative annealing ability of a single primer within a primer set to template j.

$$\varphi = (100\%)\left\{\begin{matrix}\frac{\Delta G_{F,j}}{\Delta G_F^*} \text{ where } \Delta G_{F,j} > \Delta G_{R,j} \\ \frac{\Delta G_{R,j}}{\Delta G_R^*} \text{ where } \Delta G_{F,j} < \Delta G_{R,j}\end{matrix}\right\}$$

Equation 6: Relative annealing ability of a primer set to template j.

$$\Phi = (100\%)\frac{(\Delta G_{F,j} + \Delta G_{R,j})}{(\Delta G_F^* + \Delta G_R^*)}$$

This analysis resulted in 4,217 data points. This data collection was then analyzed using Statistical Analysis System (SAS v9.4) for comparisons.

6.8.3 Results

6.8.3.1 Alignment Results

The FtsZ5 protein alignment contained 151 identity-level conserved residues and 267 conserved residues in all and the resulting CEMA contained 421 identity-level conserved residues. Next, the FtsZ20 protein alignment contained 123 identity-level conserved residues and 222 conserved residues and the respective CEMA contained 325 identity-level conserved residues. Increasing the number of sequences within the alignment showed a decrease in conservation. Finally, the FtsZ67 protein alignment contained 97 identity-level conserved residues and 202 conserved residues and the resulting CEMA contained 280 identity-level conserved residues. Again, an increase in the number of sequences within the alignment showed a decrease in conservation.

6.8.3.2 PCR Amplification Results

In order to analyze the non-degenerate consensus primer design capabilities of Primer3 via the CEMA consensus sequence, primers were designed targeting the prokaryotic *ftsZ* gene. Primers were designed both with (FtsZX.1) and without (FtsZX.0) the implementation of CEMA positional scores. Additionally, in order to test the impact of the sample size of the alignment, three alignments of five, 20, and 67 sequences were generated for each alignment method and non-degenerate primers were designed from each (Table 9). These were then tested across six prokaryotic (*ftsZ*$^{+}$) DNA templates utilizing two negative controls: eukaryotic (*ftsZ*$^{-}$) DNA template and no DNA template. Of the six prokaryotic DNA templates only *Synechocystis* sp. PCC6803 and *Escherichia coli* DH5α possessed known *ftsZ* sequences.

Table 9: Non-degenerate primer sequences designed from ftsZ MSA consensus sequences.

Primer Set	Forward Sequence	Reverse Sequence
FtsZ5.0	AAGACAAACGCTCCGAATCG	AACAGTCAAAGCGCCCATTT
FtsZ5.1	TGGTGTTGGCGGAGGTGGTGGTAATGC	ATCACCGCCCGCACGTCAGCAAAGT
FtsZ20.0	CGCATTGCTGATGATGTTCT	TCCAGCAGCGGAGAAGTAAT
FtsZ20.1	TGGTCTTTATCGCCGCTGGCATGGG	TCGGCCATCACCGAGCGCACGTC
FtsZ67.0	TCAAAAAGCAGCCGAAGAAT	TAAAGGGACGGGTGACTACG
FtsZ67.1	GCTGGCATGGGTGGCGGTACTGG	ACCGCCCGCACATCGGCAAAGT

The FtsZ5.0, FtsZ5.1, FtsZ20.0, FtsZ20.1, FtsZ67.0, and FtsZ67.1 primers correspond to positions X55034.1:(21985..22005,22200..22566), X55034.1:(21882..21909,22460..23086), X55034.1:(22381..22401,22564..23294), X55034.1:(22124..22149,22468..23102), X55034.1:(22065..22085,22221..22608), and X55034.1:(22138..22161,22460..23086) in reference to the *E. coli* accessed sequence.

PCR amplification using the FtsZ5.0 primers yielded a single product of slightly varying size, approximately 140 – 160 bp in length for all reactions. This was unexpected, as the products were relatively uniform and were present in both negative controls. To ensure the observed products were not due to contamination, the triplicate experiment was repeated with fresh reagents. The same products were observed in all cases for the repetition as well. The experiment was repeated using new reagents and fresh container sources and again yielded a uniform product. Due to the presence of the product in the no-template control reaction, it is believed to be due to auto-extension of the primers. Desired product formation, ~240 bp, was not observed for any FtsZ5.0 primer reactions.

PCR amplification using the FtsZ5.1 primers yielded desired products, ~600 bp, for five of the six prokaryotic templates, only UTEX2434 failed to amplify. Product concentrations ranged between 1.5 ng μl^{-1} and 25 ng μl^{-1}. Only a single byproduct was observed throughout all reactions (observed in UTEX2470), and neither negative control exhibited any product formation.

PCR amplification using the FtsZ20.0 primer set successfully amplified the desired product, ~200 bp, in DH5α, UTEX 1541, and UTEX2470. Product concentrations ranged between 0.7 ng μl^{-1} and 4.9 ng μl^{-1}. A single byproduct was observed for both the PCC6803 and DH5α templates, and neither negative control exhibited any product formation.

PCR amplification via the FtsZ20.1 primer set yielded the desired products, ~390 bp, in PCC6803 and UTEX2434 at concentrations ranging from 0.6 ng μl^{-1} to 22 ng μl^{-1}. No other products were observed in any other reactions.

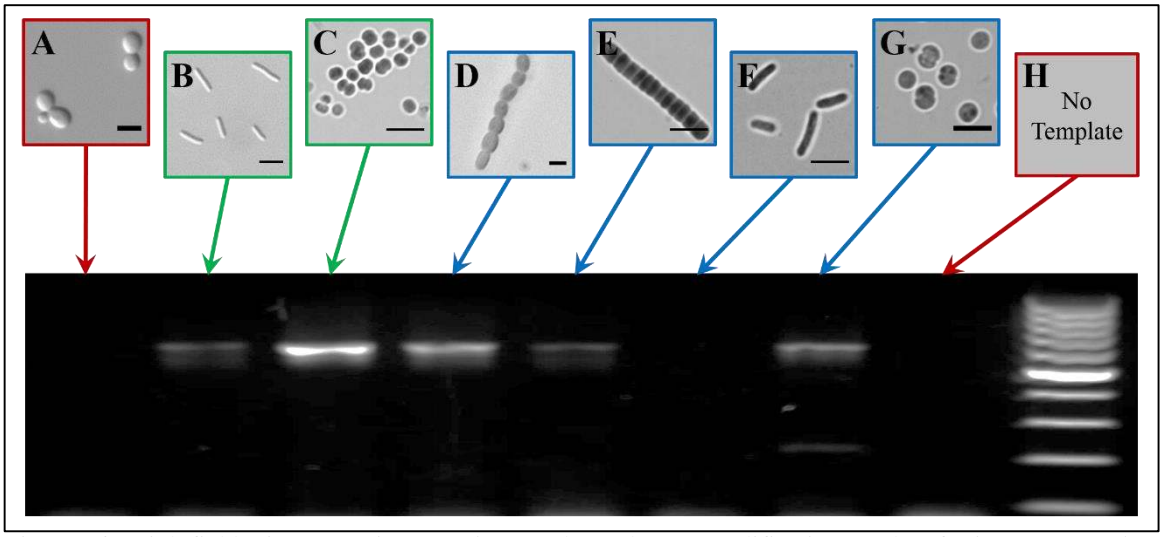


Figure 36: Brightfield microscopy images (5 µm scale) and PCR amplification results of primer set FtsZ5.1. Images depicted in green indicate organisms with known ftsZ sequences, red indicates negative control, and blue indicates a sample with unknown ftsZ sequence. Agarose gel electrophoresis was performed using 1.5% standard agarose, 0.5 µg ml^{-1} ethidium bromide stain, 0.5xTBE, 5 µL product, and 5 µL 100 bp O'GeneRuler DNA ladder.

(A) Saccharomyces cerevisiae BY4741, (B) Escherichia coli DH5α, (C) Synechocystis sp. PCC6803, (D) Nostoc muscorum UTEX1037, (E) Plectonema sp. UTEX1541, (F) Synechococcus sp. UTEX2434, (G) Synechocystis sp. UTEX2470, (H) No template control.

PCR amplification using the FtsZ67.0 primer set yielded the desired products, ~180 bp, in PCC6803, UTEX1037, UTEX2434, and UTEX2470 at concentrations ranging between < 0.1 ng μl^{-1} to 3.6 ng μl^{-1}. Byproduct formation was observed in all prokaryotic reactions at varying lengths and concentrations, and neither negative control exhibited any product formation.

PCR amplification using the FtsZ67.1 primer set yielded the desired products, ~350 bp, in PCC6803, UTEX1037, UTEX1541, and UTEX2470 at concentrations ranging between < 4.5 ng μl^{-1} to 27 ng μl^{-1}. Byproduct formation was observed in DH5α and UTEX2470, and neither negative control exhibited any product formation.

Statistical analysis was performed on the results of all experiments and can be found in Figure 37. The binomial probability of successful PCR amplification for each primer set was estimated from the detection results. Comparison of binomial proportions was performed using a chi-squared test and the more conservative Fisher's exact test. In both cases the proportion of successes in the score weighted treatment,

"+", were significantly higher than those in the unweighted treatment, "–" (χ^2: p = 0.0209, F: p = 0.0338). It was also observed that the product concentrations used in the score weighted design were significantly higher than when score weighting was not used (p < 0.0001). Additionally, when score weighting is not used, the product concentration was actually not significantly different from zero. There were significantly more undesired products formed than the score weighted treatment (p < 0.0001).

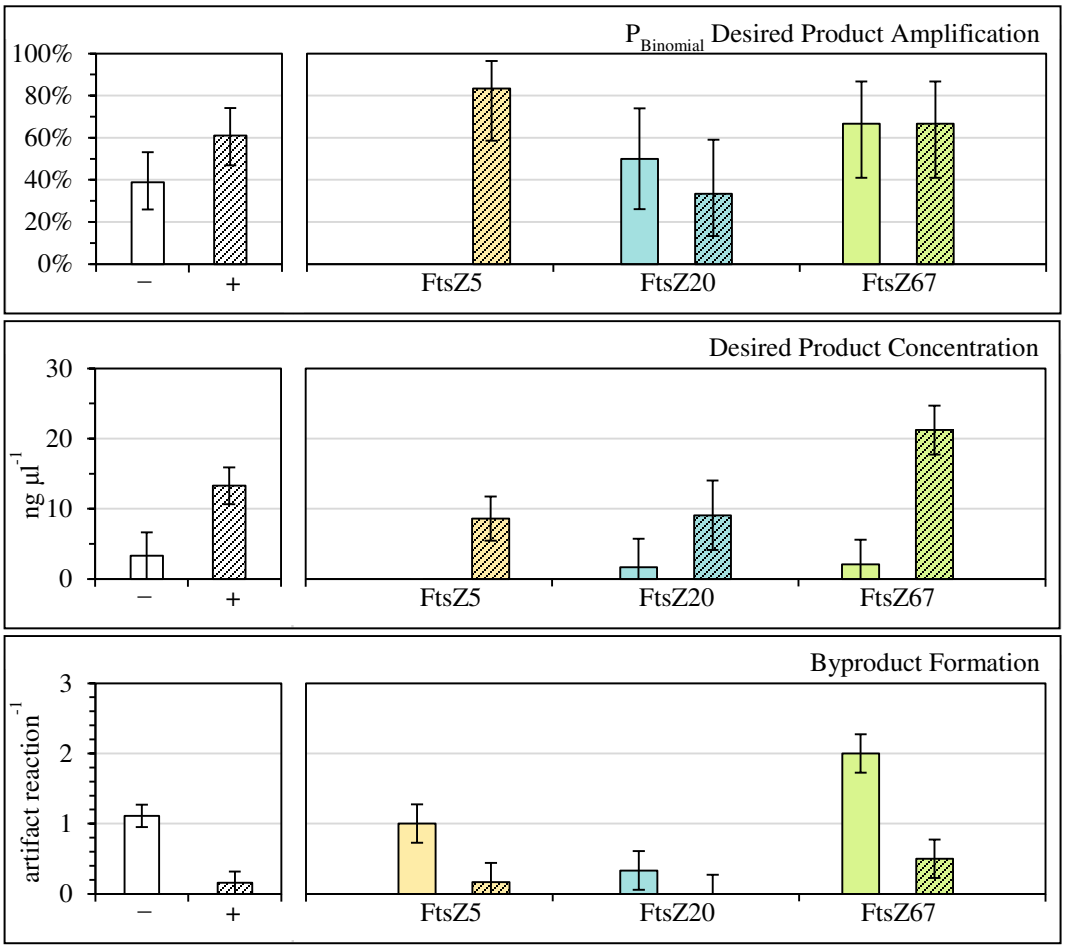


Figure 37: Descriptive statistics of the results from PCR amplification using the primers developed in this work. Exact confidence intervals are shown (α = 0.05).

Fisher's exact test comparisons accounting for alignment size indicate only FtsZ5 shows a significantly different probability of successful amplification, such that $P_{5-} < P_{5+}$ ($p_5 < 0.0001$, $p_{20} = 0.4998$, $p_{67} = 1.0$). Within the five- and 67-sequence cases, the score weighted product concentration was significantly higher than when score weighting was not implemented, such that $\mu_{5-} < \mu_{5+}$ and $\mu_{67-} < \mu_{67+}$

($p < 0.0001$ each). The 20-sequence case however showed no significant increase: $\mu_{20-} = \mu_{20+}$ ($p = 0.2075$). Again, within the five- and 67-sequence cases, significantly more byproducts were observed for no score weighting, $\mu_{5-} > \mu_{5+}$ and $\mu_{67-} > \mu_{67+}$ ($p_5 = 0.0006$, $p_{67} < 0.0001$). The 20-sequence case showed no significant difference in byproduct formation: $\mu_{20-} = \mu_{20+}$ ($p = 0.5303$).

6.8.3.3 In Silico Analysis

The relative annealing abilities of PCR primers designed in Primer3 were calculated for various inputs. First, the cases of consensus sequence gap exclusion and MSA method of generation were investigated. Using 2,412 data points, the factors of gap exclusion, number of sequences in the alignment, and the alignment method were investigated under a significance of 0.05. A depiction of the results can be observed in Figure 38.

It was observed that the minimum relative annealing ability (φ) of a single primer within a set was significantly higher in cases where gap exclusion was implemented in the consensus sequence construction, such that $\varphi_{-} < \varphi_{+}$ ($p = 0.0026$). This was also observed for the relative annealing ability of the primer set (Φ) ($\Phi_{-} < \Phi_{+}$, $p = 0.0093$). With gap exclusion, the relative annealing abilities increased by $0.8\% \leq \Delta\varphi \leq 4.0\%$ and $0.5\% \leq \Delta\Phi \leq 3.3\%$. Additionally, when a CEMA was used to construct the consensus sequence, both φ and Φ were significantly higher than when the nucleic acid sequences were aligned directly, such that $\varphi_{NAMSA} < \varphi_{CEMA}$ and $\Phi_{NAMSA} < \Phi_{CEMA}$ ($p < 0.0001$ each). With CEMA consensus sequences, the relative annealing abilities increased by $2.3\% \leq \Delta\varphi \leq 5.5\%$ and $2.2\% \leq \Delta\Phi \leq 5.0\%$ in comparison to direct NAMSA consensus sequences. It was observed that φ is exhibits two alignment sequence size groupings under the Tukey's range test: $\varphi_{20} = \varphi_{67} < \varphi_5 = \varphi_{20}$ ($p = 0.0165$); Φ showed a significant decrease for the 20 sequence treatment ($\Phi_{20} < \Phi_5 = \Phi_{67}$, $p < 0.0001$).

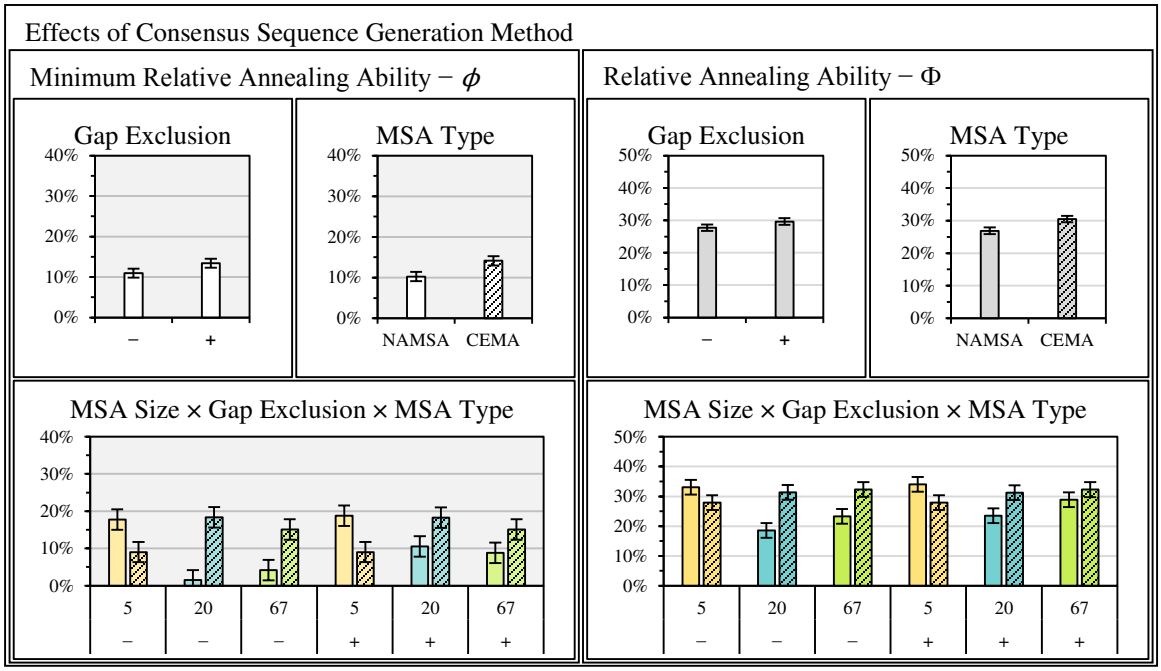


Figure 38: Effects observed from the method used to generate the consensus sequence on the relative annealing ability of a primer set. Confidence intervals are depicted (n = 2412, α = 0.05).

Next, the effects of the number of sequences and the effects of score weighting on primer design were investigated using a separate collection of 2,412 data points and a significance of 0.05. Score weighted design included equally sized samples of the three scoring algorithms stated in section 6.8.2.4. Only gap-excluded CEMA consensus sequences were used in primer design for this data due to lack of availability and/or applicability of scoring algorithms on data. A summary of this analysis is depicted in Figure 39.

It was observed that the minimum relative annealing ability was independent of the number of sequences for the pooled data ($\varphi_5 = \varphi_{20} = \varphi_{67}$, $p = 0.1064$), but Φ showed a significant increase for the highest sequence treatment ($\Phi_5 = \Phi_{20} < \Phi_{67}$, $p < 0.0001$). Both φ and Φ showed a significant increase when score weighting was implemented in the primer design ($\varphi_- < \varphi_+$ and $\Phi_- < \Phi_+$, $p < 0.0001$ each). This increase observed in the relative annealing abilities were $5.3\% \leq \Delta\varphi \leq 9.0\%$ and $3.0\% \leq \Delta\Phi \leq 6.4\%$, which is a dramatic improvement for the minimum relative annealing ability. Next the interactions of scoring and sequences were observed under a significance of 0.05.

For the five-sequence alignment, the score weighting treatment showed a significant increase in both relative annealing abilities ($p < 0.0001$ each). The increase observed in the relative annealing abilities

were $8.0\% \leq \Delta\varphi \leq 17\%$ and $3.0\% \leq \Delta\Phi \leq 12\%$, which is a substantial improvement in both cases. When score weighting is used, the effects of the number of sequences on φ are not readily observed, which is qualitatively demonstrated in Figure 39 and quantitatively described in the Tukey's range test groupings of $\varphi_5 = \varphi_{20} < \varphi_5 = \varphi_{67}$ and $\Phi_{20} < \Phi_5 = \Phi_{67}$ ($p_\varphi = 0.0228$, $p_\Phi < 0.0001$). There were no significant differences between the score weighting treatments observed for the 20-sequence alignment. For the 67-sequence alignment, both relative annealing abilities increased ($\varphi_- < \varphi_+$, $p < 0.0001$ and $\Phi_- < \Phi_+$, $p = 0.0074$). The increases observed in the relative annealing abilities were $2.6\% \leq \Delta\varphi \leq 12\%$ and $1.0\% \leq \Delta\Phi \leq 10\%$.

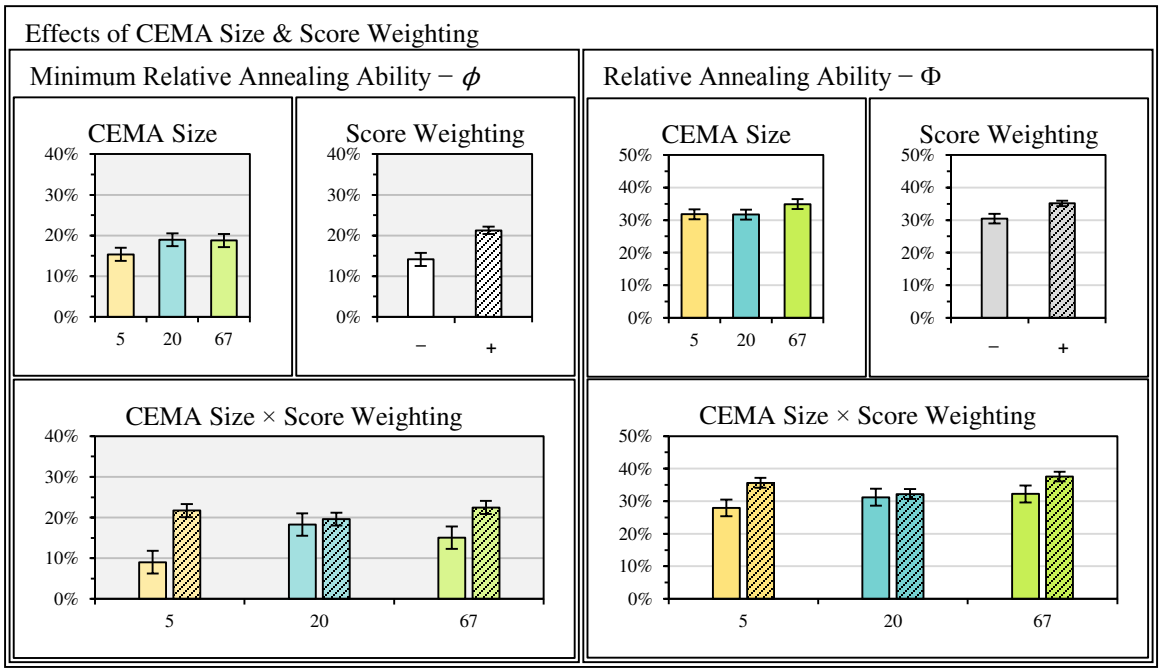


Figure 39: Effects observed from the implementation of score weighting on the relative annealing ability of a primer set. Confidence intervals are depicted ($n = 2412$, $\alpha = 0.05$).

Next, the previous dataset was analyzed accounting for the Percent Identity (ID), Identity Runs (IR), and Potential Degeneracy (PD) algorithms. This data was then subjected to statistical analysis via a two-factor ANOVA implementing Tukey's range test for multiple comparisons of specific algorithms to the no score weighting treatment. Finally the score weighting treatment subset of the previous data was reanalyzed for comparisons between scoring algorithms. The results from this analysis are depicted in Figure 40.

In comparison with the no score weighting treatment, the relative annealing abilities of the Percent Identity scoring algorithm were significantly higher ($p_\varphi < 0.0001$, $p_\Phi < 0.0001$). The respective increments

observed were $6.5\% \leq \Delta\varphi \leq 11\%$ and $4.4\% \leq \Delta\Phi \leq 8.5\%$. The Potential Degeneracy scoring algorithm showed no significant difference from the no score weighting treatment ($p_{\varphi} = 0.0785$, $p_{\Phi} = 0.4124$). The Identity Runs algorithm showed a significant increase in both relative annealing abilities ($p_{\varphi} < 0.0001$, $p_{\Phi} < 0.0001$). The respective increments observed were $8.5\% \leq \Delta\varphi \leq 13\%$ and $6.5\% \leq \Delta\Phi \leq 10\%$.

The Tukey's range test resulted in two groupings of sequence treatments for the minimum relative annealing ability, where $\varphi_5 = \varphi_{20} < \varphi_5 = \varphi_{67}$, implying resolution of these groups can be achieved through further data acquisition. For the total relative annealing ability, the test resulted in two distinct groups, where $\Phi_{20} < \Phi_5 = \Phi_{67}$ ($p < 0.0001$). Testing the effects of the algorithm showed two distinct groups for both relative annealing abilities. In both cases, the PD algorithm shows a significant decrease in the relative annealing ability in comparison with the other two algorithms ($\varphi_{PD} < \varphi_{ID} = \varphi_{IR}$ and $\Phi_{PD} < \Phi_{ID} = \Phi_{IR}$, $p < 0.0001$ each). Next the interactions of the algorithm and sequences were observed under a significance of 0.05.

For the five sequence alignment, there was no significant difference in the three scoring algorithms for the minimum relative annealing ability, but the total relative annealing ability exhibited the same trend as the overall groupings. For the 20 sequence alignment, both relative annealing abilities exhibited the same trend observed overall (PD < ID = IR). The 67 sequence alignment exhibited the most resolved results of the three sequence groups. In both cases, the relative annealing ability for the Identity Runs algorithm was higher than the other two algorithms. Additionally, the minimum relative annealing ability of the Percent Identity scoring algorithm was significantly higher than that of the Potential Degeneracy algorithm. The relationships observed for the largest alignment were $\varphi_{PD} < \varphi_{ID} < \varphi_{IR}$ and $\Phi_{PD} = \Phi_{ID} < \Phi_{IR}$.

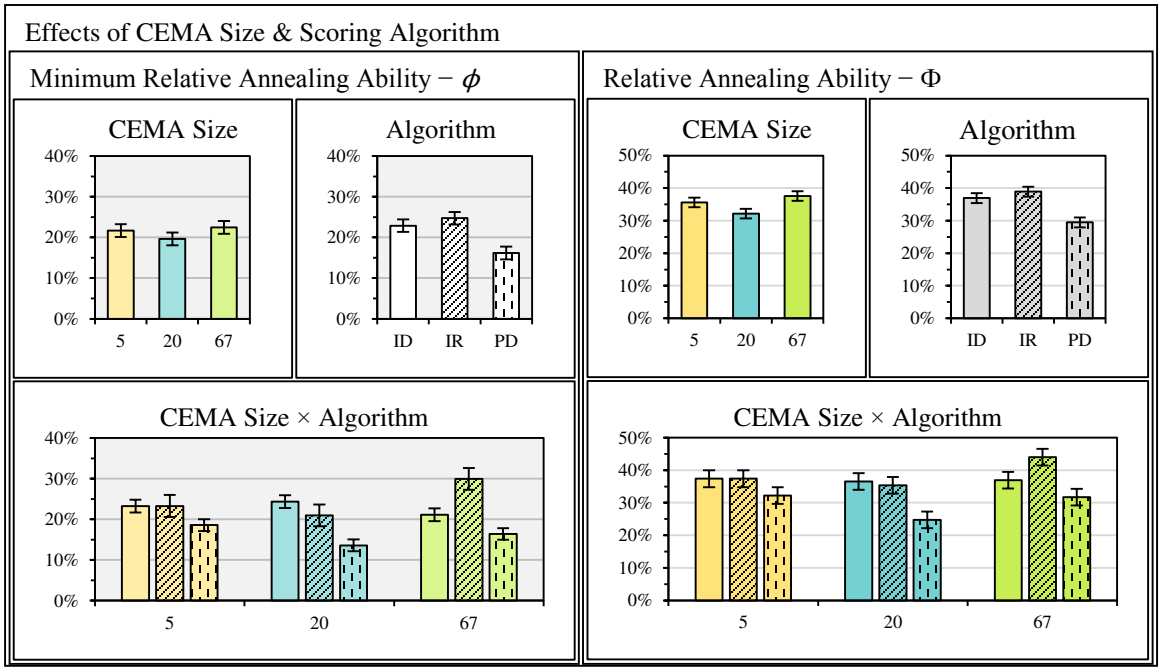


Figure 40: Effects observed from the scoring algorithm used on the relative annealing ability of a primer set. Confidence intervals are depicted (n = 1809, α = 0.05). ID – Percent Identity, IR – Identity Runs, PD – Potential Degeneracy.

6.8.4 Discussion

The consensus primer PCR amplification experiments demonstrated Primer3's efficacy as a consensus primer design tool. The experimental results also suggest that the implementation of positional scoring can play a major role in the efficacy of the returned primer set. The *in silico* analysis helped to elucidate the important influencing factors in consensus primer design using Primer3. Primers designed from of a simple nucleic acid multiple sequence alignment showed a significantly lower relative annealing ability than those designed from a CEMA. Additional evidence supporting the benefits of positional scoring when using Primer3 for consensus primer design was observed in the *in silico* analysis. The potential degeneracy scoring method proved to provide no significant benefits over the other studied design methods, suggesting it should be used as a "design by inspection" visual aid only. In contrast, the identity runs scoring method proved to return the best observed primers when the number of sequences was highest, suggesting it should be the method of choice for primer design using large CEMAs.

6.9 Summary

The methodologies behind the CEMAsuite consensus degenerate primer design algorithms were shown to increase the efficacy of a primer set through experimental and simulated results. CEMAsuite has successfully found a compromise between the two traditional consensus degenerate primer design methodologies, which was the initial goal of this project. It is important to note that even though CEMAsuite incorporates Primer3 functionality, the scoring algorithms and output graphics significantly speed up the process of primer design by inspection for those with moderate primer design experience. CEMAsuite is expected to be a valuable assay design tool for applications such as genetic/environmental screening, working with highly variable sequences, and DNA quality control via PCR.

CHAPTER 7. *IN VITRO* PHA SYNTHASE KINETICS

7.1 Chapter Preface

This chapter records the preliminary investigations of data from literature required for Objective III (section 1.4) and includes work currently submitted for publication under the title of "Novel Interpretations of *In Vitro* Polyhydroxyalkanoate Polymerization Phenomena". This chapter holds the theory, derivations, and comparisons of meso-scale *in vitro* PHA polymerization mechanics.

7.2 Introduction

Polyhydroxyalkanoates (PHAs) are a class of biologically derived polymers spanning a diverse range of polyesters capable of reducing the ecological impact of the plastics industry (Anderson & Dawes, 1990; Laycock et al., 2014; Rai et al., 2011; Rehm, 2007; Stubbe & Tian, 2003; K. Sudesh et al., 2000). These biodegradable polymers are promising replacement materials for short- to mid-life polyolefin plastics and also hold potential for biomedical applications due to their biocompatibility (Chen & Wu, 2005; D. Jendrossek et al., 1996).

PHAs are the only known polyesters existing in living organisms besides water-soluble poly(malic acid), which occurs in lower level eukaryotes, and the water-insoluble polyesters, suberin and cutin, which occur in plants (Steinbüchel & Hein, 2001). PHAs are produced as water-insoluble inclusions known as *granules* in the cytoplasm of many microorganisms. These granules are produced as a form of carbon storage during times of carbon surplus and nutrient deficiency, although this is not a requirement in all PHA producers.

The major obstacles facing industrial-scale PHA bioproduction are the costs associated with their production (Bengtsson et al., 2010; Choi & Lee, 1999). There are many logical routes to reducing the production costs, such as reducing expensive feed costs by using low-cost carbon feed stocks or photosynthetic bioproduction (Bengtsson et al., 2010; Choi & Lee, 1999; Ienczak et al., 2013; Courtney E. Lane & Benton, 2015). An increase in the speed and yield of PHA production, and subsequent

accumulation, would lead to reduced specific costs no matter the production system design. For this reason, much research is focused on the key enzyme class responsible for PHA polymerization – PHA synthases (Bhubalan et al., 2011; Fukui et al., 1976; Gerngross & Martin, 1995; Gerngross et al., 1994; Hooks & Rehm, 2015; Jia, Kappock, et al., 2000; Jia, Yuan, et al., 2000; Kikkawa et al., 2005; Liebergesell, Rahalkar, & Steinbüchel, 2000; Lu, Han, Zhou, Zhou, & Xiang, 2008; Müh et al., 1999; Numata et al., 2015; Peters & Rehm, 2005; Rehm, Antonio, Spiekermann, Amara, & Steinbüchel, 2002; Rehm & Steinbüchel, 1999; Takase, Matsumoto, Taguchi, & Doi, 2004; Ushimaru et al., 2013; B. Zhang, R. Carlson, F. Srienc, 2006; S. Zhang, Kolvek, Goodwin, & Lenz, 2004; S. Zhang, Yasuo, Lenz, & Goodwin, 2000; W. Zhang et al., 2014).

In order to tune the material properties of PHA to mimic various polyolefins, the primary structure and configuration of the polymer must be controlled. The PHA synthase dictates the range of hydroxyacyl-co-enzyme A thioester monomers that can be recognized and incorporated into the covalently-bound growing polymer. As a consequence, if PHA polymers, or co-polymers, with a specified composition are desired, then a PHA synthase capable of catalyzing all desired hydroxyacyl-CoA monomers is required. This means that the quantity of viable PHA co-polymer types that can be synthesized is limited by the current number of PHA synthases that are characterized. This also means that continual characterization of novel PHA synthases is required to take full advantage of this polymer class' versatility. Additionally, the PHA synthase is the only commonality between the wide variety of PHA metabolic routes, securing its relevance in bioplastic research (Rehm, 2007; Steinbüchel & Lütke-Eversloh, 2003). The classification, structure, and function of these enzymes is described in detail in many of the reviews mentioned previously.

In vitro PHA synthase assays, via spectroscopic quantitation of the hydoxyalkyl-CoA monomer consumption rate during the polymerization process, are used to quantify key enzyme characteristics. One approach for measuring the substrate consumption rate is to observe the decrease in the amount of unreacted substrate by analyzing its thioester bond at 232-236 nm wavelength (Fukui et al., 1976; Ushimaru et al., 2013; S. Zhang et al., 2000). A more common approach is to monitor the release of free co-enzyme A

(CoA) using 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) (Bhubalan et al., 2011; Gerngross & Martin, 1995; Gerngross et al., 1994; Müh et al., 1999; Ushimaru et al., 2013; W. Zhang et al., 2014). DNTB is a colorimetric substrate for quantitating free thiols in solution. It covalently reacts with a free thiol to produce 2-nitro-5-thiobenzoate, which can be analyzed at a wavelength of 412 nm.

Some DNTB PHA synthase enzymatic activity assays are performed continuously (Bhubalan et al., 2011). For continuous assays, the reaction solution contains DTNB before any PHA synthase is added and the DNTB reacts with CoA (a thiol polymerization byproduct) as it is generated. Müh and coworkers demonstrated that this experimental design leads to misrepresentative observations (1999). In their work, they implemented a discontinuous assay, which eliminates DTNB from the polymerization reaction, and then removed aliquots from the reaction over time to perform DTNB quantitation of free CoA. This assay design proved to have a significant effect on the observed data compared to the continuous assay (Müh et al., 1999). The substantial effect DNTB had on the apparent polymerization rate led these researchers to conclude that only discontinuous assays should be implemented in quantitative studies.

Research into *in vitro* PHA synthase enzymatic activity assays reveals that the behavior of the polymerization reaction depends heavily on the reaction design. *In vitro* PHA synthase polymerization reaction behavior can potentially exhibit multiphasic, inhibited, or a combination of these behaviors (Gerngross et al., 1994; Müh et al., 1999; S. Zhang et al., 2000). The observed kinetics are highly dependent on the initial reaction conditions, and it is vital to understand the contributing factors prior to experimental design. This work discusses observed behaviors and their associated initial conditions individually. Hypothetical mechanisms for each observed behavior are proposed and subjected to an evidentiary assessment based on available literature. Mathematical models are derived for the hypothetical mechanisms and fit to data from literature. The fundamental causes of the observed behaviors are elucidated using minimalistic kinetic models, rigorous statistical analysis, and multi-model inferences.

7.3 Modelling Theory and Procedure

7.3.1 Unprimed Polymerization

A lag phase, or delayed activity, is a multiphasic behavior typically observed in multimeric enzymes. When unassembled PHA synthase monomers are used to initialize the polymerization reaction, a lag phase is observed. The unassembled monomers require a "priming" phase, where the enzymes undergo the formation of the active complexes (Gerngross et al., 1994). This behavior has led to initially contradictory results. For example, when the PHA synthase is produced as an exogenous protein (e.g., in an expression system such as *E. coli*), the monomers will not be in the active state and a lag phase will be observed in the initial polymerization reaction. In contrast, reports of PHA synthases extracted from their native organism appear to be isolated in a primed state, meaning no lag phase is observed in the initial reaction (Gerngross et al., 1994). Several theories have been proposed involving dimerization or micelle formation, but atomic force microscopy surface reaction experiments reveal that it is very likely a combination of the two phenomena (Kikkawa et al., 2005; Sato et al., 2008). The kinetics of this priming phase are quite complex; however, some key observations have been recorded which offer a glimpse into this complicated process.

First, the duration of the lag phase has proven to be inversely dependent on the concentration of PHA synthase (Gerngross et al., 1994). This trait is evidence that traditional enzymatic multimerization governs this behavior. Secondly, it has been found that monomeric PHB synthase, whose specific substrate should be (R) β-hydroxybutyryl-CoA, can be primed using oligomeric CoA derivatives (Figure 41) (1996). These primed reactions do not proceed past initiation of the complexes because the oligo-CoA molecule is no longer recognized as substrate by the active enzymatic complex. Wodzinska et al. demonstrated that the optimal length of oligo-CoA derivatives used to prime the polymerization reaction, and subsequently reduce the observed lag phase, is a trimer primer (where n = 2 in Figure 41) (1996). This suggests that the majority of the lag observed for an unprimed reaction occurs while the polymer of a given synthase-polymer complex has less than six hydroxyacyl monomers incorporated. Because it is believed that only two substrate

molecules are required for the initial dimerization of the PHA synthase proteins, these results suggest that the complex must also become sufficiently amphipathic prior to full activation. Further evidence for this phenomena is given by unprimed reactions containing non-ionic detergents, such as TritonX-100 or Hecameg. The optimal presence of these detergents removes the lag phase entirely from unprimed reactions (Ushimaru et al., 2013). It is unclear if the presence of the detergent directly aids activation, or if it affects indirectly via alteration of the micelle formation rate.

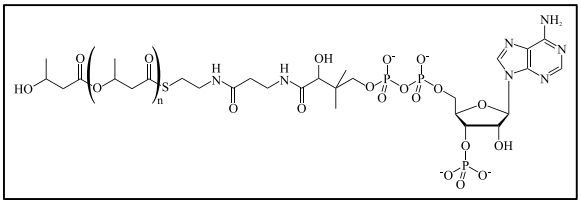


Figure 41: Oligomeric CoA derivatives used for priming PHA synthase polymerization reactions. Where n = 0, this figure depicts the PHB synthase substrate – β-hydroxybutyryl-CoA.

A simple dimerization kinetic model has been derived in an attempt to elucidate the role of dimerization kinetics in lag phase behavior as observed in *in vitro* PHA synthase polymerization reactions. Two data sets, investigating the wild-type type I PHA synthase of *Cupriavidus necator,* were used for the analysis of this model (Ushimaru et al., 2013; Yuan et al., 2001). Because lag phase behavior is commonplace for multimeric enzymes, the effects of the more physical phenomena will be supported via exclusion. In other words, if the analytical dimerization model does not represent the literature data well, the effect of the lipophilic polymer size (i.e., micelle formation) on the observed lag phase may contribute significantly to the overall lag phase behavior.

Table 10: Dimerization kinetic model derived for this work. Derivation can be found in appendix section A.13. E – monomeric synthase; E_2P_n – dimeric synthase with bound polymer of length n; S – substrate (hydroxyacyl-CoA); C – byproduct (CoA).

Mechanism	Fractional Conversion X(t)	Fractional State $\theta(t)$
$E + S \overset{k_1}{\rightarrow} \frac{1}{2}E_2P_2 + C$ $E_2P_n + S \overset{k_2}{\rightarrow} E_2P_{n+1} + C$	$1 - \exp\left[-\frac{[E]_\emptyset}{[S]_\emptyset}\left(1 - \frac{k_2}{2k_1}\right)\theta(t) - \frac{k_2[E]_\emptyset}{2}t\right]$	$1 - \exp[-k_1[S]_\emptyset t]$

7.3.2 Primed Polymerization

Definitive biphasic behavior of the PHA synthase polymerization reaction is exclusively observed for primed *in vitro* reactions and, to the authors' knowledge, there is only one published instance of this behavior. Using the primed PHA synthase of *Allochromatium vinosum*, the model type III PHA synthase, Müh et al. observed that the *in vitro* polymerization reaction undergoes a phase of high activity followed by reduced activity (1999). Because the *in vitro* reaction conditions are well controlled and the active complex assembly is complete upon initiation, the main causes of this observed phenomena are limited. For brevity, Phase I will be defined as the transient, highly active phase and Phase II will be defined as the final phase with less apparent enzymatic activity.

A logical focus would be the investigation of granule/micelle formation (Figure 42). One could hypothesize that Phase I is best represented by free-floating, soluble enzyme complexes actively polymerizing PHA with the increased enzymatic activity attributed to the high mobility/diffusivity associated with the dissolved molecules. After the polymer-synthase molecule becomes sufficiently amphipathic, it combines with other complexes and assembles into a PHA granule (Phase II). The source of the apparent decrease in the enzymatic activity could be a result of decreased diffusivity (i.e. mass-transfer limiation) as the molecule transitions from soluble to agglomerated. While not definitive, experimental results of a type I synthase contradict such a model. Gerngross and Martin successfully isolated soluble PHA synthase-polymer complexes from granule-bound PHA-synthases and determined that the soluble enzymes possessed a significantly lower activity than that of the granule-bound PHA synthases (1995). This increase in enzymatic activity for the granule-bound synthases is likely due to the fixed orientation of the catalytic site, which is normal to the granule surface and provides a higher rate of substrate/catalytic-site interaction in comparison with the more random orientation of the dissolved state.

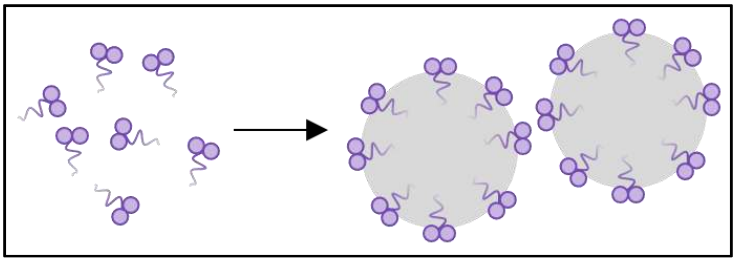


Figure 42: Formation – Hypothetical reduction in specific enzymatic activity due to the formation of the PHA granule. This hypothesis states the reduction in the apparent activity is a consequence of the relative diffusivities of the soluble (left) and granule-bound (right) enzyme states.

An alternate source of activity reduction could be granule/micelle growth. In this case, one could hypothesize that when the agglomerations are small, relative to the PHA synthase, the two subunits possess optimal interaction sterics. As the radius of the granule increases, the hydrophobic surface interacting with the hydrophobic domains of the PHA synthase multimer becomes less convex, leading to sub-optimal sterics for the interactions required for polymerization (Figure 43). Surface-binding proteins are known to be affected by the surface curvature, but experimentation is needed to investigate the hypothesis in this context (Gill et al., 2015).

