FORWARD AND INVERSE METABOLIC ENGINEERING STRATEGIES FOR IMPROVING POLYHYDROXYBYRATE PRODUCTION

by

Keith E.J. Tyo

MS CEP, Massachusetts Institute of Technology, 2003

BS ChE, West Virginia University, 2001

Submitted to the Department of Chemical Engineering in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Chemical Engineering

at the

Massachusetts Institute of Technology

September 2008

Signature of Author.....
Keith E.J. Tyo
Department of Chemical Engineering
August 27, 2008

Certified by.....
Gregory Stephanopoulos
Willard Henry Dow Professor of Chemical Engineering
Thesis Supervisor

Accepted by.....
William Deen
Dubbs Professor of Chemical Engineering
Chairman, Committee for Graduate Students

FORWARD AND INVERSE METABOLIC ENGINEERING STRATEGIES FOR IMPROVING POLYHYDROXYBYRATE PRODUCTION

by

Keith E.J. Tyo

Submitted to the Department of Chemical Engineering on August 27, 2008 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Chemical Engineering.

Abstract

Forward metabolic engineering (FME) is a rational approach to cellular engineering, relying on an understanding of the entire metabolic network to direct perturbations for phenotype improvement. Conversely, inverse metabolic engineering (IME) uses a global, combinatorial approach to identify genetic loci that are important for a given phenotype. These two approaches complement each other in a strain improvement program. FME and IME approaches were applied to poly-3-hydroxybutyrate (PHB) production in *Synechocystis* PCC6803 [IME] and recombinant *E. coli* [FME] in this thesis.

IME was appropriate for *Synechocystis*, where metabolic regulation of the native PHB pathway was not well understood. A high throughput screening method was established by developing a staining protocol that quantitatively related nile red fluorescence to PHB content, while maintaining cell viability for both organisms. This was combined with fluorescence activated cell sorting (FACS) to screen for high PHB mutants. A *Synechocystis* insertion mutagenesis library was screened to identify gene disruptions that increased PHB. Two gene disruptions in proline biosynthesis and an unknown function were identified and characterized.

An analogous IME study in *E. coli* did not find increased PHB mutants, but suggested an FME approach on the PHB pathway. Systematic overexpression of the pathway revealed *phaB*, acetoacetyl-CoA reductase, limited PHB flux. Beyond this, whole operon overexpression led to even higher PHB fluxes. In a nitrogen-limited chemostat, PHB flux did not change with dilution rate. Unlike prior pleiotropic perturbations, these systematic experiments could clearly conclude that the flux control is in the PHB pathway.

At high PHB flux, growth rate was extremely hindered and was accompanied by PHB plasmid genetic instability and rapid PHB productivity loss. Tandem gene duplication (TGD) was developed to slow productivity loss caused by "allele segregation," a fast process that propagates a DNA mutation to all copies of a plasmid. By placing the many copies in tandem, rather than on individual plasmids, allele segregation could be avoided, increasing stability significantly.

These methods and results should support PHB engineering in higher photosynthetic organisms and better *E. coli* PHB production in batch or continuous culture. TGD is a broadly applicable technique for high level recombinant expression.

Thesis Supervisor: Gregory Stephanopoulos
Title: Willard Henry Dow Professor of Chemical Engineering

Acknowledgements

During my time at MIT, I have been taught, encouraged, and inspired by many people, both inside and outside MIT. These inputs have not only shaped the quality of this thesis, but of me as a person and professional. First, thanks to my advisor, Prof. Greg Stephanopoulos, for always believing in me (even when I didn't believe in myself), giving me profound freedom and ample funding to pursue my ideas, and guidance into how my ideas fit into the larger picture of biotechnology.

Thanks to my committee for guiding and shaping my thesis with their time, effort, and expertise. Their rigorous feedback has helped solidify each of my findings along the way. The final product has been greatly improved because of it.

Thanks to many, many lab mates who have also provided technical support, encouragement, and well, a lot of fun. Curt Fischer and Dr. Hal Alper have become good colleagues and great friends, I hope that's the case for a very, very long time. I am in great debt to your wisdom, patience, and encouragement, and I'm excited to see how you will change the world. To Bill, Saliya, Angelo, Yong-su and Ilias who helped me get my start in biology. To Aji, Jason, Daniel, Mark, Joel, José, Kyle, Christine, Jamey, Brian, and Tina, thanks for making work in the

lab fun. Thank you Freddy, Hang, and George for directly measuring data in this thesis.

Thanks to my funding sources, National Science Foundation, Department of Energy, and the Singapore-MIT alliance, none of this could have happened without your financial support for what is done here.

Thanks to my many friends, inside and outside MIT, and family that have made my life worth living. MIT – Chad, Brad, Ramin, Mike, Anna, Daryl, Jason. Outside – Andrew, Heather, Judy, Adam, Ninad, Cesar. Family – You know who you are, both out-laws and in-laws. Thanks for everything.

And most of all, thanks to my wife, Kate. For being beautiful, encouraging, and helping me through the long, hard process this was. You were my greatest discovery during my time at MIT.

To God, all truth is your truth. I hope I've uncovered some of it that will help people.