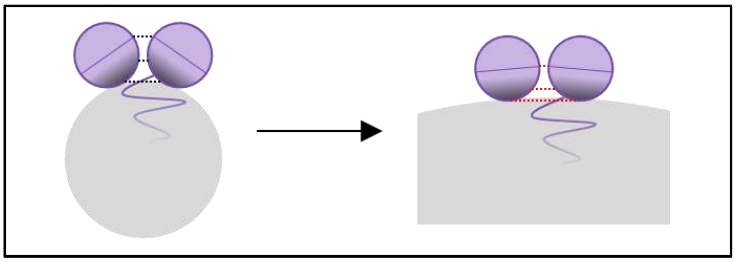


Figure 43: Growth – Hypothetical reduction in specific enzymatic activity due to the size of the PHA granule. The hydrophobic (shaded) regions of the PHA synthase possess significant interactions with the hydrophobic granule. As the granule increases in size the optimal energetic states of the protein-protein interaction sites (dashed lines) are altered.

A third possible explanation for the reduced apparent polymerization rate is not a reduction in specific activity but a loss of active enzyme. *In vitro* polymerization reactions are typically not performed with phospholipids and/or stabilization proteins, which are utilized *in vivo*, meaning the granules are consequently more suceptable to coalescence. If a loss of active enzyme were to occur when inhibitory effects are negligible, namely the presence of excess substrate and low fractional conversion, the cause

could result from the coalescence of the amorphous PHA granules. The same work that contradicts the hypothesis of Figure 42 reports observing the coalescence of PHA granules (Gerngross & Martin, 1995). Further support comes from the work of Nobes et al., who observed this phenomena exactly during transmission electron microscopy studies (2000). Nobes and coworkers noted a rapid decrease in the number of individual PHA granules in the early stages of polymerization, followed by a continual decrease in the number of PHA granules thereafter (Nobes et al., 2000). In this hypothetical system Phase I is best represented by a population of smaller and more active PHA granules. As these granules increase in size, the likelihood of coalescence also increases and rapidly leads to a population of significantly larger PHA granules. The reduction in enzymatic activity would be the result of either a lowered synthase activity on the larger granules, or the loss of surface bound enzyme during the coalescence process due to the incidental engulfment and subsequent entanglement of the polymer covalently bound to the enzyme (Figure 44). The enzyme would then work its way back to the more energetically favorable state at the granule surface, if possible.

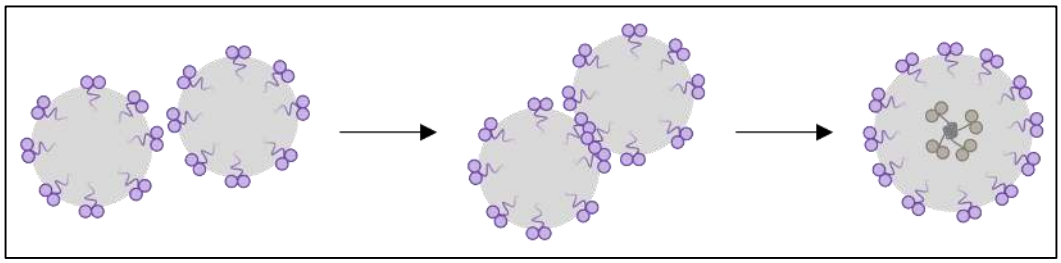


Figure 44: Coalescence – Hypothetical coalescence-mediated PHA synthase shielding. PHA synthases are incidentally engulfed within the PHA granule at the point of contact during coalescence and cannot immediately return to the surface of the PHA granule because the attached polymer has become entangled within the amorphous granule. The shielded enzyme can no longer encounter hydrophilic substrate from within the lipophilic granule.

The hypotheses discussed above are not mutually exclusive of the actual kinetic driving forces, nor are they totally inclusive when combined. The only certainty is that some fundamental state change is occurring, causing a significant decrease in the apparent enzymatic activity. An attempt to describe the primary mechanisms contributing to this phenomena has been performed by implementing four generic first-order state change models provided in Table 11. The first model is a first-order model describing only

polymerization. The second model represents the case similar to the depiction in Figure 43, where after the state change occurs, all enzymes remain capable of polymerization at a new reduced activity. The third model represents the case where an irreversible deactivation is occurring to the enzyme. The final model represents a case befitting the coalescence-mediated PHA synthase shielding (Figure 44), where a temporary/reversible deactivation occurs.

7.4 Results and Discussion

7.4.1 Unprimed Polymerization

The dimerization model was fitted to the data using residual sum of squares optimization for non-linear regression. The model significantly represented both sets of data ($p < 0.0001$ ea.) (Figure 45). The dimerization constant (k_1) varies significantly between the two sets. This result could be expected for such a comparison because these experiments were performed under different conditions, but it should be noted that statistical resolution of k_1 with this model will be rather difficult due to the nature of the kinetics being described.

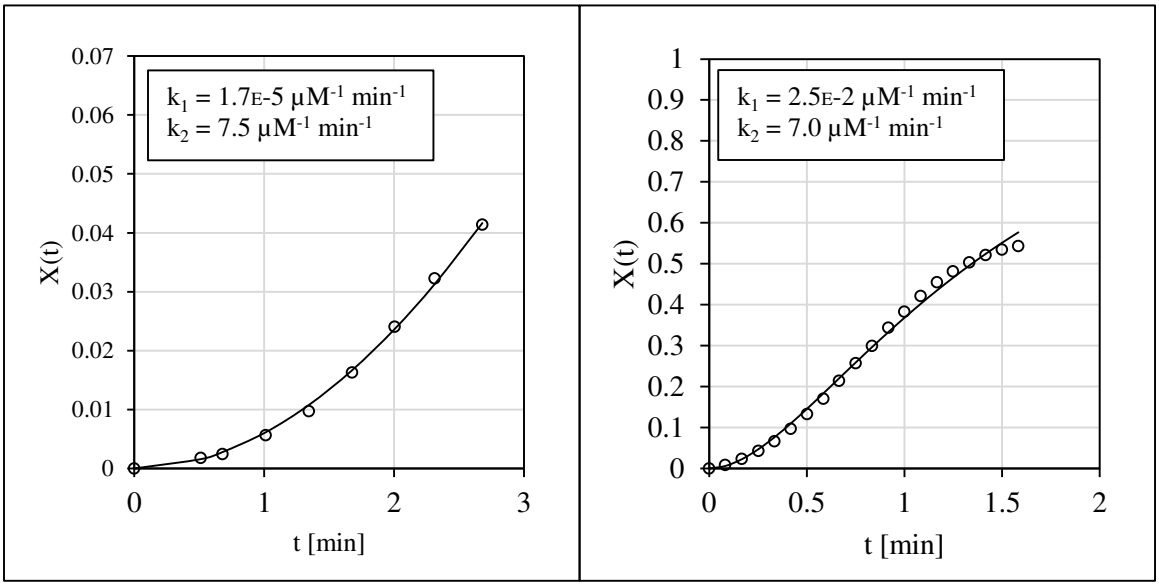


Figure 45: Dimerization model fit to data from literature. Regressor values are shown. (LEFT) Fractional conversion of the *C. necator* PHB synthase mediated polymerization reaction $[S]_\emptyset = 1{,}600$ µM, $[E]_\emptyset = 0.12$ µM (Yuan et al., 2001). (RIGHT) Transient CoA release of the *C. necator* PHB synthase mediated polymerization reaction $[S]_\emptyset = 100$ µM, $[E]_\emptyset = 0.21$ µM (Ushimaru, Sangiambut, Thomson, Sivaniah, & Tsuge, 2013).

Table 11: State change kinetic models derived for this work. The first-order model has been omitted due to its simplicity. Derivations can be found in the appendix sections A.14, A.15, and A.16. E_2P_n – Dimeric synthase with bound polymer of length n; $E_2P_n^*$ – E_2P_n in lesser active, or inactive state; S – Substrate (hydroxyacyl-CoA); C – Byproduct (CoA).

Phenomena	Mechanism	Fractional Conversion X(t)	Fractional State $\theta(t)$
Activity Reduction	$E_2P_n + S \xrightarrow{k_1} E_2P_{n+1} + C$ $E_2P_n \xrightarrow{k_2} E_2P_n^*$ $E_2P_n^* + S \xrightarrow{k_3} E_2P_{n+1}^* + C$	$1 - \exp\left[-\frac{[E]_\emptyset}{2}\left(\frac{k_1 + k_3}{k_2}\theta(t) + k_3 t\right)\right]$	$1 - \exp[-k_2 t]$
Irreversible Deactivation	$E_2P_n + S \xrightarrow{k_1} E_2P_{n+1} + C$ $E_2P_n \xrightarrow{k_2} E_2P_n^*$	$1 - \exp\left[-\frac{k_1[E]_\emptyset}{2k_2}\theta(t)\right]$	$1 - \exp[-k_2 t]$
Reversible Deactivation	$E_2P_n + S \xrightarrow{k_1} E_2P_{n+1} + C$ $E_2P_n \xleftrightarrow{k_2} E_2P_n^*$	$1 - \exp\left[-\frac{k_1[E]_\emptyset}{2}\left(\frac{\theta(t)}{k_2 + k_{-2}} + \left(1 - \frac{k_2}{k_2 + k_{-2}}\right)t\right)\right]$	$\frac{k_2}{k_2 + k_{-2}}(1 - \exp[-(k_2 + k_{-2})t])$

The regressed polymerization rate constant (k_2) from the data recorded by Yuan et al. (Yuan et al., 2001) proved to have no significant difference from the constant regressed from the data recorded by Ushimaru et al. (Ushimaru et al., 2013) (i.e., $6.2 \leq k_{2,\text{ Ushimaru et al.}} \leq 7.8\ \mu M^{-1}\ min^{-1}$). This suggests the model may allow for the elucidation of a more accurate polymerization rate constant for comparison of PHA synthase enzymatic activities.

One limitation of the dimerization model observed during the fitting of these data was that in order to produce resolved parameter values, the data should well represent the post-dimerization polymerization phase, meaning some downward curvature in the late region. The data of Yuan et al. lacks these measurements, so while the residual sum of squares value for this fit is very low, it remains difficult to resolve the fitting parameters k_1 and k_2, which is essential for experimental comparisons through traditional statistical analysis.

A second limitation of this model is the requirement of $[S]_\emptyset >> [E]_\emptyset$. The consequences of the violation of this assumption can be observed in the fit to the data of Ushimaru et al. in the right-hand panel of Figure 45. Here we can see the reaction slowing soon after dimerization is complete, which is edging on violation of the stated assumption. Because the fractional conversion remains so low as the reaction rate slows ($X_{final} = 0.55$), this effect might be caused by competitive inhibition of the free CoA (S. Zhang et al., 2000). This inference suggests that future experiments implementing this model to elucidate the polymerization rate constant should employ a substrate-to-enzyme ratio of greater than 500:1 (S:E) in order to avoid potential inhibition.

Overall these case studies provide keen insights into the factors influencing lag during PHA synthase *in vitro* polymerization. Under conditions typical of an enzymatic assay, namely substrate excess, the bulk of the complex assembly kinetic behavior can be captured through PHA synthase dimerization kinetics. These results indicate that dimerization, or an analogous first-order enzyme-dependent reaction, is the main factor influencing the lag phase kinetics in an excess of substrate. Additionally, because the two polymerization rate constant (k_2) values were comparable between discrete experiments with unique

reaction compositions, this model may be a useful tool to elucidate and compare accurate PHA synthase polymerization rate constants. The dimerization rate constant, k_1, provides qualitative insight into an enzyme's readiness to form an active complex; however, this parameter should not be acknowledged as anything more than a semi-quantitative measure because this model fails to address the dependence of the complex formation on the size of the lipophilic chain. In particular, it fails to address how the activity of the synthase-polymer complex changes as its polymer progresses between two to six monomers in length.

7.4.2 Primed Polymerization

The models listed in Table 11 were fitted to the data using residual sum of squares optimization for non-linear regression. Additionally, a simple first-order polymerization model under the same assumptions was fitted to the data for comparison to the state change model. This comparison was used to investigate if biphasic behavior is significantly different from monophasic behavior. The first-order model also proved useful in providing the initial polymerization rate constant values for nonlinear regression.

All models significantly represented the observed data ($p < 0.0001$ ea.). The fitted data can be observed in Figure 46. The Akaike's Information Criterion (AIC) for sum-of-squares likelihood was used for non-nested model comparisons. AIC is a measure of the relative quality of a statistical model (Burnham & Anderson, 2002). The calculated value depends upon the number of fitted observations (n) and the log of the residual sum of squares (RSS), so as n increases or RSS decreases (less residual error), AIC becomes more negative. The first-order model AIC value was found to be −97, while both the reduced activity and reversible deactivation models were −134, and the irreversible deactivation model was −101. The relative ability of the models to describe the data is illustrated in Figure 46. Comparing AIC, RSS, F-statistic, and graphical representation shows both the activity reduction model and the reversible deactivation models are significantly better at describing the observed phenomena than the first-order and irreversible deactivation models.

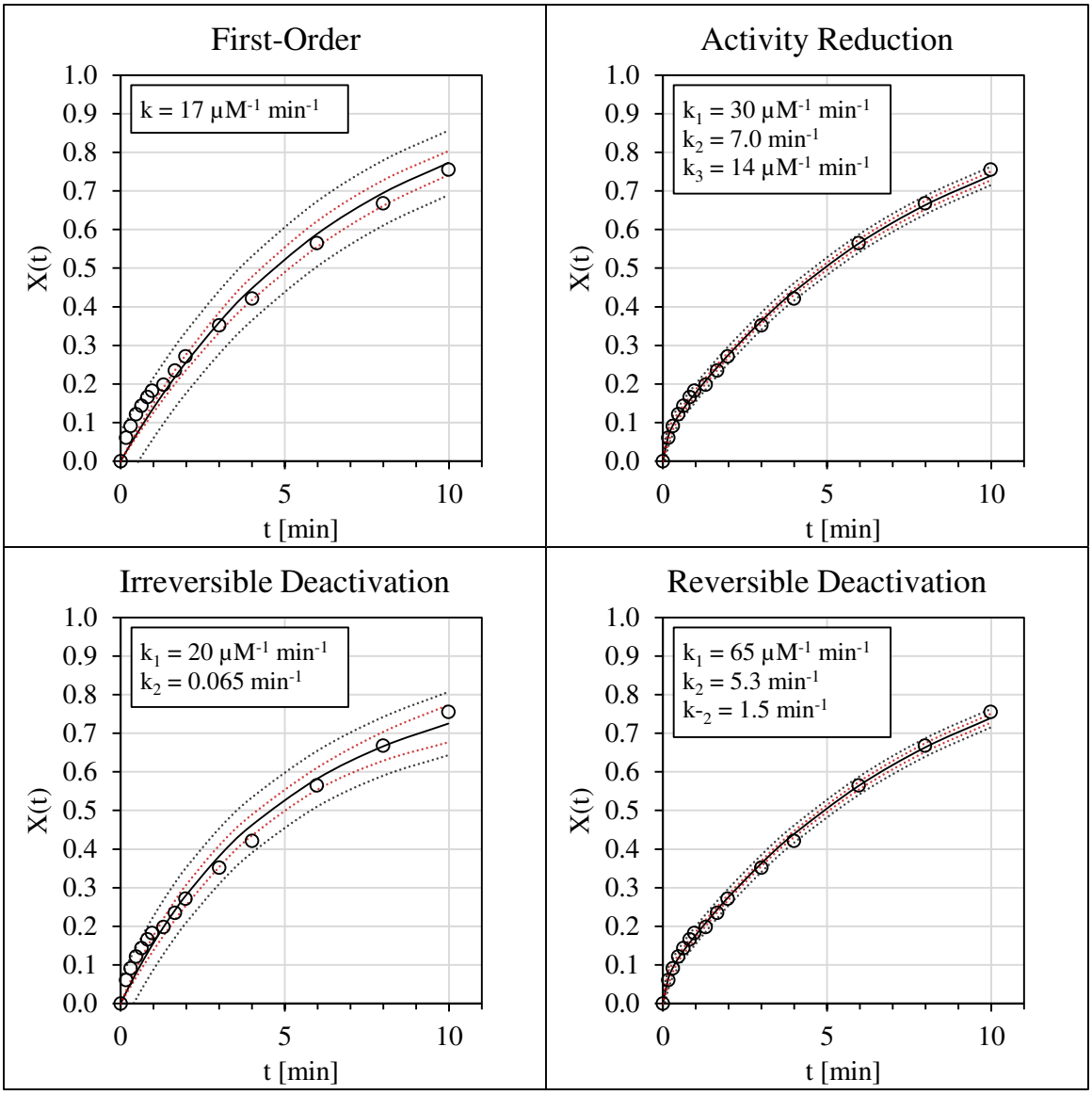


Figure 46: Fractional conversion models derived in this section and fit to the data from literature (Müh, Sinskey, Kirby, Lane, & Stubbe, 1999). Dashed lines indicate 95% approximate confidence intervals on the mean (red) and 95% approximate prediction intervals (black). Regressor expected values given for each respective model.

During analysis of the activity reduction and reversible deactivation models, it was observed that the fit descriptors were identical within the resolution of the statistical tests. In other words, the RSS for each model was 0.001144 and both F-statistics were 6691. For any given predicted value the difference between the models was negligible ($\mathcal{O}$(-6) to $\mathcal{O}$(-8)). Comparison of the two models showed that they both belong to the same function family (Equation 7), where the constants are differing functions of the model-

specific rate constants (Table 12). Because these two models both fit the same function family, it will be difficult to discern which one better represents this observed phenomena through analysis of similar kinetic experiments in the future.

Equation 7: Generic function describing the reduced activity and reversible deactivation model fractional conversions. A, B, and C are constants which are functions of the model-specific rate constants. Relationships are listed in Table 12.

$$X(t) = 1 - \exp[A(1 - \exp[Bt]) + Ct]$$

Table 12: Relationship between the model-specific rate constants and generic family constants for the reduced activity and reversible deactivation models.

Activity Reduction	Reversible Deactivation
$k_1[E_2P_n]_\emptyset = AB + C$	$k_1[E_2P_n]_\emptyset = AB - C$
$k_2 = -B$	$k_2 = \frac{-B}{1 - \frac{C}{AB}}$
$k_3[E_2P_n]_\emptyset = -C$	$k_{-2} = B\left(\frac{1}{1 - \frac{C}{AB}} - 1\right)$

The activity reduction and reversible deactivation models successfully capture the initial high activity phase region and the relative optima values of the activity reduction polymerization rate constants, k_1 and k_3 (i.e., $k_1 > k_3$). Under approximate 95% confidence intervals, the activity reduction initial polymerization rate constant k_1 ($-4.1 \leq k_1 \leq 77\ \mu M^{-1}\ min^{-1}$) and the reversible deactivation rate constant k_2 ($-0.44 \leq k_2 \leq 11\ min^{-1}$) are the only parameters not significantly different from the null. However, when Equation 7 is fit to the data, all regressors are significant. Because each rate expression possesses three rate constants that are independently related to the three significant regressors, the rate expression constants should also be significant. It was found that k_1 and k_2 were also the regressors of highest bias (17% ea.) for each respective model, suggesting the approximate standard deviation is not a good estimator of error in these cases (Box, 1971). For biased parameters, the error is not balanced above and below the expected

value, and the model requires transformation (e.g., linearization) in order to provide balanced deviations from the expected value.

Limiting the fitting of the first-order model to the data of Phase I ($n = 8$, AIC = −56), a higher polymerization rate constant was obtained ($19 \leq k \leq 27$ μM^{-1} min^{-1}, $\alpha = 0.05$). The resulting fit was poor (Figure 47) and possessed a RSS value an order of magnitude higher than the Phase I RSS of the activity reduction and reversible deactivation models. The reduction and reversible deactivation models represent the Phase I data better than the first-order approximation, and the first-order approximation of polymerization alone does not represent Phase I kinetic behavior well.

In contrast, the fitting of the first-order model to the Phase II data ($n = 7$, AIC = −54) describes the phenomena relatively well. The RSS of the Phase II first-order model (0.0017) was comparable to the Phase II RSS of the state change model (0.0011), suggesting that the AIC value suffers from a low number of data points. The only model which cannot describe the phenomena well in Phase II is the irreversible deactivation model.

The fractional state, $\theta(t)$, of the state change models can provide useful insights into the composition of the *in vitro* reaction, such as: Is this fraction bounded between zero and one? When does it significantly change? What is the final extent? All of these questions can yield useful insight into the state of the enzymes within the reaction according to the given model. The comparison of fractional states of the state change models can be observed in Figure 49.

It was found that under the activity reduction model, nearly all of the enzyme is in the lesser-active state before the end of Phase I. The reversible deactivation model suggests that even though the deactivation can be reversed, the polymerization observed in Phase II would be performed by only 25% of the initial enzyme concentration. The irreversible deactivation model shows that the fraction of enzyme deactivated is steadily increasing, as would be expected of an irreversible deactivation; however, this phenomena likely

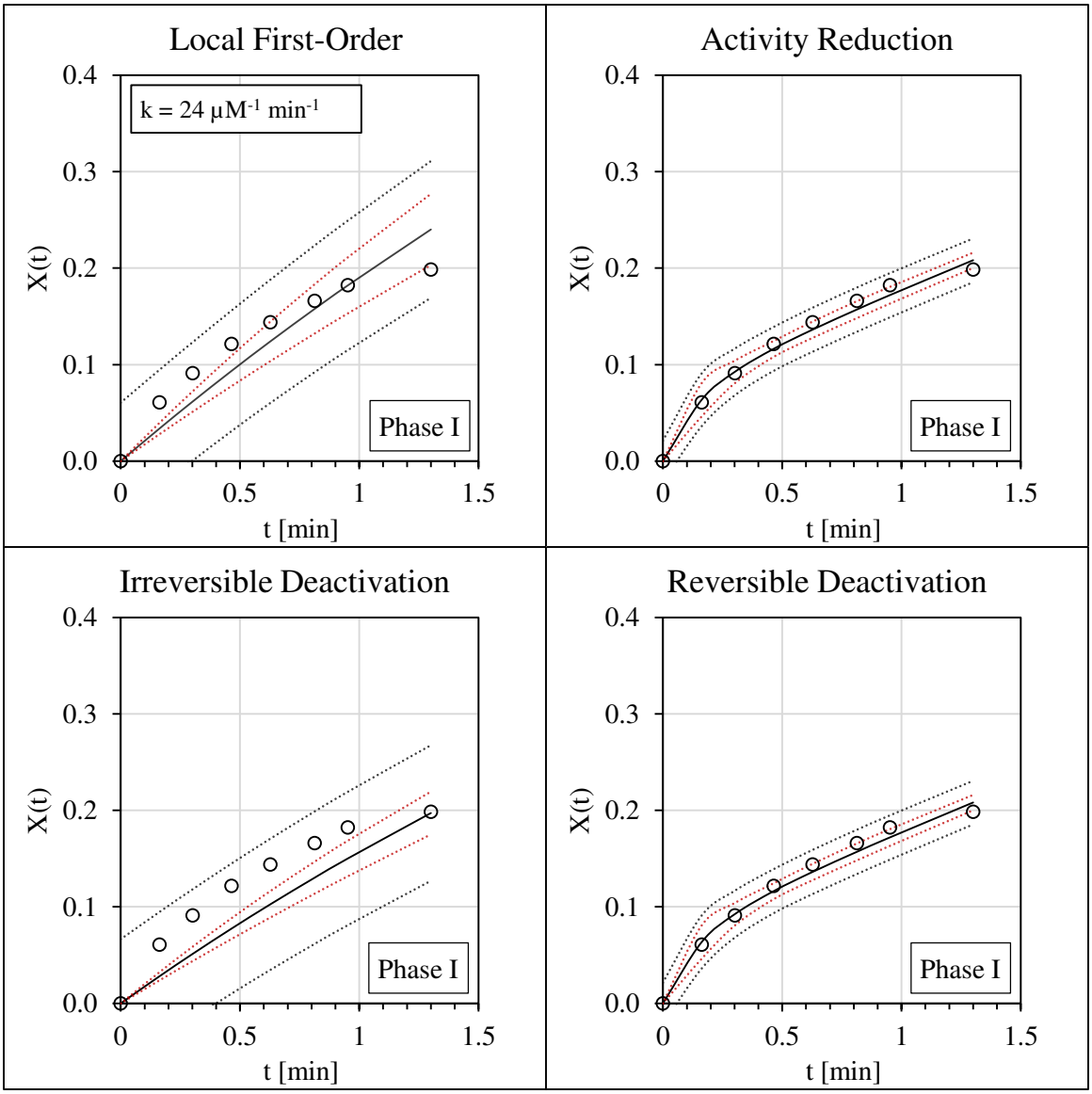


Figure 47: The first-order model fit only to points within Phase I (n = 8) of the data in Figure 46. The original fits of the activity reduction, irreversible deactivation, and reversible deactivation models are depicted in the same scale for comparison. Dashed lines indicate 95% confidence intervals on the mean (red) and 95% prediction intervals (black).

does not occur because this model does not represent the data well. This unique case study offered an interesting perspective into the *in vitro* PHA synthase polymerization kinetics. The results clearly indicate that the Phase I region is not represented well by either first-order polymerization kinetics or irreversible deactivation kinetics. The degree to which the activity reduction and reversible deactivation models describe the entirety of the data suggests that some form of state change is occurring to the active enzyme

complex throughout the polymerization process, which is not dependent on the substrate or byproduct concentrations. The work of Nobes et al. provides strong evidence that this observed state change may be caused by coalescence (2000).

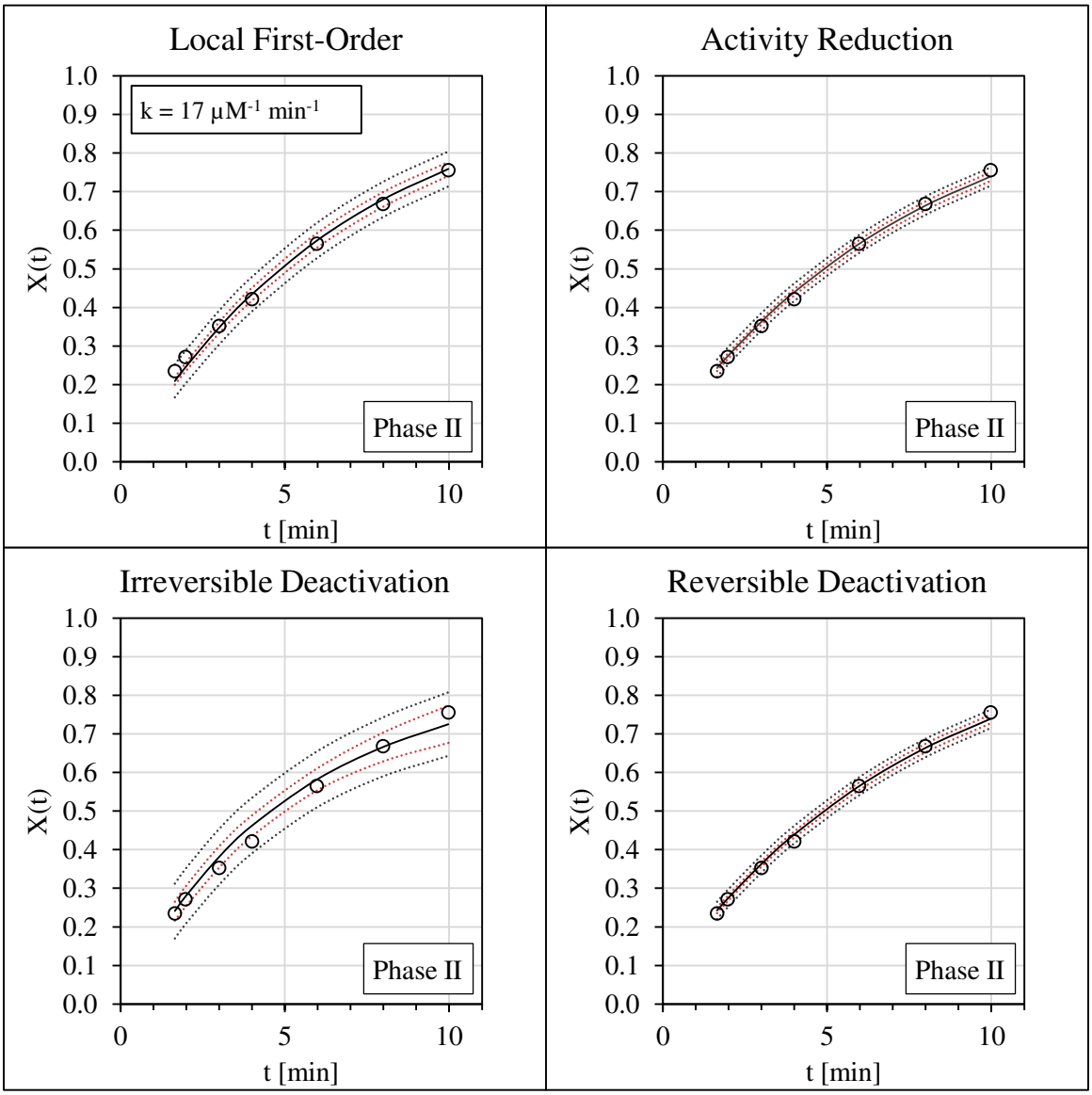


Figure 48: The first-order model fit only to points within Phase II (n = 7) of the data in Figure 46. The original fits of the activity reduction, irreversible deactivation, and reversible deactivation models are depicted in the same scale for comparison. Dashed lines indicate 95% confidence intervals on the mean (red) and 95% prediction intervals (black).

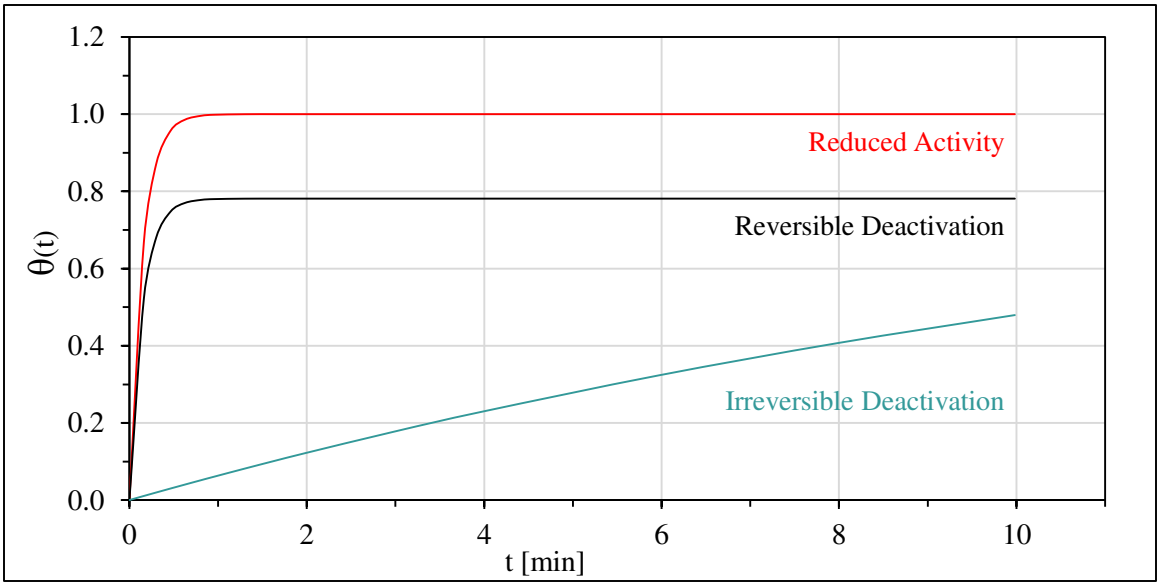


Figure 49: Transient enzymatic fractional states as defined in the models derived in this work. Reduced activity fractional state represents the fraction of enzyme in the lesser-active state. The reversible and irreversible deactivation fractional states represent the fraction of enzyme currently deactivated.

7.5 Previous Kinetic Studies

Here we examine some previously reported PHA synthase kinetic investigations. We exemplify the consequences of non-ideal *in vitro* polymerization assay design and how the undesirable effects may be avoided.

Burns et al. proposed one of the few, and rather ambitious, mathematical models attempting to describe *in vitro* PHA synthase polymerization kinetics (2007). The complex model attempts to describe the three-enzyme PHA biosynthesis pathway (PhaA, PhaB, and PhaC) of *C. necator*. Unfortunately, due to the nature of this complex experiment, they may have had to compromise on accuracy for efficiency during data acquisition. A major item called into question is the method of PHA synthase enzymatic activity quantitation. A continuous DNTB assay was utilized and corrected for background due to DNTB-PHA synthase interaction. Müh et al. showed that a continuous DNTB assay significantly underestimates the actual activity of PHA synthase (1999). Performing a continuous DNTB assay and subtracting a background value could lead to a severely underrepresented PHA synthase activity data set. A

discontinuous assay may have provided more accurate results for Burns and co-workers, if the overall experimental design allowed for such a measurement.

Nobes et al. also performed kinetic analysis on their *in vitro* PHA synthase polymerization reaction during their coalescence studies. They report using Michaelis-Menten kinetics for a single substrate reaction (Nobes et al., 2000). Unfortunately, they assayed an unprimed polymerization reaction and failed to account for the lag phase assembly kinetics, leading to initial rates that do not represent polymerization well. Fortunately, the results of the kinetic analysis have no effect on their observed coalescence results. Multimerization assembly kinetics may provide more accurate regression values for the polymerization rate constants of unprimed reactions. Alternatively, the reactions could be trimer-primed, if available. Omission of the lag phase by non-ionic detergents may not yield consistent results across PHA synthase variants, so it should be avoided when attempting to present accurate and comparable quantitative data.

7.6 Conclusions

Great precaution should be taken during the experimental design of an *in vitro* PHA synthase enzymatic activity assay in order to avoid potential misrepresentation of the active form of the enzyme. The presence of non-ionic detergents and the use of native synthase can each yield reactions lacking the lag phase behavior. The *in vitro* PHA synthase polymerization reaction possesses many complex mechanisms, which combined produce an apparent behavior that cannot (exactly) be modelled analytically. Dimerization kinetics were found to describe the majority of observed lag phase behavior well, suggesting it is a key contributor to the active complex assembly step. The biphasic behavior observed by Müh et al. was described by an enzyme-dependent first-order state change model (Müh et al., 1999). Furthermore, there is strong evidence that the state change is related to coalescence of growing PHA granules (Nobes et al., 2000).

A pictographic summary of the observed behavioral phenomena is given in Figure 50. It is very likely that a well-defined Phase I behavior is not observed in unprimed reactions because the increase in specific

enzymatic activity for granules in this state is greatly offset by the reduction in specific activity from enzymes in the Priming Phase state. It will be a difficult task to discern, through kinetic analysis, whether this state change of Phase II is actually due to reversible deactivation, which represents the hypothesized shielding discussed in Figure 44, or a reduction in the enzymatic activity, a more fundamental state change. A physical experiment could be performed to test if any PHA synthase can be located within the PHA granule after *in vitro* polymerization. This experiment could be accomplished by exposing the mature granules to a non-specific protease, followed by the recovery and detection of any proteins that may have been protected by the granule during digestion. Alternatively, one could fuse fluorescent tags to the PHA synthase and quantify the residual enzyme via epifluorescence after digestion.

The inferences of this work offer a glimpse into the complex PHA synthase polymerization mechanisms, and the results from these case studies should aid in the design of future PHA synthase enzymatic activity assays.

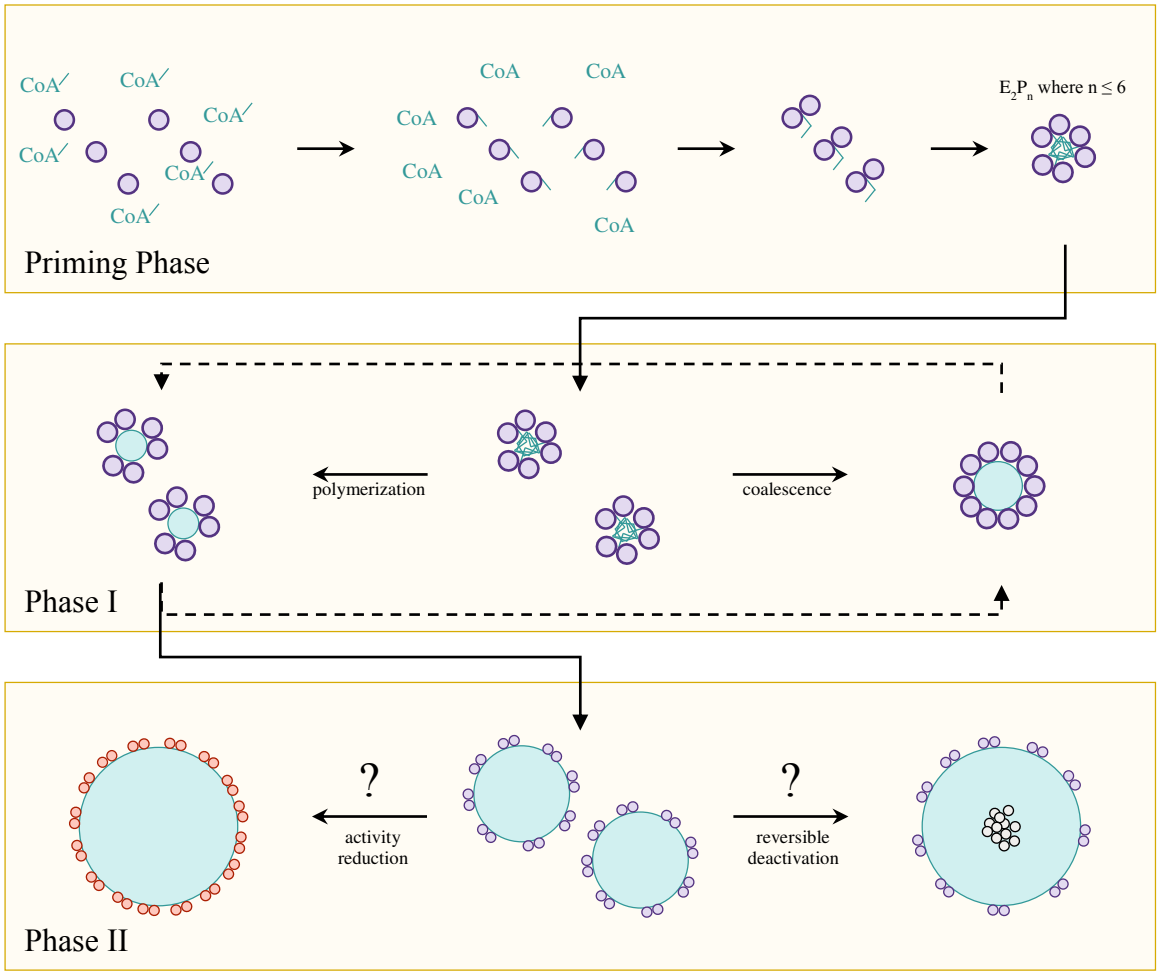


Figure 50: Depiction of major phenomena observed in *in vitro* PHA synthase polymerization kinetic behavior. The Priming, or lag, phase of the reaction describes the initial formation of the active PHA synthase-polymer complex/agglomeration. Phase I describes the transient state where the PHA synthase possess its maximum rate of substrate consumption. Granules in Phase I will proceed into Phase II once they are sufficiently large. Phase I should be most readily observed in pre-primed polymerization reactions. Phase II exhibits a reduction in the observed enzymatic activity as compared to Phase I. The mechanism of this reduction is unclear, but the phenomena is independent of substrate concentration and dependent on the concentration of active enzyme. One certainty is that the observed rate of substrate consumption is inversely proportional to the size of the PHA granules.

CHAPTER 8. PHA SYNTHASE RATIONAL MUTATION

8.1 Chapter Preface

This chapter records the progression of Objective III (section 1.4) – Investigate potential causes of low PHA accumulation in cyanobacteria via rational mutagenesis of PHA polymerization enzyme.

8.2 Introduction

Even though cyanobacteria generally show lower yields of PHA as compared to heterotrophic organisms, there is a lack of effort focused towards answering why. Instead, current research has turned its focus toward recombinant overexpression of an exogenous PHA synthase in order to increase PHA synthesis. In other words, many researchers are working to treat the symptom rather than diagnosing the cause.

Exogenous recombinant metabolic manipulation may not be ideal for industrial-scale operations, as these methods typically introduce a form of antibiotic resistance, or other metabolic selector, as a selectable marker. Looking at the case of antibiotic selection as an example, such a method of metabolic alteration requires a constant presence of the antibiotic in the cultivation media to ensure preservation of the genotype. Antibiotics in a large-scale, continuous bioprocess can increase in-house production costs, limit viable site locations, and may increase associated ecosystem contamination prevention costs. This is especially true for organisms as natively robust as cyanobacteria (Akiyama et al., 2011). Because of the many potential consequences of a "designer" cyanobacteria, a PHA bioproduction process should utilize cyanobacteria possessing a genotype as close to the wild-type as possible.

In many cases, exogenous PHA synthase expression has led to an improvement in lab-scale PHA accumulation by as much as 4-5-fold (Akiyama et al., 2011; Kumar Sudesh, Taguchi, & Doi, 2002). This suggests that the limiting factor to PHA accumulation in cyanobacteria is likely associated with the PHA synthase. The limiting factor of PHA yield could be the effects from transcription, translation, or product functionality, or any combination thereof.

After the inception of this work, Numata et al. determined that the enzymatic activity level of the PHA synthase in *Synechocystis* sp. PCC 6803 was indifferent from that of a higher PHA-accumulating heterotrophic organism (Numata et al., 2015). This observation indicates that the enzymatic activity of cyanobacterial PHA synthases does not cause the low observed yields of PHA, but it is instead more likely the transcription or translational effects (i.e., regulation) which cause the observed low PHA yields in cyanobacteria. In other words, if PHA synthase is the limiting factor of PHA accumulation in cyanobacteria, then it is likely due to the quantity of PHA synthase, rather than its aptitude.

The goal of this work is to investigate the function of cyanobacterial PHA synthase idiosyncrasies, when compared to non-cyanobacterial PHA synthases, as they relate to the apparent enzymatic activity. While evidence shows these experiments will not help to increase the PHA yields in cyanobacteria directly, they still provide key insights into the mechanics of PHA synthase catalysis in cyanobacteria.

8.3 Materials and Methods

8.3.1 Loci and Mutant Determination

The type III PHA synthase of the model cyanobacteria, *Synechocystis* sp. PCC 6803 was selected for investigation in this work. The PHA synthase conservational analysis performed in CHAPTER 5 provided an interesting conserved locus to target for mutation – the cyanobacteria-specific insertion, or cyanobacterial box. This region of the PHA synthase was selected because of its uniqueness to cyanobacteria, making it an idiosyncrasy in the PHA synthase protein family.

Following the analysis of CHAPTER 5, the cyanobacterial box is redefined in this work as the residues in the positions of the inclusive range of 205-209 in the *Synechocystis* sp. PCC 6803 accessed sequence (Appendix Table 4). Each residue of this region was mutated independently. Mutant residues were selected based off of the wild-type residue physiochemical properties. The mutant residues introduced varied the chemical properties of the residue and attempted to maintain a comparable molecular weight.

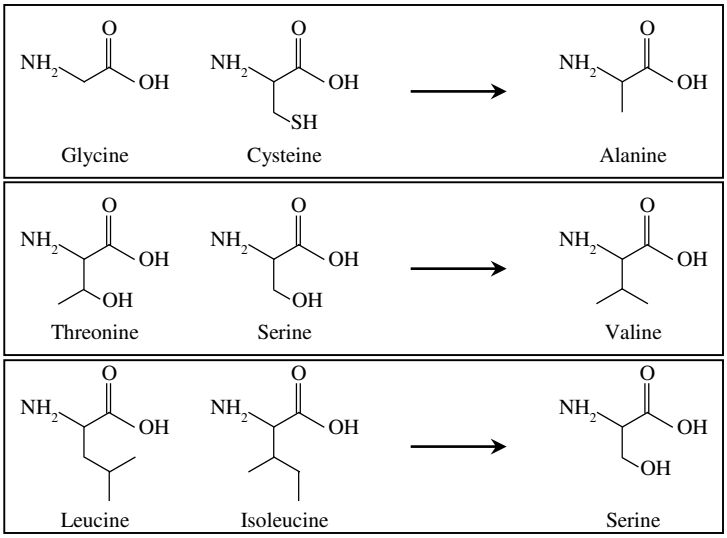


Figure 51: Molecular structure of the wild-type amino acids (left) and their respective designated mutant (right) residue used in this study.

8.3.2 Isolation and Cloning of *Synechocystis* sp. PCC6803 *phaC* and *phaE*

The wild-type strain of *Synechocystis* sp. PCC6803 was cultivated in a plant tissue culture incubator (Forma Scientific, model 3750) at 29°C under constant fluorescent lighting (60 µmol m^{-2} s^{-1}) in BG-11 medium (Rippka et al., 1979) supplemented with 100 mmol l^{-1} 2-[tris(hydroxymethyl)methylamino]-1-ethanesulfonic acid (TES) buffer (pH 8.2). Three 50 ml cultures, agitated once daily, were grown in 250 ml Erlenmeyer flasks to an optical density (λ = 730nm) of six. Next, two one milliliter samples (~5E8 cells) were collected for each culture replicate and the genomic DNA was harvested using the following protocol. Cells were pelleted via centrifugation at 3,300xg for 10 min, decanted, and centrifuged again at 9,200xg for 3 min to remove residual media. Cells were then resuspended in 100 µl lysis solution (0.5x TE buffer, pH 8.0, supplemented with 50 mmol l^{-1} sodium chloride). 15 µl of 50 mg ml^{-1} lysozyme solution (100 mmol l^{-1} Tris, pH 8.0) was then added and samples were incubated at 55°C for 30 min. Next, 5 µl of 20 mg ml^{-1} proteinase K and 20 µl of 10% sodium dodecyl sulfate was added and the incubation was repeated for 20 min. Following lysis, two equal volume organic separations using phenol:chloroform:isoamyl alcohol (25:24:1) were performed. Finally, ethanol precipitation was

performed using a final concentration of 0.2 mol l^{-1} sodium chloride. Samples were resuspended in TE buffer, analyzed via spectroscopy, and stored at -20°C.

Table 13: Table of primer sequences used in this work. Relevant restriction enzyme recognition sites are listed and cleavage sites are denoted by '^'. Strand is in reference to coding sequence as strand 1.

Name	Sequence	Strand	Sites
1829.1	ATTG^GATC^CATGGAATCGACAAATAAAACCTGG	1	BamHI, NcoI
1829.2	AAGG^TCGACTTAGCCTGGGTTTGCTTCTG	2	SalI
1829.3	TACT^GTACACAACATGGAATCGACAAATAAAACCTGG	1	Bsp1407I
1829.4	CAGA^AGCTTTTAGCCTGGGTTTGCTTCTG	2	HindIII
1829.6a	AAGC^TCGAGGCCTGGGTTTGCTTCTG	2	XhoI
1829.7	CCGA^GATCTCATGGAATCGACAAATAAAACCTGG	1	BglII
1830.1	TCAG^GATCCATGTTTTTACTATTTTTTATCGTTCATTGGT	1	BamHI
1830.2	GCAG^TCGACTCACTGTCGTTCCGATAGC	2	SalI
1830.3	TACT^GTACACAACATGTTTTTACTATTTTTTATCGTTCATTGGT	1	Bsp1407I
1830.4	CAGA^AGCTTTCACTGTCGTTCCGATAGC	2	HindIII
1830.6a	CAGC^TCGAGTCACTGTCGTTCCGATAGC	2	XhoI

Once genomic DNA was isolated, high-fidelity PCR amplification was performed using Phusion High-Fidelity DNA Polymerase (New England Biolabs, #F530S) for each biological replicate (n = 3). The 1829.1 and 1829.2 (*see* Table 13) oligonucleotides were used to amplify the wild-type *phaE* gene (slr1829), and the 1830.1 and 1830.2 primers were used to amplify the wild-type *phaC* gene (slr1830). Each 50 µl reaction consisted of approximately 200 ng DNA template, 0.5 µmol l^{-1} each oligonucleotide primer, 1x Phusion HF Buffer (New England Biolabs), 0.2 mmol l^{-1} dNTPs, and 1 U polymerase. These reactions were subjected to [98.0°C for 00:30], followed by 32 cycles of [98.0°C for 00:10, T_A for 00:20, 72.0°C for 00:30] (where T_A was 67.5°C for *phaC* and 65.0°C for *phaE*), then [72.0°C for 10:00] and immediately stored at four centigrade. Initial target validation of amplification products was performed through partial restriction digestion. The combined products were subjected to 1 hour of digestion with SmaI (Thermo Scientific, FD0663). Products and digests were visualized using a one percent agarose gel (0.5xTBE) with 0.5 µg ml ethidium bromide staining, subjected to an electric tension of 5 V cm^{-1} for 90 min.

The PCR amplification products were then independently inserted into the pUC18 vector (Agilent Technologies, 200231-42). This was performed through one hour digestions of both products and vector with BamHI (Thermo Scientific, FD0054) and SalI (Thermo Scientific, FD0644), followed by a 3:1

(insert:vector) ligation using T4 DNA Ligase (Thermo Scientific, EL0014), and finally, transformation into *E. coli* XL1-Blue (Agilent Technologies, 200130). Traditional blue-white screening was implemented for colony selection for subsequent colony PCR.

Colony PCR was performed using the universal M13 primers and DreamTaq (Thermo Scientific, EP0702). Colony PCR products were subjected to SmaI digestion again in order to confirm the presence of the high fidelity PCR product insert. Confirmed strains were then stored in 30% glycerol selective medium at -80°C. Next, reserves were continued plasmid for plasmid isolation using a miniprep kit (Axygen Biosciences, AP-MN-P-50). Isolated plasmids were then subjected to DNA sequencing (Applied BioSystems BigDye Terminator (v3.1), ABI Prism 3130) using the M13F primer for sequence confirmation. These plasmids were deemed pCL0001 (pUC18(*phaE*)) and pCL0002 (pUC18(*phaC*)).

8.3.3 Site-Directed Mutagenesis of *phaC*

Site-directed mutagenesis of the cyanobacteria-specific insertion region of *phaC* was performed using the phosphorylated primers depicted in Table 13. High-fidelity PCR was performed on pCL0002 using the two of the primers shown, under the same reaction conditions used for previous high-fidelity PCR amplification. These reactions then underwent the following thermal cycles: [98.0°C for 00:30], followed by 24 cycles of [98.0°C for 00:10, 72.0°C for 02:00], and then [72.0°C for 10:00], reactions were then immediately stored at four centigrade. PCR products were then re-circularized using T4 DNA ligase and transformed into *E. coli* XL1-Blue.

The 1830.205.R primer was used in conjunction with 1830.G205A.F, 1830.C206A.F, and 1830.T207V.F primers to create the pCL0002(G205A), pCL0002(C206A), and pCL0002(T207V) mutant vectors, respectively. The 1830.209.F primer was used with the 1830.L208S.R and 1830.G209A.R primers to create the pCL0002(L208S) and pCL0002(G209A) mutant *phaC* vectors, respectively. The deletion mutant, pCL0002(Δ205-209), was the re-circularized product of 1830.209.F and 1830.205.R. Mutant pCL0002 derivatives were isolated and subjected to DNA sequencing using the M13F universal primer.

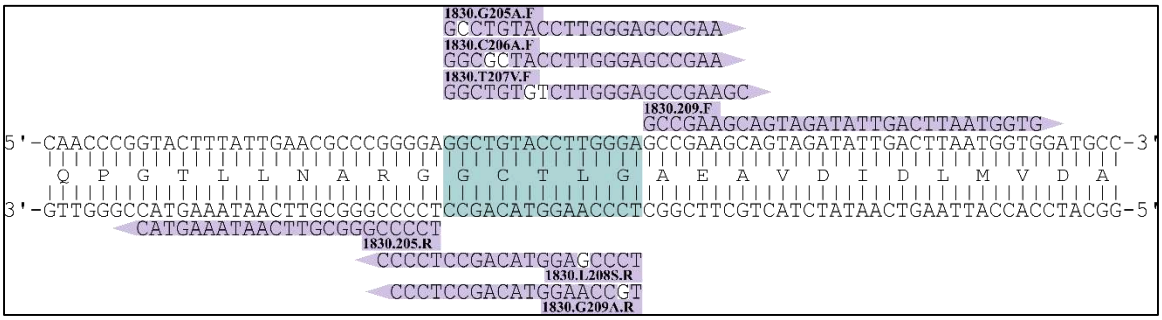


Figure 52: Primer sequences of mutational insertion primers used in this study. The cyanobacteria-specific insertion region of the *phaC* template sequence is highlighted.

8.3.4 Dual ELP-tagged Split Intein Expression Vector Construction

In order to express and purify an untagged form of the recombinant enzymatic subunits, it was decided to use the dual ELP-tagged split-intein system described by Shi *et al.* (Shi et al., 2013). For this system, a mutant intein has been split into two functional motifs – N-terminus and C-terminus. The target protein is fused to the partial intein C-terminus fragment and can be cleaved selectively through N-terminus complementation. Additionally, both intein fragments are fused to elastin-like proteins for non-chromatographic separations/purification. *E. coli* BLR(DE3) strains harboring the following vectors were a gift from the lab of David Wood at Ohio State University: pE/E_{11}I0N, pE/E_{11}I0C(GFP), and pE/E_{11}I0C(MBP).

pE/E_{11}I0C(GFP) was isolated, digested with Bsp1407I (Thermo Scientific, FD0934) and HindIII (Thermo Scientific, FD0504), and gel purified in order to obtain the vector with which to insert the PHA synthase subunit coding sequences. The pCL0005 (pE/E_{11}I0C(*phaC*)) plasmid inserts were obtained through the high-fidelity PCR amplification of the respective pCL0002 vector using primers 1830.3 and 1830.4 (*see* Table 13). The pCL0006 (pE/E_{11}I0C(*phaE*)) plasmid was constructed similarly using primers 1829.3 and 1829.4 and amplifying the pCL0001 template. PCR products were cleaned up via spin column (Thermo Scientific GeneJet, K0691), digested, ligated into the pE/E_{11}I0C vector, and then transformed into *E. coli* XL1-Blue. The insert's location and orientation were confirmed via DNA sequencing using the

upstream primer poly(Asn).Ck.F (TAACAACAACCTCGGGATCG) and downstream universal T7terminator primer. Once sequenced, the same plasmid isolates were transformed into *E. coli* BLR(DE3) using the one-step transformation protocol described by Chung and Miller (Chung & Miller, 1993).

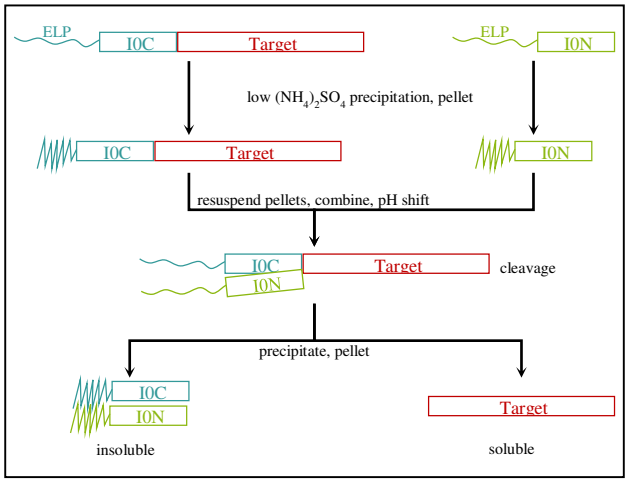


Figure 53: Flow diagram depicting purification of total cell lysate using the dual ELP-tagged split intein expression system (Shi, Meng, & Wood, 2013).

8.3.5 Hexahistidine-Tagged Expression Vector Construction

As an alternate method of expression, it was decided to perform traditional hexahistidine-tagging on the enzymatic subunits. The pET28a vector was selected because it allowed for N-terminal and C-terminal tagging of the expressed proteins.

The pCL0007 (pET28a(*phaE*)) plasmid insert was obtained through high-fidelity PCR amplification of pCL0001 using primers 1829.1 and 1829.6a. The product was then purified and digested using NcoI (Thermo Scientific, FD0574) and XhoI (Thermo Scientific, FD0695). Because the *phaE* insert possessed an internal NcoI recognition site, the reaction was first digested to completion using XhoI (one hour), then one-quarter of the recommended amount of NcoI was added and digested for 15 min before being denatured at 80°C for 20 min. This digest was then subjected to gel purification (Thermo Scientific, GeneJET K0831). The pET28a vector was also digested and gel purified with the same enzymes. The vector and insert were ligated and transformed into *E. coli* XL1-Blue. The insert presence and orientation were confirmed through DNA sequencing using T7promoter and T7terminator universal primers. The

exact plasmid isolates used for sequencing were then used to transform *E. coli* BLR(DE3), again using the one-step transformation protocol (Chung & Miller, 1993).

The pCL0008 (pET28a(*phaC*)) plasmids were derived in an analogous fashion to the pCL0007 vector. The primers 1830.1 and 1830.6a were used for PCR amplification of the respective pCL0002 vector and the BamHI and XhoI restriction enzymes were used for digestion of both the vector and insert.

8.3.6 Hexahistidine-Tagged Co-Expression Vector Construction

As an alternate method of expression, vectors for the co-expression of the PhaC and PhaE subunits were designed using the pETDuet-1 vector (Novagen, 71146). The *phaC* gene, and respective mutants, were inserted into the vector first. The pCL0010 vectors (pETDuet-1(*phaC*)) were constructed by digesting the respective pCL0008 vectors with XbaI (Thermo Scientific, FD0684) and SalI followed by the digestion of pETDuet-1 with BamHI, XbaI, and XhoI. Next, the 1.3 kb fragment of each pCL0008 digestion was gel purified while the 50 bp MCS regions of the pETDuet-1 vector were excluded via column purification. The ligation utilized the XbaI sites and exploited the XhoI/SalI compatible sticky ends.

Next, the insertion of the PhaE subunit was performed in order to create the pCL0011 (pETDuet-1(*phaC*, *phaE*)) vectors. As before, PCR amplification of pCL0001 was performed using the primers 1829.7 and 1829.2 which produced the insert, this product was then digested with BglII (ThermoScientific, FD0083) and SalI. The pCL0010 vectors were digested with BglII and XhoI. Once again, the ligation exploited the XhoI/SalI compatible sticky ends.

8.3.7 Expression and Purification of PHA Synthase Subunits

For the protein expression, an overnight culture of selective LB (100 $\mu g\ ml^{-1}$ ampicillin or 50 $\mu g\ ml^{-1}$ kanamycin, final) was seeded 1:100 in similarly selective TB medium. Cultures were continued at 37°C with rotation at 150 rpm until an optical density of 0.8 was observed at 600 nm wavelength. At this point, a 0.5 ml sample was extracted for expression analysis, pelleted, and stored at −20°C.

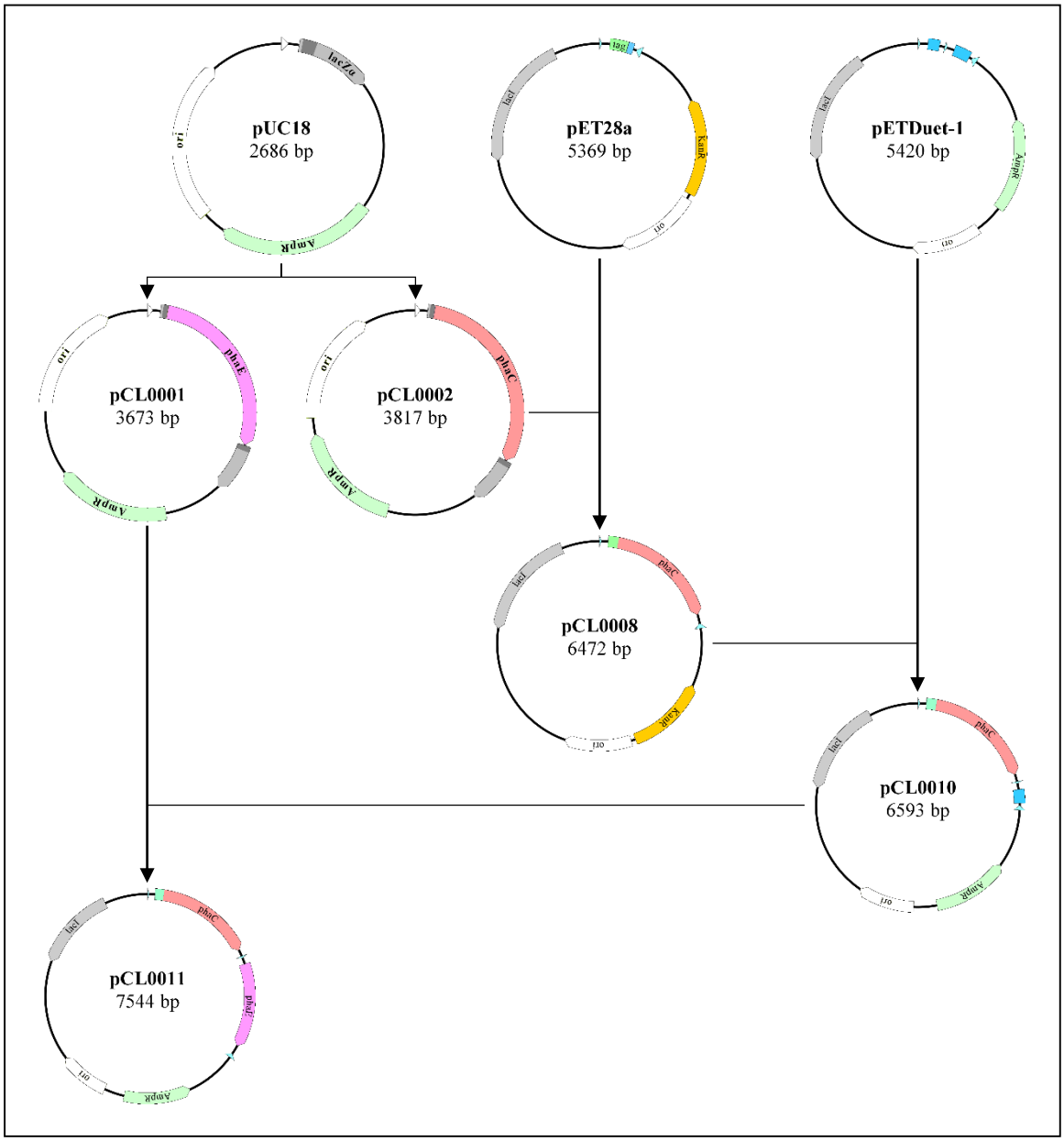


Figure 54: Flow diagram depicting construction of co-expression vectors. This process was repeated for every mutant derivative. Blue boxes represent multiple cloning sites. Promoters and terminators are depicted, but not labelled.

Next, the cultures were induced to a final concentration of 0.1 mM IPTG and allowed to express overnight. Once expression was complete, a sample was diluted to an optical density of 0.8, 0.5 ml of it was pelleted, and stored. Cells were then harvested via centrifugation into 50 ml centrifuge tubes at 3,000xg for 10 minutes at four centigrade. Volumes greater than 50 ml were harvested sequentially into the same tube for a given cultivar.

Recombinant protein expression capabilities of the cell lines were analyzed through whole-cell/total protein visualization using denaturing gel electrophoresis with coomassie staining. Stored sample pellets, obtained in the previous paragraph, were resuspended in 0.25 ml deionized water, aspirated via micropipette to aid osmolysis, and used directly as sample to mix with 4x laemmli buffer. 12 µl of sample (16 µl total) was loaded for SDS-PAGE visualization.

Hexahistidine-tagged protein purification and consultation was performed by Drs. Octavia Goodwin and Megan Macnaughtan from the Louisiana State University Department of Chemistry. Briefly, harvested cells were resuspended in BugBuster protein extraction reagent and subjected to ultrasonication. Lysates then underwent centrifugation at 30,000xg for 30 minutes at four centigrade, where the supernatants were deemed cleared lysates. Cleared lysates were loaded onto nickel-affinity gel media equilibrated to a pH value of 8.0 with 50 mM sodium phosphate supplemented with 300 mM sodium chloride, then set up for ambient gravity flow. Once complete, the affinity media was washed with buffer containing 10 mM imidazole and subsequently eluted with 300 mM imidazole. Purified products were visualized via typical SDS-PAGE.

8.3.8 Preparation of Clarified Lysate

The clarified lysate used in the enzymatic activity experiments was produced from 200 mg of wet cells from pellet reserves stored at −80°C. All of the following steps were performed on ice unless otherwise stated. The wet cells were resuspended in one milliliter of 130 mM potassium phosphate (pH 7.0) supplemented with 0.1% (v/v) TritonX-100 nonionic detergent. Cells were lysed via six pulses of sonication at 25% for 10 seconds with intermediate breaks of five seconds. Next, the lysates were

centrifuged at 21,000xg for 10 minutes at four centigrade. One milliliter of the supernatant was then transferred to a fresh 1.5 ml polypropylene tube. The total protein content was estimated using a bicinchonic acid colormetric assay (Pierce, 23225). Lysates were then standardized to five milligrams of total protein per milliliter.

8.3.9 Clarified Lysate Enzymatic Activity Assays

To observe early-phase phenomena of the PHA synthase polymerization reaction, the consumption of 3-hydroxybutyryl-CoA (HBCoA) was monitored spectroscopically at a wavelength of 232 nm during the first 15 minutes (Fukui et al., 1976). Clarified lysates were obtained within a 24 hour period. The HBCoA solution consisted of 130 mM potassium phosphate (pH 5.8) with 50% (v/v) glycerol and was stored at −20°C to reduce the amount of incidental hydrolysis.

At 30°C, each 200 µl early-phase reaction consisted of 179 µl of 130 mM potassium phosphate (pH 7.0) and 20 µl 10 mM HBCoA (Sigma-Aldrich, H0261) solution. The reaction was initiated by the addition of one microliter of the standardized lysate. Reactions were performed in triplicate.

Additionally, an end-point reaction was performed in order to test the case of a complete loss of functionality. These reactions were carried out simultaneously in ambient conditions for one hour. Each 100 µl end-point reaction consisted of 59 µl of 130 mM potassium phosphate (pH 7.0) and 40 µl 10 mM HBCoA solution. The reaction was initiated by the addition of one microliter of the standardized lysate. End-point reactions were performed in triplicate.

8.4 Results

8.4.1 Hypothetical Structural Analysis

The conserved 5 amino acid region (`GC[S|T][L|I]G`) shows some similarities to a characterized helical motif (`GxxLG`), suggesting that this region may play a role in protein secondary structure. Additionally, this region contains a conserved cysteine residue (similar to the active catalytic site) meaning its thiol group may affect substrate binding in some way, or may assist in disulfide bridges.

Because there is no information currently available on the protein tertiary structure of type III PhaC, the closest structural analogue is the lipase. Using the cited sequence accessions, a multiple sequence alignment similar to Jia *et. al.* was created to approximate the spatial location of the insertion (2000). An annotated rendering of the aligned lipase can be observed in Figure 55. It was found the insertion may fall within an alpha helix (α4) of the α/β-hydrolase domain neighboring the catalytic region. It is possible that this insertion could play a major role in the structure and/or function of cyanobacterial PHA synthases.

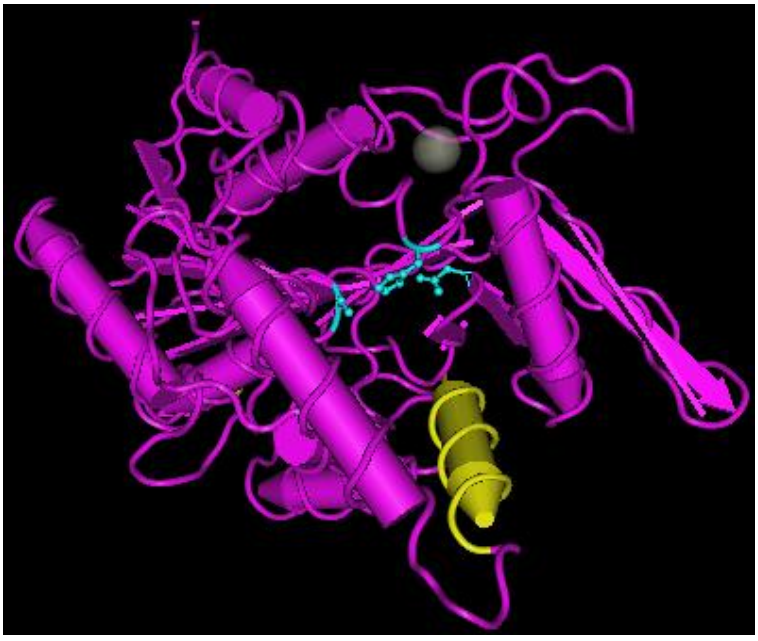


Figure 55: Three dimensional rendering of the *Pseudomonas* lipase crystal structure. The three catalytic sites are highlighted in cyan, while the alpha helix suspected to house the analogous cyanobacterial box is highlighted in yellow.

8.4.2 Isolation, Cloning, and Mutagenesis of *phaC* and *phaE*

The *Synechocystis* sp. PCC6803 genomic DNA isolation protocol yielded 40(±20) ng of nucleic acids per million cells. The high fidelity PCR amplification of *phaC* and *phaE* genes yielded approximately 30 ng μl^{-1} of a unique amplification product for each gene. The presence and orientation of *phaC* in pCL0002 was confirmed by the DNA sequencing results. Likewise, the analogous results were obtained for *phaE* in pCL0001.

The site-directed mutagenesis PCR amplifications of pCL0002 yielded 25(±5) ng μl^{-1} on average, and all reactions contained a single detectable linear amplification product. Once self-circularized, these mutant vectors were transformed into *E. coli* XL1-Blue with an average efficiency ranging from 1×10^6 - 5×10^6 CFU μg^{-1}.

The PHA synthase subunit expression vectors each required significant optimization for their construction. Once successfully constructed, the pCL0005 and pCL0006 vectors transformed into *E. coli* BLR(DE3) with an average efficiency of 700 CFU μg^{-1}. Similarly, the constructed pCL0007 and pCL0008 vectors transformed into *E. coli* BLR(DE3) with an average efficiency of 1000 CFU μg^{-1} and the pCL0011 vectors transformed with an average efficiency of 400 CFU μg^{-1}. It was found that the transformation efficiency of a similarly obtained and quantified pUC18 isolate into BLR(DE3) was $>3\times10^5$ CFU μg^{-1}, suggesting the low efficiency observed in the PHA synthase subunit expression system was likely not due to the procedure itself.

8.4.3 Dual ELP-tagged Split Intein Expression/Purification

Development and optimization of the purification process was performed using the GFP expression system in communication with Dr. David Wood and Dr. Steven Shi, authors from the original publication. In order to test this expression system for the production of PHA synthase, the wild-type subunits were used as a test case (pCL0005 & pCL0006). Both recombinant PHA synthase subunit genes successfully expressed when bound to the ELP-split-intein tag (EI0C-PhaC & EI0C-PhaE), as can be observed in lanes five and six of Figure 56. Unfortunately, detectable cleavage was not observed, even after extending the seven hour cleavage time to 13 hours. Figure 56 depicts the results of the 13 hour cleavage time. In Figure 56, negligible amounts of detectable protein mass occur at the expected cleaved EI0C molecular weight. Additionally, the majority of the EI0C-PhaC and EI0C-PhaE remain in the cleavage reaction insoluble phase after purification, which is further evidence of a highly inefficient cleavage (lanes 8 & 9, Figure 56). Upon comparison of lanes 5-to-8 and 6-to-9, it is important to note that the observed difference in band

mass is due to the 3-fold dilution of the EI0C-tagged product, as per the purification protocol, and not due to cleavage.

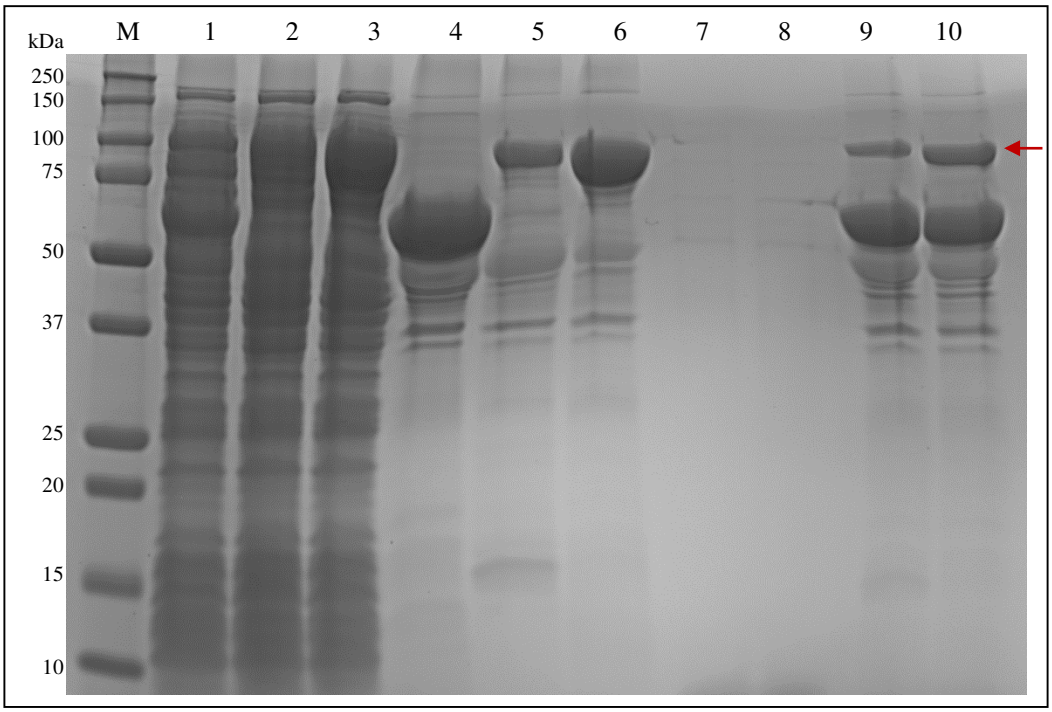


Figure 56: SDS-PAGE results of the *phaC* and *phaE* dual ELP-tagged split-intein expression systems with 13 hour cleavage time. The red arrow indicates the EI0C-PhaC/EI0C-PhaE expected migration point.
Lanes: (M) Protein standard,(1) EI0N clarified lysate, (2) EI0C-PhaC clarified lysate, (3) EI0C-PhaE clarified lysate, (4) EI0N purified, (5) EI0C-PhaC purified, (6) EI0C-PhaE purified, (7) PhaC cleavage soluble phase, (8) PhaE cleavage soluble phase, (9) PhaC cleavage insoluble phase, (10) PhaE cleavage insoluble phase.
Note: EI0C-PhaC and EI0C-PhaE purified protein samples are diluted 3-fold for cleavage.

Additionally, once it was determined that the interaction of these proteins was important for desirable expression, purification, and handling purposes, a co-cleavage reaction was attempted. In other words, EI0C-PhaC and EI0C-PhaE were both added to the same reaction and allowed to interact in hopes of promoting cleavage efficiency. Again, no detectable cleavage was observed for either protein.

8.4.4 Hexahistidine-Tagged Expression/Purification

Development and optimization of the His_6-tagged recombinant proteins was again performed using the wild-type *phaC* and *phaE* strains (pCL0008 & pCL0007 respectively). It was found that when compared to an empty-vector control (lane 1, Figure 57), the expression of the His_6-PhaC subunit was

unresolvable via SDS-PAGE visualization (lane 3, Figure 57). In contrast, the PhaE-His_6 protein overexpressed quite efficiently (lane 2, Figure 57).

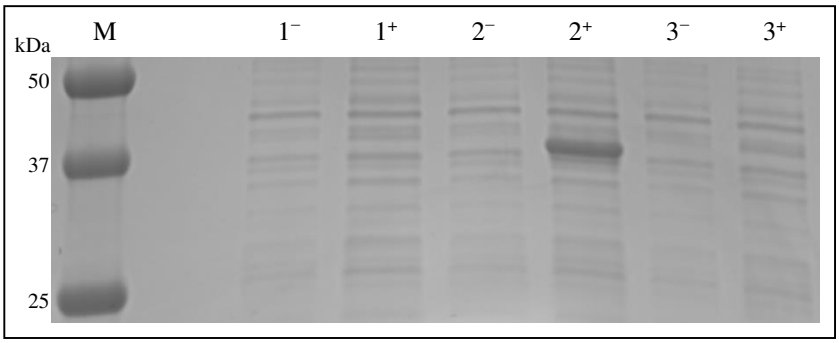


Figure 57: Total cellular protein SDS-PAGE profiles of the His_6-PhaC (3) and PhaE-His_6 (2) expression strains before (−) and after (+) addition of IPTG to induce recombinant protein expression.
Lanes: (M) Protein standard, (1) BLR(DE3, pET28a) control, (2) BLR(DE3, pCL0007) strain, (3) BLR(DE3, pCL0008) strain

In hopes that His_6-PhaC was present in low quantities, 250 ml of expression culture for each mutant strain was subjected to column purification. None of these purification products indicated successful overexpression of the His_6-PhaC subunit via SDS-PAGE visualization. The expression experiment was repeated using a Rosetta (DE3, pLysS) host strain, which possesses the genetic machinery required to efficiently utilize codons rarely observed in *E. coli*. It was found that the cause of the inefficient His_6-PhaC production was not due to difficulties related to the presence rare codons within the *phaC* coding sequence.

Interestingly, the purification of the PhaE-His_6 expression sample showed negligible yield as well. PhaE-His_6 remained insoluble during the lysis process and the bulk of which was found in the clarified lysate pellet during purification.

8.4.5 Hexahistidine-Tagged Co-Expression/Co-Purification

Because the previous purification attempt indicated that PhaE was insoluble, it was decided to attempt co-expression of the His_6-PhaC and untagged PhaE subunits, allowing for interaction within the cell during production. The goal was to attempt to exploit the PhaC-PhaE interaction to co-purify the proteins while keeping both subunits soluble.

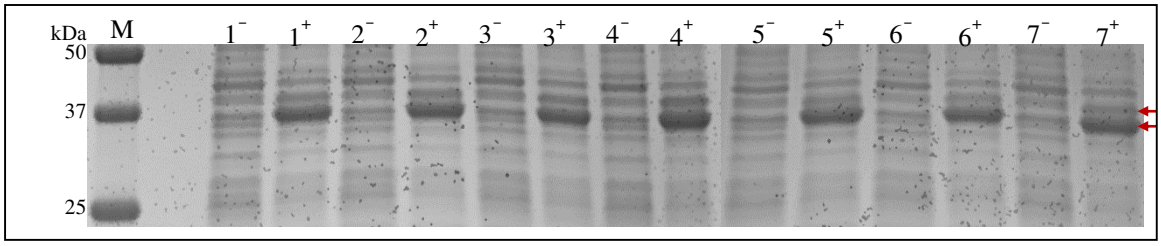


Figure 58: Total cellular protein SDS-PAGE profiles of the His_6-PhaC and PhaE co-expression strains before (−) and after (+) addition of IPTG to induce recombinant protein expression. Top and bottom arrows indicate perceived induced expression of PhaC and PhaE respectively.
Lanes: (M) Protein standard, (1) PhaC&PhaE, (2) PhaC(G205A)&PhaE, (3) PhaC(C206A)&PhaE, (4) PhaC(T207V)&PhaE, (5) PhaC(L208S)&PhaE, (6) PhaC(G209A)&PhaE, (7) PhaC(205-209Δ)&PhaE.

The results of the His_6-PhaC\PhaE co-expression can be observed in Figure 58. Comparing Figure 57 and Figure 58, the His_6-PhaC expression level showed an apparent increase during co-expression as compared to singular expression. At this point, the expression levels were believed to be sufficient for purification processing. The purification resulted in dilute recovery of the expected proteins as can be observed near the red arrows in Figure 59. Similar to Numata et al. (2015), a dilute 66 kDa band was also observed for our purification (black arrow, Figure 59).

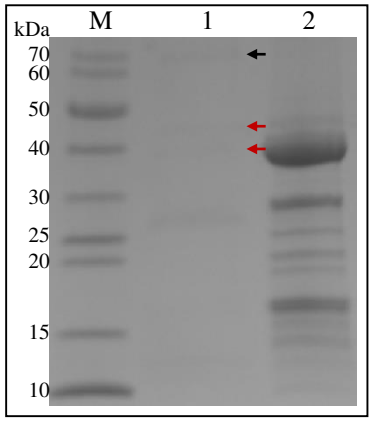


Figure 59: Purification products of co-expression strains. Red arrows indicate perceived His_6-PhaC and PhaE proteins. Black arrow indicates the 66 kDa band observed by Numata et al. (2015).
Lanes: (M) Protein standard, (1) purified sample, (2) lysate pellet.

8.4.6 Clarified Lysate Enzymatic Activity Assays

Because the proteins had proven too difficult to purify efficiently using the methods above, enzymatic activity experiments were carried out using clarified lysate. Lysates were expressed and isolated using a standardized protocol to ensure similar treatments to better control relative expression levels

between groups. The clarified lysates were all also normalized to total protein content. Short-time kinetics were observed by monitoring the consumption of HBCoA spectroscopically at intervals for times less than 15 minutes. The results from these experiments can be found in Figure 60.

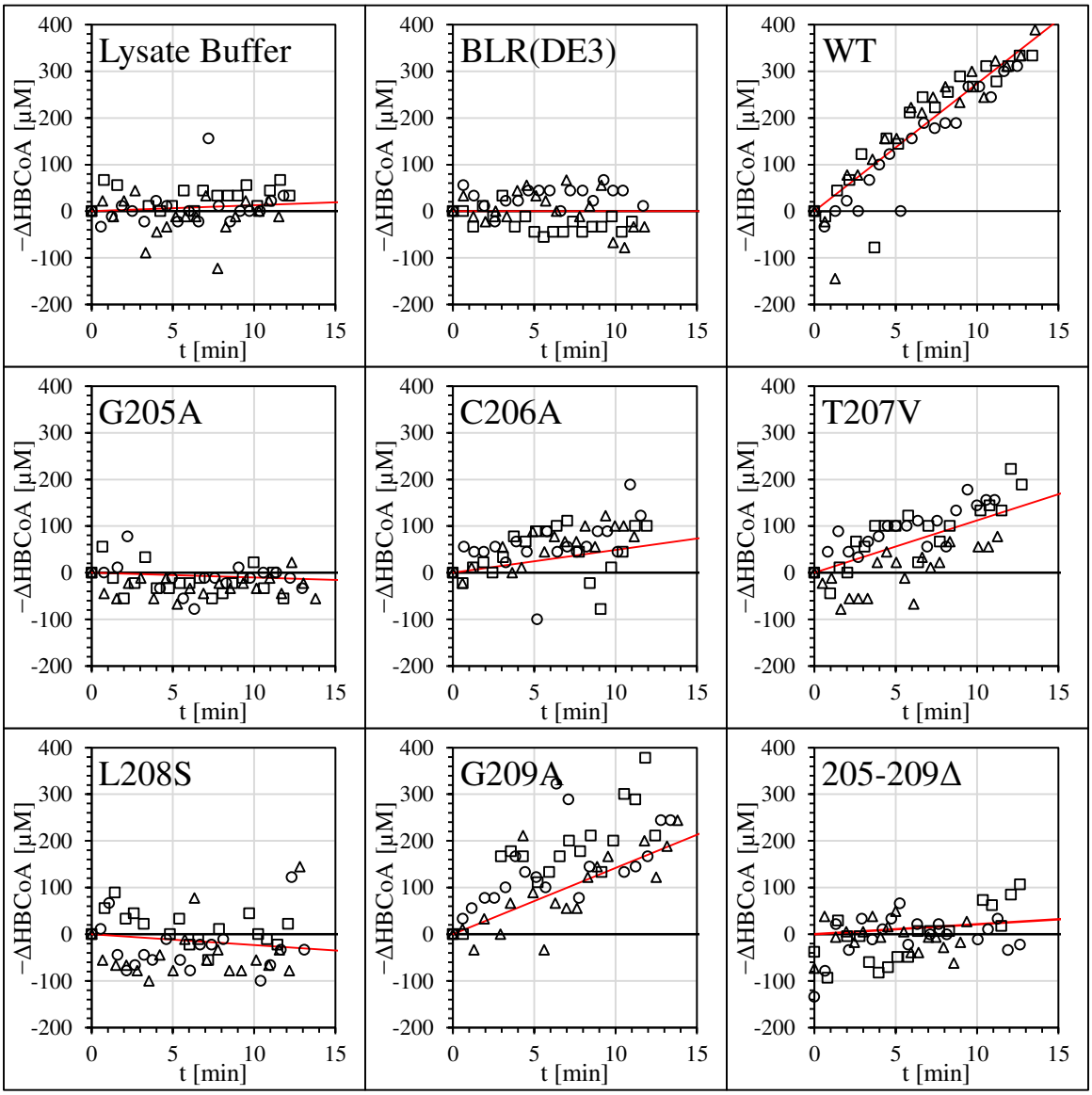


Figure 60: Short-time consumption of HBCoA from PHA synthase in clarified lysate. Each experiment performed in triplicate, independent samples are designated by a shape (square, circle, or triangle). Linear regression fit depicted as red line.

It was found that no significant HBCoA hydrolysis rate occurred automatically within the reaction ($p = 0.3680$), as evidenced by the Lysate Buffer graph in Figure 60. Furthermore, no observable consumption of substrate could be attributed to the *E. coli* endogenous proteins present in the clarified

lysate ($p = 0.5885$), as evidenced by the BLR(DE3) graph in Figure 60. In contrast, the wild-type PHA synthase lysate exhibited a significant HBCoA rate of consumption within the first 15 minutes ($p < 0.0001$). The C206A ($p = 0.0727$), T207V ($p < 0.0001$), and G209A ($p < 0.0001$) mutants exhibit significant rates of consumption; albeit, significantly less than the wild-type ($p < 0.0001$ ea.). The remainder of the mutants, G205A, L208S, and 205-209Δ, exhibited a lack of any significant observable consumption of HBCoA.

In order to determine the extent of the lost functionality for the mutants, an end-point reaction was performed. The reaction was allowed to proceed one hour at room temperature. The results from these end-point reactions can be observed in Figure 61.