Table of contents

Abstract....3
Acknowledgements....5
Table of contents....7
List of figures....11
List of tables....13

1 Introduction 15
1.1. Thesis objectives 19
1.2. Specific aims 19
1.3. Thesis organization 20

2 Background 23
2.1. Forward Metabolic Engineering (FME) 23
2.2. Inverse Metabolic Engineering (IME) 29
2.3. Synergies between FME and IME 32
2.4. Polyhydroxybutyrate 33
Material properties 34
Applications 35
Biological occurrence 36
PHB metabolic pathway 37
2.5. Previous metabolic engineering 40
Recombinant *Escherichia coli* 40
Synechocystis PCC 6803 48

3 Development of a high-throughput screen for PHB 53
3.1. Nile red fluorescence screen for PHB 53
3.1.1. Introduction 53
3.1.2. Materials and methods 55
Bacterial strains and growth medium 55
Staining and fluorescence activated cell sorting (FACS) 57
Chemical PHB analysis 58
Synechocystis staining optimization 58
E. coli staining optimization 59
3.1.3. Results 61
Synechocystis protocol development 61
E. coli protocol development 65
3.1.4. Discussion 71
3.2. Dielectrophoresis screen for PHB 75

4 Identification of gene disruptions for increased poly-3-hydroxybutyrate accumulation in *Synechocystis* PCC 6803 81
4.1. Introduction 81
4.2. Materials and methods 82
4.2.1. Strains, culture conditions, reagents 82
4.2.2. Development of transposon library in *Synechocystis* 83
4.2.3. Flourescence-based PHB measurement and screening 84
4.2.4. Transferring transposon insertions to parental strain 86
4.2.5. PHB analysis 87
4.3. Results 87
4.3.1. Library development 87
4.3.2. Library screening 89
4.3.3. Location of transposon insertions 93
4.3.4. PHB accumulation in improved strains 94
4.4. Discussion 97

5 FME perturbations for PHB productivity in *E. coli* 103
5.1. Introduction 103
5.2. Materials and methods 106
5.2.1. Strains, plasmids, primers and media 106
5.2.2. Systematic overexpression studies 109
5.2.3. Chemostats 110
5.2.4. Analytical methods 111
5.2.5. Transcriptional analysis 111
5.2.6. Flux balance analysis 112
5.3. Results 113
5.3.1. Validating conditions for balanced growth 113
5.3.2. Model estimates of PHB pathway flux control 116
5.3.3. Systematic overexpression of PHB pathway 117
5.3.4. Nitrogen-limited chemostat PHB production 121
5.3.5. Comparison of experimental observations to a stoichiometric maximal model 124
5.4. Discussion 127

6 Tandem Gene Duplication 133
6.1. Introduction 133
6.2. Materials and Methods 139
6.2.1. Strains and media 139
6.2.2. Construction of TGD strain 140
6.2.3. Subculturing genetic stability assay 143
6.2.4. Exponential phase genetic stability assay 143
6.2.5. qPCR measurement of copy number 144
6.2.6. Poly-3-hydroxybutyrate Assay 144
6.2.7. FACS Assay 144

6.3. Results 145
6.4. Discussion 152

7 Conclusions and Recommendations 157
7.1. Conclusions 157
7.2. Recommendations 161

References..... 165

Appendix I: Kinetic model of PHB pathway..... 175
Appendix II: Plasmid propagation model..... 179

List of figures

Figure 2-1: Approach of Forward Metabolic Engineering..... 28

Figure 2-2: Structure of polyhydroxyalkanoates..... 34

Figure 2-3: PHB metabolic pathway..... 38

Figure 2-4: Metabolic pathways at acetyl-CoA node..... 42

Figure 3-1: *Synechocystis* protocol optimization..... 62

Figure 3-2: PHB staining effectiveness for *Synechocystis*..... 64

Figure 3-3: *E. coli* protocol optimization..... 67

Figure 3-4: PHB staining effectiveness for *E. coli*..... 70

Figure 3-5: Dielectrophoretic force of *E. coli*..... 77

Figure 3-6: Schematic representation of DEP-FFF..... 78

Figure 4-1: Library development strategy..... 89

Figure 4-2: Library screening..... 91

Figure 4-3 Quantitative analysis of isolated clones and transposons systematically introduced to WT by HPLC 92

Figure 4-4 Context of transposon insertions..... 94

Figure 4-5 Growth time course data for improved mutants in different growth media. 96

Figure 4-6 PHB accumulation of *Synechocystis* transposons..... 97

Figure 5-1: PHB accumulation during batch growth..... 115

Figure 5-2: Predicted enzyme activity effects on PHB productivity. 117

Figure 5-3: Systematic overexpression of steps in

PHB pathway. 119
Figure 5-4: Overexpression of entire PHB pathway..... 120
Figure 5-5: C-mol fluxes in PHB nitrogen-limited chemostat... 122
Figure 5-6: Comparison of overexpression and chemostat study to stoichiometric model for PHB in *E. coli*. 125
Figure 6-1: Tandem gene duplication: construction, amplification, and stabilization. 138
Figure 6-2: Plasmid map for pTGD..... 142
Figure 6-3: *recA* is required for copy number amplification... 146
Figure 6-4: TGD *recA*- is more stable than plasmids or TGD *recA*+ without antibiotics. 148
Figure 6-5: Long term performance of tandem PHB operon..... 150
Figure 6-6: Copy number of TGD can be varied fined-tuned expression. 152

List of tables

Table 2-1: Properties of PHAs and polypropylene..... 35

Table 4-1: Oligonucleotides used in *Synechocystis* study..... 86

Table 5-1: Strains and plasmids used in *E. coli* study..... 107

Table 5-2: Oligonucleotides used in *E. coli* study..... 108

Table 5-3: PHB production in chemostat..... 122

Table 6-1: Oligonucleotides used in TGD study..... 141

1 Introduction

Strain engineering and optimization is a central problem in biotech applications. Genetic manipulation of an organism to yield a strain that is superior to the wild type with regard to a particular phenotype is the goal of strain improvement programs. Retrofitting evolution to give us more of a particular metabolite/protein or resistance to a higher temperature/pH allows us to use living organisms in specialized applications to benefit mankind. A variety of tactics have been employed to optimize strains for the production of compound of interest.

It has been shown time and time again that for recombinant production of a compound of interest, the productivity of an organism is not only determined by the activity of the product-forming step, but is a function of the pathway in its entirety (Niederberger, Prasad R et al. 1992; Fell and Thomas 1995; Sudesh, Taguchi et al. 2002). In the cited examples, increased activity in individual steps of the pathway was not able to increase the flux through the pathway. As well, many distal

factors may affect the flux through the product-forming pathway: competing pathways, physiological effects of the accumulation of the product or intermediate metabolites, redox state, co-factor availability, and more (Burgard, Pharkya et al. 2003; Alper, Miyaoku et al. 2005; Alper, Moxley et al. 2006). These factors must be taken into consideration when trying to improve the productivity of an organism.

In practice, engineering cells can be limited because the aforementioned effects may be difficult to identify. Random approaches, such as genome shuffling, chemical mutagenesis, and unnatural evolution, have been used to generate genetic diversity that may contain improved strains (Crameri, Raillard et al. 1998; Parekh, Vinci et al. 2000). These combinatorial methods in conjunction with an effective selection/screen are a powerful way to improve cell performance. However, these techniques often do not provide insight into the cell biology. Although an improved strain is isolated, further development of the strain is difficult because the genetic changes may not be obvious. Strategies that attempt to close this loop by employing tools that control the number of mutations and make the locations more readily identified is therefore necessary.

A two-pronged complementary approach of (a) forward, or traditional, metabolic engineering (FME) that measures and analyzes what is known about the metabolic pathways and the

system as a whole to direct genetic changes, combined with (b) inverse metabolic engineering (IME), which seeks to identify previously unknown mechanisms for improving flux to the desired product, offers a way forward that leverages current knowledge of the system, while seeking to add relevant knowledge.

The challenge then is to develop methods for each of these two avenues of strain development. For FME, more (and specifically more relevant) knowledge obtained through the systematic analysis of metabolic pathways will move our knowledge beyond which enzymes are involved in the pathway to which enzymes play the most dominant role in controlling the flux to the desired product. For IME, experimental methods to introduce genetic diversity into the desired organism and high throughput methods for identifying clones from the diversity that result in improved phenotypes will lead to the identification of unpredicted genetic loci that will expand our understanding of the system.

Another key challenge is the dynamic nature of DNA, and the loss of productivity of the compound of interest that accompanies DNA mutations in engineered organisms. Organisms which grow faster will overtake a given population. Typically production of a desired molecule results in reduced fitness for an organism. Therefore, mutants that evolve in a way that mitigates the product formation will grow faster and overtake

the culture. In batch processes, these problems can be reduced by minimizing the number of generations that are allowed to grow. However, this is not the case in production schemes that require many cell divisions during production, such as chemostats.

These questions can not be addresses in the abstract, and therefore, a model system must be used that is appropriate to explore these ideas and develop methods for gathering the desired information and achieving the desired strain properties. Improving the accumulation of the biopolymer poly-3-hydroxybutyrate (PHB) by the cyanobacterium *Synechocystis* PCC680 and *Escherichia coli* would have significant ecological and economical implications, that will be discussed in Chapter 2. *Synechocystis* naturally synthesizes the biopolymer from carbon dioxide and light, allowing direct conversion to the desired product. PHB production in *Synechocystis* uses native genes and is subject to internal regulation that has not been elucidated. PHB production in *E. coli* is based on the heterologous expression of the PHB pathway and is not subject to native regulation. Although it requires sugars or fatty acids (and thereby another photosynthetic organism to provide these materials) to generate PHB, it has much higher productivities in general than photosynthetic organisms and has much better understood metabolism.

1.1. Thesis objectives

The objective of this thesis is to identify genetic loci that affect the production of poly-3-hydroxybutyrate in *Synechocystis* and *E. coli* by developing the necessary methods to obtain this information and improve the productivity and genetic stability of PHB production using the paradigms of forward and inverse metabolic engineering.

1.2. Specific aims

In pursuance of the above mentioned general goals, the follows specific aims were pursued based on two bioengineering approaches, (1) inverse metabolic engineering and (2) forward (or traditional) metabolic engineering.

- ☐ Inverse Metabolic Engineering
    a. Develop libraries of genetic diversity in *Synechocystis* and *E. coli*.
    b. Develop a high throughput screen for PHB accumulation for *Synechocystis* and *E. coli.*
    c. Identify genetic loci that contribute to PHB production by isolating and characterizing library mutants from Aim (a) using the high throughput screen developed in Aim (b).
- ☐ Forward Metabolic Engineering

a. Analyze PHB production limitations by systematically perturbing the enzymes in the PHB pathway in *E. coli.*
b. Analyze PHB pathway precursor availability and genetic stability by perturbing the overall PHB pathway in *E. coli.*
c. Analyze PHB productivity by varying growth rates in a nitrogen-limited chemostat.
d. Develop methods for the genetically stable, high expression of the PHB operon in *E. coli.*

1.3. Thesis organization

A background on metabolic engineering principles, both forward and inverse are presented in Chapter 2, followed by a review of PHB focusing on prior efforts to improve PHB production in autotrophic and heterotrophic systems.