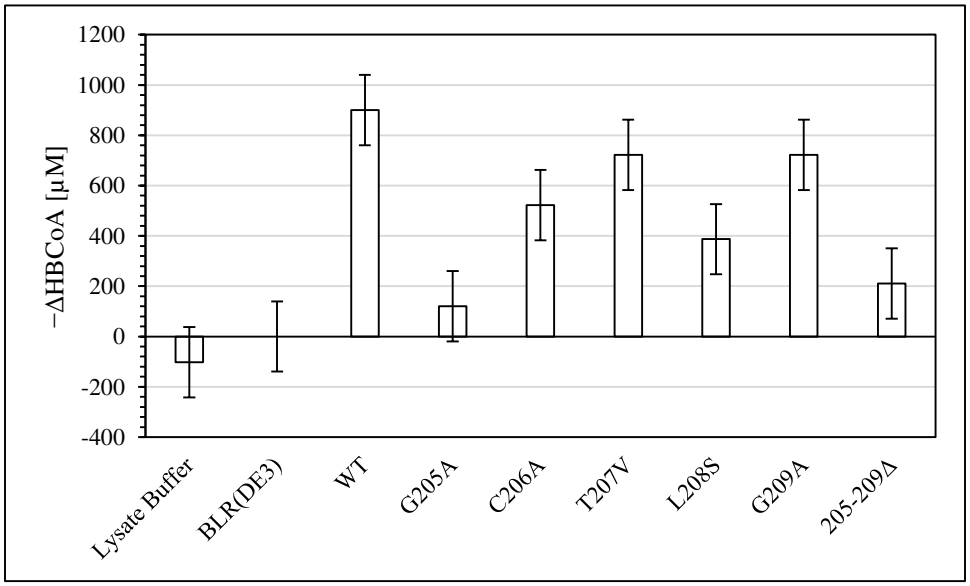


Figure 61: Consumption of HBCoA from PHA synthase in clarified lysate after one hour at ambient temperature. Each experiment performed in triplicate with duplicate technical repetitions.

Using the traditional null hypothesis, the end-point reactions showed that nearly all of the mutants retained some level of enzymatic activity, with the exception of the G205A mutant, which may require further data to resolve it from the null. The detectable enzymatic activity of the L208S mutant showed that 15 minutes was insufficient time to detect HBCoA consumption for these lesser-active mutants. In both cases where the G_{205} position is mutated (the G205A and the 205-209Δ mutants), the enzymatic

activity is severely hindered. In fact, these points are not significantly different from either negative control. To put it another way, the amount of substrate consumed by the 205-209Δ mutant is significantly detectable, but this amount is negligible within the error of the measurements. Similar to the short-time assay, the remainder of the mutants showed significantly lower rates of substrate consumption when compared to the wild-type PHA synthase.

8.5 Discussion

Because the clarified lysate preparation was strictly standardized, there are only two major assumptions from which the conclusions of this work are based. The first assumption is that the relative expression levels were comparable between mutants. In other words, the ability of the host cell to express the PhaC protein did not dramatically change in comparison to the wild-type with the alteration of a single codon. This assumption is supported qualitatively by the results of Figure 58, where the expression levels appear quite comparable across all strains. Additionally, the mutations introduced do not incorporate any codons typically identified as rare, or inefficient, in *E. coli* (AGA, AGG, ATA, CCC, CGG, CTA, and GGA). The second assumption is that the interaction between the contaminant *E. coli* proteins and the exogenous PHA synthase proteins was not significantly different throughout all of the cases. In other words, the alteration of a single amino acid in the PHA synthase made a mutant neither more nor less prone to interacting with the various contaminant proteins in comparison with the wild-type. With these assumptions in place, it is understood that each reaction received the same amount of PHA synthase and each was subjected to the same likelihood of interaction due to the contaminant proteins.

The results of the enzymatic activity assays provide evidence that the G_{205} residue is necessary for efficient enzymatic activity in the *Synechocystis* sp. PCC 6803 PHA synthase. Glycine residues are not commonly associated with protein-protein interaction due to the lack of a side chain (confer Figure 51). This suggests that the effect of this mutation is likely the disruption of a significant structural element within the PHA synthase. This is an interesting finding because the entire conserved loci is specific to

cyanobacteria and this specific residue shows identity-level conservation throughout all cyanobacterial PHA synthases (Appendix Figure 1). This suggests that the cyanobacterial PHA synthase possesses at least one structural motif critical to its proper function which does not occur within any other known PHA synthases.

CHAPTER 9. EXPANSION OF THESIS

9.1 Chapter Preface

The motive of this chapter is to expand upon each of the major thesis objectives listed within the introductory chapter. Direction, insight, and hypotheses are provided and assessed, which allows this work to be carried forward in the most efficient manner possible. The potential impact of each future work relative to its parent objective is graded under the author's opinion as a *minor*, *moderate*, or *major* level.

9.2 PCR-based detection of *phaC* in Cyanobacteria

Objective I provided a rapid and efficient method of detecting cyanobacteria which possess the gene for the active subunit of PHA synthase, *phaC*. The assay was developed on diverse cyanobacteria; however, only two true implementations of the assay were tested during development – the two unknown cases. Because the goal of this objective was to provide a tool which could be used to screen environmental isolates, it naturally follows that one should implement this assay for its intended purpose. As argued in CHAPTER 4, a strain of cyanobacteria with the native desired PHA accumulation abilities would be a novel subject for a photosynthetic PHA biosynthesis process. The testing of more cyanobacteria using the PCR-based assay alone possesses a *minor* impact level; however, implementing the assay and further assessment of the PHA production capabilities to potentially discover a high-yielding cyanobacterium receives a *major* impact level.

Because it was determined that PHA yields in cyanobacteria are limited upstream of the PHA synthase functionality, it would be beneficial to design PCR primers for the additional regulatory elements associated with PHA accumulation. This would allow for a more thorough and potentially more selective screening method. Such a screening process would likely present new insights into regulator/PHA correlation. The impact level for this project follows that of the previous paragraph.

Lastly, if a more robust primer set is desired for the detection of *phaC*, one could begin the development of an assay implementing the PCR primers developed at the end of CHAPTER 4. These

primers have greater degeneracy than the original set, leaning the balance of sensitivity and specificity more towards the former. This will be useful if the PhaC subunit shows less conservation than originally thought as more sequence information presents itself. Because the implementation of this primer set not definite, this project receives a *minor* potential impact level.

9.3 CEMAsuite

As with most computer programs, there will always be algorithm and interface optimization to be performed for this package. Additionally, an interesting new feature to include in the near future would be the ability to detect common restriction endonuclease sites between the PCR products within the alignment for a given primer set. The new feature tab could present the PCR product length for each accession within the sequence alignment, followed by the lengths for full and/or partial digestion products for a given restriction enzyme. This would allow for streamlined PCR and restriction enzyme confirmation of the desired PCR product and would be relatively simple to implement. Such a project would return a *minor* impact because the functionality added to the CEMAsuite package is not sufficient material for anything more than a short communication.

A second future work would be to separate the hybridization algorithms into its own executable for more modular implementation. This setup would allow for implementation of this algorithm within other similar software packages. Such a project would allow for easier construction of novel bioinformatic programs requiring DNA oligonucleotide hybridization, but by itself results in no apparent change of CEMAsuite to the user. Therefore, this project possesses a *minor* impact level.

Finally an ambitious goal would be to implement CEMAsuite as a software as a service (SaaS). This would broaden the user-base and make CEMAsuite truly platform-independent with very limited dependencies. Additionally, regulation of updates and error reporting would be simplified. This method of distribution could even allow for the development of a macro/scripting language which could allow users

to author their own custom scoring algorithms. This project is a large undertaking with high-risk, high-reward odds; thus, it possesses the potential for a *major* impact.

9.4 *In Vitro* PHA Synthase Polymerization

As discussed in CHAPTER 7, the *in vitro* PHA synthase polymerization reaction is highly dependent on the initial conditions. Furthermore, investigators are left with incomparable enzymatic kinetic data when researching across authors and publications. For this reason, a standardized *in vitro* PHA synthase kinetic activity assay should be developed. In theory, a standardized assay would remove all idiosyncrasies of the authors and present truly comparable results. The review of literature in CHAPTER 7 provides a starting place for the development of such an assay. The experiments would entail the optimization and testing PHA synthases from various model organisms. Recombinant expression strains can be obtained from the authors of many of the works in CHAPTER 8 and purification can be followed identical to the published works. This project would provide a standardized activity assay from which insightful model parameters can be inferred using the models derived in CHAPTER 7. This project would provide a *major* impact to the field of PHA synthase enzymatic activity because it would provide a comparative look at the various PHA synthase kinetics and present the methods for prospective investigators to follow suit.

In order to test the existence of the hypothetical coalescence-mediated PHA synthase shielding phenomena proposed in CHAPTER 7, one could perform a set of experiments which could be done alongside the experiments of the previous paragraph. This set of experiments would require an *in vitro* PHA synthase polymerization reaction in a late stage of polymerization. The granules could then be isolated via centrifugation, washed to remove soluble protein, and then subjected to non-specific enzymatic protein degradation. Precaution should be taken to avoid disturbing the granule state, so steps like sonication should be avoided. In theory, this step should remove all surface-bound PHA synthase. Once the naked granules have been washed of the protease, the shielded PHA synthase (if any) can likely be freed using a nonionic detergent and mechanical mixing (e.g., sonication). Protein which has been freed can be quantitated through any number of means, such as SDS-PAGE, a secondary enzymatic reaction, or

spectrophotometry. The most convenient method may be to fuse a fluorescent label onto one of the PHA synthase subunits and quantitate it via epifluorescence. This project could potentially elucidate the cause of the significant decrease in the enzymatic activity of PHA synthase and provide meaningful insights into the design of *in vitro* enzymatic activity assays, therefore it is declared a *moderate* priority level.

The purification of the *Synechocystis* sp. PCC 6803 PHA synthase can be further optimized for recombinant expression in *Escherichia coli*. There are many difficulties to overcome, as evidenced by CHAPTER 8, and little potential reward. The enzyme has been characterized and proven to possess no lesser functionality than other PHA synthases (Numata et al., 2015). Further investigation of this enzyme will likely present *minor* impact on the field.

9.5 Continuous *In Vitro* PHA Production Process Development

A continuous *in vitro* PHA synthesis process possesses many benefits, including reduced purification costs, a lack of biological limitations (e.g.: toxic or metabolic), and increased process and quality control. A novel hypothetical high-efficiency, high-production, and continuous *in vitro* PHA synthesis process is outlined using information gathered throughout the entirety of this author's studies. Development and assessment of a successful lab-scale process of this kind could potentially have a *major* impact on the PHA production industry.

The ELP-tagged split-intein protein expression and purification system is a novel vector system for continuous recombinant protein production (Shi et al., 2013). This process is considered readily adaptable to a continuous process and would prove quite useful for the purposes of enzyme production. Because the PHA synthase associated genes of *Cupriavidus necator* possess a single subunit PHA synthase which has been well studied, it would be the metabolic system of choice for this process. An additional benefit to the *C. necator* system is that it expresses efficiently within *E. coli* systems. The *phaA*, *phaB*, and *phaC* genes of *C. necator* should each be independently incorporated into the EI0C split-intein vector. The single-step purification process described in the related publication should be sufficient for this process.

The hydroxyacyl-CoA monomer will be a major contributor to the production cost and its synthesis should be rigorously optimized. Recombinant *Saccharomyces cerevisiae* has proven capable of yielding cytosolic acetyl-CoA at an unprecedented rate and efficiency when the pyruvate dehydrogenase from *Enterococcus faecalis*, which bypasses the native adenosine-triphosphate-dependent pathway, has been incorporated (Kozak et al., 2014). This should provide for an excellent source of acetyl-CoA.

Spent CoA can be recycled via the acetyl-CoA synthetase, or Acs (Jossek & Steinbüchel, 1998). In theory, CoA is generated in a 1:1 ratio (CoA:AACoA) with respect to the acetoacetyl-CoA product (Reactor A, Figure 62). CoA is also released in a 1:1 ratio as a hydroxyacyl-CoA monomer is incorporated into the PHA polymer (Reactor C, Figure 62). In order to synthesize acetyl-CoA from CoA, acetate must be supplied; however, this is a simple hydrocarbon and can come from any number of relatively inexpensive sources. The source of acetate is left ambiguous to allow for future cost optimization. This should dramatically reduce the input rate of acetyl-CoA production required from the *S. cerevisiae* source.

For separations it may be ideal to take advantage of the dual-phase system and allow the granules to separate passively. A settling tank would be sufficient for this purpose. A multi-chamber settling tank would allow for further design optimization. The smaller PHA granules could be recirculated as a potential source of PHA synthase nucleation and/or coalescence sites. This would also allow for some level of size exclusion for product control, leaving less potential process disturbances for the downstream PHA purification. Additionally, because PHA granule size is not limited by a cell envelope in *in vitro* synthesis, a direct recycle from the polymerization reactor could also be implemented to influence a larger granule size.

Lifecycle and economical assessments of this process will yield important information on the potential applicability of it in industry. These assessments should be evaluated throughout the design process.

This design exploits characteristics of *in vitro* PHA synthesis inferred during CHAPTER 7 and utilizes recombinant protein expression systems used in CHAPTER 8 to produce a potentially high-efficiency PHA synthesis process. This process was designed using the skillset unique to a chemical engineer. The development and optimization of this process could potentially be a major step towards an environmentally-friendly bioplastic production process.

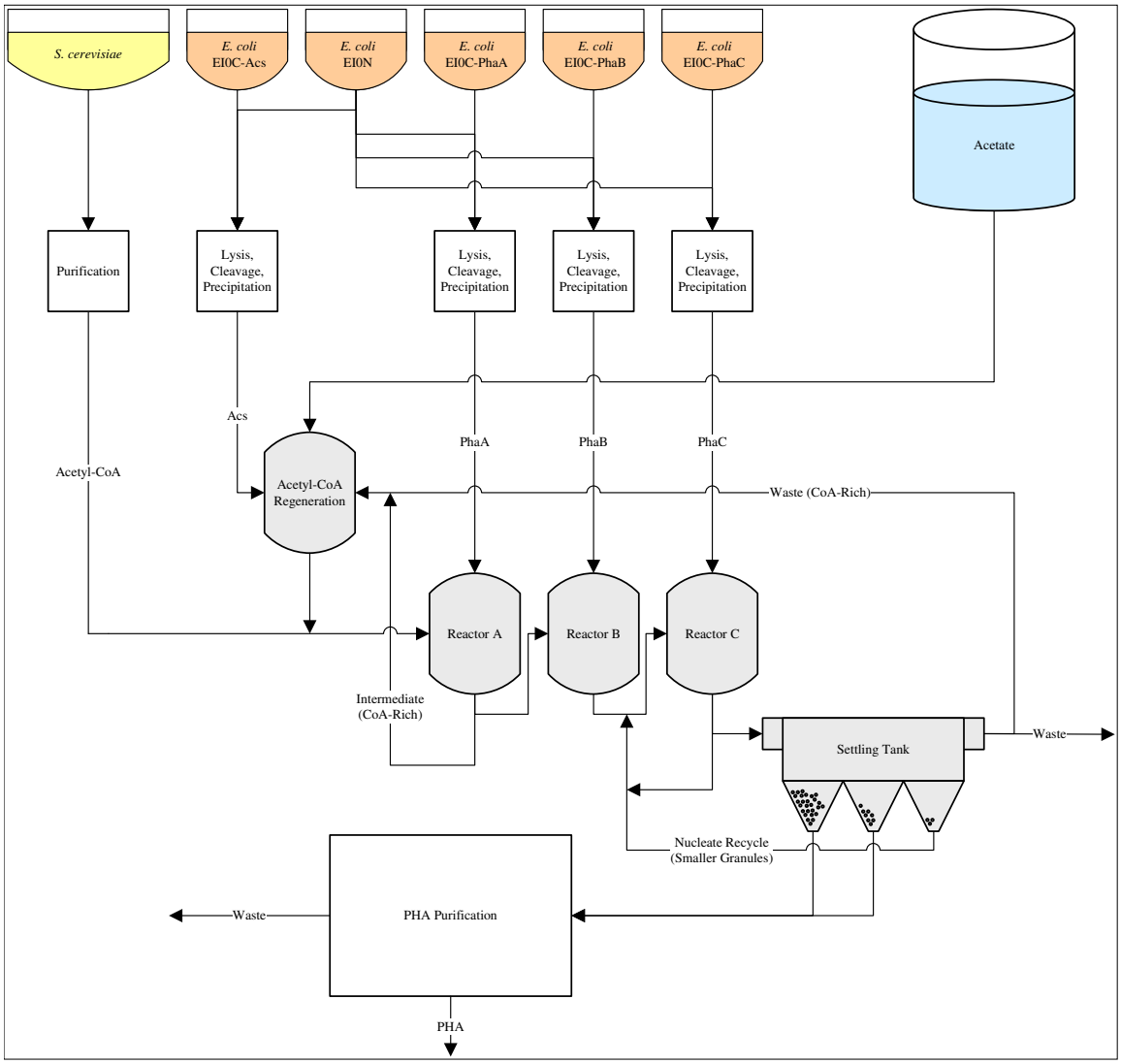


Figure 62: Prototypical continuous *in vitro* PHA synthesis process.

9.6 Conclusion

In conclusion, there are many ways this work can be continued. Each objective of this project has led to multiple continuations spanning a wide range of potential importance. The field of PHA production still requires attention now more than ever in order for these ecologically-friendly polymers to compete with their petroleum-based rivals. The ideas of this chapter are documented in hopes of further contributing to the goal of the sustainable and economically-viable bioplastic production process our ecosystem desperately needs. Thank you for your interest in this work.

REFERENCES

Abed, R. M. M., Dobretsov, S., & Sudesh, K. (2009). Applications of cyanobacteria in biotechnology. *Journal of Applied Microbiology, 106*(1), 1-12.

ACC. (2007). Plastic Packaging Resins (pp. Resin Codes).

ACC. (2013). Plastic Resins in the United States: American Chemistry Council: Economics & Statistics Department.

ACC. (2014). Impact of Plastics Packaging on Life Cycle Energy Consumption & Greenhouse Gas Emissions in the United States and Canada Substitution Analysis American Chemistry Council.

Agnew, D. E., Stevermer, A. K., Youngquist, J. T., & Pfleger, B. F. (2012). Engineering *Escherichia coli* for production of C12–C14 polyhydroxyalkanoate from glucose. *Metabolic Engineering, 14*(6), 705-713. doi: http://dx.doi.org/10.1016/j.ymben.2012.08.003.

Akiyama, H., Okuhata, H., Onizuka, T., Kanai, S., Hirano, M., Tanaka, S., . . . Miyasaka, H. (2011). Antibiotics-free stable polyhydroxyalkanoate (PHA) production from carbon dioxide by recombinant cyanobacteria. *Bioresource Technology, 102*(23), 11039-11042. doi: 10.1016/j.biortech.2011.09.058.

Allawi, H. T., & SantaLucia, J. (1997). Thermodynamics and NMR of Internal G·T Mismatches in DNA. *Biochemistry, 36*(34), 10581-10594. doi: 10.1021/bi962590c.

Allawi, H. T., & SantaLucia, J. (1998a). Nearest-Neighbor Thermodynamics of Internal A·C Mismatches in DNA: Sequence Dependence and pH Effects. *Biochemistry, 37*(26), 9435-9444. doi: 10.1021/bi9803729.

Allawi, H. T., & SantaLucia, J. (1998c). Nearest Neighbor Thermodynamic Parameters for Internal G·A Mismatches in DNA. *Biochemistry, 37*(8), 2170-2179. doi: 10.1021/bi9724873.

Allawi, H. T., & SantaLucia, J. (1998). Thermodynamics of internal C.T mismatches in DNA. *Nucleic Acids Research, 26*(11), 2694-2701.

Allcock, H., Lampe, F., & Mark, J. (2003). *Contemporary Polymer Chemistry* (3 ed.): Pearson Education, Inc.

Anderson, A. J., & Dawes, E. A. (1990). Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiology and Molecular Biology Reviews, 54*(4), 450-472.

Andrady, A. L. (1989). Environmental degradation of plastics under land and marine exposure conditions. *Proceedings of the Second International Conference on Marine Debris*.

Andrady, A. L., & Neal, M. A. (2009). Applications and societal benefits of plastics. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1526), 1977-1984. doi: 10.1098/rstb.2008.0304.

Asada, Y., Miyake, M., Miyake, J., Kurane, R., & Tokiwa, Y. (1999). Photosynthetic accumulation of poly-(hydroxybutyrate) by cyanobacteria—the metabolism and potential for CO_2 recycling. *International Journal of Biological Macromolecules, 25*(1–3), 37-42. doi: http://dx.doi.org/10.1016/S0141-8130(99)00013-6.

Atlić, A., Koller, M., Scherzer, D., Kutschera, C., Grillo-Fernandes, E., Horvat, P., . . . Braunegg, G. (2011). Continuous production of poly([R]-3-hydroxybutyrate) by *Cupriavidus necator* in a multistage bioreactor cascade. *Applied Microbiology and Biotechnology, 91*(2), 295-304. doi: 10.1007/s00253-011-3260-0.

Barnes, D. K. A. (2002). Biodiversity: Invasions by marine life on plastic debris. *Nature, 416*(6883), 808-809.

Bengtsson, S., Pisco, A. R., Johansson, P., Lemos, P. C., & Reis, M. A. M. (2010). Molecular weight and thermal properties of polyhydroxyalkanoates produced from fermented sugar molasses by open mixed cultures. *Journal of Biotechnology, 147*(3–4), 172-179. doi: http://dx.doi.org/10.1016/j.jbiotec.2010.03.022.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2009). GenBank. *Nucleic Acids Research, 37*(suppl 1), D26-D31. doi: 10.1093/nar/gkn723.

Bhaskar, T., Uddin, M. A., Murai, K., Kaneko, J., Hamano, K., Kusaba, T., . . . Sakata, Y. (2003). Comparison of thermal degradation products from real municipal waste plastic and model mixed plastics. *Journal of Analytical and Applied Pyrolysis, 70*(2), 579-587. doi: http://dx.doi.org/10.1016/S0165-2370(03)00027-5.

Bhati, R., Samantaray, S., Sharma, L., & Mallick, N. (2010). Poly-β-hydroxybutyrate accumulation in cyanobacteria under photoautotrophy. *Biotechnology Journal, 5*(11), 1181-1185. doi: 10.1002/biot.201000252.

Bhaya, D. (2004). Light matters: phototaxis and signal transduction in unicellular cyanobacteria. *Molecular Microbiology, 53*(3), 745-754. doi: 10.1111/j.1365-2958.2004.04160.x.

Bhubalan, K., Chuah, J.-A., Shozui, F., Brigham, C. J., Taguchi, S., Sinskey, A. J., . . . Sudesh, K. (2011). Characterization of the Highly Active Polyhydroxyalkanoate Synthase of *Chromobacterium* sp. Strain USM2. *Applied and Environmental Microbiology, 77*(9), 2926-2933. doi: 10.1128/aem.01997-10.

Bommarito, S., Peyret, N., & Jr, J. S. (2000). Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Research, 28*(9), 1929-1934.

Booth, C. S., Pienaar, E., Termaat, J. R., Whitney, S. E., Louw, T. M., & Viljoen, H. J. (2010). Efficiency of the Polymerase Chain Reaction. *Chemical Engineering Science, 65*(17), 4996-5006. doi: 10.1016/j.ces.2010.05.046.

Boutte, C., Grubisic, S., Balthasart, P., & Wilmotte, A. (2006). Testing of primers for the study of cyanobacterial molecular diversity by DGGE. *Journal of Microbiological Methods, 65*(3), 542-550. doi: 10.1016/j.mimet.2005.09.017.

Box, M. J. (1971). Bias in Nonlinear Estimation. *Journal of the Royal Statistical Society. Series B (Methodological), 33*(2), 171-201. doi: 10.2307/2985002.

Breuer, U., Terentiev, Y., Kunze, G., & Babel, W. (2002). Yeasts as Producers of Polyhydroxyalkanoates: Genetic Engineering of *Saccharomyces cerevisiae*. *Macromolecular Bioscience, 2*(8), 380-386. doi: 10.1002/1616-5195(200211)2:8<380::AID-MABI380>3.0.CO;2-X.

Buick, R. (2008). When did oxygenic photosynthesis evolve? *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1504), 2731-2743. doi: 10.1098/rstb.2008.0041.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference : a practical information-theoretic approach* (2nd ed.): Springer-Verlag New York, Inc.

Burns, K. L., Oldham, C. D., Thompson, J. R., Lubarsky, M., & May, S. W. (2007). Analysis of the in vitro biocatalytic production of poly-(β)-hydroxybutyric acid. *Enzyme and Microbial Technology, 41*(5), 591-599. doi: http://dx.doi.org/10.1016/j.enzmictec.2007.05.003.

Chen, G.-Q., & Wu, Q. (2005). The application of polyhydroxyalkanoates as tissue engineering materials. *Biomaterials, 26*(33), 6565-6578. doi: DOI: 10.1016/j.biomaterials.2005.04.036.

Choi, J., & Lee, S. Y. (1999). Factors affecting the economics of polyhydroxyalkanoate production by bacterial fermentation. *Applied Microbiology and Biotechnology, 51*(1), 13-21. doi: 10.1007/s002530051357.

Chung, C. T., & Miller, R. H. (1993). [43] Preparation and storage of competent *Escherichia coli* cells. In W. Ray (Ed.), *Methods in Enzymology* (Vol. Volume 218, pp. 621-627): Academic Press.

DaniMer scaling up production of bio-based PHA resins. (2011). *Plastics News.* 2014, from http://www.plasticsnews.com/article/20110705/NEWS/307059985/danimer-scaling-up-production-of-bio-based-pha-resins.

de Roda Husman, A.-M., Walboomers, J. M. M., van den Brule, A. J. C., Meijer, C. J. L. M., & Snijders, P. J. F. (1995). The use of general primers GP5 and GP6 elongated at their 3′ ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. *Journal of General Virology, 76*(4), 1057-1062. doi: 10.1099/0022-1317-76-4-1057.

Doi, Y., Kitamura, S., & Abe, H. (1995). Microbial Synthesis and Characterization of Poly(3-hydroxybutyrate-co-3-hydroxyhexanoate). *Macromolecules, 28*(14), 4822-4828. doi: 10.1021/ma00118a007.

Ducat, D. C., Way, J. C., & Silver, P. A. (2011). Engineering cyanobacteria to generate high-value products. *Trends in Biotechnology, 29*(2), 95-103.

Ebewele, R. O. (2000). *Polymer Science and Technology*. 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431: CRC Press LLC.

Erickson, H. P., Anderson, D. E., & Osawa, M. (2010). FtsZ in Bacterial Cytokinesis: Cytoskeleton and Force Generator All in One. *Microbiology and Molecular Biology Reviews, 74*(4), 504-528. doi: 10.1128/mmbr.00021-10.

Eriksen, N. T. (2008). Production of phycocyanin—a pigment with applications in biology, biotechnology, foods and medicine. *Applied Microbiology and Biotechnology, 80*(1), 1-14. doi: 10.1007/s00253-008-1542-y.

Filby, A., & Tyler, C. (2007). Appropriate 'housekeeping' genes for use in expression profiling the effects of environmental estrogens in fish. *BMC Molecular Biology, 8*(1), 10.

Flores, E., & Herrero, A. (2010). Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nature Reviews Microbiology, 8*(1), 39-50.

Fredslund, J., Schauser, L., Madsen, L. H., Sandal, N., & Stougaard, J. (2005). PriFi: using a multiple alignment of related sequences to find primers for amplification of homologs. *Nucleic Acids Research, 33*(suppl 2), W516-W520. doi: 10.1093/nar/gki425.

Fukui, T., Yoshimoto, A., Matsumoto, M., Hosokawa, S., Saito, T., Nishikawa, H., & Tomita, K. (1976). Enzymatic synthesis of poly-β-hydroxybutyrate inZoogloea ramigera. *Archives of Microbiology, 110*(2-3), 149-156. doi: 10.1007/BF00690222.

Gadberry, M. D., Malcomber, S. T., Doust, A. N., & Kellogg, E. A. (2005). Primaclade—a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics, 21*(7), 1263-1264. doi: 10.1093/bioinformatics/bti134.

Garcia-Pichel, F., Nübel, U., Muyzer, G., & Kühl, M. (1999). *On cyanobacterial community diversity and its quantification*. Paper presented at the Microbial Biosystems: New Frontiers Proceedings of the 8th International Symposium on Microbial Ecology Atlantic Canada Society for Microbial Ecology, Halifax, Canada.

Garg, R., Sahoo, A., Tyagi, A. K., & Jain, M. (2010). Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.). *Biochemical and Biophysical Research Communications, 396*(2), 283-288. doi: 10.1016/j.bbrc.2010.04.079.

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., . . . Bryant, S. H. (2010). The NCBI BioSystems database. *Nucleic Acids Res, 38*(Database issue), D492-496. doi: 10.1093/nar/gkp858.

Gerngross, T. U., & Martin, D. P. (1995). Enzyme-catalyzed synthesis of poly[(R)-(-)-3-hydroxybutyrate]: formation of macroscopic granules in vitro. *Proceedings of the National Academy of Sciences of the United States of America, 92*(14), 6279-6283.

Gerngross, T. U., Snell, K. D., Peoples, O. P., Sinskey, A. J., Csuhai, E., Masamune, S., & Stubbe, J. (1994). Overexpression and Purification of the Soluble Polyhydroxyalkanoate Synthase from Alcaligenes eutrophus: Evidence for a Required Posttranslational Modification for Catalytic Activity. *Biochemistry, 33*(31), 9311-9320. doi: 10.1021/bi00197a035.

Gill, R. L., Castaing, J.-P., Hsin, J., Tan, I. S., Wang, X., Huang, K. C., . . . Ramamurthi, K. S. (2015). Structural basis for the geometry-driven localization of a small protein. *Proceedings of the National Academy of Sciences, 112*(15), E1908-E1915. doi: 10.1073/pnas.1423868112.

Gonnet, G. H., Cohen, M. A., & Benner, S. A. (1992). Exhaustive matching of the entire protein sequence database. *Science, 256*(5062), 1443-1445.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Research, 38*(suppl 2), W695-W699.

Grothe, E., Moo-Young, M., & Chisti, Y. (1999). Fermentation optimization for the production of poly(β-hydroxybutyric acid) microbial thermoplastic. *Enzyme and Microbial Technology, 25*(1–2), 132-141. doi: http://dx.doi.org/10.1016/S0141-0229(99)00023-X.

Hahn, S., Zhong, X. Y., Troeger, C., Burgemeister, R., Gloning, K., & Holzgreve, W. (2000). Current applications of single-cell PCR. *Cellular and Molecular Life Sciences CMLS, 57*(1), 96-105. doi: 10.1007/s000180050501.

Hai, T., Hein, S., & Steinbuchel, A. (2001). Multiple evidence for widespread and general occurrence of type-III PHA synthases in cyanobacteria and molecular characterization of the PHA synthases from two thermophilic cyanobacteria: *Chlorogloeopsis fritschii* PCC 6912 and *Synechococcus* sp. strain MA19. *Microbiology, 147*(11), 3047-3060.

Harju, S., Fedosyuk, H., & Peterson, K. (2004). Rapid isolation of yeast genomic DNA: Bust n' Grab. *BMC Biotechnology, 4*(1), 8.

Harper, C. A., & Baker, A.-M. (2000). *Modern plastics handbook*: McGraw-Hill New York.

Hauf, W., Schlebusch, M., Hüge, J., Kopka, J., Hagemann, M., & Forchhammer, K. (2013). Metabolic Changes in *Synechocystis* PCC6803 upon Nitrogen-Starvation: Excess NADPH Sustains Polyhydroxybutyrate Accumulation. *Metabolites, 3*(1), 17.

Hein, S., Tran, H., & Steinbüchel, A. (1998). *Synechocystis* sp. PCC6803 possesses a two-component polyhydroxyalkanoic acid synthase similar to that of anoxygenic purple sulfur bacteria. *Archives of Microbiology, 170*(3), 162-170. doi: 10.1007/s002030050629.

Hoiczyk, E., & Hansel, A. (2000). Cyanobacterial Cell Walls: News from an Unusual Prokaryotic Envelope. *Journal of Bacteriology, 182*(5), 1191-1199.

Holmes, P. A. (1985). Applications of PHB - a microbially produced biodegradable thermoplastic. *Physics in Technology, 16*(1), 32.

Hooks, D., & Rehm, B. A. (2015). Insights into the surface topology of polyhydroxyalkanoate synthase: self-assembly of functionalized inclusions. *Applied Microbiology and Biotechnology*, 1-9. doi: 10.1007/s00253-015-6719-6.

Hughes, J., Lamparter, T., Mittmann, F., Hartmann, E., Gartner, W., Wilde, A., & Borner, T. (1997). A prokaryotic phytochrome. *Nature, 386*(6626), 663-663.

Ienczak, J., Schmidell, W., & Aragão, G. (2013). High-cell-density culture strategies for polyhydroxyalkanoate production: a review. *Journal of Industrial Microbiology & Biotechnology, 40*(3-4), 275-286. doi: 10.1007/s10295-013-1236-z.

Ishii, K., & Fukui, M. (2001). Optimization of Annealing Temperature To Reduce Bias Caused by a Primer Mismatch in Multitemplate PCR. *Applied and Environmental Microbiology, 67*(8), 3753-3755. doi: 10.1128/aem.67.8.3753-3755.2001.

Jain, M., Nijhawan, A., Tyagi, A. K., & Khurana, J. P. (2006). Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications, 345*(2), 646-651. doi: 10.1016/j.bbrc.2006.04.140.

Jau, M.-H., Yew, S.-P., Toh, P. S. Y., Chong, A. S. C., Chu, W.-L., Phang, S.-M., . . . Sudesh, K. (2005). Biosynthesis and mobilization of poly(3-hydroxybutyrate) [P(3HB)] by *Spirulina platensis*. *International Journal of Biological Macromolecules, 36*(3), 144-151. doi: DOI: 10.1016/j.ijbiomac.2005.05.002.

Jendrossek, D. (2007). Peculiarities of PHA granules preparation and PHA depolymerase activity determination. *Applied Microbiology and Biotechnology, 74*(6), 1186-1196. doi: 10.1007/s00253-007-0860-9.

Jendrossek, D., Schirmer, A., & Schlegel, H. G. (1996). Biodegradation of polyhydroxyalkanoic acids. *Applied Microbiology and Biotechnology, 46*(5-6), 451-463. doi: 10.1007/s002530050844.

Jia, Y., Kappock, T. J., Frick, T., Sinskey, A. J., & Stubbe, J. (2000). Lipases Provide a New Mechanistic Model for Polyhydroxybutyrate (PHB) Synthases: Characterization of the Functional Residues in *Chromatium vinosum* PHB Synthase†. *Biochemistry, 39*(14), 3927-3936. doi: 10.1021/bi9928086.

Jia, Y., Yuan, W., Wodzinska, J., Park, C., Sinskey, A. J., & Stubbe, J. (2000). Mechanistic Studies on Class I Polyhydroxybutyrate (PHB) Synthase from *Ralstonia eutropha*: Class I and III Synthases Share a Similar Catalytic Mechanism†. *Biochemistry, 40*(4), 1011-1019. doi: 10.1021/bi002219w.

Jossek, R., & Steinbüchel, A. (1998). In vitro synthesis of poly(3-hydroxybutyric acid) by using an enzymatic coenzyme A recycling system. *FEMS Microbiology Letters, 168*(2), 319-324. doi: 10.1111/j.1574-6968.1998.tb13290.x.

Kahar, P., Tsuge, T., Taguchi, K., & Doi, Y. (2004). High yield production of polyhydroxyalkanoates from soybean oil by *Ralstonia eutropha* and its recombinant strain. *Polymer Degradation and Stability, 83*(1), 79-86. doi: http://dx.doi.org/10.1016/S0141-3910(03)00227-1.

Kessler, B., & Witholt, B. (2001). Factors involved in the regulatory network of polyhydroxyalkanoate metabolism. *Journal of Biotechnology, 86*(2), 97-104. doi: Doi: 10.1016/s0168-1656(00)00404-1.

Kikkawa, Y., Narike, M., Hiraishi, T., Kanesato, M., Sudesh, K., Doi, Y., & Tsuge, T. (2005). Organization of Polyhydroxyalkanoate Synthase for In Vitro Polymerization as Revealed by Atomic Force Microscopy. *Macromolecular Bioscience, 5*(10), 929-935. doi: 10.1002/mabi.200500115.

Kozak, B. U., van Rossum, H. M., Luttik, M. A. H., Akeroyd, M., Benjamin, K. R., Wu, L., . . . van Maris, A. J. A. (2014). Engineering Acetyl Coenzyme A Supply: Functional Expression of a Bacterial Pyruvate Dehydrogenase Complex in the Cytosol of Saccharomyces cerevisiae. *mBio, 5*(5). doi: 10.1128/mBio.01696-14.

Kufryk, G. I., Sachet, M., Schmetterer, G., & Vermaas, W. F. J. (2002). Transformation of the cyanobacterium *Synechocystis* sp. PCC 6803 as a tool for genetic mapping: optimization of efficiency. *FEMS Microbiology Letters, 206*(2), 215-219. doi: 10.1111/j.1574-6968.2002.tb11012.x.

Lane, C. E., & Benton, M. G. (2015). Detection of the enzymatically-active polyhydroxyalkanoate synthase subunit gene, phaC, in cyanobacteria via colony PCR. *Molecular and Cellular Probes*. doi: http://dx.doi.org/10.1016/j.mcp.2015.07.001.

Lane, C. E., Gutierrez-Wing, M. T., Rusch, K. A., & Benton, M. G. (2012). Homogeneous detection of cyanobacterial DNA via polymerase chain reaction. *Letters in Applied Microbiology, 55*(5), 376-383. doi: 10.1111/j.1472-765X.2012.03304.x.

Lane, C. E., Hulgan, D., O'Quinn, K., & Benton, M. G. (2015). CEMAsuite: open source degenerate PCR primer design. *Bioinformatics*. doi: 10.1093/bioinformatics/btv420.

Lantz, P.-G., Stålhandske, F. I. T., Lundahl, K., & Rådström, P. (1999). Detection of yeast by PCR in sucrose solutions using a sample preparation method based on filtration. *World Journal of Microbiology and Biotechnology, 15*(3), 345-348. doi: 10.1023/a:1008923302360.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England), 23*(21), 2947-2948. doi: citeulike-article-id:1931121.

Laycock, B., Halley, P., Pratt, S., Werker, A., & Lant, P. (2014). The chemomechanical properties of microbial polyhydroxyalkanoates. *Progress in Polymer Science, 39*(2), 397-442. doi: http://dx.doi.org/10.1016/j.progpolymsci.2013.06.008.

Liebergesell, M., Rahalkar, S., & Steinbüchel, A. (2000). Analysis of the *Thiocapsa pfennigii* polyhydroxyalkanoate synthase: subcloning, molecular characterization and generation of hybrid synthases with the corresponding *Chromatium vinosum* enzyme. *Applied Microbiology and Biotechnology, 54*(2), 186-194. doi: 10.1007/s002530000375.

Lu, Q., Han, J., Zhou, L., Zhou, J., & Xiang, H. (2008). Genetic and Biochemical Characterization of the Poly(3-Hydroxybutyrate-co-3-Hydroxyvalerate) Synthase in *Haloferax mediterranei*. *Journal of Bacteriology, 190*(12), 4173-4180.

Lütke-Eversloh, T., Bergander, K., Luftmann, H., & Steinbüchel, A. (2001). Identification of a new class of biopolymer: bacterial synthesis of a sulfur-containing polymer with thioester linkages. *Microbiology, 147*(1), 11-19.

Lymperopoulou, D., Kormas, K., Moustaka-Gouni, M., & Karagouni, A. (2011). Diversity of cyanobacterial phylotypes in a Mediterranean drinking water reservoir (Marathonas, Greece). *Environmental Monitoring and Assessment, 173*(1), 155-165. doi: 10.1007/s10661-010-1378-7.

MacColl, R. (1998). Cyanobacterial Phycobilisomes. *Journal of Structural Biology, 124*(2–3), 311-334. doi: 10.1006/jsbi.1998.4062.

MacColl, R., Eisele, L. E., Dhar, M., Ecuyer, J.-P., Hopkins, S., Marrone, J., . . . Lewitus, A. J. (1999). Bilin Organization in Cryptomonad Biliproteins. *Biochemistry, 38*(13), 4097-4105. doi: 10.1021/bi982059c.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., . . . Bryant, S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research, 43*(D1), D222-D226. doi: 10.1093/nar/gku1221.

McDaniels, A. E., Rice, E. W., Reyes, A. L., Johnson, C. H., Haugland, R. A., & Stelma, G. N., Jr. (1996). Confirmational identification of *Escherichia coli*, a comparison of genotypic and phenotypic assays for glutamate decarboxylase and β-D- glucuronidase. *Applied and Environmental Microbiology, 62*(9), 3350-3354.

McGregor, G. B., & Rasmussen, J. P. (2008). Cyanobacterial composition of microbial mats from an Australian thermal spring: a polyphasic evaluation. *FEMS Microbiology Ecology, 63*(1), 23-35.

Mecking, S. (2004). Nature or Petrochemistry?—Biologically Degradable Materials. *Angewandte Chemie International Edition, 43*(9), 1078-1085. doi: 10.1002/anie.200301655.

Michal, J., Mitera, J., & Tardon, S. (1976). Toxicity of thermal degradation products of polyethylene and polypropylene. *Fire and Materials, 1*(4), 160-168. doi: 10.1002/fam.810010407.

Miller, K. (2014). Green Fence boon for China, group says. *Plastics News*. Retrieved from Plastics News website: http://www.plasticsnews.com/article/20141114/NEWS/141119952/green-fence-boon-for-china-group-says

Morin, N., Vallaeys, T., Hendrickx, L., Natalie, L., & Wilmotte, A. (2010). An efficient DNA isolation protocol for filamentous cyanobacteria of the genus *Arthrospira*. *Journal of Microbiological Methods, 80*(2), 148-154. doi: 10.1016/j.mimet.2009.11.012.

Müh, U., Sinskey, A. J., Kirby, D. P., Lane, W. S., & Stubbe, J. (1999). PHA Synthase from *Chromatium vinosum*: Cysteine 149 Is Involved in Covalent Catalysis†. *Biochemistry, 38*(2), 826-837. doi: 10.1021/bi9818319.

Neilan, B. A. (2002). The Molecular Evolution and DNA Profiling of Toxic Cyanobacteria. *Current Issues in Molecular Biology, 4*, 1-11.

Neilan, B. A., Jacobs, D., & Goodman, A. E. (1995). Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphisms within the phycocyanin locus. *Applied and Environmental Microbiology, 61*(11), 3875-3883.

Nicholas, K. B., & Nicholas, H. B. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments. *Distributed by the author*. doi: citeulike-article-id:3198041.

Nishioka, M., Nakai, K., Miyake, M., Asada, Y., & Taya, M. (2001). Production of poly-β-hydroxybutyrate by thermophilic cyanobacterium, *Synechococcus* sp. MA19, under phosphate-limited conditions. *Biotechnology Letters, 23*(14), 1095-1099. doi: 10.1023/A:1010551614648.

Nobes, G. A. R., Jurasek, L., Marchessault, R. H., Martin, D. P., Putaux, J. L., & Chanzy, H. (2000). Growth and kinetics of in vitro poly([R]-(–)-3-hydroxybutyrate) granules interpreted as particulate polymerization with coalescence. *Macromolecular Rapid Communications, 21*(2), 77-84. doi: 10.1002/(SICI)1521-3927(20000201)21:2<77::AID-MARC77>3.0.CO;2-6.

Nomura, C., & Taguchi, S. (2007). PHA synthase engineering toward superbiocatalysts for custom-made biopolymers. *Applied Microbiology and Biotechnology, 73*(5), 969-979. doi: 10.1007/s00253-006-0566-4.

Nübel, U., Garcia-Pichel, F., & Muyzer, G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology, 63*(8), 3327-3332.

Numata, K., Motoda, Y., Watanabe, S., Osanai, T., & Kigawa, T. (2015). Co-expression of Two Polyhydroxyalkanoate Synthase Subunits from Synechocystis sp. PCC 6803 by Cell-Free Synthesis and Their Specific Activity for Polymerization of 3-Hydroxybutyryl-Coenzyme A. *Biochemistry, 54*(6), 1401-1407. doi: 10.1021/bi501560b.

Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y., & Matsubara, K. (1985). An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *Journal of Biological Chemistry, 260*(5), 2605-2608.

Osanai, T., Numata, K., Oikawa, A., Kuwahara, A., Iijima, H., Doi, Y., . . . Hirai, M. Y. (2013). Increased Bioplastic Production with an RNA Polymerase Sigma Factor SigE during Nitrogen Starvation in Synechocystis sp. PCC 6803. *DNA Research: An International*

Journal for Rapid Publication of Reports on Genes and Genomes, 20(6), 525-535. doi: 10.1093/dnares/dst028.

Ostle, A. G., & Holt, J. G. (1982). Nile blue A as a fluorescent stain for poly-β-hydroxybutyrate. *Applied and Environmental Microbiology, 44*(1), 238-241.

Owczarzy, R., Moreira, B. G., You, Y., Behlke, M. A., & Walder, J. A. (2008). Predicting Stability of DNA Duplexes in Solutions Containing Magnesium and Monovalent Cations. *Biochemistry, 47*(19), 5336-5353. doi: 10.1021/bi702363u.

Peters, V., & Rehm, B. H. A. (2005). In vivo monitoring of PHA granule formation using GFP-labeled PHA synthases. *FEMS Microbiology Letters, 248*(1), 93-100.

Peyret, N., Seneviratne, P. A., Allawi, H. T., & SantaLucia, J. (1999). Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A·A, C·C, G·G, and T·T Mismatches. *Biochemistry, 38*(12), 3468-3477. doi: 10.1021/bi9825091.

Philip, S., Keshavarz, T., & Roy, I. (2007). Polyhydroxyalkanoates: biodegradable polymers with a range of applications. *Journal of Chemical Technology and Biotechnology, 82*(3), 233-247. doi: 10.1002/jctb.1667.

Price, G. D., Sültemeyer, D., Klughammer, B., Ludwig, M., & Badger, M. R. (1998). The functioning of the CO2 concentrating mechanism in several cyanobacterial strains: a review of general physiological characteristics, genes, proteins, and recent advances. *Canadian Journal of Botany, 76*(6), 973-1002. doi: 10.1139/b98-081.

Rai, R., Keshavarz, T., Roether, J. A., Boccaccini, A. R., & Roy, I. (2011). Medium chain length polyhydroxyalkanoates, promising new biomedical materials for the future. *Materials Science and Engineering: R: Reports, 72*(3), 29-47. doi: http://dx.doi.org/10.1016/j.mser.2010.11.002.

RamKumar Pandian, S., Deepak, V., Kalishwaralal, K., Rameshkumar, N., Jeyaraj, M., & Gurunathan, S. (2010). Optimization and fed-batch production of PHB utilizing dairy waste and sea water as nutrient sources by *Bacillus megaterium* SRKP-3. *Bioresource Technology, 101*(2), 705-711. doi: http://dx.doi.org/10.1016/j.biortech.2009.08.040.

Reddy, C. S. K., Ghai, R., Rashmi, & Kalia, V. C. (2003). Polyhydroxyalkanoates: an overview. *Bioresource Technology, 87*(2), 137-146. doi: Doi: 10.1016/s0960-8524(02)00212-2.

Rehm, B. H. A. (2003). Polyester synthases: natural catalysts for plastics. *Biochemical Journal, 376*(1), 15-33. doi: 10.1042/bj20031254.

Rehm, B. H. A. (2007). Biogenesis of microbial polyhydroxyalkanoate granules: a platform technology for the production of tailor-made bioparticles. *Current Issues in Molecular Biology, 9*(1), 41-62.

Rehm, B. H. A., Antonio, R. V., Spiekermann, P., Amara, A. A., & Steinbüchel, A. (2002). Molecular characterization of the poly(3-hydroxybutyrate) (PHB) synthase from Ralstonia eutropha: in vitro evolution, site-specific mutagenesis and development of a PHB synthase protein model. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1594*(1), 178-190. doi: http://dx.doi.org/10.1016/S0167-4838(01)00299-0.

Rehm, B. H. A., & Steinbüchel, A. (1999). Biochemical and genetic analysis of PHA synthases and other proteins required for PHA synthesis. *International Journal of Biological Macromolecules, 25*(1-3), 3-19. doi: Doi: 10.1016/s0141-8130(99)00010-0.

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., & Stainer, R. Y. (1979). Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Journal of General Microbiology, 111*, 1-61.

Rose, T. M., Henikoff, J. G., & Henikoff, S. (2003). CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res, 31*(13), 3763-3766.

Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology (Clifton, N.J.), 132*, 365-386. doi: citeulike-article-id:568051.

Rudi, K., Skulberg, O. M., Larsen, F., & Jakobsen, K. S. (1997). Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA sequences from the variable regions V6, V7, and V8. *Applied and Environmental Microbiology, 63*(7), 2593-2599.

Saker, M., Welker, M., & Vasconcelos, V. (2007). Multiplex PCR for the detection of toxigenic cyanobacteria in dietary supplements produced for human consumption. *Applied Microbiology and Biotechnology, 73*(5), 1136-1142. doi: 10.1007/s00253-006-0565-5.

Salehizadeh, H., & Van Loosdrecht, M. C. M. (2004). Production of polyhydroxyalkanoates by mixed culture: recent trends and biotechnological importance. *Biotechnology Advances, 22*(3), 261-279. doi: http://dx.doi.org/10.1016/j.biotechadv.2003.09.003.

Sambrook, J. (2001). *Molecular cloning : a laboratory manual 3rd ed.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Sambrook, J., & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual* (3 ed.). Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

SantaLucia, J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences, 95*(4), 1460-1465.

SantaLucia, J., & Hicks, D. (2004). THE THERMODYNAMICS OF DNA STRUCTURAL MOTIFS. *Annual Review of Biophysics and Biomolecular Structure, 33*(1), 415-440. doi: doi:10.1146/annurev.biophys.32.110601.141800.

SantaLucia J Jr, & N., P. (2001). Method and system for predicting nucleic acid hybridization thermodynamics and computer-readable storage medium for use therein. *World Intellectual Property Organization, WO 01/94611*, Appendix.

Sato, S., Ono, Y., Mochiyama, Y., Sivaniah, E., Kikkawa, Y., Sudesh, K., . . . Tsuge, T. (2008). Polyhydroxyalkanoate Film Formation and Synthase Activity During In Vitro and In Situ Polymerization on Hydrophobic Surfaces. *Biomacromolecules, 9*(10), 2811-2818. doi: 10.1021/bm800566s.

Sayers, E. W., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., . . . Ye, J. (2009). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research, 37*(suppl 1), D5-D15. doi: 10.1093/nar/gkn741.

Schirrmeister, B., Antonelli, A., & Bagheri, H. (2011). The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology, 11*(1), 45.

Schlebusch, M., & Forchhammer, K. (2010). Requirement of the Nitrogen Starvation-Induced Protein Sll0783 for Polyhydroxybutyrate Accumulation in *Synechocystis* sp. Strain PCC 6803. *Applied and Environmental Microbiology, 76*(18), 6101-6107. doi: 10.1128/aem.00484-10.

Sharma, L., Kumar Singh, A., Panda, B., & Mallick, N. (2007). Process optimization for poly-β-hydroxybutyrate production in a nitrogen fixing cyanobacterium, *Nostoc muscorum* using response surface methodology. *Bioresource Technology, 98*(5), 987-993. doi: DOI: 10.1016/j.biortech.2006.04.016.

Shi, C., Meng, Q., & Wood, D. (2013). A dual ELP-tagged split intein system for non-chromatographic recombinant protein purification. *Applied Microbiology and Biotechnology, 97*(2), 829-835. doi: 10.1007/s00253-012-4601-3.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology, 7*. doi: http://www.nature.com/msb/journal/v7/n1/suppinfo/msb201175_S1.html.

Simmons, T. L., Andrianasolo, E., McPhail, K., Flatt, P., & Gerwick, W. H. (2005). Marine natural products as anticancer drugs. *Molecular Cancer Therapeutics, 4*(2), 333-342.

Singh, A., & Mallick, N. (2009). Exploitation of inexpensive substrates for production of a novel SCL–LCL-PHA co-polymer by *Pseudomonas aeruginosa* MTCC 7925. *Journal of Industrial Microbiology and Biotechnology, 36*(3), 347-354. doi: 10.1007/s10295-008-0503-x.

Singh, A. K., & Mallick, N. (2009). SCL-LCL-PHA copolymer production by a local isolate, *Pseudomonas aeruginosa* MTCC 7925. *Biotechnology Journal, 4*(5), 703-711. doi: 10.1002/biot.200800307.

Singh, B., & Sharma, N. (2008). Mechanistic implications of plastic degradation. *Polymer Degradation and Stability, 93*(3), 561-584. doi: http://dx.doi.org/10.1016/j.polymdegradstab.2007.11.008.

Snijders, P. J. F., van den Brule, A. J. C., Schrijnemakers, H. F. J., Snow, G., Meijer, C. J. L. M., & Walboomers, J. M. M. (1990). The Use of General Primers in the Polymerase Chain Reaction Permits the Detection of a Broad Spectrum of Human Papillomavirus Genotypes. *Journal of General Virology, 71*(1), 173-181. doi: 10.1099/0022-1317-71-1-173.

Södergård, A., & Stolt, M. (2002). Properties of lactic acid based polymers and their correlation with composition. *Progress in Polymer Science, 27*(6), 1123-1163. doi: http://dx.doi.org/10.1016/S0079-6700(02)00012-6.

Stal, L. J., & Moezelaar, R. (1997). *Fermentation in cyanobacteria* (Vol. 21).

Steinbüchel, A. (2005). Non-biodegradable biopolymers from renewable resources: perspectives and impacts. *Current Opinion in Biotechnology, 16*(6), 607-613. doi: http://dx.doi.org/10.1016/j.copbio.2005.10.011.

Steinbüchel, A., & Hein, S. (2001). Biochemical and Molecular Basis of Microbial Synthesis of Polyhydroxyalkanoates in Microorganisms. In W. Babel & A. Steinbüchel (Eds.), *Biopolyesters* (Vol. 71, pp. 81-123): Springer Berlin Heidelberg.

Steinbüchel, A., & Lütke-Eversloh, T. (2003). Metabolic engineering and pathway construction for biotechnological production of relevant polyhydroxyalkanoates in microorganisms. *Biochemical Engineering Journal, 16*(2), 81-96. doi: http://dx.doi.org/10.1016/S1369-703X(03)00036-6.

Stubbe, J., & Tian, J. (2003). Polyhydroxyalkanoate (PHA) homeostasis: the role of the PHA synthase. *Natural Product Reports, 20*(5), 445-457.

Sudesh, K., Abe, H., & Doi, Y. (2000). Synthesis, structure and properties of polyhydroxyalkanoates: biological polyesters. *Progress in Polymer Science, 25*(10), 1503-1555. doi: Doi: 10.1016/s0079-6700(00)00035-6.

Sudesh, K., Taguchi, K., & Doi, Y. (2002). Effect of increased PHA synthase activity on polyhydroxyalkanoates biosynthesis in *Synechocystis* sp. PCC6803. *International Journal of Biological Macromolecules, 30*(2), 97-104. doi: http://dx.doi.org/10.1016/S0141-8130(02)00010-7.

Suriyamongkol, P., Weselake, R., Narine, S., Moloney, M., & Shah, S. (2007). Biotechnological approaches for the production of polyhydroxyalkanoates in microorganisms and plants -- A review. *Biotechnology Advances, 25*(2), 148-175. doi: DOI: 10.1016/j.biotechadv.2006.11.007.

Takase, K., Matsumoto, K. i., Taguchi, S., & Doi, Y. (2004). Alteration of Substrate Chain-Length Specificity of Type II Synthase for Polyhydroxyalkanoate Biosynthesis by in Vitro Evolution: in Vivo and in Vitro Enzyme Assays. *Biomacromolecules, 5*(2), 480-485. doi: 10.1021/bm034323+.

Tan, G.-Y. A., Chen, C.-L., Ge, L., Li, L., Wang, L., Zhao, L., . . . Wang, J.-Y. (2014). Enhanced gas chromatography-mass spectrometry method for bacterial polyhydroxyalkanoates analysis. *Journal of Bioscience and Bioengineering, 117*(3), 379-382. doi: http://dx.doi.org/10.1016/j.jbiosc.2013.08.020.

Tan, L. T. (2007). Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry, 68*(7), 954-979. doi: http://dx.doi.org/10.1016/j.phytochem.2007.01.012.

Tian, J., Sinskey, A. J., & Stubbe, J. (2005). Kinetic Studies of Polyhydroxybutyrate Granule Formation in *Wautersia eutropha* H16 by Transmission Electron Microscopy. *Journal of Bacteriology, 187*(11), 3814-3824. doi: 10.1128/jb.187.11.3814-3824.2005.

Tripathi, L., Wu, L.-P., Meng, D., Chen, J., & Chen, G.-Q. (2013). Biosynthesis and Characterization of Diblock Copolymer of P(3-Hydroxypropionate)-block-P(4-hydroxybutyrate) from Recombinant *Escherichia coli*. *Biomacromolecules, 14*(3), 862-870. doi: 10.1021/bm3019517.

Ughy, B., & Ajlani, G. (2004). Phycobilisome rod mutants in *Synechocystis* sp. strain PCC6803. *Microbiology, 150*(12), 4147-4156. doi: 10.1099/mic.0.27498-0.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research, 40*(15), e115-e115. doi: 10.1093/nar/gks596.

US-BEA. (2014). *Real Value Added by Industry*. Retrieved from http://www.bea.gov/iTable/iTable.cfm?ReqID=51&step=1#reqid=51&step=51&isuri=1&5114=a&5102=10.

US-EIA. (2014a). *Frequently Asked Questions: How much oil is used to make plastic?* : Retrieved from http://www.eia.gov/tools/faqs/faq.cfm?id=34&t=6.

US-EIA. (2014c). *Petroleum & Other Liquids: U.S. Net Imports by Country*. Retrieved from http://www.eia.gov/dnav/pet/pet_move_impcus_d_nus_Z00_mbbl_a.htm.

US-EPA. (2014). *Municipal Solid Waste Generation, Recycling, and Disposal in the United States Tables and Figures for 2012*.

Ushimaru, K., Sangiambut, S., Thomson, N., Sivaniah, E., & Tsuge, T. (2013). New insights into activation and substrate recognition of polyhydroxyalkanoate synthase from Ralstonia eutropha. *Applied Microbiology and Biotechnology, 97*(3), 1175-1182. doi: 10.1007/s00253-012-4089-x.

Vaitomaa, J., Rantala, A., Halinen, K., Rouhiainen, L., Tallberg, P., Mokelke, L., & Sivonen, K. (2003). Quantitative Real-Time PCR for Determination of Microcystin Synthetase E Copy Numbers for *Microcystis* and *Anabaena* in Lakes. *Applied and Environmental Microbiology, 69*(12), 7289-7297. doi: 10.1128/aem.69.12.7289-7297.2003.

Vermaas, W. (1996). Molecular genetics of the cyanobacterium *Synechocystis* sp. PCC 6803: Principles and possible biotechnology applications. *Journal of Applied Phycology, 8*(4-5), 263-273. doi: 10.1007/BF02178569.

Webb , H. K., Arnott, J., Crawford, R. J., & Ivanova, E. P. (2013). Plastic Degradation and Its Environmental Implications with Special Reference to Poly(ethylene terephthalate). *Polymers, 5*(1), 1-18. doi: 10.3390/polym5010001.

Wodzinska, J., Snell, K. D., Rhomberg, A., Sinskey, A. J., Biemann, K., & Stubbe, J. (1996). Polyhydroxybutyrate Synthase: Evidence for Covalent Catalysis. *Journal of the American Chemical Society, 118*, 6319-6320.

Yamamoto, S., & Harayama, S. (1995). PCR amplification and direct sequencing of gyrB genes with universal primers and their application to the detection and taxonomic analysis of Pseudomonas putida strains. *Applied and Environmental Microbiology, 61*(3), 1104-1109.

Yang, T. H., Kim, T. W., Kang, H. O., Lee, S. H., Lee, E. J., Lim, S. C., . . . Lee, S. Y. (2010). Biosynthesis of polylactic acid and its copolymers using evolved propionate CoA transferase and PHA synthase. *Biotechnology and Bioengineering, 105*(1), 150-160.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics, 13*(1), 134.

Yeesang, C., & Cheirsilp, B. (2011). Effect of nitrogen, salt, and iron content in the growth medium and light intensity on lipid production by microalgae isolated from freshwater sources in Thailand. *Bioresource Technology, 102*(3), 3034-3040. doi: http://dx.doi.org/10.1016/j.biortech.2010.10.013.

Yuan, W., Jia, Y., Tian, J., Snell, K. D., Müh, U., Sinskey, A. J., . . . Stubbe, J. (2001). Class I and III Polyhydroxyalkanoate Synthases from Ralstonia eutropha and Allochromatium vinosum: Characterization and Substrate Specificity Studies. *Archives of Biochemistry and Biophysics, 394*(1), 87-98. doi: http://dx.doi.org/10.1006/abbi.2001.2522.

Zhang, B., R. Carlson, F. Srienc. (2006). Engineering the Monomer Composition of Polyhydroxyalkanoates Synthesized in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology, 72*(1), 536-543.

Zhang, C.-C., Laurent, S., Sakr, S., Peng, L., & Bédu, S. (2006). Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. *Molecular Microbiology, 59*(2), 367-375. doi: 10.1111/j.1365-2958.2005.04979.x.

Zhang, S., Kolvek, S., Goodwin, S., & Lenz, R. W. (2004). Poly(hydroxyalkanoic acid) Biosynthesis in Ectothiorhodospira shaposhnikovii: Characterization and Reactivity of a Type III PHA Synthase. *Biomacromolecules, 5*(1), 40-48. doi: 10.1021/bm034171i.

Zhang, S., Yasuo, T., Lenz, R. W., & Goodwin, S. (2000). Kinetic and Mechanistic Characterization of the Polyhydroxybutyrate Synthase from Ralstonia eutropha†. *Biomacromolecules, 1*(2), 244-251. doi: 10.1021/bm005513c.

Zhang, W., Shrestha, R., Buckley, R. M., Jewell, J., Bossmann, S. H., Stubbe, J., & Li, P. (2014). Mechanistic Insight with HBCH(2)CoA as a Probe to Polyhydroxybutyrate (PHB) Synthases. *ACS Chemical Biology, 9*(8), 1773-1779. doi: 10.1021/cb5002735.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., & van 't Riet, K. (1990). Modeling of the Bacterial Growth Curve. *Applied and Environmental Microbiology, 56*(6), 1875-1881.

APPENDICES

A.1 Polymerase Chain Reaction Variants

Appendix Table 1: List of PCR variations and their applications.

Variant	Description
Asymmetric PCR	Preferentially amplifies one strand of target sequence over the other by limiting one primer, method to amplify single stranded nucleic acids.
COLD-PCR	CO-amplification at Lower Denaturation temperature PCR. PCR which exploits decreased stability of dsDNA containing mismatches by denaturing at lower temperatures to avoid complementary dsDNA denaturing.
Colony PCR	Application of PCR directly on bacterial colonies, omits DNA isolation steps.
dPCR	Digital PCR. Real-Time PCR variant which divides a single reaction into many aliquots and analyzes each partition individually.
Hot-start PCR	Method controlling non-specific amplification by either omitting or inhibiting the polymerase until after the initial denaturation step.
Inverse PCR	Allows the amplification of sequences surrounding a known region of DNA by cleavage and circularization of genomic DNA before amplification with primers extending outward from the known sequence.
ISSR-PCR	InterSequence-Specific PCR. Opposite of VNTR-PCR, this method designs primers within regions of repeated segments to analyze the variations of the sequences between these repeat regions.
LATE-PCR	Linear-After-The-Exponential PCR. Asymmetric PCR variant with more stable limiting primer.
Ligation-mediated PCR	PCR with primers targeting small oligonucleotide 'linkers' which are ligated to the target DNA before amplification.
long PCR	Amplification of high molecular weight (20-50 kbp) nucleic acids.
MSP	Methylation-specific PCR. Utilization of PCR to identify patterns of DNA methylation.
Multiplex-PCR	Targets multiple sequences with multiple primer sets in a single reaction.
Nested PCR	Use of successive PCRs. First PCR to increase initial template concentration of the subsequent reaction, reduces non-specific binding of primers.
OE PCR	Overlap Extension PCR. Application of primers with overhanging sequences used for mutation insertion and splicing of nucleic acids.
PCA	Polymerase Cycling Assembly, Assembly PCR. Use of multiple PCRs, each with overlapping products to assemble a large overall product.
qPCR	Quantitative PCR. Real-Time PCR variant used to measure the initial concentration of nucleic acid template.
Real-time PCR	Allows the measurement of double stranded nucleic acid concentrations in the reaction in real-time.
RT-PCR	Reverse Transcription PCR. Amplification of dsDNA products from RNA template.

Variant	Description
Suicide PCR	Implements multiple PCR primers targeting the same target and only allows the use of a single forward/reverse combination once before it cannot be used again, thus reducing the likelihood of false-positive results from contaminating DNA.
TAIL-PCR	Thermal Asymmetric InterLaced PCR. Allows the amplification of unknown sequences flanking a known region through the use of a nested pair of primers with differing annealing temperatures and a degenerate primer to amplify in the other direction from the unknown sequence.
Touchdown PCR	Method controlling non-specific amplification by initially applying high annealing temperature during early product formation, then gradually decreasing annealing temperature to increase amplification efficiency.
VNTR-PCR	Variable Number of Tandem Repeat PCR. Analysis of PCR products in regions of short tandem repeats.

A.2 Derivation of DNA Melting Temperature

Two-state theory assumes DNA can only exist as either dsDNA or ssDNA, no intermediates at equilibrium.

Dissociation.

Reaction: (*r*) $D \leftrightarrow S_1 + S_2 \qquad K_d = \frac{[S_1][S_2]}{[D]}$

Material Balance: (*i*) $[D]_\emptyset - [D] = [S_1] - [S_1]_\emptyset = [S_2] - [S_2]_\emptyset$

Define dissociation fraction, $\theta_D = \frac{[D]_\emptyset - [D]}{[D]_\emptyset}$, and substitute (*i*) into K_d:

(*ii*) $K_d = \frac{[D]_\emptyset\left(\theta_D + \frac{[S_1]_\emptyset}{[D]_\emptyset}\right)\left(\theta_D + \frac{[S_2]_\emptyset}{[D]_\emptyset}\right)}{1-\theta_D}$

Definition 1: (*iii*) $\Delta G^\ominus = \Delta H^\ominus - T\Delta S^\ominus$

Definition 2: (*iv*) $\Delta G^\ominus = -RT\ln(K_d)$

Substitution of (*iii*) into (*iv*) and solving for temperature (T):

$$(v) \qquad T = \frac{\Delta H^{\ominus}}{\Delta S^{\ominus} - R\ln(K_d)}$$

Definition 3: $$(vi) \qquad \theta_D|_{T=T_m} = \frac{1}{2}$$

Find (*v*) at (*vi*):

$$\boxed{T_m = \frac{\Delta H^{\ominus}}{\Delta S^{\ominus} - R\ln\left(2[D]_{\emptyset}\left(\frac{1}{2} + \frac{[S_1]_{\emptyset}}{[D]_{\emptyset}}\right)\left(\frac{1}{2} + \frac{[S_2]_{\emptyset}}{[D]_{\emptyset}}\right)\right)}}$$

Note: for self-complementary dsDNA: $S_1 = S_2 = S \quad \rightarrow K_d = \frac{[S]^2}{[D]} = \frac{[D]_{\emptyset}\left(\frac{\theta_D}{2} + \frac{[S]_{\emptyset}}{[D]_{\emptyset}}\right)^2}{(1-\theta_D)}$

$$\boxed{T_m = \frac{\Delta H^{\ominus}}{\Delta S^{\ominus} - R\ln\left(2[D]_{\emptyset}\left(\frac{1}{4} + \frac{[S]_{\emptyset}}{[D]_{\emptyset}}\right)^2\right)}}$$

Association.

Assuming $[S_1]_{\emptyset} \geq [S_2]_{\emptyset}$ – S_2 is limiting.

Reaction: $$(r) \qquad S_1 + S_2 \leftrightarrow D \qquad K_a = \frac{[D]}{[S_1][S_2]}$$

Material Balance: $$(i) \qquad [D] - [D]_{\emptyset} = [S_1]_{\emptyset} - [S_1] = [S_2]_{\emptyset} - [S_2]$$

Define association fraction, $\theta_A = \frac{[D]-[D]_{\emptyset}}{[S_2]_{\emptyset}}$, and substitute (*i*) into K_a:

$$(ii) \quad K_a = \frac{\frac{[D]_{\emptyset}}{[S_2]_{\emptyset}} + \theta_A}{[S_2]_{\emptyset}\left(\frac{[S_1]_{\emptyset}}{[S_2]_{\emptyset}} - \theta_A\right)(1-\theta_A)}$$

$$T_m = \frac{\Delta H^{\ominus}}{\Delta S^{\ominus} - R\ln\left(\frac{\frac{2[D]_{\emptyset}}{[S_2]_{\emptyset}} + 1}{\left([S_1]_{\emptyset} - \frac{[S_2]_{\emptyset}}{2}\right)}\right)} \quad where\ [S_1]_{\emptyset} \geq [S_2]_{\emptyset}$$

Note: for self-complementary dsDNA: $S_1 = S_2 = S \rightarrow K_a = \frac{[D]}{[S]^2} = \frac{\frac{[D]_{\emptyset}}{[S]_{\emptyset}} + \frac{\theta_A}{2}}{[S]_{\emptyset}\left(1 - \frac{\theta_A}{4}\right)^2}$

$$T_m = \frac{\Delta H^{\ominus}}{\Delta S^{\ominus} - R\ln\left(\frac{64}{49[S]_{\emptyset}}\left(\frac{[D]_{\emptyset}}{[S]_{\emptyset}} + \frac{1}{4}\right)\right)} \ where\ \theta_A = \frac{2([D] - [D]_{\emptyset})}{[S]_{\emptyset}}$$

A.3 Accession Tables

Appendix Table 2: The primary protein structure (αPC) and genetic coding sequence (*cpcA*) of the cyanobacterial phycocyanin alpha subunit analyzed via multiple sequence alignments for design of PCR primers to act as a cyanobacteria genomic DNA quality assurance.

Taxonomy	αPC Accession	*cpcA* Accession
Acaryochloris marina MBIC11017	YP_001521631.1	gi|158340280:c148530-148042
Arthrospira maxima CS-328	ZP_03271568.1	gi|209522890:c153475-152987
Arthrospira platensis str. Paraca	ZP_06380686.1	gi|254349541:1530-2018
Crocosphaera watsonii WH 8501	ZP_00516609.1	gi|67923114:27350-27838
Cyanobium sp. PCC 7001	ZP_05045216.1	gi|254430111:1503852-1504340
Cyanothece sp. PCC 7424	YP_002375498.1	gi|218437013:c178991-178503
Cyanothece sp. PCC 7425	YP_002482426.1	gi|220905643:c1537801-1537313
Cyanothece sp. PCC 7822	YP_003886916.1	gi|307149945:c1808486-1807998
Cylindrospermopsis raciborskii CS-505	ZP_06308539.1	gi|282900552:43296-43784
Fischerella sp. JSC-11	ZP_08984589.1	gi|354565113:c350240-349752
Gloeobacter violaceus PCC 7421	NP_926164.1	gi|37519569:3425312-3425800
Lyngbya majuscula 3L	ZP_08428233.1	gi|332708240:c11382-10894
Microcoleus vaginatus FGP-2	ZP_08490947.1	gi|334116516:c429233-428745
Microcystis aeruginosa NIES-843	YP_001657460.1	gi|166362741:2210230-2210718
Nostoc azollae 0708	YP_003722228.1	gi|298489614:c3558921-3558433
Nostoc punctiforme PCC 73102	YP_001868554.1	gi|186680550:6544141-6544632
Raphidiopsis brookii D9	ZP_06304364.1	gi|282896246:102383-102871
Synechococcus elongatus PCC 6301	YP_171210.1	gi|56750010:c559913-559422
Synechococcus sp. PCC 7002	YP_001735446.1	gi|170076636:2301231-2301719
Synechocystis sp. PCC 6803	AAA91033.1	gb|U34930.1|SPU34930:856-1344
Thermosynechococcus elongatus BP-1	NP_682748.1	gi|22297544:2042263-2042751
Trichodesmium erythraeum IMS101	YP_724429.1	gi|113473942:c7709337-7708849

Appendix Table 3: 18S ribosomal RNA small subunit coding sequences obtained and used in the multiple sequence alignment used in the design of a PCR primer set intended for eukaryotic DNA detection within a sample to help prevent false positives from cryptophyta and rhodophyta plastids.

Taxonomy	18S rDNA Accession
Actinastrum hantzschii	FM205884.1
Archaeospora leptoticha	AB047306.1
Basidiobolus haptosporus	AF368504.1
Cercosporella virgaureae	GU214658.1
Chlorella sorokiniana	FM205860.1
Chlorella sp. CB 2008/73	HQ111435.1
Chlorella vulgaris strain CCAP 211/11F	AY591515.1
Coronastrum ellipsoideum strain UTEX LB1382	GQ507370.1
Cryptophyceae sp. CCMP2293	GQ375265.1
Diacanthos belenophorus	AY323837.1
Dictyosphaerium sp. CB 2008/108	GQ507371.1
Dothistroma pini strain CBS 116487	GU214532.1
Endogone lactiflua isolate AFTOL-ID 45	DQ536471.1
Endogone pisiformis strain DAOM 233144	NG_017181.1
Hemiselmis virescens	AJ007284.1
Hindakia fallax strain CCAP 222/30	GQ487224.1
Hindakia tetrachotoma strain CCAP 222/78	GQ487240.1
Komma caudata	U53122.1
Lobosphaeropsis lobophora	FM205833.1
Marvania coccoides	FR865696.1
Meliniomyces variabilis strain UAMH 8861	AY762619.1
Micractinium pusillum	FM205873.1
Mortierellaceae sp. LN07-7-4	EU688964.1
Mycosphaerella graminicola strain CBS 115943	GU214540.1
Neochloris aquatica	FR865697.1
Passalora fulva strain STE-U 3688	AY251109.2
Proteomonas sulcata	AJ007285.1
Pseudocercosporella sp. CPC 10050	GU214685.1
Ramichloridium cerophilum strain CBS 103.59	EU041798.2
Rhizophlyctis rosea strain JEL 318	NG_017175.1
Rhodomonas sp. M1480	AJ007286.1
Saccharomyces cerevisiae	Z75578.1
Taphrina alni	AJ495831.1
Zasmidium anthuriicola strain CBS 118742	GU214595.1
Zygomycete sp. AM-2008a isolate 105	EU428770.1

Appendix Table 4: Accession table for the 29 cyanobacteria type III PHA synthase PhaC subunit primary protein structures (PhaC) and respective coding sequences (*phaC*) incorporated in the multiple sequence alignment used in the design of a PCR primer set intended for categorizing cyanobacteria as potential PHA producers and non-producers.

Strain	PhaC Accession	*phaC* Accession
Arthrospira maxima CS-328	EDZ97226.1	ABYK01000001.1:311680..312774
Arthrospira platensis C1	ZP_17052183.1	NZ_CM001632.1:1565798..1566892
Arthrospira platensis NIES-39	BAI94014.1	AP011615.1:6284046..6285140
Arthrospira platensis str. *Paraca*	ZP_11274489.1	NZ_ACSK02000570.1:6062..7156
Arthrospira sp. PCC 8005	ZP_09784510.1	NZ_CAFN01000673.1:31405..32526
Chlorogloeopsis fritschii	AAL76316.1	AF371369.1:84..1184
Cyanothece sp. PCC 7424	YP_002375830.1	NC_011729.1:549481..550575
Cyanothece sp. PCC 7425	YP_002484732.1	NC_011884.1:4089283..4090383
Cyanothece sp. PCC 7822	YP_003886606.1	NC_014501.1:1444440..1445534
Gloeocapsa sp. PCC 73106	ZP_21050948.1	NZ_ALVY01000193.1:19970..21070
Gloeocapsa sp. PCC 7428	YP_007128633.1	NC_019745.1:3330613..3331650
Microcoleus vaginatus FGP-2	ZP_08493657.1	NZ_AFJC01000008.1:44091..45143
Microcystis aeruginosa DIANCHI905	ZP_21132094.1	NZ_AOCI01000120.1:7889..8983
Microcystis aeruginosa NIES-843	YP_001660017.1	NC_010296.1:4581257..4582351
Microcystis aeruginosa PCC 7806	CAO90143.1	AM778949.1:31605..32699
Microcystis aeruginosa PCC 7941	ZP_18828651.1	NZ_HE973143.1:188862..189956
Microcystis aeruginosa PCC 9432	ZP_18815344.1	NZ_HE972538.1:194502..195596
Microcystis aeruginosa PCC 9443	ZP_18828448.1	NZ_HE973089.1:234..1328
Microcystis aeruginosa PCC 9701	ZP_18848456.1	NZ_CAIQ01000501.1:1660..2754
Microcystis aeruginosa PCC 9717	ZP_18823283.1	NZ_HE972766.1:79055..80149
Microcystis aeruginosa PCC 9806	ZP_16392283.1	NZ_HE973252.1:48273..49367
Microcystis aeruginosa PCC 9807	ZP_18837223.1	NZ_HE973368.1:103427..104521
Microcystis aeruginosa PCC 9808	ZP_18838494.1	NZ_HE973582.1:217172..218266
Microcystis aeruginosa PCC 9809	ZP_18843815.1	NZ_HE973750.1:126579..127673
Microcystis aeruginosa TAIHU98	ZP_20934181.1	NZ_ANKQ01000002.1:1395147..1396241
Microcystis sp. T1-4	ZP_10230362.1	NZ_CAIP01000427.1:5011..6105
Pleurocapsa sp. PCC 7327	YP_007082717.1	NC_019689.1:4290686..4291846
Synechococcus sp. MA19	AAK38139.1	AY030295.1:1..1095
Synechocystis sp. PCC 6803	AGF51119.1	CP003265.1:932812..933948

Appendix Table 5: Accession table for 67 tubulin-like cell division protein sequences (FtsZ) and their respective coding sequences (*ftsZ*). Accessions were implemented in the design of primers using a CEMA consensus sequence and a simple nucleic acid MSA (FtsZ67). [20] – Denotes use in alignment of 20 sequences (FtsZ20), [5] – Denotes use in alignment of 5 sequences (FtsZ5).

Taxon	FtsZ Accession	*ftsZ* Accession
Acaryochloris marina MBIC11017	YP_001515164.1	NC_009925.1:782340..783467
Anabaena cylindrica PCC 7122	YP_007154641.1	NC_019771.1:134488..135774
Anabaena sp. 90	YP_006998153.1	NC_019427.1:4157260..4158567
Anabaena variabilis ATCC 29413 [20]	YP_322354.1	NC_007413.1:2284186..2285472
Arthrospira platensis NIES-39 [5]	YP_005071738.1	NC_016640.1:5511062..5512342
Calothrix sp. PCC 6303	YP_007138691.1	NC_019751.1:4514286..4515578
Calothrix sp. PCC 7507	YP_007064972.1	NC_019682.1:1815489..1816775
Chamaesiphon minutus PCC 6605	YP_007095100.1	NC_019697.1:300466..301704
Chroococcidiopsis thermalis PCC 7203	YP_007091931.1	NC_019695.1:2885696..2886961
Crinalium epipsammum PCC 9333	YP_007140600.1	NC_019753.1:119446..120702
Cyanobacterium aponinum PCC 10605	YP_007161808.1	NC_019776.1:1984252..1985514
Cyanobacterium stanieri PCC 7202	YP_007164315.1	NC_019778.1:775215..776471
Cyanobium gracile PCC 6307	YP_007045866.1	NC_019675.1:1361486..1362574
Cyanothece sp. ATCC 51142	YP_001802730.1	NC_010546.1:1312605..1313864
Cyanothece sp. PCC 7424	YP_002380244.1	NC_011729.1:5582031..5583287
Cyanothece sp. PCC 7425	YP_002485398.1	NC_011884.1:4819739..4821103
Cyanothece sp. PCC 7822	YP_003887567.1	NC_014501.1:2581819..2583075
Cyanothece sp. PCC 8801	YP_002374333.1	NC_011726.1:4465231..4466508
Cyanothece sp. PCC 8802	YP_003139935.1	NC_013161.1:4458669..4459946
Cylindrospermum stagnale PCC 7417	YP_007149035.1	NC_019757.1:4819737..4821026
Dactylococcopsis salina PCC 8305	YP_007171918.1	NC_019780.1:1937757..1938848
Escherichia coli [5, 20]	CAA38872.1	X55034.1:21835..22986
Geitlerinema sp. PCC 7407	YP_007111279.1	NC_019703.1:4579574..4580857
Gloeobacter violaceus PCC 7421 [20]	NP_923244.1	NC_005125.1:306122..307381
Gloeocapsa sp. PCC 7428	YP_007126104.1	NC_019745.1:378413..379684
Halothece sp. PCC 7418	YP_007168336.1	NC_019779.1:2144129..2145391
Leptolyngbya sp. PCC 7376 [5]	YP_007069850.1	NC_019683.1:674794..676014
Microcoleus sp. PCC 7113	YP_007120243.1	NC_019738.1:1089558..1090829
Microcystis aeruginosa NIES-843	YP_001657656.1	NC_010296.1:2391964..2393211
Nostoc azollae 0708	YP_003721117.1	NC_014248.1:1975022..1976311
Nostoc punctiforme PCC 73102	YP_001868096.1	NC_010628.1:5952853..5954169
Nostoc sp. PCC 7107	YP_007052223.1	NC_019676.1:5294895..5296289
Nostoc sp. PCC 7120 [5, 20]	NP_487898.1	NC_003272.1:4655902..4657188

Continued on next page...

Appendix Table 5 continued...

Taxon	FtsZ Accession	*ftsZ* Accession
Nostoc sp. PCC 7524	YP_007076972.1	NC_019684.1:4374268..4375557
Oscillatoria acuminata PCC 6304	YP_007085256.1	NC_019693.1:2029648..2030931
Pleurocapsa sp. PCC 7327	YP_007079753.1	NC_019689.1:816903..818153
Prochlorococcus marinus str. AS9601 [20]	YP_001009898.1	NC_008816.1:1291754..1292869
Prochlorococcus marinus str. MIT 9215	YP_001484737.1	NC_009840.1:1333132..1334241
Prochlorococcus marinus str. MIT 9301	YP_001091719.1	NC_009091.1:1264859..1265974
Prochlorococcus marinus str. MIT 9303 [20]	YP_001018004.1	NC_008820.1:1755151..1756314
Prochlorococcus marinus str. MIT 9312 [20]	YP_397901.1	NC_007577.1:1321567..1322682
Prochlorococcus marinus str. MIT 9313 [20]	NP_894152.1	NC_005071.1:365392..366555
Prochlorococcus marinus str. MIT 9515 [20]	YP_001011784.1	NC_008817.1:1307400..1308515
Prochlorococcus marinus str. NATL1A	YP_001015549.1	NC_008819.1:1415572..1416669
Prochlorococcus marinus str. NATL2A [20]	YP_292069.1	NC_007335.2:1381065..1382162
Pseudanabaena sp. PCC 7367	YP_007101792.1	NC_019701.1:1369244..1370530
Rivularia sp. PCC 7116	YP_007054089.1	NC_019678.1:1229735..1231057
Stanieria cyanosphaera PCC 7437	YP_007133498.1	NC_019748.1:3422182..3423435
Synechococcus elongatus PCC 6301 [20]	YP_172437.1	NC_006576.1:1870459..1871640
Synechococcus elongatus PCC 7942 [20]	YP_401395.1	NC_007604.1:2445387..2446568
Synechococcus sp. CC9311	YP_729948.1	NC_008319.1:684252..685349
Synechococcus sp. CC9605 [20]	YP_381169.1	NC_007516.1:815055..816164
Synechococcus sp. CC9902 [20]	YP_377546.1	NC_007513.1:1493985..1495130
Synechococcus sp. JA-2-3B'a(2-13) [20]	YP_478319.1	NC_007776.1:2203612..2204727
Synechococcus sp. JA-3-3Ab [20]	YP_473602.1	NC_007775.1:104219..105340
Synechococcus sp. PCC 6312	YP_007062673.1	NC_019680.1:3005181..3006272
Synechococcus sp. PCC 7002	YP_001733298.1	NC_010475.1:21115..22362
Synechococcus sp. PCC 7502	YP_007104976.1	NC_019702.1:715339..716562
Synechococcus sp. RCC307	YP_001226856.1	NC_009482.1:534057..535229
Synechococcus sp. WH 7803	YP_001225482.1	NC_009481.1:1612267..1613388
Synechococcus sp. WH 8102 [20]	NP_897737.1	NC_005070.1:1581620..1582765
Synechocystis sp. PCC 6803 [5, 20]	NP_440816.1	NC_000911.1:1013045..1014337
Synechocystis sp. PCC 6803 substr. GT-I	YP_005382684.1	NC_017038.1:1013147..1014439
Synechocystis sp. PCC 6803 substr. PCC-N	YP_005408560.1	NC_017052.1:1013288..1014580
Synechocystis sp. PCC 6803 substr. PCC-P	YP_005385853.1	NC_017039.1:1013300..1014592
Thermosynechococcus elongatus BP-1 [20]	NP_683172.1	NC_004113.1:2491962..2493218
Trichodesmium erythraeum IMS101 [20]	YP_723288.1	NC_008312.1:5788222..5789493

A.4 Derivation of Saturation Model Biologically Relevant Parameters

Saturation model: (*i*) $\ln\left(\frac{C(t)}{C(t=\lambda)}\right) = y(t) = A(1 - \exp[B - Ct])$

Subject to the conditions: (*ii*) $y(t = \lambda) = \ln(1) = 0$

(*iii*) $\mu_m = \left.\frac{dy}{dt}\right|_{t=\lambda}$

(*iv*) $y_\infty = \lim_{t\to\infty} y(t)$

Application of (*ii*) yields: (*v*) $B = C\lambda$

Application of (*iii*) yields: (*vi*) $AC = \mu_m$

Application of (*iv*) yields: (*vii*) $A = y_\infty$

Solving system of (*iv*), (*v*), & (*vi*): $\boxed{A = y_\infty}$ $\boxed{B = \frac{\mu_m \lambda}{y_\infty}}$ $\boxed{C = \frac{\mu_m}{y_\infty}}$

Substitution into (*i*) yields:

$$\boxed{\ln\left(\frac{C(t)}{C(t=\lambda)}\right) = y(t) = y_\infty\left(1 - \exp\left[\frac{\mu_m}{y_\infty}(\lambda - t)\right]\right)}$$

A.5 Standard Genetic Code

AAT N Asn	CAT H His	GAT D Asp	TAT Y Tyr
AAC N Asn	CAC H His	GAC D Asp	TAC Y Tyr
AAA K Lys	CAA Q Gln	GAA E Glu	TAA * Ter
AAG K Lys	CAG Q Gln	GAG E Glu	TAG * Ter
ACT T Thr	CCT P Pro	GCT A Ala	TCT S Ser
ACC T Thr	CCC P Pro	GCC A Ala	TCC S Ser
ACA T Thr	CCA P Pro	GCA A Ala	TCA S Ser
ACG T Thr	CCG P Pro	GCG A Ala	TCG S Ser
AGT S Ser	CGT R Arg	GGT G Gly	TGT C Cys
AGC S Ser	CGC R Arg	GGC G Gly	TGC C Cys
AGA R Arg	CGA R Arg	GGA G Gly	TGA * Ter
AGG R Arg	CGG R Arg	GGG G Gly	TGG W Trp
ATT I Ile	CTT L Leu	GTT V Val	TTT F Phe
ATC I Ile	CTC L Leu	GTC V Val	TTC F Phe
ATA I Ile	CTA L Leu	GTA V Val	TTA L Leu
ATG M Met	CTG L Leu	GTG V Val	TTG L Leu

X – initiation. (Sayers et al., 2009).