While it may seem more natural to discuss forward metabolic engineering before inverse metabolic engineering, this thesis will be motivated by showing some of the limitations of inverse metabolic engineering first that motivated a moved to forward approaches.

Chapters 3 and 4 will focus on the development of tools for inverse metabolic engineering and their deployment to identifying genetic nodes in *Synechocystis*. Chapter 5 will

focus on the forward metabolic engineering analysis of *E. coli*, followed by Chapter 6 which will discuss the development of tandem gene duplication and its application to improving genetic stability. Chapter 7 will offer concluding thoughts on the thesis and next steps that are motivated by the conclusions of this work.

2 Background

This chapter will cover the background both of the methodologies pursued in this thesis as well as technical details and prior work on engineering poly-3-hydroxybutyrate (PHB) production in *Synechocystis* and *E. coli*. Forward and inverse metabolic engineering principles will be discussed. These principles will motivate the methods developed and the data measured in the study, as well as why these approaches are appropriate. Technical details of the PHB pathway and prior engineering efforts will reveal what is known, what has been attempted, and will reveal the niches of inquiry that will be most appropriate.

2.1. Forward Metabolic Engineering (FME)

Metabolic engineering is the systematic analysis and alteration of metabolic networks, the series of enzymatic reactions of the cell that are responsible for the chemistry of the cell. This will be called "forward" metabolic engineering, because of the knowledge-driven approach to modifying the metabolic network and contrasts well with inverse metabolic

engineering, a random approach that generates knowledge, which will be discussed later. Metabolic engineering emerged with the advent of recombinant DNA technology (Stephanopoulos 1999). For the first time it was possible to recombine genes from one organism with those of another, opening the door to a realm of possibilities not yet explored. While the initial applications of genetic engineering were simply producing human proteins in bacteria for therapeutic treatment of specific protein deficiencies, engineers quickly realized the vast potential of using multiple genes to create entirely new pathways that could produce a wide range of compounds from a diverse substrate portfolio (Bailey 1991; Stephanopoulos and Vallino 1991). Aided by advanced methods for the analysis of biochemical systems, metabolic engineering sets out to create new industrial innovations based on recombinant DNA technology.

Metabolic engineering is different from other cellular engineering strategies because its systematic approach focuses on understanding the larger metabolic network in the cell. In contrast, genetic engineering approaches often only consider narrow phenotypic improvements resulting from the manipulation of genes directly involved in creating the product of interest. The need for a systematic approach to cellular engineering has been demonstrated by several vivid examples in which choices for improving product formation, such as increasing the activity of

the product-forming enzyme, have only resulted in incremental improvements in output (Niederberger, Prasad R et al. 1992; Fell and Thomas 1995; Sudesh, Taguchi et al. 2002). Intuitively, this makes sense. A typical cell has evolved to catalyze thousands of reactions that serve a multitude of purposes critical for maintaining cellular physiology and fitness within its environment. Thus changing pathways that do not improve fitness, or even detract from fitness within a population, often causes the cell's regulatory network to adjust by diverting resources back to processes that optimize cellular fitness. This may lead to relatively small improvements in product formation despite large increases in specific enzymatic activities. Without a good understanding of the metabolic network, further progress is often difficult to achieve and must rely on other time consuming methodologies based on rounds of screening for the phenotype of interest. Classical strain improvement (CSI) relies on random mutagenesis to accumulate genomic alterations that improve the phenotype (Parekh, Vinci et al. 2000). This method typically has diminishing returns for a variety of reasons: 1) it does not extract information about the location or nature of the mutagenesis; 2) it often results in deleterious mutations alongside advantageous mutations and therefore is less efficient, and; 3) it does not harness the power of nature's biodiversity by mixing specialized genes

between organisms. Gene shuffling approaches attempt to correct 2) and 3) by swapping large pieces of DNA between different parental strains to eliminate deleterious mutations or incorporate genes from other organisms (Crameri, Raillard et al. 1998). In contrast, metabolic engineering approaches embrace techniques that fill the gaps left by CSI and gene shuffling methodologies by placing an emphasis on understanding the mechanistic features that genetic modifications confer, thereby adding knowledge that can be used for rational approaches while searching the metabolic landscape.

Metabolic engineering overcomes the shortcomings of alternative approaches by considering both the regulatory and intracellular reaction networks in detail. Research on the metabolic pathways has primarily focused on the effect of substrate uptake, byproduct formation, and other genetic manipulations that affect the distribution of intracellular chemical reactions (flux). Because many of the desired products are organic molecules, metabolic engineers often concentrate their efforts on carbon flow through the metabolic network. In diagnosing the metabolic network, engineers rely on intracellular flux measurements conducted *in vivo* using isotopic tracers as opposed to simply using macroscopic variables such as growth rate and metabolite exchange rates (Stephanopoulos 1999). The latter measurements contain less information about the

intracellular reaction network and therefore give a very limited perception of the phenotype of the cell. *In vitro* enzymatic assays can also provide helpful, but potentially misleading, information about the behavior of an enzyme *in vivo* and cannot be used to calculate individual fluxes, which also depend upon the size of the metabolite pools and other intracellular environmental factors. Research on regulatory networks has ranged widely from engineering allosteric regulation, to constructing new genetic regulator elements such as promoters, activators and repressors that influence the reaction network (Lu and Liao 1997; Farmer and Liao 2000; Ostergaard, Olsson et al. 2000; Alper, Moxley et al. 2006). By understanding the systemic features of the network, metabolic engineering can identify rational gene targets that may not be intuitive when relying upon extracellular or activity measurements alone. In practice metabolic engineering studies proceed through a cycle of perturbation, measurement, and analysis (Figure 2-1).