A.6 IUPAC Degenerate Nucleotide Nomenclature

A - Adenine	
C - Cytosine	
G - Guanine	
T - Thymine	
W - Weak	A,T
S - Strong	C,G
M - aMino	A,C
K - Keto	G,T
R - puRine	A,G
Y - pYrimidine	C,T
B - not A	C,G,T
D - not C	A,G,T
H - not G	A,C,T
V - not T	A,C,G
N - Any	A,C,G,T

A.7 PHA Synthase Protein MSA (29 sequences)

Appendix Figure 1: Cyanobacteria (Type III) PHA synthase multiple sequence alignment of sequences in Appendix Table 4. Analogous primer regions are depicted with ">>" and "<<". Under/over lines depict the "Cyanobacterial box" described by Hai *et al.* (2001). Catalytic-triad residues are highlighted (Rehm, 2007).

```
CLUSTAL O(1.1.0) multiple sequence alignment


gi|209496929|gb|EDZ97226.1|           ----------------------MLPFALQMGLEDLTQEYADLTEKIVHGMDNLSSLREEE
gi|423063393|ref|ZP_17052183.1|       ----------------------MLPFALQMGLEDLTQEYADLTEKIVHGMDNLSSLREEE
gi|291571742|dbj|BAI94014.1|          ----------------------MLPFALQMGLEDLTEEYADLTEKIVHGMDNLSSLREEE
gi|409991206|ref|ZP_11274489.1|       ----------------------MLPFALQMGLEDLTEEYADLTEKIVHGMDNLSSLREEE
gi|376007312|ref|ZP_09784510.1|       ------------MSLLGGHQAMLPFALQMGLEDLTQEYADLTEKIVHGMDNLSSLREEE
gi|18644660|gb|AAL76316.1|            ----------------------MLPFLLQIHLEEATHESAQLTHKLVKGMENLSQLREED
gi|218437501|ref|YP_002375830.1|      ----------------------MLPFLDQIRLEDAVHEYTEITKKMIKGLDNLSRLREED
gi|220909421|ref|YP_002484732.1|      ----------------------MLPFLLQIHLEEAAHESAQLTHKLVKGMENLSQLREED
gi|307151222|ref|YP_003886606.1|      ----------------------MLPFLDQIRLEDAVHEYTEITKKMLKGLDNLSRLREED
gi|443321910|ref|ZP_21050948.1|       ----------------------MLPFWTQISIEDTTCEYIELTKKLLKGIQNLRELRESD
gi|434393686|ref|YP_007128633.1|      ------------------------------------MLTDL-AELWQRGEVFSRLREED
gi|334119572|ref|ZP_08493657.1|       ---------------------------------MVTTELTQKLVKGVEIFTRLREED
gi|443658276|ref|ZP_21132094.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|166367744|ref|YP_001660017.1|      ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|159029277|emb|CAO90143.1|          ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|425448807|ref|ZP_18828651.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|425434880|ref|ZP_18815344.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|425448547|ref|ZP_18828448.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|425469530|ref|ZP_18848456.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|425443051|ref|ZP_18823283.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|422304946|ref|ZP_16392283.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|425457520|ref|ZP_18837223.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|425459008|ref|ZP_18838494.1|       ----------------------MWSFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|425464502|ref|ZP_18843815.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|440754979|ref|ZP_20934181.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|390442356|ref|ZP_10230362.1|       ----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEED
gi|428204128|ref|YP_007082717.1|      MHFDSPTRKIHLAQLDKEETKDMLPFLTQIRLEDVSHEYTELTKKVLQGIENLSRLREED
gi|18642978|gb|AAK38139.1|            ----------------------MLPFLMQMRLDDATEEYTELIKKIVKGIENLSRLREED
gi|451780150|gb|AGF51119.1|           ---------MFLLFFIVHWLKIMLPFFAQVGLEENLHETLDFTEKFLSGLENLQGLNEDD
                                                                                ::  :     : :  :.*.:


gi|209496929|gb|EDZ97226.1|           IIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVPLLIVYALVNRPFMVDLQEGRSLVAN
gi|423063393|ref|ZP_17052183.1|       IIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVPLLIVYALVNRPFMVDLQEGRSLVAN
gi|291571742|dbj|BAI94014.1|          IIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVPLLIVYALVNRPFMVDLQEGRSLVAN
gi|409991206|ref|ZP_11274489.1|       IIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVPLLIVYALVNRPFMVDLQEGRSLVAN
gi|376007312|ref|ZP_09784510.1|       IIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVPLLIVYALVNRPFMVDLQEGRSLVAN
gi|18644660|gb|AAL76316.1|            IEVGATPREVVFQEDKVKLYRFKSPVDQKKTVKTPILMVYALVNRPFMVDLQEDRSLVAN
gi|218437501|ref|YP_002375830.1|      IESGVSPKEVVYQEDKVVLYRFKSQVEH--PLPIPLLMVYALVNRPFMVDLQEGRSLVAN
gi|220909421|ref|YP_002484732.1|      IEVGSTPREVVYQEDKVKLYRFKAPANQGKTVQTPILMVYALVNRPFMVDLQEDRSLVAN
gi|307151222|ref|YP_003886606.1|      IQSGVSAKEAVYKEDKVILYRFTPQVAQ--PLHIPLLMVYALVNRPFMVDLQEGRSLVAN
gi|443321910|ref|ZP_21050948.1|       IEIGITPKEVIYQEDKMLLYRFKPMVEN--PLTIPLLIVYALVNRPFMVDLQENRSLVAN
gi|434393686|ref|YP_007128633.1|      NLIGVTPKEEIYREDKVVLYRFTPQVKN--LLNTPILIVYALVNRPYIVDLQAKRSLVAN
gi|334119572|ref|ZP_08493657.1|       IQVGVTPKEEVYREDKVLLYHFSPKVEH--SLNIPILIVYALVNRPYIVDLQEGRSLVAN
gi|443658276|ref|ZP_21132094.1|       IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|166367744|ref|YP_001660017.1|      IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|159029277|emb|CAO90143.1|          IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|425448807|ref|ZP_18828651.1|       IECGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|425434880|ref|ZP_18815344.1|       IECGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|425448547|ref|ZP_18828448.1|       IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|425469530|ref|ZP_18848456.1|       IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|425443051|ref|ZP_18823283.1|       IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|422304946|ref|ZP_16392283.1|       IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|425457520|ref|ZP_18837223.1|       IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|425459008|ref|ZP_18838494.1|       IECGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|425464502|ref|ZP_18843815.1|       IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|440754979|ref|ZP_20934181.1|       IECGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|390442356|ref|ZP_10230362.1|       IQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIPLLMVYALVNRPYMVDLQEGRSLVAN
gi|428204128|ref|YP_007082717.1|      IEIGATPKEAVYKEDKVILYRFKPMVEQ--PLSIPLLIVYALVNRPYMVDLQEDRSLVAN
gi|18642978|gb|AAK38139.1|            IEIGVTPKEAVYREEKLTLYHFQSTVQK--QLRTPVLIVYALVNRPFMVDLQEDRSLVAN
gi|451780150|gb|AGF51119.1|           IQVGFTPKEAVYQEDKVILYRFQPVVEN--PLPIPVLIVYALVNRPYMVDLQEGRSLVAN
                                         * : :* :::*:*: **:*   . .   .  *:*:********::****  ******
```

```
                                        >>>>>>>>
gi|209496929|gb|EDZ97226.1|          LLSLGLDVYLIDWGYPTRSDRWLTLDDYINGYINNCVDFLRDHYELDKINLLGVCQGGTF
gi|423063393|ref|ZP_17052183.1|      LLSLGLDVYLIDWGYPTRSDRWLTLDDYINGYINNCVDFLRDHYELDKINLLGVCQGGTF
gi|291571742|dbj|BAI94014.1|         LLSLGLDVYLIDWGYPTRSDRWLTLDDYINGYINNCVDFLRDHYELDKINLLGVCQGGTF
gi|409991206|ref|ZP_11274489.1|      LLSLGLDVYLIDWGYPTRSDRWLTLDDYINGYINNCVDFLRDHYELDKINLLGVCQGGTF
gi|376007312|ref|ZP_09784510.1|      LLSLGLDVYLIDWGYPTRSDRWLTLDDYINGYINNCVDFLRDHYELDKINLLGVCQGGTF
gi|18644660|gb|AAL76316.1|           LLKLGLDIYLIDWGYPGRGDRWLTLDDYINGYLNNCVDFIRTSHQLDKVNLLGICQGGTF
gi|218437501|ref|YP_002375830.1|     LLKLGLDVYLIDWGYPTRADRWLTLDDYINGYINNCVDFIRKQHNLDKINLLGICQGGTF
gi|220909421|ref|YP_002484732.1|     LLKLGLDIYLIDWGYPGRGDRWLTLDDYINGYLNNCVDFIRASHQLDKVNLLGICQGGTF
gi|307151222|ref|YP_003886606.1|     LLKLGLDVYLIDWGYPTRADRWLTLDDYINGYIDNCVDYIRKTHNIDKVNLLGICQGGTF
gi|443321910|ref|ZP_21050948.1|      LLKLGLDIYLIDWGYPTRADRWMNLDDYINGYINNCVEVVRKRHGLEKINLLGICQGGAF
gi|434393686|ref|YP_007128633.1|     LLKLGVDVYLIDWGYPSRIDRWLTLDDYINGYINNCIDVVCDRHNLAQINLLGICQGGTF
gi|334119572|ref|ZP_08493657.1|      LLELGLDVYLIDWGYPSRGDRWLTLDDYINGYINNCVDVVRDRHNLEQINLLGICQGGTF
gi|443658276|ref|ZP_21132094.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|166367744|ref|YP_001660017.1|     LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|159029277|emb|CAO90143.1|         LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|425448807|ref|ZP_18828651.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|425434880|ref|ZP_18815344.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|425448547|ref|ZP_18828448.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|425469530|ref|ZP_18848456.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|425443051|ref|ZP_18823283.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|422304946|ref|ZP_16392283.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|425457520|ref|ZP_18837223.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|425459008|ref|ZP_18838494.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|425464502|ref|ZP_18843815.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|440754979|ref|ZP_20934181.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVDNCVDFIRQSHHLDKINLLGICQGGTF
gi|390442356|ref|ZP_10230362.1|      LLKLGLDIYLIDWGYPTRSDRWLTLDDYINGYVNNCVDFIRQSHHLDKINLLGICQGGTF
gi|428204128|ref|YP_007082717.1|     LLKLGLDVYLIDWGYPSRADRWLTLDDYINGYINNCVDFIREKHGLEKINLLGICQGGAF
gi|18642978|gb|AAK38139.1|           LLKLGLDIYLIDWGYPTRADRWLTLDDYINGYINNCVDFIRKKHDLDKINLLGICQGGTF
gi|451780150|gb|AGF51119.1|          LLKLGLDVYLIDWGYPSRGDRWLTLEDYLSGYLNNCVDIICQRSQQEKITLLGVCQGGTF
                                     **.**:*:******** * ***:.*:**:.**::**:: :       ::.***:****:*


gi|209496929|gb|EDZ97226.1|          SLCYSSLYPEKVQNLITMVAPVNFDMPNTLLNARGGCTLGPEAIDVDLMVEALGNIPGDY
gi|423063393|ref|ZP_17052183.1|      SLCYSSLYPEKVQNLITMVAPVNFDMPNTLLNARGGCTLGPEAIDVDLMVEALGNIPGDY
gi|291571742|dbj|BAI94014.1|         SLCYSSLYPEKVQNLITMVAPVNFDMPNTLLNARGGCTLGPEAVDIDLMVEALGNIPGDY
gi|409991206|ref|ZP_11274489.1|      SLCYSSLYPEKVQNLITMVAPVNFDMPNTLLNARGGCTLGPEAVDIDLMVEALGNIPGDY
gi|376007312|ref|ZP_09784510.1|      SLCYSSLYPEKVQNLITMVAPVNFDMPNTLLNARGGCTLGPEAIDVDLMVEALGNIPGDY
gi|18644660|gb|AAL76316.1|           SLCYSSLYPDKVNNLVVMVAPVDFHQPETLLNMRGGCTLGAEAIDVDLMVDALGNIPGDF
gi|218437501|ref|YP_002375830.1|     SVCYSAIYPEKVKNLIVMVAPIDFRMPGTLLNMRGGCTIGAEALDVDLMIDSMGNVPGDY
gi|220909421|ref|YP_002484732.1|     SLCYSSLYPDKVNNLVVMVAPVDFHQPETLLNMRGGCTLGAEAIDVDLMVDALGNIPGDF
gi|307151222|ref|YP_003886606.1|     SVCYSALHPEKVKNLIVMVAPIDFRMPGTLLNMRGGCTIGNEALDVDLMIEAMGNVPGDY
gi|443321910|ref|ZP_21050948.1|      SLCYSAIYPEKVKNLIVMVTPVDFHIPNAFLNIRGGCSLGKDALDVDLMVDALGNIPGDW
gi|434393686|ref|YP_007128633.1|     SLCYSALYPAKVKNLIVMVTPVDFHTQEGLLNVWSGCTLGAKALDVDLAIDTLGNVPGDW
gi|334119572|ref|ZP_08493657.1|      SLCYSSLYPEKVKNLITMVTPVDFHINEGLLNVWGGCTLGSKAVDIDLMVDTLGNIPGDF
gi|443658276|ref|ZP_21132094.1|      SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGAEALDIDLMVDTMGNIPGDF
gi|166367744|ref|YP_001660017.1|     SLCYSSLYPDKIKNLVTMVTPVDFYQTETLLNMRGGCSLGAEALDIDLMVDTMGNIPGDF
gi|159029277|emb|CAO90143.1|         SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGAEALDIDLMVDTMGNIPGDF
gi|425448807|ref|ZP_18828651.1|      SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGAEALDIDLMVDTMGNIPGDF
gi|425434880|ref|ZP_18815344.1|      SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGAEALDIDLMVDTMGNIPGDF
gi|425448547|ref|ZP_18828448.1|      SLCYSSLYPDKIKNLVTMVTPVDFYQTETLLNMRGGCSLGSEALDIDLMVDTMGNIPGDF
gi|425469530|ref|ZP_18848456.1|      SLCYSSLYPDKIKNLVTMVTPVDFYQTETLLNMRGGCSLGSEALDIDLMVDAMGNIPGDF
gi|425443051|ref|ZP_18823283.1|      SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGAEALDIDLMVDTMGNIPGDF
gi|422304946|ref|ZP_16392283.1|      SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGSEALDIDLMVDAMGNIPGDF
gi|425457520|ref|ZP_18837223.1|      SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGSEALDIDLMVDTMGNIPGDF
gi|425459008|ref|ZP_18838494.1|      SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGAEALDIDLMVDTMGNIPGDF
gi|425464502|ref|ZP_18843815.1|      SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGAEALDIDLMVDTMGNIPGDF
gi|440754979|ref|ZP_20934181.1|      SLCYSSLYPDKVKNLVTMVTPVDFYQTETLLNMRGGCSLGAEALDIDLMVDTMGNIPGDF
gi|390442356|ref|ZP_10230362.1|      SLCYSSLYPDKIKNLVTMVTPVDFYQTETLLNMRGGCSLGSEALDIDLMVDAMGNIPGDF
gi|428204128|ref|YP_007082717.1|     SLCYSSLYPEKVKNLIVMVAPVDFNMPNTLLNMRGGCTLGAEALDVDLMVKSLGNIPGDF
gi|18642978|gb|AAK38139.1|           SLCYSAIYPEKVKNLIVMVTPVDFQISDSLLYMRGGCTLGAEALDIDLMVDCLGNIPGDF
gi|451780150|gb|AGF51119.1|          SLCYASLFPDKVKNLVVMVAPVDFEQPGTLLNARGGCTLGAEAVDIDLMVDAMGNIPGDY
                                     *:**:::.* *:***:.**:*::*     :*   .**::* .*:*:** :. :**:***:
```

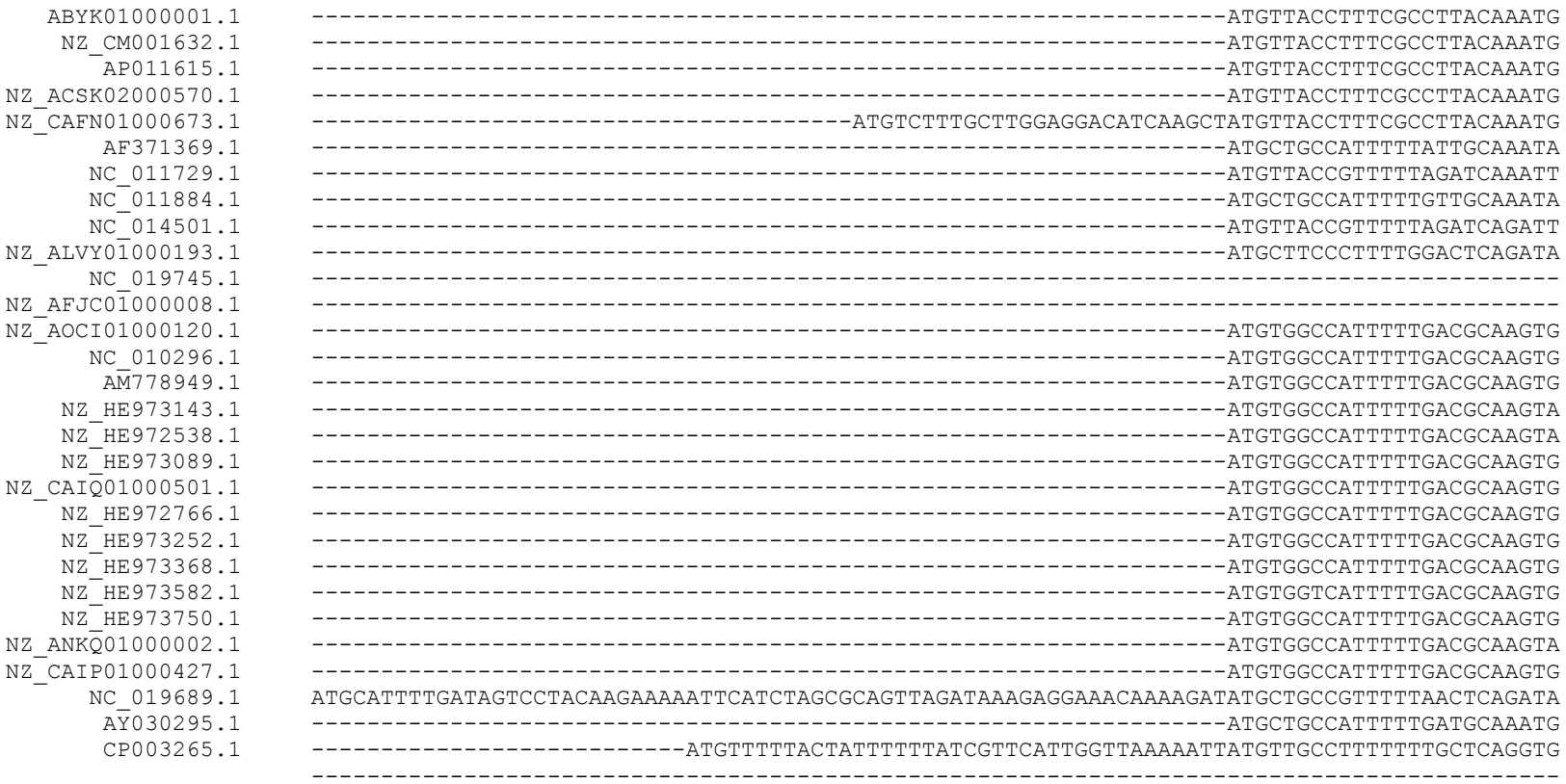
```
                                                                                                <<<<<<<
gi|209496929|gb|EDZ97226.1|          LNIEFLMLKPLQLGYQKYLDLPEIMGSRDKLLNFLRMEKWIFDSPDQAGETYRQFLKDFY
gi|423063393|ref|ZP_17052183.1|      LNIEFLMLKPLQLGYQKYLDLPEIMGSRDKLLNFLRMEKWIFDSPDQAGETYRQFLKDFY
gi|291571742|dbj|BAI94014.1|         LNIEFLMLKPLQLGYQKYLDLPEIMGSRDKLLNFLRMEKWIFDSPDQAGETYRQFLKDFY
gi|409991206|ref|ZP_11274489.1|      LNIEFLMLKPLQLGYQKYLDLPEIMGSRDKLLNFLRMEKWIFDSPDQAGETYRQFLKDFY
gi|376007312|ref|ZP_09784510.1|      LNIEFLMLKPLQLGYQKYLDLPEIMGSRDKLLNFLRMEKWIFDSPDQAGETYRQFLKDFY
gi|18644660|gb|AAL76316.1|           LNLEFLMLKPQQLGIQEYLDVPDLMDSPEKLLNFLRMEKWIFDSPDQAGETYRQFMKDFY
gi|218437501|ref|YP_002375830.1|     LNLEFLMLKPLQLGYQKYLDFPDIMENESKLANFMRMEKWIFDSPDQAGEAYRQFMKDFY
gi|220909421|ref|YP_002484732.1|     LNLEFLMLKPQQLGIQKYLDVPDLMDSPEKLLNFLRMEKWIFDSPDQAGETYRQFMKDFY
gi|307151222|ref|YP_003886606.1|     LNLEFLMLKPLQLGYQKYLDFPDIMENEDKLTNFMRMEKWIFDSPDQAGEAYRQFMKDFY
gi|443321910|ref|ZP_21050948.1|      LNWEFLMLKPYQLGIQKYVDFFNIMENKEQMLNFLRMEKWIFDSPEQVGEAYRQFLKDFY
gi|434393686|ref|YP_007128633.1|     LNFQFLMLKPFQLGVEKYIKFLESSDSEEKIINFFRMEKWIFDSPDLAGEAFRQYMKDFY
gi|334119572|ref|ZP_08493657.1|      LNLEFLMLKPFQLGVQKYIDLLENIDCESKLINFLRMEKWIFDSPDQAGEAYRQFMKDFY
gi|443658276|ref|ZP_21132094.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|166367744|ref|YP_001660017.1|     LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|159029277|emb|CAO90143.1|         LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|425448807|ref|ZP_18828651.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|425434880|ref|ZP_18815344.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|425448547|ref|ZP_18828448.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|425469530|ref|ZP_18848456.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|425443051|ref|ZP_18823283.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|422304946|ref|ZP_16392283.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|425457520|ref|ZP_18837223.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|425459008|ref|ZP_18838494.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|425464502|ref|ZP_18843815.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|440754979|ref|ZP_20934181.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|390442356|ref|ZP_10230362.1|      LNLEFLELKPLQLGYQKYLDFPDIMEDESKLVNFLRMEKWIFDSPDQAGESYRQFLKDFY
gi|428204128|ref|YP_007082717.1|     LNLEFLMLKPQQLGIQKYLDFPEVMTSEDKLLNFMRMEKWIFDSPDQAGEAYRQFMKDFY
gi|18642978|gb|AAK38139.1|           LNFEFLMLKPRQLGIQKYLDFPEIMHSEDKLLNFLRMEKWIFDSPDQAGEAYRQFLKDFY
gi|451780150|gb|AGF51119.1|          LNLEFLMLKPLQLGYQKYLDVPDIMGDEAKLLNFLRMEKWIFDSPDQAGETYRQFLKDFY
                                     ** :** *** *** ::*:.. :      :: **:**********: .**::**::****


gi|209496929|gb|EDZ97226.1|          QENKLIKGEVMIGDSRVDLSNITMPVLNLYAEKDHLVPPSSSLALEEYIS-SEDYTAKSF
gi|423063393|ref|ZP_17052183.1|      QENKLIKGEVMIGDSRVDLSNITMPVLNLYAEKDHLVPPSSSLALEEYIS-SEDYTAKSF
gi|291571742|dbj|BAI94014.1|         QENKLIKGEVMIGDSRVDLSNITMPVLNLYAEKDHLVPPSSSLALEEYIS-SEDYTAKSF
gi|409991206|ref|ZP_11274489.1|      QENKLIKGEVMIGDSRVDLSNITMPVLNLYAEKDHLVPPSSSLALEEYIS-SEDYTAKSF
gi|376007312|ref|ZP_09784510.1|      QENKLIKGEVMIGDSRVDLSNITMPVLNLYAEKDHLVPPSSSLALEEYIS-SEDYTAKSF
gi|18644660|gb|AAL76316.1|           QGNKLIKNQVKIGDRQVNLLNLTMPILNLYAEKDHLVPPASSLALAKYID-TQDYTAKGF
gi|218437501|ref|YP_002375830.1|     QSNKLIKNEVVIGNKPVNLQNLTMPILNLYAELDHLVDPASSKALEKYVN-TTDYIVQSF
gi|220909421|ref|YP_002484732.1|     QGNKLIKNQVKIGDQLVNLLNLTMPILNLYAEKDHLVPPASSVALAKYIG-TQDYTAKGF
gi|307151222|ref|YP_003886606.1|     QGNKLIKNEVVIGDQRVNLQNLTMPILNLYAEQDHLVDPVSSKALEKYVN-SSDYTLKSF
gi|443321910|ref|ZP_21050948.1|      QENKLIHNEIQIGDKRVDLGQILMPVLNLYAEKDHLVPPLSSLALEKYVG-TQDYTTQSF
gi|434393686|ref|YP_007128633.1|     QENKLIKGQLEIGGKRVHLEKIRIPIFNIYAEQDHLVPPASSLALEKYVA-SSEYTVRSF
gi|334119572|ref|ZP_08493657.1|      QGNKLIQGQVEIGNKRVDLGNIRIPILNIYAEQDHLVAPASSLALKTYIA-SEDYTLRSF
gi|443658276|ref|ZP_21132094.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|166367744|ref|YP_001660017.1|     QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|159029277|emb|CAO90143.1|         QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|425448807|ref|ZP_18828651.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|425434880|ref|ZP_18815344.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|425448547|ref|ZP_18828448.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|425469530|ref|ZP_18848456.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|425443051|ref|ZP_18823283.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|422304946|ref|ZP_16392283.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|425457520|ref|ZP_18837223.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|425459008|ref|ZP_18838494.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|425464502|ref|ZP_18843815.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|440754979|ref|ZP_20934181.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|390442356|ref|ZP_10230362.1|      QQNKLIKGEVMLGDKRVDLHNLTMPILNLYADKDHLVPPASSLALGNYIG-TSDYTACAF
gi|428204128|ref|YP_007082717.1|     QENKLIKGEVMLGDKRVDLKNVRMPVLNLYAEKDHLVDPESSKALEKYVG-TDDYTVRSF
gi|18642978|gb|AAK38139.1|           QANKLIKGEVTIGDKQVNLGNIRMPVLNLYAEKDHLVPPRSSIALERYIG-TTDYTVRSF
gi|451780150|gb|AGF51119.1|          QQNKLIKGEVMIGDRLVDLHNLTMPILNLYAEKDHLVAPASSLALGDYLPENCDYTVQSF
                                     * ****: :: :*   *.* :: :*::*:**: **** * ** **  *:  . :*   .*
```

```
gi|209496929|gb|EDZ97226.1|          PVGHIGMYVSGKVQRDLPPTIVDWLKVRE--
gi|423063393|ref|ZP_17052183.1|      PVGHIGMYVSGKVQRDLPPTIVDWLKVRE--
gi|291571742|dbj|BAI94014.1|         PVGHIGMYVSGKVQRDLPPTIVDWLKVRE--
gi|409991206|ref|ZP_11274489.1|      PVGHIGMYVSGKVQRDLPPTIVDWLKVRE--
gi|376007312|ref|ZP_09784510.1|      PVGHIGMYVSGKVQRDLPPTIVDWLKVRE--
gi|18644660|gb|AAL76316.1|           PVGHIGMYVSGKVQRDLPPVIADWLRNRD--
gi|218437501|ref|YP_002375830.1|     PVGHIGMYVSGKVQATLPPTIVEWLTARA--
gi|220909421|ref|YP_002484732.1|     PVGHIGMYVSGKVQQDLPPVIADWLRNRD--
gi|307151222|ref|YP_003886606.1|     PVGHIGMYVSGKVQKDLPPTIVDWLKARS--
gi|443321910|ref|ZP_21050948.1|      PVGHIGMYVSSKVQRDLPQIIVNWIKARSLN
gi|434393686|ref|YP_007128633.1|     PVGHIGMYVSRKVQKDLPEAIADWLK-----
gi|334119572|ref|ZP_08493657.1|      PVGHIGMYVSSKVQRDLPPTIVDWLKMRA--
gi|443658276|ref|ZP_21132094.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARG--
gi|166367744|ref|YP_001660017.1|     PVGHIGMYVSGKVQRDLPPAISDWLKARA--
gi|159029277|emb|CAO90143.1|         PVGHIGMYVSGKVQRDLPPAISDWLKARG--
gi|425448807|ref|ZP_18828651.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARA--
gi|425434880|ref|ZP_18815344.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARA--
gi|425448547|ref|ZP_18828448.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARA--
gi|425469530|ref|ZP_18848456.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARG--
gi|425443051|ref|ZP_18823283.1|      PVGHIGMYVSGKVQRDLPPAITDWLKARA--
gi|422304946|ref|ZP_16392283.1|      PVGHIGMYVSGKVQRDLPPAITDWLKARG--
gi|425457520|ref|ZP_18837223.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARG--
gi|425459008|ref|ZP_18838494.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARA--
gi|425464502|ref|ZP_18843815.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARA--
gi|440754979|ref|ZP_20934181.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARA--
gi|390442356|ref|ZP_10230362.1|      PVGHIGMYVSGKVQRDLPPAISDWLKARA--
gi|428204128|ref|YP_007082717.1|     PVGHIGMYVSGKVQRDLPPTIVDWLKARM--
gi|18642978|gb|AAK38139.1|           PVGHIGIYVSSKVQRDLPPIIANWLNARE--
gi|451780150|gb|AGF51119.1|          PVGHIGMYVSGKVQRDLPPAIAHWLSERQ--
                                     ******:*** ***  **  * .*:
```

A.8 PHA Synthase CEMA (29 sequences)

Appendix Figure 2: Alignment of *phaC* coding sequences generated from the analogous coding product MSA returned from Clustal Ω. Top row depicts conserved amino acid residues from clustal output. The bottom row depicts consensus residues. Primer annealing regions are highlighted. See Appendix Table 4 for sequence accession table.



```
                                                e__l__       k__                d__    l__       l__k__E__e__d__
   ABYK01000001.1     GGTTTAGAAGACTTAACCCAGGAATATGCAGACCTCACCGAAAAAATTGTTCATGGTATGGACAACCTTAGCAGTTTACGGGAGGAAGAA
    NZ_CM001632.1     GGTTTAGAAGACTTAACCCAGGAATATGCAGACCTCACCGAAAAAATTGTTCATGGTATGGACAACCTTAGCAGTTTACGGGAGGAAGAA
       AP011615.1     GGTTTAGAAGACTTAACCGAGGAATATGCAGACCTCACCGAAAAAATTGTTCATGGTATGGACAACCTTAGTAGTTTACGGGAGGAAGAA
NZ_ACSK02000570.1     GGTTTAGAAGACTTAACCGAGGAATATGCAGACCTCACCGAAAAAATTGTTCATGGTATGGACAACCTTAGTAGTTTACGGGAGGAAGAA
NZ_CAFN01000673.1     GGTTTAGAAGACTTAACCCAGGAATATGCAGACCTCACCGAAAAAATTGTTCATGGTATGGACAACCTTAGCAGTTTACGGGAGGAAGAA
       AF371369.1     CATCTGGAAGAGGCCACGCACGAATCCGCACAGCTCACCCACAAACTGGTGAAGGGCATGGAAAACCTCAGCCAGCTCCGTGAGGAAGAC
      NC_011729.1     CGTTTAGAAGATGCAGTCCACGAATACACCGAAATCACCAAAAAGATGATTAAAGGGCTAGATAATTTGAGCCGTTTACGAGAAGAAGAT
      NC_011884.1     CACCTGGAAGAGGCCGCCCATGAATCCGCACAGCTCACCCACAAACTGGTGAAGGGCATGGAAAACCTCAGCCAGCTCCGTGAGGAAGAC
      NC_014501.1     CGTTTAGAAGATGCCGTCCACGAATATACTGAAATCACCAAAAAGATGCTCAAAGGGCTGGATAATTTAAGCCGCTTGCGGGAAGAAGAT
NZ_ALVY01000193.1     AGTATTGAAGATACTACCTGTGAGTATATTGAGCTAACTAAAAAATTACTTAAAGGTATTCAAAATTTAAGGGAGTTGAGAGAAAGTGAC
      NC_019745.1     ----------------------ATGCTTACAGATCTA---GCCGAACTTTGGCAACGTGGTGAAGTTTTTAGCCGTTTGCGCGAAGAAGAC
NZ_AFJC01000008.1     ------------------ATGGTAACGACAGAGCTAACCCAAAAACTGGTCAAGGGCGTCGAGATTTTCACCCGCCTGCGGGAGGAAGAT
NZ_AOCI01000120.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGGTTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
      NC_010296.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGGTTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
       AM778949.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGGTTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
    NZ_HE973143.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGTTTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
    NZ_HE972538.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGTTTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
    NZ_HE973089.1     AAACTGGAAGACTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGTTTGGACAATCTCAAACGAGTTAAAGAAGAAGAT
NZ_CAIQ01000501.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTTAAAGGTCTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
    NZ_HE972766.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAAAAAAATCTCAAAGGTTTGGACAATCTCAAACGAGTTAAAGAAGAAGAT
    NZ_HE973252.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGTTTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
    NZ_HE973368.1     AAACTGGAAGACTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGTTTGGACAATCTCAAACGAGTTAAAGAAGAAGAT
    NZ_HE973582.1     AAACTGGAAGACTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGTTTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
    NZ_HE973750.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAAAAAAATCTCAAAGGTTTGGACAATCTCAAACGAGTTAAAGAAGAAGAT
NZ_ANKQ01000002.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGTTTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
NZ_CAIP01000427.1     AAACTGGAAGATTTTACCCAAGATTATCTAGAATTAACTCAGAAAAATCTCAAAGGTTTAGACAATCTCAAACGAGTTAAAGAAGAAGAT
      NC_019689.1     CGCCTTGAAGATGTGAGCCACGAGTACACCGAACTAACTAAAAAAGTCCTTCAAGGCATTGAAAATCTAAGTCGCTTGCGAGAGGAAGAT
       AY030295.1     CGCCTTGATGACGCCACCGAGGAGTATACCGAACTTATTAAAAAGATTGTCAAAGGAATTGAAAATTTAAGTCGCCTGCGAGAAGAAGAC
       CP003265.1     GGGTTAGAAGAAAATCTCCATGAAACCCTAGATTTTACTGAAAAATTTCTCTCTGGCTTGGAAAATTTGCAGGGTTTGAATGAAGATGAC
                      ----------------------gattatctagAatTa---caaaAaattctcaaagGtttggAcaatcTcaaacgagTtaaaGAagaaGAt
```

```
                          G__   s__   k__E__   v__y__r__E__d__K__i__   L__Y__h__F__         v__   k__
   ABYK01000001.1     ATTATCGTTGGGGTCACACCCAAAGAAGCAGTTTATCAGGAAGATAAAGTCACCCTTTATCGTTTTGAACCCAAAGTCAAAAAA------
    NZ_CM001632.1     ATTATCGTTGGGGTCACACCCAAAGAAGCAGTTTATCAGGAAGATAAAGTCACCCTTTATCGTTTTGAACCCAAAGTCAAAAAA------
       AP011615.1     ATTATCGTTGGGGTCACACCCAAAGAAGCAGTTTACCAGGAAGATAAAGTCACCCTTTATCGGTTTGAACCCAAAGTCAAAAAA------
NZ_ACSK02000570.1     ATTATCGTTGGGGTCACACCCAAAGAAGCAGTTTACCAGGAAGATAAAGTCACCCTTTATCGGTTTGAACCCAAAGTCAAAAAA------
NZ_CAFN01000673.1     ATTATCGTTGGGGTCACACCCAAAGAAGCAGTTTATCAGGAAGATAAAGTCACCCTTTATCGTTTTGAACCCAAAGTCAAAAAA------
      AF371369.1     ATTGAGGTGGGGGCCACTCCCAGGGAGGTGGTTTTCCAGGAGGATAAGGTCAAACTCTATCGCTTCAAATCACCCGTTGATCAGAAAAAG
      NC_011729.1     ATTGAAAGCGGTGTATCTCCCAAAGAAGTAGTTTATCAAGAGGATAAAGTTGTCCTCTATCGGTTTAAATCTCAAGTTGAACAT------
      NC_011884.1     ATTGAGGTGGGGTCCACTCCCAGGGAGGTGGTTTACCAGGAGGATAAGGTTAAACTCTATCGATTTAAAGCTCCAGCTAACCAGGGAAAA
      NC_014501.1     ATTCAAAGCGGCGTATCAGCTAAAGAAGCCGTTTATAAAGAGGATAAAGTCATCCTCTATCGGTTTACCCCTCAAGTGGCGCAA------
NZ_ALVY01000193.1     ATTGAAATTGGTATAACTCCTAAAGAAGTAATTTACCAAGAGGACAAAATGTTACTCTATCGCTTTAAGCCAATGGTAGAAAAC------
      NC_019745.1     AATTTAATAGGAGTCACGCCGAAGGAAGAAATCTACCGCGAAGATAAGGTGGTGTTGTATCGCTTTACTCCACAAGTGAAAAAT------
NZ_AFJC01000008.1     ATTCAAGTCGGGGTGACTCCCAAGGAAGAAGTTTACCGGGAAGATAAGGTACTGCTGTACCACTTCTCGCCGAAAGTTGAGCAT------
NZ_AOCI01000120.1     ATTCAGTGTGGAGTATCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTCAAACCCGTGGTCGAAAAA------
      NC_010296.1     ATTCAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTCGAAAAA------
       AM778949.1     ATTCAGTGTGGAGTATCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTCAAACCCGTGGTCGAAAAA------
    NZ_HE973143.1     ATTGAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTCGAAAAA------
    NZ_HE972538.1     ATTGAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTCGAAAAA------
    NZ_HE973089.1     ATTCAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTTGAAAAA------
NZ_CAIQ01000501.1     ATTCAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTCGAAAAA------
    NZ_HE972766.1     ATTCAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTCGAAAAA------
    NZ_HE973252.1     ATTCAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTTGAAAAA------
    NZ_HE973368.1     ATTCAGTGTGGAGTCTCAGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTTGAAAAA------
    NZ_HE973582.1     ATTGAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTTGAAAAA------
    NZ_HE973750.1     ATTCAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTCGAAAAA------
NZ_ANKQ01000002.1     ATTGAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTCGAAAAA------
NZ_CAIP01000427.1     ATTCAGTGTGGAGTCTCGGAAAAAGAAGCAGTTTATCGGGAAGATAAAATCATTCTCTACCACTTTAAACCCGTGGTTGAAAAA------
      NC_019689.1     ATTGAAATCGGCGCTACGCCTAAAGAAGCGGTTTACAAAGAGGATAAAGTGATTCTATACCGCTTCAAGCCGATGGTCGAGCAG------
       AY030295.1     ATCGAAATTGGTGTCACTCCCAAGGAAGCTGTCTATCGCGAGGAAAAATTGACTTTGTACCACTTTCAATCAACGGTACAGAAG------
       CP003265.1     ATCCAGGTGGGCTTTACCCCCAAAGAAGCAGTTTACCAGGAAGATAAGGTTATTCTTTACCGTTTCCAACCGGTGGTGGAAAAT------
                      AttcagtgtGGagtctCggcaAaaGAaGcagTtTatcggGAaGAtAAaaTcattcTcTAcCacTTtaaacCcgtgGtcgaaaAa------
```

```
                        f__       P__l__L__m__V__Y__A__L__V__N__R__P__y__m__V__D__L__Q__       R__S__L__V__A__N__
   ABYK01000001.1     ACCCTTTCTGTACCCCTGCTAATTGTTTATGCTTTAGTCAATCGTCCCTTTATGGTAGATTTGCAAGAGGGTCGTTCCTTAGTTGCTAAT
    NZ_CM001632.1     ACCCTTTCTGTACCCCTGCTAATTGTTTATGCTTTAGTCAATCGTCCCTTTATGGTAGATTTGCAAGAGGGTCGTTCCTTAGTTGCTAAT
       AP011615.1     ACCCTTTCTGTACCTCTGCTAATTGTCTATGCTTTAGTCAATCGTCCCTTTATGGTAGATTTGCAAGAGGGTCGTTCCTTAGTTGCTAAT
NZ_ACSK02000570.1     ACCCTTTCTGTACCTCTGCTAATTGTCTATGCTTTAGTCAATCGTCCCTTTATGGTAGATTTGCAAGAGGGTCGTTCCTTAGTTGCTAAT
NZ_CAFN01000673.1     ACCCTTTCTGTACCCCTGCTAATTGTTTATGCTTTAGTCAATCGTCCCTTTATGGTAGATTTGCAAGAGGGTCGTTCCTTAGTTGCTAAT
       AF371369.1     ACAGTCAAAACACCTATTCTGATGGTCTACGCCCTGGTAAACCGCCCCTTTATGGTGGACTTACAGGAAGATCGCTCCCTAGTGGCTAAT
      NC_011729.1     CCTCTGCCGATTCCTTTATTGATGGTTTATGCGTTGGTAAATCGCCCTTTTATGGTAGATTTACAGGAAGGACGCTCTTTAGTCGCTAAT
      NC_011884.1     ACGGTGCAAACACCGATACTGATGGTCTACGCCCTGGTGAACCGCCCCTTTATGGTGGACTTACAGGAAGATCGCTCCCTGGTGGCTAAT
      NC_014501.1     CCGTTACATATCCCGTTATTGATGGTTTATGCTTTGGTAAACCGTCCTTTTATGGTGGATTTGCAGGAAGGACGCTCTTTAGTCGCTAAT
NZ_ALVY01000193.1     CCTCTAACTATTCCCTTATTGATTGTCTACGCTTTGGTCAATCGTCCTTTTATGGTTGATCTACAGGAAAATCGTTCCTTAGTGGCTAAT
      NC_019745.1     TTACTCAACACTCCGATTTTGATTGTTTATGCCTTAGTGAATCGTCCTTATATTGTTGATTTACAAGCAAAGCGATCGCTTGTTGCTAAT
NZ_AFJC01000008.1     TCGCTGAATATTCCCATACTCATCGTTTACGCCCTGGTTAATCGTCCCTACATAGTCGATTTACAGGAGGGGCGATCGCTCGTTGCGAAT
NZ_AOCI01000120.1     CCCTTCGAGATTCCCTTGCTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTAGCAAAT
      NC_010296.1     CCCTTCGAGATTCCCTTGTTGATGGTATATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTGGCCAAT
       AM778949.1     CCCTTCGAGATTCCCTTGCTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTAGCAAAT
    NZ_HE973143.1     CCCTTCGAGATTCCCTTGCTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTAGCAAAT
    NZ_HE972538.1     CCCTTCGAGATTCCCTTGCTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTGGCCAAT
    NZ_HE973089.1     CCCTTCGAGATTCCCTTGCTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTGGCCAAT
NZ_CAIQ01000501.1     CCCTTCGAGATTCCCTTGCTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTAGCAAAT
    NZ_HE972766.1     CCCTTCGAGATTCCCTTGTTGATGGTTTATGCCTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTGGCAAAT
    NZ_HE973252.1     CCTTTCGAGATTCCCTTGTTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTGGCCAAT
    NZ_HE973368.1     CCCTTCGAGATTCCCTTGCTGATGGTTTATGCTTTGGTTAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTGGCCAAT
    NZ_HE973582.1     CCTTTCGAGATTCCCTTGTTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTAGCAAAT
    NZ_HE973750.1     CCCTTCGAGATTCCCTTGTTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTGGCAAAT
NZ_ANKQ01000002.1     CCCTTCGAGATTCCCTTGCTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTAGCAAAT
NZ_CAIP01000427.1     CCCTTCGAGATTCCCTTGCTGATGGTTTATGCTTTGGTCAATCGTCCCTACATGGTAGATTTACAGGAAGGGCGTTCTTTAGTGGCCAAT
      NC_019689.1     CCTTTGAGCATTCCCCTCCTAATTGTTTATGCTTTGGTTAACCGTCCCTATATGGTCGATCTGCAAGAGGATCGATCCCTGGTTGCCAAT
       AY030295.1     CAATTGAGAACTCCTGTTCTCATCGTCTACGCCTTGGTAAACCGCCCTTTTATGGTCGATTTGCAAGAAGATCGATCGCTGGTTGCTAAC
       CP003265.1     CCCTTACCTATCCCGGTTTTAATTGTTTACGCCCTGGTAAATCGCCCCTACATGGTGGATTTGCAGGAAGGACGCTCCCTGGTGGCCAAC
                      ccctTcgagattCCctTgcTgATgGTtTAtGCttTgGTcAAtCGtCCcTacATgGTaGAttTaCAgGaagggCGtTCttTaGTgGCtAAt
```