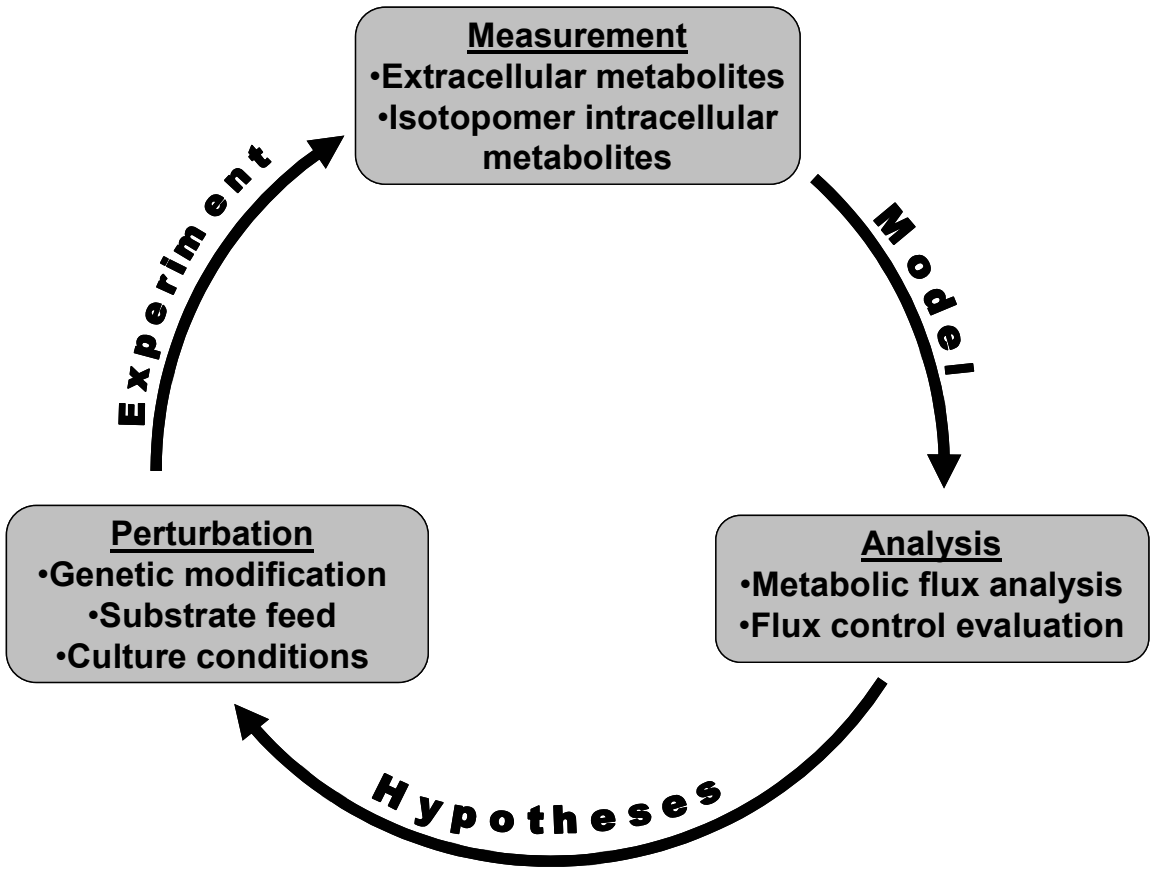


Figure 2-1: Approach of Forward Metabolic Engineering.

Metabolic engineering is an information-driven approach to phenotype improvement that involves (1) measurement, (2) analysis, and (3) perturbation. Data from measurements can be used to formulate models. These models can then be analyzed to generate new targets for manipulation (hypotheses). After performing the genetic manipulations, experiments must be formulated to determine how the metabolic network has adjusted to each genetic manipulation. The cycle can then continue, providing more information with each round.

Measurement requires the ability to assay large parts of the network to extract as much information about the effect of an imposed network perturbation as possible. GC-MS and NMR are commonly used to measure metabolite pools and the rates of chemical reactions within cells (Stephanopoulos 1999; Fischer, Zamboni et al. 2004; Styczynski, Moxley et al. 2007).

Microarrays have been developed, and new proteomic tools are evolving to monitor the response of gene expression to different perturbations (Bro, Regenberg et al. 2003; Ong and Mann 2005).

Finally, to complete the cycle before proceeding to the next iteration, robust analyses are necessary to determine which portions of the network are the most sensitive or amenable to genetic manipulation and generate meaningful hypotheses from the vast quantities of data that can be gathered. By analyzing the differences in the metabolic fluxes following a perturbation, new targets can be identified that are most likely to improve the phenotype. The new targets set the foundation for hypotheses, leading to another perturbation of the network. Such perturbations are followed by another round of measurement and analysis and may include: increasing the activity of desirable enzymes within a pathway either by overexpression or deregulation, deleting enzymes that divert carbon to undesired by-products, using different substrates, or changing the overall state of the cell to favor certain pathways.

2.2. Inverse Metabolic Engineering (IME)

When engineering cellular phenotypes such as product formation, substrate utilization, product tolerance, etc., rational approaches to genetic modification can be limited by the prior knowledge of the system. While many metabolic

pathways have been identified or predicted by comparative genomics, missing regulatory information can limit how far the organism can be reprogrammed. As well, knowledge of the metabolic pathways does not reveal the concentrations of intermediates and products of a pathway, or at what threshold concentrations these molecules will begin to have toxic effects on the cell. Sometimes combinatorial or evolutionary approaches can be effective in eliciting these desired phenotypes by making perturbations that would not be considered in a rational approach (Santos and Stephanopoulos 2008).

Inverse metabolic engineering is a combinatorial approach to identifying genotypes that elicit desired phenotypes (Bailey, Sburlati et al. 1996). As mentioned, CSI approaches use chemical mutagenesis which can cause many mutations throughout the genome. After screening for improved mutants, some of these mutations can be beneficial and some detrimental, and most often it is very difficult to locate the mutations. IME differentiates itself from CSI approaches by using methods that introduce one genetic change at a time that can be readily identified in the clone. Identifying the location and nature of the genotype can be very advantageous for further engineering the phenotype. Even if the exact mechanism is not apparent, by identifying the genetic locus, further hypothesis-driven experiments can uncover the mechanism, as well as further

improve the desired phenotype by more refined changes around the genetic perturbation. An additional advantage to this approach is it allows screening to be done in easy-to-work-with lab strains, while analogous directed genetic changes can be made in other organisms, where such screening would be impossible.

IME has been successfully used recently in several systems. Jin has successfully used IME to identify genetic loci that are relevant in lycopene production by overexpressing random gene fragments from an *E. coli* library. By doing this, he confirmed targets that were already known in the isoprenoid pathway, as well as new genes of unknown function. Lycopene production was increased by X% (Jin and Stephanopoulos 2007). Jin also used a combinatorial IME approach to optimize the levels of XYLA, XYLB, and XYLC in a recombinant pathway, toward improving xylose metabolism and ethanol production in *Saccromyces cerevisiae*, while reducing the byproduct xylitol, which had been a consistent problem in this area (Jin, Alper et al. 2005). (Weikert, Canonaco et al. 2000; Gill 2003; Sauer and Schlattner 2004).

The IME strategy involves three steps of perturbation, selection/screening, and genotype characterization. Perturbation is generating a library of mutants with different genotypes. The type of genetic variation that can be introduced includes: single gene mutagenesis carried on a plasmid,

transposon insertion mutagenesis, plasmid library overexpression, and others. An important aspect of the perturbation is that the genetic change can be easily characterized in the mutant. This allows for routine identification of the genetic changes in mutants that show improved phenotype. Step two employs an effective selection or high throughput screen to isolate mutants exhibiting desired phenotypes from the library. Cells can be grown under a selection pressure that favors mutants with improved phenotype of interest. If growth selection can not be used, as is often true when it is desired to increase the product, single cell measurement or microtiter plate screening can be employed for screening of desired mutants. In step three, the mutants that have been isolated by screening as being of interest are characterized, and the genotypes underlying the desired behaviors of the selected cells are determined.

2.3. Synergies between FME and IME

IME is a natural complement to the approaches of forward metabolic engineering, and some natural synergies exist between the two approaches. Forward metabolic engineering is an approach that utilizes existing information about the cell in a holistic, model-driven approach. Existing data is integrated to create hypothesizes that will improve production characteristics. It requires fewer experimental methods than IME by virtue of

prior data minimizes the number of hypotheses that need to be evaluated. However, forward approaches are limited by the available prior data. These gaps in information will eventually limit the improvements that can be made by forward approaches. At this point, IME strategies can uncover new mechanisms that effect product accumulation in a methodical way. These new mechanisms not only improve the productivity in their own right, most importantly, they add knowledge about the system. This knowledge can be fed back to improve forward metabolic engineering models. This can, and in most circumstances must, proceed iteratively due to the non-linear effects of many cellular perturbations.

2.4. Polyhydroxybutyrate

Poly-3-hydroxybutyrate (PHB) is a biopolymer from a class of naturally occurring polyesters called polyhyroxyalkanoates (PHAs) (Figure 2-2(a)). Figure 2-2(b) shows the structural unit of PHB. PHB has been studied extensively and is produced commercially on a small scale in bacteria (Kennedy 2007). The polymer is easily biodegraded by a variety of bacteria that can depolymerize it for use as a carbon source (Anderson and Dawes 1990).

Figure 2-2: Structure of polyhydroxyalkanoates.

(a) Structure of a general poly-3-hydroxyalkanoate. R position can be substituted with many possible groups. 4-hydroxy- and 5-hydroxyalkanoates have also been observed. (b) Structure of poly-3-hydroxybutyrate. This is the most abundant and most studied PHA.