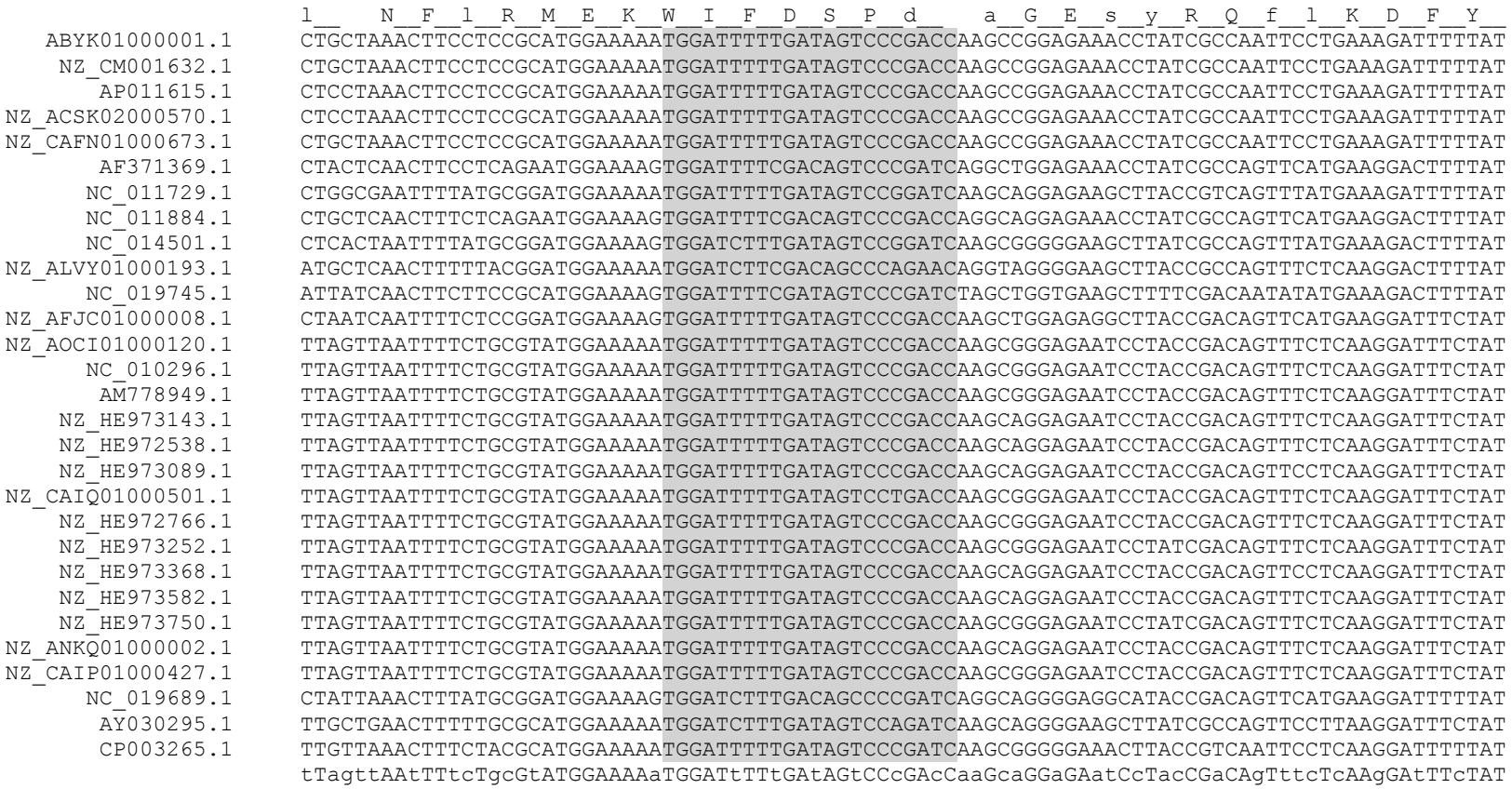
```
                    L__L__k__L__G__l__D__i__Y__L__I__D__W__G__Y__P__    R__    D__R__W__l__t__L__d__D__Y__i__n__
   ABYK01000001.1   TTACTCAGTTTGGGATTAGATGTCTATTTGATTGACTGGGGATATCCTACCCGTAGCGATCGCTGGTTAACCTTAGATGATTACATCAAC
    NZ_CM001632.1   TTACTCAGTTTGGGATTAGATGTCTATTTGATTGACTGGGGATATCCTACCCGTAGCGATCGCTGGTTAACCTTAGATGATTACATCAAC
      AP011615.1    TTACTCAGTTTGGGATTAGATGTCTATTTGATTGATTGGGGATATCCTACCCGTAGCGATCGCTGGTTAACCTTAGATGATTATATCAAC
NZ_ACSK02000570.1   TTACTCAGTTTGGGATTAGATGTCTATTTGATTGATTGGGGATATCCTACCCGTAGCGATCGCTGGTTAACCTTAGATGATTATATCAAC
NZ_CAFN01000673.1   TTACTCAGTTTGGGATTAGATGTCTATTTGATTGACTGGGGATATCCTACCCGTAGCGATCGCTGGTTAACCTTAGATGATTACATCAAC
      AF371369.1    CTGCTCAAGTTGGGCCTGGATATCTATCTGATCGATTGGGGCTATCCCGGCCGGGGCGATCGCTGGTTGACCCTGGACGATTACATCAAT
      NC_011729.1   TTGCTCAAATTAGGCTTAGATGTCTATTTAATTGATTGGGGTTATCCGACAAGGGCAGACCGATGGTTAACCTTGGATGATTATATCAAT
      NC_011884.1   CTGCTCAAGTTGGGTCTGGATATCTATCTGATCGATTGGGGCTATCCCGGCCGGGGCGATCGCTGGTTGACCCTGGACGATTACATCAAT
      NC_014501.1   TTGCTCAAATTGGGCTTAGATGTTTATTTGATTGACTGGGGATATCCTACCCGTGCCGATCGCTGGCTGACATTAGATGATTATATCAAC
NZ_ALVY01000193.1   TTACTCAAATTAGGATTAGATATTTATTTAATTGATTGGGGTTATCCTACTCGCGCCGATCGCTGGATGAATCTTGACGATTATATCAAT
      NC_019745.1   TTGCTCAAACTCGGTGTTGATGTTTACTTAATCGATTGGGGTTATCCAAGTCGAATTGATCGTTGGCTAACGCTTGATGATTACATTAAT
NZ_AFJC01000008.1   TTGCTAGAGCTTGGTTTGGATGTTTACCTGATTGACTGGGGCTATCCGAGTCGCGGCGATCGCTGGTTAACTCTTGACGACTACATTAAT
NZ_AOCI01000120.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGTGATCGCTGGTTAACCCTTGATGATTATATCAAT
      NC_010296.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
       AM778949.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGTGATCGCTGGTTAACCCTTGATGATTATATCAAT
    NZ_HE973143.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
    NZ_HE972538.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
    NZ_HE973089.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
NZ_CAIQ01000501.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
    NZ_HE972766.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGCTATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
    NZ_HE973252.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
    NZ_HE973368.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
    NZ_HE973582.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
    NZ_HE973750.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGCTATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
NZ_ANKQ01000002.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
NZ_CAIP01000427.1   CTGCTGAAATTAGGCTTAGATATCTACTTAATTGATTGGGGATATCCCACCAGAAGCGATCGCTGGTTAACCCTTGATGATTATATCAAT
      NC_019689.1   TTGCTCAAACTGGGTTTGGATGTCTATTTGATTGACTGGGGATACCCCAGCAGAGCCGATCGCTGGTTAACTCTCGACGATTACATTAAT
       AY030295.1   TTGCTGAAATTAGGTTTGGATATTTATTTGATTGATTGGGGTTACCCTACCAGAGCCGATCGCTGGCTGACTCTAGATGACTACATTAAC
       CP003265.1   CTCCTCAAACTGGGTTTGGACGTGTATTTAATTGATTGGGGTTATCCCTCCCGGGGCGATCGTTGGTTGACCCTAGAAGATTATTTGTCT
                    cTgCTgaaatTaGGctTaGAtaTcTActTaATtGAtTGGGGaTAtCCcaccaGaagcGAtCGcTGGtTaAcccTtGAtGAtTAtaTcaat
```

```
                    G__Y__v__n__N__C__v__d__    i__                        k__i__n__L__L__G__i__C__Q__G__G__t__F__
   ABYK01000001.1   GGTTATATTAATAACTGTGTTGATTTTCTGCGCGATCACTATGAACTCGATAAAATCAACCTCCTAGGAGTTTGTCAGGGAGGAACCTTT
    NZ_CM001632.1   GGTTATATTAATAACTGTGTTGATTTTCTGCGCGATCACTATGAACTCGATAAAATCAACCTCCTAGGAGTTTGTCAGGGAGGAACCTTT
       AP011615.1   GGTTATATTAATAACTGTGTTGATTTTCTTCGCGATCACTATGAACTCGATAAAATCAACCTCCTAGGAGTTTGTCAGGGAGGAACCTTT
NZ_ACSK02000570.1   GGTTATATTAATAACTGTGTTGATTTTCTTCGCGATCACTATGAACTCGATAAAATCAACCTCCTAGGAGTTTGTCAGGGAGGAACCTTT
NZ_CAFN01000673.1   GGTTATATTAATAACTGTGTTGATTTTCTGCGCGATCACTATGAACTCGATAAAATCAACCTCCTAGGAGTTTGTCAGGGAGGAACCTTT
       AF371369.1   GGTTATCTGAACAATTGCGTCGATTTTATCCGCACCAGCCATCAACTGGACAAGGTGAACCTGCTGGGCATTTGTCAGGGTGGCACCTTC
       NC_011729.1   GGTTACATCAATAATTGTGTTGATTTTATCCGCAAACAACATAATTTAGACAAAATCAATTTATTGGGCATTTGTCAAGGGGGAACATTT
       NC_011884.1   GGTTATCTGAACAATTGCGTCGATTTTATCCGTGCCAGCCATCAACTGGACAAGGTGAACCTGCTGGGCATTTGTCAGGGTGGCACCTTC
       NC_014501.1   GGCTACATTGATAATTGTGTGGATTATATCCGCAAAACGCACAATATCGATAAAGTTAATCTGTTAGGCATCTGTCAAGGGGGAACTTTT
NZ_ALVY01000193.1   GGTTACATTAATAACTGTGTCGAGGTAGTGCGAAAAAGGCATGGTTTAGAAAAGATTAATCTTTTAGGAATTTGTCAGGGAGGAGCTTTT
       NC_019745.1   GGCTATATTAATAACTGCATCGATGTTGTCTGCGATCGCCACAACCTTGCGCAAATTAATCTTTTAGGCATTTGTCAGGGAGGAACCTTT
NZ_AFJC01000008.1   GGCTATATCAACAACTGTGTGGATGTGGTACGCGATCGTCACAACTTAGAGCAAATCAACCTGTTAGGTATTTGTCAGGGGGGAACCTTC
NZ_AOCI01000120.1   GGTTATGTGGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGTTAGGAATCTGTCAGGGGGGAACCTTT
       NC_010296.1   GGTTATGTCGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGTTAGGAATCTGTCAGGGGGGAACCTTT
        AM778949.1   GGTTATGTGGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGTTAGGAATCTGTCAGGGGGGAACCTTT
      NZ_HE973143.1   GGTTATGTGGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGCTAGGAATCTGTCAGGGGGGAACCTTT
      NZ_HE972538.1   GGTTATGTGGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGTTAGGAATCTGTCAGGGGGGAACCTTT
      NZ_HE973089.1   GGTTATGTCGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAGATTAATCTCTTAGGAATCTGTCAGGGAGGAACCTTT
NZ_CAIQ01000501.1   GGTTATGTGGATAATTGCGTCGATTTTATTCGTCAAAGTCATCATCTCGACAAAATTAATCTGTTAGGAATCTGTCAGGGGGGAACCTTT
      NZ_HE972766.1   GGTTATGTGGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGTTAGGAATCTGTCAGGGAGGAACCTTT
      NZ_HE973252.1   GGTTATGTCGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGTTAGGAATCTGTCAGGGAGGAACCTTT
      NZ_HE973368.1   GGTTATGTGGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAGATTAATCTCTTAGGAATCTGTCAGGGAGGAACCTTT
      NZ_HE973582.1   GGTTATGTGGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGTTAGGAATCTGTCAGGGGGGAACCTTT
      NZ_HE973750.1   GGTTATGTGGATAATTGTGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGCTAGGGATCTGTCAGGGAGGAACCTTT
NZ_ANKQ01000002.1   GGTTATGTGGATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGCTAGGAATCTGTCAGGGGGGAACCTTT
NZ_CAIP01000427.1   GGTTATGTCAATAATTGCGTCGATTTTATTCGTCAAAGTCACCATCTCGACAAAATTAATCTGCTAGGAATCTGTCAGGGGGGAACCTTT
       NC_019689.1   GGCTATATCAATAACTGTGTCGATTTCATCAGAGAAAAGCACGGTTTGGAGAAAATTAACCTTTTAGGGATTTGTCAAGGGGGAGCTTTC
        AY030295.1   GGCTACATCAATAACTGTGTAGATTTCATCAGGAAAAAGCATGATTTAGACAAAATAAACCTACTGGGAATTTGTCAGGGTGGAACTTTT
        CP003265.1   GGATATTTGAACAACTGTGTCGATATTATTTGTCAACGCTCCCAGCAAGAAAAAATTACGTTGTTAGGAGTTTGTCAGGGGGGCACATTT
                    GGtTAtgTgaAtAAtTGcgTcGAttttaTtcGtcaaagtcaccatctcGacaAaaTtAatcTgtTaGGaaTcTGTCAgGGgGGaaCcTTt
```

```
                  S__l__C__Y__s__s__l__y__P__   K__v__k__N__L__v__t__M__V__t__P__v__d__F__              l__
   ABYK01000001.1 AGCCTCTGCTACAGTTCCCTATATCCCGAAAAAGTGCAAAACCTCATCACCATGGTTGCGCCAGTCAACTTTGATATGCCAAATACCCTG
    NZ_CM001632.1 AGCCTCTGCTACAGTTCCCTATATCCCGAAAAAGTGCAAAACCTCATCACCATGGTTGCGCCAGTCAACTTTGATATGCCAAATACCCTG
      AP011615.1 AGCCTCTGCTACAGTTCCCTATATCCCGAAAAAGTGCAAAACCTAATCACCATGGTTGCGCCAGTCAACTTTGATATGCCAAATACCCTC
NZ_ACSK02000570.1 AGCCTCTGCTACAGTTCCCTATATCCCGAAAAAGTGCAAAACCTAATCACCATGGTTGCGCCAGTCAACTTTGATATGCCAAATACCCTC
NZ_CAFN01000673.1 AGCCTCTGCTACAGTTCCCTATATCCCGAAAAAGTGCAAAACCTCATCACCATGGTTGCGCCAGTCAACTTTGATATGCCAAATACCCTG
      AF371369.1 AGCCTGTGCTACAGCTCCCTCTATCCGGATAAGGTGAACAATCTGGTCGTGATGGTGGCCCCCGTGGACTTTCATCAACCCGAAACCCTG
      NC_011729.1 AGCGTTTGCTACAGTGCAATTTACCCCGAAAAGGTGAAAAATCTCATCGTCATGGTTGCTCCCATTGATTTTCGGATGCCCGGCACGTTA
      NC_011884.1 AGCCTTTGCTACAGTTCCCTCTATCCGGATAAGGTGAACAATCTGGTCGTGATGGTGGCCCCCGTGGACTTTCATCAACCCGAAACCCTG
      NC_014501.1 AGCGTTTGTTATAGTGCCCTTCACCCTGAAAAGGTGAAAAATCTAATTGTGATGGTTGCTCCCATTGATTTTCGGATGCCGGGTACGCTG
NZ_ALVY01000193.1 AGTCTTTGTTATAGCGCTATTTACCCAGAAAAGGTTAAAAATCTCATTGTCATGGTTACTCCCGTAGATTTCCATATTCCTAATGCTTTT
      NC_019745.1 AGCCTTTGCTACAGTGCGCTTTACCCAGCGAAGGTAAAAAACCTGATTGTGATGGTTACGCCTGTTGATTTCCATACTCAAGAAGGGCTG
NZ_AFJC01000008.1 AGCCTCTGCTACAGTTCCCTTTACCCCGAGAAGGTAAAAAACCTGATTACTATGGTCACCCCCGTCGATTTTCACATCAATGAGGGACTC
NZ_AOCI01000120.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGGTAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
      NC_010296.1 AGTTTATGCTATAGTTCCCTCTACCCCGATAAGATAAAAAATCTGGTGACAATGGTGACACCAGTGGACTTTTATCAAACCGAGACCCTC
       AM778949.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGGTAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
    NZ_HE973143.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGGTAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
    NZ_HE972538.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGGTAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
    NZ_HE973089.1 AGTTTATGCTATAGTTCCCTCTATCCTGATAAGATAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
NZ_CAIQ01000501.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGATAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
    NZ_HE972766.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGGTAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
    NZ_HE973252.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGGTAAAAAATCTGGTGACAATGGTGACACCAGTGGACTTTTATCAAACCGAGACCCTC
    NZ_HE973368.1 AGTTTATGCTATAGTTCCCTCTATCCTGATAAGGTAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
    NZ_HE973582.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGGTAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
    NZ_HE973750.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGGTAAAAAATCTGGTGACAATGGTGACACCAGTGGACTTTTATCAAACCGAGACCCTC
NZ_ANKQ01000002.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGGTAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
NZ_CAIP01000427.1 AGTTTATGCTATAGTTCCCTCTATCCCGATAAGATAAAAAATCTGGTGACAATGGTGACACCGGTGGACTTTTATCAAACCGAGACCCTC
      NC_019689.1 AGTCTATGCTACTCTTCCCTTTATCCCGAAAAGGTGAAAAATTTAATCGTCATGGTTGCGCCAGTAGATTTTAACATGCCCAACACCTTG
       AY030295.1 AGCCTTTGCTATAGCGCTATCTATCCCGAAAAGGTCAAAAACCTGATCGTAATGGTTACGCCTGTTGATTTTCAAATATCAGATTCACTG
       CP003265.1 AGCCTGTGTTACGCTTCTCTATTCCCGGATAAGGTTAAAAATTTGGTGGTGATGGTGGCTCCGGTGGACTTTGAACAACCCGGTACTTTA
                  AGttTaTGcTAtagttCccTctatCCcGatAAggTaaAaAAtcTggTgacaATGGTgaCaCCggTggAcTTttatcaaaccgagacccTc
```

```
                    L__          g__G__C__s__l__G__    e__A__l__D__i__D__L__    v__d__    m__G__N__i__P__G__D__f__
   ABYK01000001.1   TTAAATGCGCGGGGAGGCTGTACATTGGGACCGGAAGCCATAGATGTTGACCTCATGGTTGAGGCTTTAGGTAACATTCCCGGCGACTAT
    NZ_CM001632.1   TTAAATGCGCGGGGAGGCTGTACATTGGGACCGGAAGCCATAGATGTTGACCTCATGGTTGAGGCTTTAGGTAACATTCCCGGCGACTAT
       AP011615.1   TTAAATGCGCGGGGAGGCTGCACATTAGGACCCGAAGCCGTAGATATTGACCTCATGGTTGAGGCTTTAGGTAACATTCCCGGCGACTAT
NZ_ACSK02000570.1   TTAAATGCGCGGGGAGGCTGCACATTAGGACCCGAAGCCGTAGATATTGACCTCATGGTTGAGGCTTTAGGTAACATTCCCGGCGACTAT
NZ_CAFN01000673.1   TTAAATGCGCGGGGAGGCTGTACATTGGGACCGGAAGCCATAGATGTTGACCTCATGGTTGAGGCTTTAGGTAACATTCCCGGCGACTAT
      AF371369.1    TTGAATATGCGTGGTGGTTGCACCCTGGGGGCGGAAGCGATCGATGTGGATTTGATGGTGGATGCCCTGGGCAATATTCCGGGTGATTTT
      NC_011729.1   TTAAATATGCGAGGAGGCTGTACTATAGGCGCAGAAGCCCTAGATGTGGATTTAATGATAGATTCAATGGGGAATGTGCCGGGAGATTAC
      NC_011884.1   TTGAATATGCGCGGTGGTTGCACCCTGGGGGCAGAAGCGATCGATGTGGATTTGATGGTGGATGCCCTGGGCAATATTCCGGGTGATTTT
      NC_014501.1   TTAAATATGCGGGGAGGATGTACCATCGGCAACGAGGCGCTTGATGTGGATCTGATGATCGAGGCGATGGGAAATGTGCCCGGGGATTAC
NZ_ALVY01000193.1   CTCAACATTCGTGGAGGCTGTAGTCTAGGTAAGGATGCTCTGGATGTAGATTTAATGGTGGATGCTTTGGGAAATATTCCCGGAGATTGG
      NC_019745.1   CTCAACGTTTGGAGTGGCTGTACGCTAGGCGCAAAAGCTCTAGATGTGGATCTAGCGATTGATACTTTGGGGAATGTTCCTGGCGACTGG
NZ_AFJC01000008.1   CTCAATGTTTGGGGTGGATGCACTCTGGGGTCAAAGGCTGTAGATATCGATTTAATGGTGGATACTCTGGGGAACATTCCCGGCGACTTC
NZ_AOCI01000120.1   TTAAATATGCGCGGGGGGTGTTCCTTGGGAGCCGAAGCATTAGACATCGATTTAATGGTAGATACTATGGGCAATATCCCGGGAGATTTT
      NC_010296.1   TTAAATATGCGCGGGGGGTGTTCCTTGGGAGCCGAAGCATTAGACATCGATTTAATGGTAGATACTATGGGCAATATCCCGGGAGATTTT
       AM778949.1   TTAAATATGCGCGGGGGGTGTTCCTTGGGAGCCGAAGCATTAGACATCGATTTAATGGTAGATACTATGGGCAATATCCCGGGAGATTTT
    NZ_HE973143.1   TTAAATATGCGCGGGGGGTGTTCCTTGGGAGCCGAAGCATTAGACATCGATTTAATGGTAGATACCATGGGCAATATTCCGGGAGATTTT
    NZ_HE972538.1   TTAAATATGCGCGGGGGGTGTTCCTTGGGAGCCGAAGCATTAGACATCGATTTAATGGTAGATACTATGGGCAATATCCCGGGAGATTTT
    NZ_HE973089.1   TTAAATATGCGCGGGGGATGTTCCCTTGGTTCCGAAGCGTTAGACATCGATTTAATGGTAGATACCATGGGCAATATCCCGGGAGATTTT
NZ_CAIQ01000501.1   TTAAATATGCGGGGGGGATGTTCCTTGGGTTCCGAAGCATTAGACATCGATTTAATGGTAGATGCTATGGGCAATATCCCGGGAGATTTT
    NZ_HE972766.1   TTAAATATGCGCGGGGGGTGTTCCTTGGGAGCCGAAGCATTAGACATCGATTTAATGGTAGATACTATGGGCAATATCCCGGGAGATTTT
    NZ_HE973252.1   TTAAATATGCGGGGGGGATGTTCCTTGGGTTCCGAAGCATTAGACATCGATTTAATGGTAGATGCTATGGGCAATATCCCGGGAGATTTT
    NZ_HE973368.1   TTAAATATGCGCGGGGGGTGTTCCCTGGGTTCCGAAGCGTTAGACATCGATTTAATGGTAGATACCATGGGCAATATCCCGGGAGATTTT
    NZ_HE973582.1   TTAAATATGCGCGGGGGGTGTTCCTTGGGAGCCGAAGCATTAGACATCGATTTAATGGTAGATACTATGGGCAATATCCCGGGAGATTTT
    NZ_HE973750.1   TTAAATATGCGCGGGGGGTGTTCCTTAGGAGCCGAAGCATTAGACATCGATTTAATGGTAGATACTATGGGCAATATCCCGGGAGATTTT
NZ_ANKQ01000002.1   TTAAATATGCGCGGGGGGTGTTCCTTGGGAGCCGAAGCATTAGACATCGATTTAATGGTAGATACCATGGGCAATATTCCGGGAGATTTT
NZ_CAIP01000427.1   TTAAATATGCGCGGGGGGTGTTCCTTGGGGTCCGAAGCATTAGACATCGATTTAATGGTAGATGCTATGGGCAATATCCCGGGAGATTTT
      NC_019689.1   CTCAACATGCGCGGAGGCTGCACCCTTGGGGCAGAAGCTTTGGATGTGGATCTGATGGTCAAGAGCCTGGGCAATATTCCTGGCGATTTT
       AY030295.1   TTGTACATGCGCGGCGGCTGCACTCTCGGAGCGGAAGCTTTAGATATTGATTTGATGGTAGATTGTTTGGGCAATATTCCTGGCGATTTC
       CP003265.1   TTGAACGCCCGGGGAGGCTGTACCTTGGGAGCCGAAGCAGTAGATATTGACTTAATGGTGGATGCCATGGGCAATATTCCAGGGGATTAT
                    tTaaAtatgcGcgGgGGcTGtacctTgGGagccgAaGCatTaGAtaTcGAttTaatGgTagAtactaTgGGcAAtaTtCCgGGaGAtTtt
```

```
                    L__N__    e__F__L__    L__K__P__    Q__L__G__    q__k__Y__l__d__f__    d__                  k__
   ABYK01000001.1   TTAAACATCGAGTTTCTGATGTTAAAACCCCTACAATTAGGATATCAAAAATATCTCGATTTACCCGAAATCATGGGAAGTCGCGACAAA
    NZ_CM001632.1   TTAAACATCGAGTTTCTGATGTTAAAACCCCTACAATTAGGATATCAAAAATATCTCGATTTACCCGAAATCATGGGAAGTCGCGACAAA
       AP011615.1   TTAAACATCGAGTTTTTGATGTTAAAACCCCTACAATTAGGATATCAAAAATATCTCGATTTACCCGAAATCATGGGAAGTCGCGACAAA
NZ_ACSK02000570.1   TTAAACATCGAGTTTTTGATGTTAAAACCCCTACAATTAGGATATCAAAAATATCTCGATTTACCCGAAATCATGGGAAGTCGCGACAAA
NZ_CAFN01000673.1   TTAAACATCGAGTTTCTGATGTTAAAACCCCTACAATTAGGATATCAAAAATATCTCGATTTACCCGAAATCATGGGAAGTCGCGACAAA
       AF371369.1   CTCAACCTGGAATTTCTGATGTTAAAACCGCAGCAGTTGGGCATTCAGGAATACCTGGATGTACCGGATCTGATGGACAGCCCGGAAAAA
      NC_011729.1   CTCAATTTAGAGTTTTTGATGCTCAAACCTTTACAATTGGGTTATCAAAAGTATCTTGATTTTCCAGATATTATGGAAAATGAAAGTAAA
      NC_011884.1   CTCAACCTGGAATTTCTGATGTTAAAACCGCAGCAGTTGGGCATTCAGAAATACCTGGATGTACCGGATCTGATGGACAGCCCGGAAAAA
      NC_014501.1   CTGAATTTAGAGTTTTTGATGCTGAAACCGTTACAGTTGGGCTATCAAAAGTATCTCGATTTTCCCGATATTATGGAAAATGAGGATAAG
NZ_ALVY01000193.1   CTCAATTGGGAATTTTTAATGCTTAAACCCTATCAACTAGGCATTCAGAAATACGTTGATTTTTTCAACATTATGGAAAATAAGGAACAA
      NC_019745.1   CTCAATTTTCAATTTCTGATGCTTAAACCTTTTCAATTAGGAGTTGAAAAGTATATTAAGTTTTTAGAAAGTAGTGATTCTGAGGAAAAA
NZ_AFJC01000008.1   CTGAATTTGGAGTTCTTGATGCTGAAGCCTTTTCAGTTAGGAGTTCAGAAGTATATTGACCTTCTGGAGAACATCGATTGTGAAAGCAAA
NZ_AOCI01000120.1   CTCAACTTAGAGTTTCTAGAATTGAAACCTTTGCAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
      NC_010296.1   CTCAACTTAGAATTTCTGGAATTGAAACCTTTGCAGTTAGGTTATCAGAAATACCTTGATTTTCCTGACATCATGGAAGACGAATCAAAA
       AM778949.1   CTCAACTTAGAGTTTCTAGAATTGAAACCTTTGCAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
    NZ_HE973143.1   CTCAACTTAGAGTTTCTGGAATTGAAACCTTTACAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
    NZ_HE972538.1   CTCAACTTAGAGTTTCTGGAATTGAAACCTTTGCAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
    NZ_HE973089.1   CTCAACTTAGAGTTTCTGGAATTGAAACCTTTACAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
NZ_CAIQ01000501.1   CTCAACTTAGAGTTTCTGGAATTGAAACCTTTACAGTTAGGTTATCAGAAATACCTTGATTTTCCTGACATCATGGAAGACGAATCAAAA
    NZ_HE972766.1   CTCAACTTAGAGTTTCTGGAATTGAAACCTTTACAGTTAGGTTATCAGAAATACCTTGATTTTCCTGACATCATGGAAGACGAATCAAAA
    NZ_HE973252.1   CTTAACTTAGAGTTTCTGGAATTGAAACCTTTACAGTTAGGTTATCAGAAATACCTTGATTTTCCTGACATCATGGAAGACGAATCAAAA
    NZ_HE973368.1   CTCAACTTAGAGTTTCTGGAATTGAAACCTTTGCAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
    NZ_HE973582.1   CTCAACTTAGAGTTTCTGGAATTGAAACCTTTGCAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
    NZ_HE973750.1   CTCAACTTAGAGTTTCTGGAATTGAAACCTTTGCAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
NZ_ANKQ01000002.1   CTCAACTTAGAGTTTCTGGAATTGAAACCTTTGCAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
NZ_CAIP01000427.1   CTTAACTTAGAGTTTCTGGAATTGAAACCTTTGCAGTTAGGTTATCAGAAATACCTCGATTTTCCTGACATCATGGAAGACGAATCAAAA
      NC_019689.1   CTTAACCTTGAGTTTTTGATGCTCAAACCCCAGCAGTTAGGAATTCAAAAATACCTTGACTTTCCAGAAGTCATGACCAGCGAAGACAAG
       AY030295.1   CTCAATTTTGAGTTTTTAATGCTCAAACCCCGACAACTAGGAATTCAAAAATATCTAGACTTTCCCGAGATCATGCACAGCGAAGACAAG
       CP003265.1   CTTAACCTAGAATTTCTCATGCTTAAACCCCTGCAATTAGGTTACCAAAAGTATCTTGATGTGCCCGATATTATGGGGGATGAAGCGAAA
                    cTcAActtagAgTTtcTgatgtTgAAaCCtttaCAgtTaGGttatcAgaAaTAccTcgAttTtcctgAcatcAtggaagacgaatcaaAa
```


l N F l R M E K W I F D S P d a G E s y R Q f l K D F Y
ABYK01000001.1 CTGCTAAACTTCCTCCGCATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCCGGAGAAACCTATCGCCAATTCCTGAAAGATTTTTAT
NZ_CM001632.1 CTGCTAAACTTCCTCCGCATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCCGGAGAAACCTATCGCCAATTCCTGAAAGATTTTTAT
AP011615.1 CTCCTAAACTTCCTCCGCATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCCGGAGAAACCTATCGCCAATTCCTGAAAGATTTTTAT
NZ_ACSK02000570.1 CTCCTAAACTTCCTCCGCATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCCGGAGAAACCTATCGCCAATTCCTGAAAGATTTTTAT
NZ_CAFN01000673.1 CTGCTAAACTTCCTCCGCATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCCGGAGAAACCTATCGCCAATTCCTGAAAGATTTTTAT
AF371369.1 CTACTCAACTTCCTCAGAATGGAAAAGTGGATTTTCGACAGTCCCGATCAGGCTGGAGAAACCTATCGCCAGTTCATGAAGGACTTTTAT
NC_011729.1 CTGGCGAATTTTATGCGGATGGAAAAATGGATTTTTGATAGTCCGGATCAAGCAGGAGAAGCTTACCGTCAGTTTATGAAAGATTTTTAT
NC_011884.1 CTGCTCAACTTTCTCAGAATGGAAAAGTGGATTTTCGACAGTCCCGACCAGGCAGGAGAAACCTATCGCCAGTTCATGAAGGACTTTTAT
NC_014501.1 CTCACTAATTTTATGCGGATGGAAAAGTGGATCTTTGATAGTCCGGATCAAGCGGGGGAAGCTTATCGCCAGTTTATGAAAGACTTTTAT
NZ_ALVY01000193.1 ATGCTCAACTTTTTACGGATGGAAAAATGGATCTTCGACAGCCCAGAACAGGTAGGGGAAGCTTACCGCCAGTTTCTCAAGGACTTTTAT
NC_019745.1 ATTATCAACTTCTTCCGCATGGAAAAGTGGATTTTCGATAGTCCCGATCTAGCTGGTGAAGCTTTTCGACAATATATGAAAGACTTTTAT
NZ_AFJC01000008.1 CTAATCAATTTTCTCCGGATGGAAAAGTGGATTTTTGATAGTCCCGACCAAGCTGGAGAGGCTTACCGACAGTTCATGAAGGATTTCTAT
NZ_AOCI01000120.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCGGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
NC_010296.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCGGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
AM778949.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCGGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
NZ_HE973143.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCAGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
NZ_HE972538.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCAGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
NZ_HE973089.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCAGGAGAATCCTACCGACAGTTCCTCAAGGATTTCTAT
NZ_CAIQ01000501.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCTGACCAAGCGGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
NZ_HE972766.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCGGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
NZ_HE973252.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCGGGAGAATCCTATCGACAGTTTCTCAAGGATTTCTAT
NZ_HE973368.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCAGGAGAATCCTACCGACAGTTCCTCAAGGATTTCTAT
NZ_HE973582.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCAGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
NZ_HE973750.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCGGGAGAATCCTATCGACAGTTTCTCAAGGATTTCTAT
NZ_ANKQ01000002.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCAGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
NZ_CAIP01000427.1 TTAGTTAATTTTCTGCGTATGGAAAAATGGATTTTTGATAGTCCCGACCAAGCGGGAGAATCCTACCGACAGTTTCTCAAGGATTTCTAT
NC_019689.1 CTATTAAACTTTATGCGGATGGAAAAGTGGATCTTTGACAGCCCCGATCAGGCAGGGGAGGCATACCGACAGTTCATGAAGGATTTTTAT
AY030295.1 TTGCTGAACTTTTTGCGCATGGAAAAATGGATCTTTGATAGTCCAGATCAAGCAGGGGAAGCTTATCGCCAGTTCCTTAAGGATTTCTAT
CP003265.1 TTGTTAAACTTTCTACGCATGGAAAAATGGATTTTTGATAGTCCCGATCAAGCGGGGGAAACTTACCGTCAATTCCTCAAGGATTTTTAT
tTagttAAtTTtcTgcGtATGGAAAAaTGGATtTTtGAtAGtCCcGAcCaaGcaGGaGAatCcTacCGaCAgTttcTcAAgGAtTTcTAT

```
                    Q__   N__K__L__I__k__   e__v__   l__G__         V__d__L__   n__l__   m__P__i__l__N__l__Y__
   ABYK01000001.1   CAGGAAAACAAACTAATCAAAGGCGAAGTAATGATTGGTGATTCTCGGGTAGATTTAAGCAATATTACCATGCCAGTTCTCAACCTCTAC
    NZ_CM001632.1   CAGGAAAACAAACTAATCAAAGGCGAAGTAATGATTGGTGATTCTCGGGTAGATTTAAGCAATATTACCATGCCAGTTCTCAACCTCTAC
       AP011615.1   CAGGAAAACAAACTAATCAAAGGCGAAGTAATGATTGGTGATTCTCGGGTAGATTTAAGTAATATTACCATGCCAGTTCTCAACCTCTAC
NZ_ACSK02000570.1   CAGGAAAACAAACTAATCAAAGGCGAAGTAATGATTGGTGATTCTCGGGTAGATTTAAGTAATATTACCATGCCAGTTCTCAACCTCTAC
NZ_CAFN01000673.1   CAGGAAAACAAACTAATCAAAGGCGAAGTAATGATTGGTGATTCTCGGGTAGATTTAAGCAATATTACCATGCCAGTTCTCAACCTCTAC
       AF371369.1   CAGGGCAATAAGCTGATCAAAAACCAGGTCAAAATTGGGGATCGGCAGGTAAATCTACTCAATCTGACCATGCCGATTCTCAACCTCTAT
      NC_011729.1   CAAAGTAATAAATTGATTAAAAATGAGGTGGTTATTGGCAATAAACCGGTTAATTTACAAAATCTAACGATGCCGATTTTAAACCTTTAC
      NC_011884.1   CAGGGCAATAAGCTGATCAAAAACCAGGTCAAGATTGGGGATCAGCTAGTAAATCTACTCAATCTGACCATGCCGATTCTTAACCTCTAT
      NC_014501.1   CAAGGCAATAAGCTGATTAAAAATGAGGTGGTGATCGGCGATCAACGAGTAAATTTACAAAACTTAACCATGCCGATTTTAAACCTTTAT
NZ_ALVY01000193.1   CAGGAAAACAAACTAATTCACAACGAGATCCAAATTGGCGATAAACGAGTGGATTTGGGACAGATACTTATGCCAGTGCTAAACTTATAC
      NC_019745.1   CAAGAAAATAAACTTATAAAAGGTCAACTGGAGATAGGAGGAAAACGAGTACATTTAGAGAAGATTCGTATTCCAATTTTTAATATATAT
NZ_AFJC01000008.1   CAGGGAAACAAACTGATTCAAGGTCAGGTCGAGATTGGCAACAAGCGAGTGGATCTGGGAAACATCCGCATCCCGATTTTGAACATTTAC
NZ_AOCI01000120.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACCATGCCAATTCTCAATCTTTAC
      NC_010296.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACCATGCCAATTCTCAATCTTTAT
       AM778949.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACCATGCCAATTCTCAATCTTTAC
    NZ_HE973143.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACCATGCCAATTCTCAATCTTTAC
    NZ_HE972538.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACTATGCCAATTCTCAATCTTTAC
    NZ_HE973089.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACTATGCCAATTCTCAATCTTTAT
NZ_CAIQ01000501.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACCATGCCAATTCTCAATCTTTAT
    NZ_HE972766.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACCATGCCAATTCTCAATCTTTAT
    NZ_HE973252.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACTATGCCAATTCTCAATCTTTAT
    NZ_HE973368.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACTATGCCAATTCTCAATCTTTAT
    NZ_HE973582.1   CAGCAAAATAAACTGATTAAAGGAGAAGTAATGTTAGGAGATAAACGGGTAGATTTACATAATCTGACCATGCCAATTCTCAATCTTTAT
    NZ_HE973750.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACTATGCCAATTCTCAATCTTTAT
NZ_ANKQ01000002.1   CAGCAAAATAAACTGATTAAAGGAGAAGTAATGTTAGGAGATAAACGGGTAGATTTACATAATCTGACTATGCCAATTCTCAATCTTTAC
NZ_CAIP01000427.1   CAGCAAAATAAACTGATTAAAGGGGAAGTGATGTTAGGAGATAAACGGGTAGATTTACACAATCTGACCATGCCAATTCTCAATCTTTAT
      NC_019689.1   CAGGAAAATAAACTCATTAAAGGGGAAGTGATGCTCGGCGATAAGCGAGTGGATCTAAAAAACGTGCGGATGCCAGTCTTGAACCTTTAC
       AY030295.1   CAAGCAAATAAACTTATCAAGGGAGAGGTAACCATCGGAGATAAACAAGTCAATTTAGGTAACATTCGCATGCCTGTACTGAATCTTTAC
       CP003265.1   CAACAAAATAAATTGATCAAAGGGGAAGTGATGATTGGCGATCGCCTGGTGGATCTGCATAATTTGACCATGCCCATATTGAATTTATAT
                    CAgcaaAAtAAacTgATtaAaggggAagTgatgaTaGGagataaaCggGTagATtTacacaAtcTgaccATgCCaaTtcTcAAtcTtTAc
```