Material properties

PHB is a white crystalline polymer and can vary in molecular weight from 50,000 to 1,000,000 Da. It is a straight chain monomer of 3-hydroxybutyrate. PHB exists amorphously in the bacterial cell, but becomes highly crystalline, and therefore brittle upon extraction (Madison and Huisman 1999). Copolymers of 3-hydroxybutyrate and 3-hydroxyvalerate have shown to be less brittle and have properties similar to polyethylene and polypropylene in thermoplastic applications. Table 2-1 shows a comparison of material properties of PHAs to polypropylene.

Table 2-1. Properties of PHAs and polypropylene

Parameter	Value for*:				
	P(3HB)	P(3HB-3HV)	P(3HB-4HB)	P(3HO-3HH)	Polypropylene
T_m (^oC)	177	145	150	61	176
T_g (^oC)	2	-1	-7	-36	-10
Crystallinity (%)	70	56	45	30	60
Extension to break (%)	5	50	444	300	400

*Data from (Madison and Huisman 1999)
T_m - melting temperature, T_g - glass transition temperature, P(3HB) is poly(3-hydroxybutyrate), P(3HB-3HV) is poly(3-hydroxybutyrate-co-3-hydroxyvalerate) containing 20% 3HV, P(3HB-4HB) is poly(3-hydroxybutyrate-co-4-hydroxybutyrate) containing 16% 4HB, P(3HO-3HH) is poly(3-hydroxyoctanoate-co-3-hydroxyhexanoate) containing 11% 3HH.

Applications

PHB is a biodegradable thermoplastic with the majority of applications being to replace current petrochemical polymers. From an ecology standpoint, PHAs offer two advantages over traditional plastics. PHB is biodegradeable, in that many microorganisms are able to depolymerize PHB in the environment to recover the carbon on the order of months to years (Jendrossek, Schirmer et al. 1996). (It is important to note that for commercial applications, PHB is not biodegraded in more sterile environments such as houses and offices.) Also, PHB is made from renewable resources, typically carbohydrates or fatty acids. These are in contrast to our current bulk plastics which will persist for long times and are synthesized from petroleum feedstocks.

For this reason industrial fermentation of PHB has been the focus of a number of studies (Abe, Taima et al. 1990; de Koning, Bilsen et al. 1994; de Koning, Kellerhals et al. 1997), while other studies have been directed towards the use of PHB-producing organisms in the breakdown of wastewater (Doi 1990; Curley, Hazer et al. 1996), just to name a few. (Engineering strategies will be discussed later.)

PHB has been used in molding applications such as bottles, cosmetic containers, golf tees, and pens (Baptist 1963; Baptist 1963; Webb 1990). Polyvinyl alcohol/PHB blends can be extruded as sheets, for plastic film applications (Holmes 1986). High value applications of PHB are uses as biomaterials for the medical industry, such as absorbable surgical suture, prosthesis or tissue engineering material (Williams SF, Martin DP et al. 1999) and use in controlled drug release (Steinbüchel and Füchtenbusch 1998).

Biological occurrence

PHB is produced naturally by a variety of organisms. Bacteria including *Ralstonia eutropha*, *Methylobacterium*, *Pseudonomas alcaligenes latus*, and *Azobacter vinelandii* have been reported to produce PHB (Madison and Huisman 1999). *R. eutropha* has been the most extensively studied native PHB producer and has been examined for use in industrial production.

The biological function of PHB is as a carbon and electron sink when there is a nutrient imbalance and nitrogen or phosphate is scarce. There are many advantages for the organism to polymerize the excess carbon. The insoluble PHB granules can be stored at high concentrations (up to 90% dry cell weight), does not affect the osmotic state of the liquid phase, and does not leak out. When other nutrients become available, the PHB granules can be depolymerized and the carbon/reducing equivalents are recovered (Anderson and Dawes 1990).

PHB metabolic pathway

The genes for the PHB pathway have been identified and been expressed in *Escherichia coli* (Peoples and Sinskey 1989; Peoples and Sinskey 1989). PHB is produced by a 3 step pathway from the central carbon metabolic intermediate, acetyl-CoA, and the reducing equivalent, NADPH (Figure 2-3). The first step is the condensation of two acetyl-CoA moieties to acetoacetyl-CoA with the release of a CoA. This step is catalyzed by β-ketothiolase (*phaA*) and is believed to be near equilibrium in most organisms (Masamune, Walsh et al. 1989). The equilibrium favors the two acetyl-CoA molecules, but under PHB producing conditions depletion of acetoacetyl-CoA causes a thermodynamic shift allowing the reaction to proceed from acetyl-CoA to acetoacetyl-CoA. The second step is the NADPH-dependent reduction of

acetoacetyl-CoA to the monomer 3-hydroxybutyryl-CoA by the acetoacetyl-CoA reductase, *phaB*. The availability of NADPH reducing equivalents are therefore necessary to drive the pathway forward. Finally, the 3-hydroxybutyryl-CoA is polymerized to PHB by the PHB synthase, which can either be a one or two subunit protein depending on the organism. The synthase typically sits on the cytosolic/PHB interface and extends the polymer chain by adding 3-hydoxybutyrate molecules to the end of the growing PHB chain after releasing the CoA. In the systems explored in this thesis, only two component synthases (*phaEC*) were involved in the pathway.