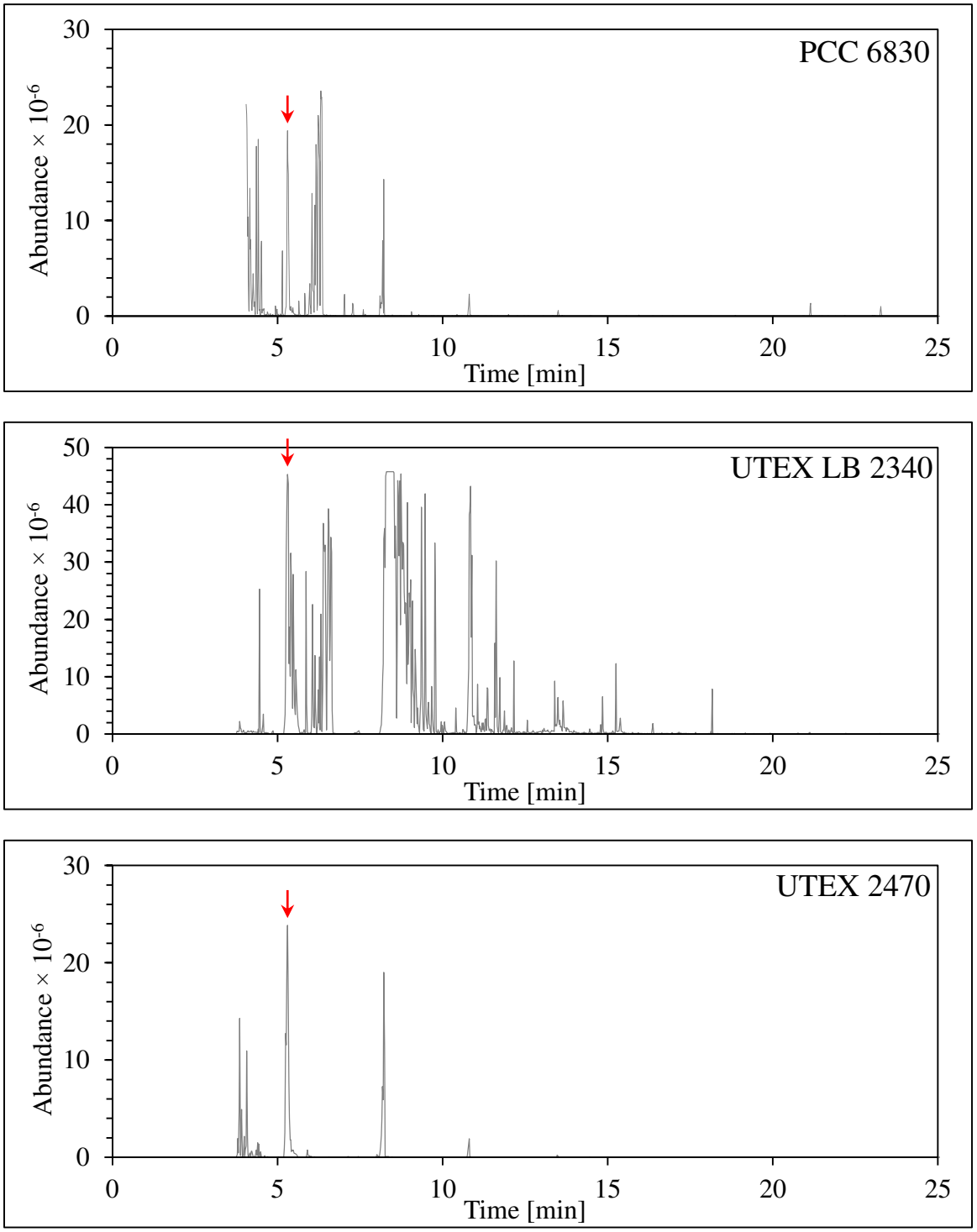
```
                  A__e__    D__H__L__V__    P__    S__S__    A__L__       Y__i__       t__    d__Y__          a__F__
   ABYK01000001.1 GCCGAGAAAGATCACCTAGTCCCCCCTTCCTCTTCCCTAGCCTTAGAGGAATACATCAGC---AGTGAGGACTACACCGCCAAATCCTTC
    NZ_CM001632.1 GCCGAGAAAGATCACCTAGTCCCCCCTTCCTCTTCCCTAGCCTTAGAGGAATACATCAGC---AGTGAGGACTACACCGCCAAATCCTTC
       AP011615.1 GCCGAAAAAGATCACCTAGTCCCCCCTTCCTCTTCCCTCGCACTAGAGGAATACATCAGC---AGTGAAGACTACACCGCCAAATCCTTC
NZ_ACSK02000570.1 GCCGAAAAAGATCACCTAGTCCCCCCTTCCTCTTCCCTCGCACTAGAGGAATACATCAGC---AGTGAAGACTACACCGCCAAATCCTTC
NZ_CAFN01000673.1 GCCGAGAAAGATCACCTAGTCCCCCCTTCCTCTTCCCTAGCCTTAGAGGAATACATCAGC---AGTGAGGACTACACCGCCAAATCCTTC
       AF371369.1 GCTGAGAAAGACCATCTCGTTCCCCCCGCTTCTTCCCTGGCTCTAGCCAAATACATCGAC---ACCCAGGATTACACGGCTAAAGGCTTC
      NC_011729.1 GCAGAATTAGATCATTTAGTCGATCCGGCTTCGTCTAAAGCCTTAGAAAAATACGTTAAT---ACGACGGATTATATAGTTCAGTCTTTC
      NC_011884.1 GCTGAGAAAGACCATCTCGTTCCCCCCGCTTCTTCCGTGGCTCTAGCCAAGTACATCGGC---ACCCAGGATTACACGGCTAAAGGCTTC
      NC_014501.1 GCTGAACAGGATCATTTAGTTGATCCCGTGTCTTCTAAGGCTTTAGAAAAATATGTCAAC---AGCAGTGATTATACGCTTAAGTCTTTC
NZ_ALVY01000193.1 GCCGAAAAAGATCATCTGGTACCACCGTTATCTTCCTTAGCCCTAGAAAAATATGTGGGT---ACTCAAGATTATACAACGCAATCGTTC
      NC_019745.1 GCCGAGCAGGATCATTTAGTACCTCCTGCATCTTCGTTAGCGTTAGAGAAGTACGTTGCT---AGCAGCGAGTACACGGTACGTTCTTTT
NZ_AFJC01000008.1 GCCGAGCAAGATCATCTCGTTGCTCCAGCCTCTTCCTTAGCTCTTAAGACATACATTGCT---AGCGAAGACTACACCTTGCGCTCCTTC
NZ_AOCI01000120.1 GCAGATAAAGATCACCTTGTGCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
      NC_010296.1 GCGGATAAAGATCACCTTGTCCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
       AM778949.1 GCAGATAAAGATCACCTTGTGCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
    NZ_HE973143.1 GCAGATAAAGATCACCTTGTGCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
    NZ_HE972538.1 GCAGATAAAGATCACCTTGTGCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
    NZ_HE973089.1 GCGGATAAAGATCACCTTGTCCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
NZ_CAIQ01000501.1 GCGGATAAAGATCACCTTGTACCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
    NZ_HE972766.1 GCGGATAAAGATCACCTTGTACCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
    NZ_HE973252.1 GCGGATAAAGATCACCTTGTACCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCCTTC
    NZ_HE973368.1 GCGGATAAAGATCACCTTGTCCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
    NZ_HE973582.1 GCGGATAAAGATCACCTTGTCCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
    NZ_HE973750.1 GCGGATAAAGATCACCTTGTCCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
NZ_ANKQ01000002.1 GCAGATAAAGATCATCTTGTCCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
NZ_CAIP01000427.1 GCGGATAAAGATCACCTTGTCCCTCCCGCTTCTTCCCTCGCTTTAGGGAATTATATCGGT---ACTTCTGACTATACCGCTTGTGCTTTC
      NC_019689.1 GCCGAAAAGGATCATCTGGTAGATCCCGAATCTTCTAAAGCACTAGAAAAATACGTGGGA---ACAGACGATTATACGGTGCGCTCCTTC
       AY030295.1 GCAGAAAAGGATCATTTGGTACCGCCTCGGTCTTCTATCGCCCTGGAAAGGTACATTGGT---ACAACCGATTATACTGTGCGCTCTTTT
       CP003265.1 GCGGAAAAAGACCACTTGGTGGCCCCTGCTTCTTCCCTAGCTTTGGGGGACTATTTGCCGGAAAACTGTGACTACACCGTCCAATCTTTC
                  GCgGAtaaaGAtCAccTtGTccctCCcgctTCtTCcctcGCttTagggaatTAtaTcggt---ActtctGAcTAtAccgcttgtgctTTc
```

```
                    P__V__G__H__I__G__m__Y__V__S__    K__V__Q__       L__P__       I__    d__W__l__
   ABYK01000001.1   CCTGTAGGTCATATCGGTATGTATGTCAGTGGTAAAGTACAGCGAGACCTACCCCCAACCATTGTTGATTGGTTAAAAGTGCGAGAG---
    NZ_CM001632.1   CCTGTAGGTCATATCGGTATGTATGTCAGTGGTAAAGTACAGCGAGACCTACCCCCAACCATTGTTGATTGGTTAAAAGTGCGAGAG---
       AP011615.1   CCCGTAGGTCATATCGGTATGTATGTCAGTGGTAAAGTACAGCGAGATTTACCCCCAACCATTGTCGATTGGTTAAAAGTGCGAGAG---
NZ_ACSK02000570.1   CCCGTAGGTCATATCGGTATGTATGTCAGTGGTAAAGTACAGCGAGATTTACCCCCAACCATTGTCGATTGGTTAAAAGTGCGAGAG---
NZ_CAFN01000673.1   CCTGTAGGTCATATCGGTATGTATGTCAGTGGTAAAGTACAGCGAGACCTACCCCCAACCATTGTTGATTGGTTAAAAGTGCGAGAG---
       AF371369.1   CCCGTGGGACACATCGGTATGTACGTTAGCGGCAAGGTGCAACGAGATCTGCCCCCGGTGATTGCGGACTGGCTCAGGAATCGGGAT---
      NC_011729.1   CCAGTTGGACATATTGGAATGTATGTCAGTGGAAAGGTACAAGCCACTTTACCGCCGACAATTGTAGAATGGCTCACCGCTAGAGCT---
      NC_011884.1   CCGGTAGGACACATCGGTATGTACGTCAGTGGCAAGGTGCAACAAGATCTGCCCCCGGTGATTGCGGACTGGCTCAGGAACCGGGAT---
      NC_014501.1   CCGGTGGGACATATCGGAATGTATGTGAGTGGGAAAGTTCAAAAGGATTTACCGCCAACCATTGTAGATTGGCTTAAAGCTAGGTCT---
NZ_ALVY01000193.1   CCTGTAGGACATATCGGAATGTATGTCAGTAGTAAGGTACAGCGAGACTTACCACAGATAATCGTAAATTGGATTAAGGCGCGATCGCTA
      NC_019745.1   CCGGTTGGGCATATTGGTATGTATGTGAGTCGTAAAGTTCAAAAGGATCTACCTGAGGCGATCGCTGATTGGTTGAAG------------
NZ_AFJC01000008.1   CCAGTGGGGCACATCGGGATGTACGTCAGCAGTAAAGTGCAGCGAGACCTTCCTCCAACTATTGTCGATTGGCTTAAGATGCGGGCG---
NZ_AOCI01000120.1   CCCGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGGT---
      NC_010296.1   CCCGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGCT---
       AM778949.1   CCCGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGGT---
    NZ_HE973143.1   CCGGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGCT---
    NZ_HE972538.1   CCGGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGCT---
    NZ_HE973089.1   CCCGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTAAAAGCAAGAGCT---
NZ_CAIQ01000501.1   CCCGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGGT---
    NZ_HE972766.1   CCGGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTACTGATTGGTTGAAAGCAAGAGCT---
    NZ_HE973252.1   CCGGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTACTGATTGGTTAAAAGCAAGAGGT---
    NZ_HE973368.1   CCCGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGGT---
    NZ_HE973582.1   CCGGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGCT---
    NZ_HE973750.1   CCCGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGCT---
NZ_ANKQ01000002.1   CCGGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGCT---
NZ_CAIP01000427.1   CCCGTCGGACATATCGGAATGTATGTCAGTGGCAAAGTGCAACGGGATTTACCCCCCGCTATTAGCGATTGGTTGAAAGCAAGAGCT---
      NC_019689.1   CCAGTCGGTCATATCGGCATGTATGTCAGCGGCAAGGTACAGCGAGATTTGCCGCCTACAATCGTCGATTGGTTAAAGGCACGGATG---
       AY030295.1   CCTGTCGGTCACATTGGCATATATGTCAGCAGTAAAGTACAGCGAGATTTACCACCTATAATTGCAAACTGGTTGAATGCGCGCGAA---
       CP003265.1   CCCGTGGGTCATATTGGCATGTATGTCAGTGGTAAAGTACAACGGGATCTGCCCCCGGCGATCGCCCATTGGCTATCGGAACGACAG---
                    CCcGTcGGaCAtATcGGaATgTAtGTcAGtgGcAAaGTgCAacgggattTaCCccccgctATtggcgAtTGGtTgaaa------------
```

```
   ABYK01000001.1     ---TAA
    NZ_CM001632.1     ---TAA
       AP011615.1     ---TAA
NZ_ACSK02000570.1     ---TAA
NZ_CAFN01000673.1     ---TAA
       AF371369.1     ---TAA
      NC_011729.1     ---TAA
      NC_011884.1     ---TAA
      NC_014501.1     ---TAA
NZ_ALVY01000193.1     AATTAA
      NC_019745.1     ---TAA
NZ_AFJC01000008.1     ---TAG
NZ_AOCI01000120.1     ---TAA
      NC_010296.1     ---TAA
       AM778949.1     ---TAA
    NZ_HE973143.1     ---TAA
    NZ_HE972538.1     ---TAA
    NZ_HE973089.1     ---TAA
NZ_CAIQ01000501.1     ---TAA
    NZ_HE972766.1     ---TAA
    NZ_HE973252.1     ---TAA
    NZ_HE973368.1     ---TAA
    NZ_HE973582.1     ---TAA
    NZ_HE973750.1     ---TAA
NZ_ANKQ01000002.1     ---TAA
NZ_CAIP01000427.1     ---TAA
      NC_019689.1     ---TGA
       AY030295.1     ---TAA
       CP003265.1     ---TGA
                      ---Taa
```

A.9 Cyanobacteria PHA GC-MS Data

Appendix Figure 3: GC-MS total ion chromatograms of cells grown under nutrient limiting conditions to induce PHA accumulation. Cells were then acetone dried and subjected to methanolysis. The hydroxyalkanoate methylester content was assessed as described in Tan *et al.* Journal of Bioscience and Bioengineering. 2014;117:379-82. Hydroxybutanoate methyl esters highlighted.



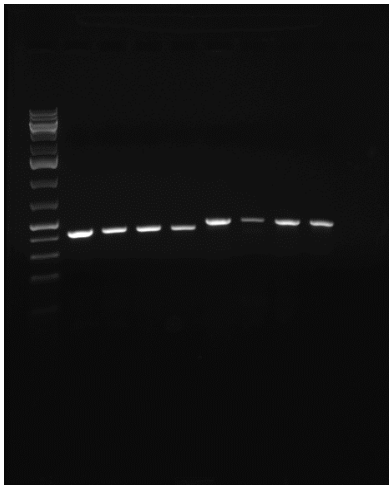
A.10 Electrophoresis Images of *phaC* PCR Detection

Appendix Figure 4: Agarose gel electrophoresis images of *cpcA* and *phaC* colony PCR unpurified/raw reactions. One percent standard agarose, 0.5x TBE running buffer, 0.5 µg ml^{-1} ethidium bromide, Thermo Scientific OWL Easycast B1A, 120V, 40 min. Images are by strain, lanes (from left to right) are as follows for each image:

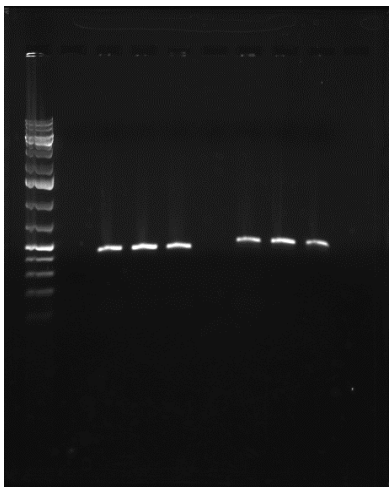
1. 5 µl GeneRuler 1kb Plus (Life Technologies, #SM1333)
2. 10 µl cpcA amplification of whole cells
3. 10 µl cpcA amplification of 1 µl DMSO/TE clarified lysate
4. 10 µl cpcA amplification of 5 µl DMSO/TE clarified lysate
5. 10 µl cpcA amplification of 10 µl DMSO/TE clarified lysate
6. 10 µl phaC(3.1) amplification of whole cells
7. 10 µl phaC(3.1) amplification of 1 µl DMSO/TE clarified lysate
8. 10 µl phaC(3.1) amplification of 5 µl DMSO/TE clarified lysate
9. 10 µl phaC(3.1) amplification of 10 µl DMSO/TE clarified lysate
10. EMPTY

Note: For proper quantitation, exposure times for images were adjusted to just above saturation for the strongest signal. For qualitative mass comparisons across strains observe the intensity relative to the standard in each case.

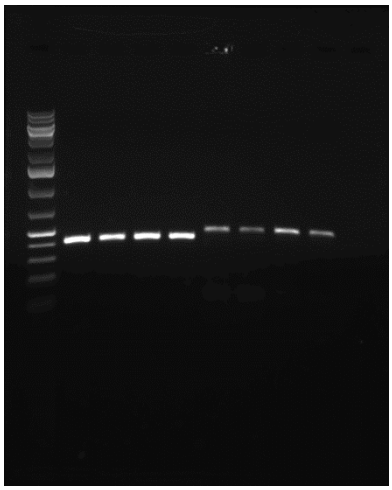
Arthrospira maxima UTEX LB2342 (CS-328)



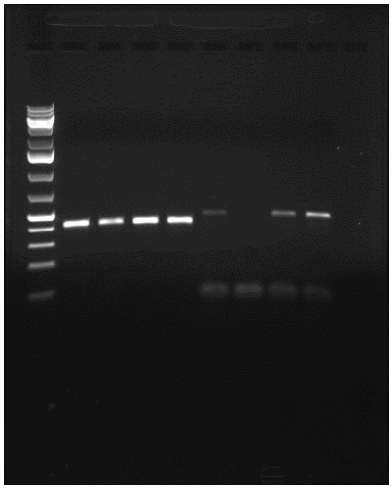
Arthrospira platensis UTEX LB2340



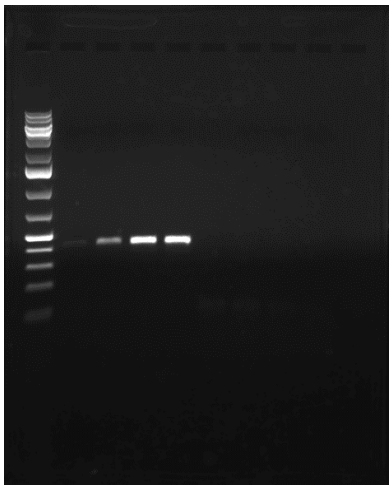
Microcoleus vaginatus PCC 9802 (FGP-2)



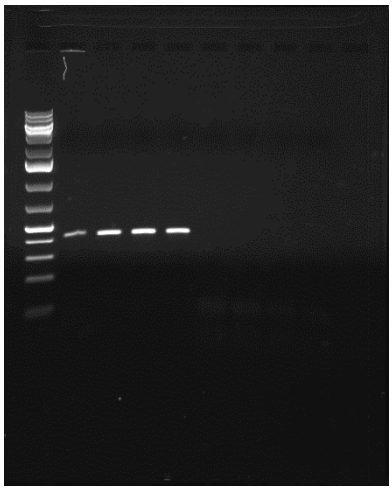
Microcystis aeruginosa NIES-843



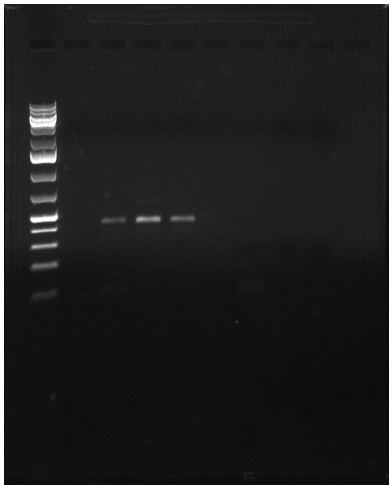
Plectonema sp. UTEX 1541



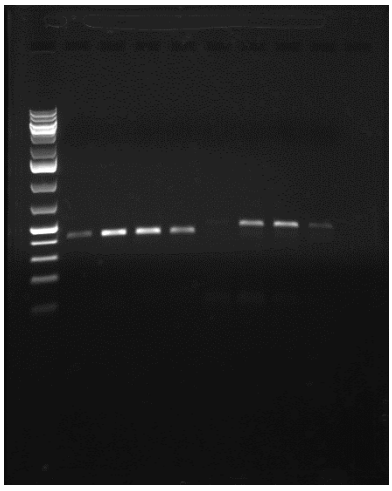
Synechococcus elongatus PCC 7942



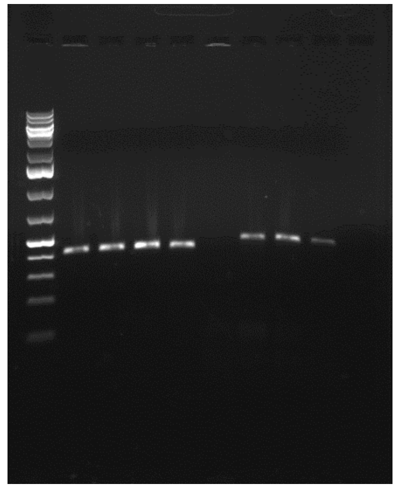
Synechococcus leopoliensis UTEX 2434



Synechocystis sp. UTEX 2470



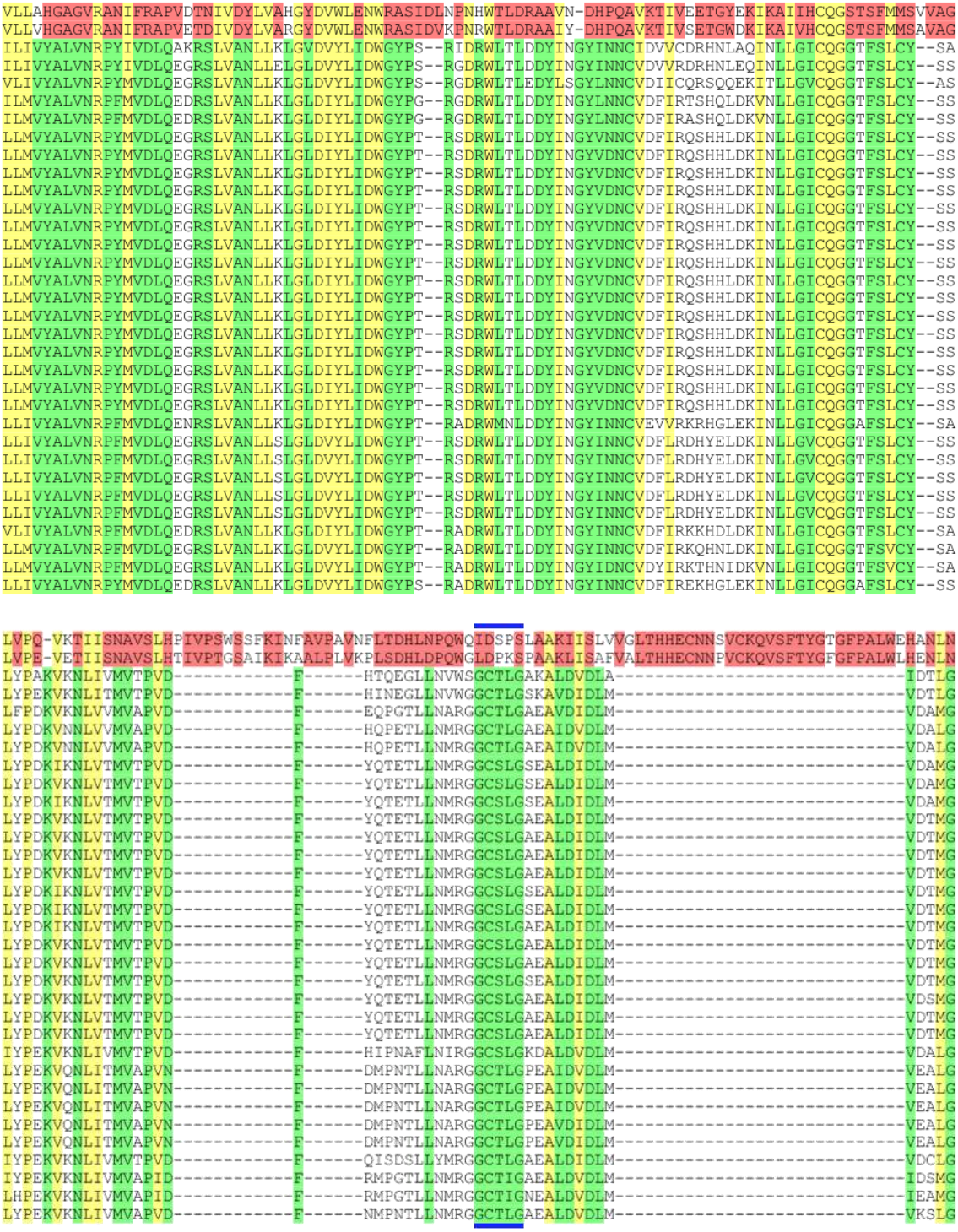
Synechocystis sp. PCC 6803

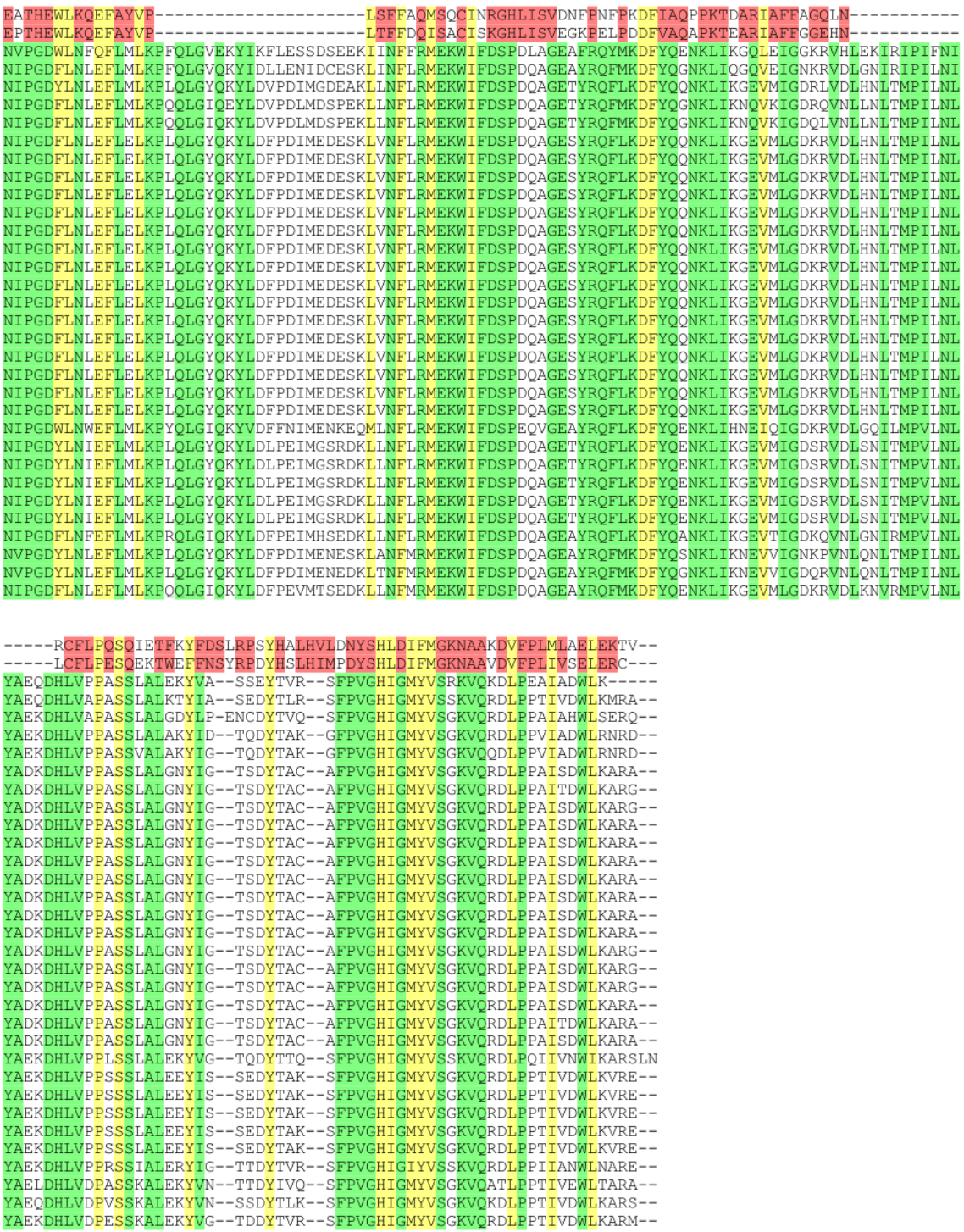


A.11 PHA Synthase Protein MSA (33 sequences)

Appendix Figure 5: Multiple sequence alignment of the PHA synthase PhaC subunit of 33 cyanobacteria. Red color denotes conservation between *Nostoc punctiforme* PCC 73102 and *Xenococcus* sp. PCC 7305 accessed sequences. Green color denotes conservation between all other sequences. Yellow color depicts conservation between groups. Blue bars depict “Cyanobacterial-box” region.

```
---------MNTTTYTQSLQERIVPFTSKDGF-----------------------------E-----------------CNLINVRGEKPPSKSP
---------MLEVNPNKNIKETIFPFTTQDGI-----------------------------E-----------------CNLIRVQGANPPDKAP
-------------------------------------MLTDL-AELWQRGEVFSRLREEDNLIGVTPKEEIYREDKVVLYRFTPQVKN--LLNTP
------------------------------------MVTTELTQKLVKGVEIFTRLREEDIQVGVTPKEEVYREDKVLLYHFSPKVEH--SLNIP
---------MFLLFFIVHWLKIMLPFFAQVGLEENLHETLDFTEKFLSGLENLQGLNEDDIQVGFTPKEAVYQEDKVILYRFQPVVEN--PLPIP
----------------------MLPFLLQIHLEEATHESAQLTHKLVKGMENLSQLREEDIEVGATPREVVFQEDKVKLYRFKSPVDQKKTVKTP
----------------------MLPFLLQIHLEEAAHESAQLTHKLVKGMENLSQLREEDIEVGSTPREVVYQEDKVKLYRFKAPANQGKTVQTP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIECGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIECGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIECGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWSFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIECGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MWPFLTQVKLEDFTQDYLELTQKNLKGLDNLKRVKEEDIQCGVSEKEAVYREDKIILYHFKPVVEK--PFEIP
----------------------MLPFWTQISIEDTTCEYIELTKKLLKGIQNLRELRESDIEIGITPKEVIYQEDKMLLYRFKPMVEN--PLTIP
----------------------MLPFALQMGLEDLTQEYADLTEKIVHGMDNLSSLREEEIIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVP
----------------------MLPFALQMGLEDLTQEYADLTEKIVHGMDNLSSLREEEIIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVP
-------------MSLLGGHQAMLPFALQMGLEDLTQEYADLTEKIVHGMDNLSSLREEEIIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVP
----------------------MLPFALQMGLEDLTEEYADLTEKIVHGMDNLSSLREEEIIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVP
----------------------MLPFALQMGLEDLTEEYADLTEKIVHGMDNLSSLREEEIIVGVTPKEAVYQEDKVTLYRFEPKVKK--TLSVP
----------------------MLPFLMQMRLDDATEEYTELIKKIVKGIENLSRLREEDIEIGVTPKEAVYREEKLTLYHFQSTVQK--QLRTP
----------------------MLPFLDQIRLEDAVHEYTEITKKMIKGLDNLSRLREEDIESGVSPKEVVYQEDKVVLYRFKSQVEH--PLPIP
----------------------MLPFLDQIRLEDAVHEYTEITKKMLKGLDNLSRLREEDIQSGVSAKEAVYKEDKVILYRFTPQVAQ--PLHIP
MHFDSPTRKIHLAQLDKEETKDMLPFLTQIRLEDVSHEYTELTKKVLQGIENLSRLREEDIEIGATPKEAVYKEDKVILYRFKPMVEQ--PLSIP
```





A.12 Definition of Variables for *In Vitro* PHA Polymerization Models

- **C** – Co-enzyme A
- **E** – PHA synthase enzyme
  - Assembled PhaC·PhaE complex for Type III synthases
  - Assembled PhaC·PhaR complex for Type IV synthases
- **S** – Substrate (hydroxyalkyl-coenzyme-A)
- **ES** – Hydroxyalkyl-PHA synthase
- **E_2P_n** – PHA synthase dimer covalently bound to PHA polymer of length n
- **$E_2P_n^*$** – Lesser-active state of E_2P_n (deactivated or reduced activity)
- **[…]** – Transient quanta-based concentration (e.g., molar at time *t*)
- **[…]ø** - Initial concentration (i.e., […](t = 0))

A.13 PHA Synthase Dimerization Model Derivation

This section details the derivation of the dimerization model used to describe lag phase behavior in PHA synthase unprimed *in vitro* polymerization models. See section A.12 on page 206 for the definition of variables.

A.13 Proposed Mechanism

$$\left.\begin{array}{c} E + S \rightarrow ES + C \\ 2ES \rightleftharpoons (ES)_2 \rightarrow E_2P_2 \\ E_2P_n + S \rightarrow E_2P_{n+1} + C \end{array}\right\} \rightarrow \begin{array}{c} E + S \overset{k_1}{\rightarrow} \frac{1}{2}E_2P_2 + C \\ E_2P_n + S \overset{k_2}{\rightarrow} E_2P_{n+1} + C \end{array}$$

A.13 Species Balances

$$[E]_{\emptyset} = [E] + 2[E_2P_n] \qquad \textit{(Enzyme)}$$

$$[S]_{\emptyset} = [S] + [C] \qquad \textit{(Substrate)}$$

$$[C]_{\emptyset} = 0 \qquad \textit{(Byproduct)}$$

A.13 Defined Dimensionless Variables

$$X(t) = \frac{[S]_{\emptyset} - [S]}{[S]_{\emptyset}} \textit{ (Fractional Conversion)}$$

$$\theta(t) = \frac{[E]_{\emptyset} - [E]}{[E]_{\emptyset}} \textit{ (Fractional Dimerization)}$$

A.13 Rate of Measurable Product Formation

$$\frac{d[C]}{dt} = k_1[E][S] + k_2[E_2P_n][S] \quad \rightarrow \quad \frac{dX}{dt} = \left(k_1[E]_{\emptyset}(1-\theta) + \frac{k_2[E]_{\emptyset}}{2}\theta\right)(1-X)$$

A.13 Rate of Dimerization

$$\frac{d[E_2P_n]}{dt} = k_1[E][S] \quad \rightarrow \quad \frac{d\theta}{dt} = k_1[S]_{\emptyset}(1-\theta)(1-X)$$

A.13 Determination of Co-Dependence

$$\frac{dX}{dt} = \frac{dX}{d\theta}\frac{d\theta}{dt} \quad \rightarrow \quad \frac{dX}{d\theta} = \frac{[E]_{\emptyset}}{[S]_{\emptyset}}\left(1 + \frac{k_2}{2k_1}\left(\frac{\theta}{1-\theta}\right)\right)$$

$$X(\theta) = \frac{[E]_{\emptyset}}{[S]_{\emptyset}}\left(\left(1 - \frac{k_2}{2k_1}\right)\theta - \frac{k_2}{2k_1}\ln(1-\theta)\right)$$

A.13 Transient Fractional Dimerization

$$\frac{d\theta}{dt} = k_1[S]_\emptyset(1-\theta)(1-X) \cong k_1[S]_\emptyset(1-\theta) \quad \text{where } [E]_\emptyset \ll [S]_\emptyset$$

$$\boxed{\theta(t) = 1 - \exp[-k_1[S]_\emptyset t]}$$

This step of the analysis assumes that the initial enzyme concentration is much less than the initial substrate concentration. This assumption removes dependence of the fractional dimerization on the fractional conversion and allows for an analytical solution to this model. See *Determination of Co-Dependence* section above for dependence of fractional conversion on fractional dimerization.

A.13 Fractional Conversion Model

$$\frac{dX}{dt} = \left(k_1[E]_\emptyset\big(1-\theta(t)\big) + \frac{k_2[E]_\emptyset}{2}\theta(t)\right)(1-X)$$

$$\boxed{X(t) = 1 - \exp\left[-\frac{[E]_\emptyset}{[S]_\emptyset}\left(1 - \frac{k_2}{2k_1}\right)\theta(t) - \frac{k_2[E]_\emptyset}{2}t\right]}$$

A.14 PHA Synthase Activity Reduction Model Derivation

This section details the derivation of the activity reduction model used to describe biphasic behavior in PHA synthase primed *in vitro* polymerization models. See section A.12 on page 206 for the definition of variables.

A.14 Proposed Mechanism

The proposed mechanism assumes PHA synthase complexes have formed initially. In other words, this mechanism assumes the reaction has been completely primed (as was the case in the recorded data) using an oligo-CoA derivative prior to the polymerization reaction. First polymerization occurs (r_1), followed by an irreversible state change (r_2), finally polymerization continues in the new enzymatic state (r_3).

$$\boxed{\begin{gathered} E_2P_n + S \xrightarrow{k_1} E_2P_{n+1} + C \\ E_2P_n \xrightarrow{k_2} E_2P_n^* \\ E_2P_n^* + S \xrightarrow{k_3} E_2P_{n+1}^* + C \end{gathered}}$$

A.14 Species Balances

$$[E_2P_n]_\emptyset = \frac{[E]_\emptyset}{2} = [E_2P_n] + [E_2P_n^*] \qquad \textit{(Enzyme)}$$

$$[S]_\emptyset = [S] + [C] \qquad \textit{(Substrate)}$$

A.14 Defined Dimensionless Variables

$$X(t) = \frac{[S]_\emptyset - [S]}{[S]_\emptyset} \textit{ (Fractional Conversion)}$$

$$\theta(t) = \frac{[E_2P_n]_\emptyset - [E_2P_n]}{[E_2P_n]_\emptyset} \qquad \textit{(Fractional Reduction)}$$

A.14 Fractional Reduction Model

$$\frac{d[E_2P_n^*]}{dt} = k_2[E_2P_n] \quad \rightarrow \quad \frac{d\theta}{dt} = k_2(1-\theta)$$

$$\boxed{\theta(t) = 1 - \exp[-k_2 t]}$$

A.14 Fractional Conversion Model

$$\frac{d[C]}{dt} = k_1[E_2P_n][S] + k_3[E_2P_n^*][S] \quad \rightarrow \quad \frac{dX}{dt} = [E_2P_n]_\emptyset(k_1(1-\theta) + k_3\theta)(1-X)$$

$$\boxed{X(t) = 1 - \exp\left[-[E_2P_n]_\emptyset\left(\frac{k_1 + k_3}{k_2}\theta(t) + k_3 t\right)\right]}$$

A.15 PHA Synthase Irreversible Deactivation Model Derivation

This section details the derivation of the irreversible deactivation model used to describe biphasic behavior in PHA synthase primed *in vitro* polymerization models. See section A.12 on page 206 for the definition of variables.

A.15 Proposed Mechanism

The proposed mechanism assumes PHA synthase complexes have formed initially. In other words, this mechanism assumes the reaction has been completely primed (as was the case in the recorded data) using an oligo-CoA derivative prior to the polymerization reaction. This model describes polymerization (r_1) with simultaneous and irreversible deactivation of the enzyme (r_2).

$$\boxed{\begin{gathered} E_2P_n + S \overset{k_1}{\rightarrow} E_2P_{n+1} + C \\ E_2P_n \overset{k_2}{\rightarrow} E_2P_n^* \end{gathered}}$$

A.15 Species Balances

$$[E_2P_n]_\emptyset = \frac{[E]_\emptyset}{2} = [E_2P_n] + [E_2P_n^*] \qquad \textit{(Enzyme)}$$

$$[S]_\emptyset = [S] + [C] \qquad \textit{(Substrate)}$$

A.15 Defined Dimensionless Variables

$$X(t) = \frac{[S]_\emptyset - [S]}{[S]_\emptyset} \textit{ (Fractional Conversion)}$$

$$\theta(t) = \frac{[E_2P_n]_\emptyset - [E_2P_n]}{[E_2P_n]_\emptyset} \qquad \textit{(Fractional Deactivation)}$$

A.15 Fractional Deactivation Model

$$\frac{d[E_2P_n^*]}{dt} = k_2[E_2P_n] \quad \rightarrow \quad \frac{d\theta}{dt} = k_2(1-\theta)$$

$$\boxed{\theta(t) = 1 - \exp[-k_2 t]}$$

A.15 Fractional Conversion Model

$$\frac{d[C]}{dt} = k_1[E_2P_n][S] \quad \rightarrow \quad \frac{dX}{dt} = k_1[E_2P_n]_\emptyset(1-\theta)(1-X)$$

$$\boxed{X(t) = 1 - \exp\left[-\frac{k_1[E_2P_n]_\emptyset}{k_2}\theta(t)\right]}$$

A.16 PHA Synthase Reversible Deactivation Model Derivation

This section details the derivation of the reversible deactivation model used to describe biphasic behavior in PHA synthase primed *in vitro* polymerization models. See section A.12 on page 206 for the definition of variables.

A.16 Proposed Mechanism

The proposed mechanism assumes PHA synthase complexes have formed initially. In other words, this mechanism assumes the reaction has been completely primed (as was the case in the recorded data) using an oligo-CoA derivative prior to the polymerization reaction. This model describes polymerization (r_1), accounting for reversible deactivation of the enzyme (r_2).

$$\boxed{\begin{gathered} E_2P_n + S \overset{k_1}{\rightarrow} E_2P_{n+1} + C \\ E_2P_n \overset{k_2}{\leftrightarrow} E_2P_n^* \end{gathered}}$$

A.16 Species Balances

$$[E_2P_n]_\emptyset = \frac{[E]_\emptyset}{2} = [E_2P_n] + [E_2P_n^*] \qquad \textit{(Enzyme)}$$

$$[S]_\emptyset = [S] + [C] \qquad \textit{(Substrate)}$$

A.16 Defined Dimensionless Variables

$$X(t) = \frac{[S]_\emptyset - [S]}{[S]_\emptyset} \textit{ (Fractional Conversion)}$$

$$\theta(t) = \frac{[E_2P_n]_\emptyset - [E_2P_n]}{[E_2P_n]_\emptyset} \qquad \textit{(Fractional Deactivation)}$$

A.16 Fractional Deactivation Model

$$\frac{d[E_2P_n^*]}{dt} = k_2[E_2P_n] - k_{-2}[E_2P_n^*] \quad \rightarrow \quad \frac{d\theta}{dt} = k_2(1-\theta) - k_{-2}\theta$$

$$\boxed{\theta(t) = \frac{k_2}{k_2 + k_{-2}}(1 - \exp[-(k_2 + k_{-2})t])}$$

A.16 Fractional Conversion Model

$$\frac{d[C]}{dt} = k_1[E_2P_n][S] \quad \rightarrow \quad \frac{dX}{dt} = k_1[E_2P_n]_{\emptyset}(1-\theta)(1-X)$$

$$\boxed{X(t) = 1 - \exp\left[-k_1[E_2P_n]_{\emptyset}\left(\frac{\theta(t)}{k_2 + k_{-2}} + \left(1 - \frac{k_2}{k_2 + k_{-2}}\right)t\right)\right]}$$

VITA

Courtney Edward (Eddie) Lane graduated from Fontainebleau High School in 2005. He graduated from Louisiana State University in 2010 with a bachelor's degree in chemical engineering and was fortunate enough to be a member of their competitive cheerleading squad during his bachelor studies. As his understanding of chemical engineering progressed, he began to realize the importance of sustainability and the ecological impact the field of chemical engineering can have on future generations. It occurred to him that if he wanted to see paradigm-shifting advances in sustainability and reduce the environmental impact of his field, then he needed to actively contribute to the cause. This revelation sparked his interest in graduate-level study and prompted him to begin researching sustainable alternative processes. He expects to receive his doctorate degree in December of 2015 after the successful defense of his dissertation entitled "BIOPLASTIC PRODUCTION IN CYANOBACTERIA AND CONSENSUS DEGENERATE PCR PROBE DESIGN".