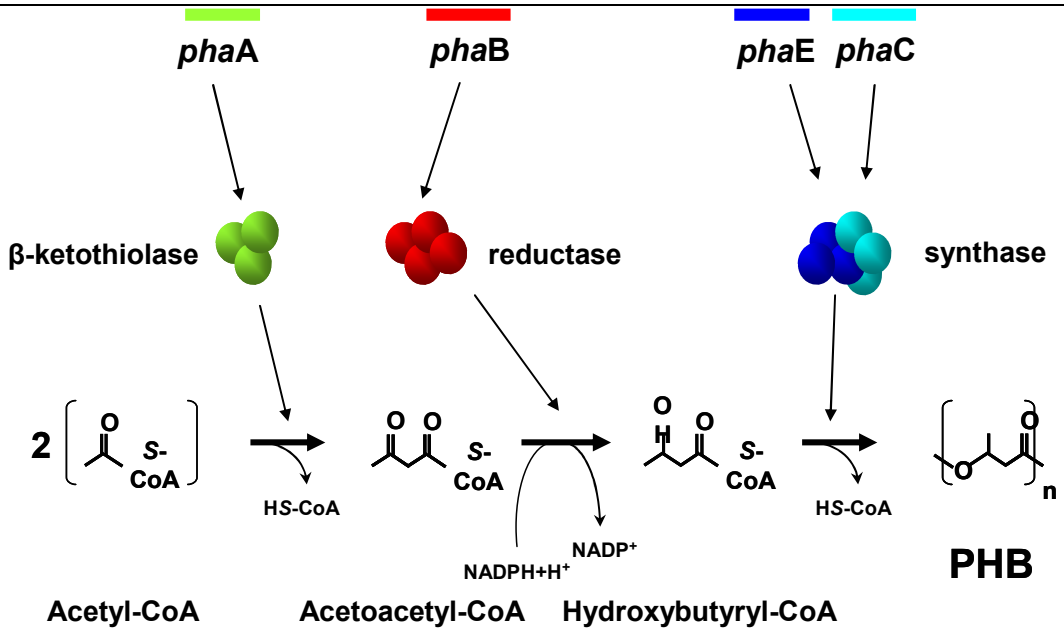


Figure 2-3: PHB metabolic pathway.
The PHB metabolic pathway starts at the common precursor acetyl-CoA. Acetyl-CoA and a reducing equivalent are consumed in the pathway, and PHB and 2 CoA's are released. Top level – genes involved in the pathway. Second level – enzymes, Third Level – metabolites. Vertical arrows show the enzyme that is coded by the gene, or the reaction that is catalyzed by the enzyme. Horizontal and curved arrows connect the metabolites involved in each reaction.

Several other proteins play roles in the depolymerization and regulation of PHB. *phaZ* is the depolymerase responsible for breaking down PHB granules into shorter, soluble polymers from the PHB granule and an oligomer hydrolase converts the short chain PHB to (R)-(-)-hydroxybutyric acid which can be further metabolized by the cell (Huisman, Wonink et al. 1991). PHB levels are regulated to manage the nutrient resources that are available to a cell at a given time. Specifically, if the cell is growing in a carbon rich medium that is lacking in other necessary elements (nitrogen, phophorus, oxygen, etc.), the pathway can be upregulated. *phaP* acts as to regulate the size and number of PHB granules (Wieczorek, Pries et al. 1995). *phaP* binds at the interface of the PHB granule and the cytoplasm, and the concentration of *phaP* in the cell is inversely proportional to the number of granules in a cell. *phaR* is a regulatory protein that reduces *phaP* levels, presumably to manage granule size and number (York, Stubbe et al. 2002). The levels of the polymerase (*phaEC*) have been shown to affect the molecular weight of the polymer (Sim, Snell et al. 1997). Sim et. al. showed that overexpression of the *phaC* gene resulted in shorter chain length PHB, consistent with the hypothesis that a small number of polymerases will generate a few long chain length PHB, while many polymerases will generate lots of short chain length

PHB. A model for PHB granule formation is presented by Madison and Huisman that is helpful, but beyond the scope of this review (Madison and Huisman 1999).

Plants such as *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), and *Zea mays*(corn) have been engineered to make PHB (Madison and Huisman 1999).

2.5. Previous metabolic engineering

As there is a large opportunity for PHAs to make in-roads to the commodity polymers markets, much effort has been undertaken to better characterize and improve PHB production. While the molecular weight of the polymer is very important in affecting the material properties of the polymer, in this review we will focus only on efforts to increase yield, titer, and productivity, the traditional objectives of metabolic engineering. Additionally, we will focus on the engineering studies in recombinant *E. coli* and *Synechocystis* in this review.

Recombinant *Escherichia coli*

Following the expression of the PHB biosynthetic pathway from *R. eutrophus* into *E. coli* (Schubert, Steinbüchel et al. 1988; Peoples and Sinskey 1989), much work has taken place to improve the production in this system. The overall PHB reaction from glucose is shown in Equation 2-1. Two carbon atoms are

lossed as two pyruvates are converted to two acetyl-CoAs in the biochemical pathway. Three reducing equivalents are also produced, and must be oxidized elsewhere (shown as H_2). A maximum theoretical yield of 0.48 g-PHB/g-glucose can be achieved.

Equation 2-1: Chemical conversion of glucose to PHB

$$C_6H_{12}O_6 \rightarrow (C_4H_6O_2)_n + 2CO_2 + 3H_2$$

The route in which glucose is metabolized to PHB can vary. Equation 2-2, Equation 2-3, and Equation 2-4 below are different routes to pyruvate, an immediate precursor to the acetyl-CoA used for PHB synthesis, that give different forms of reducing equivalents and energy. Figure 2-4 shows the final conversion from pyruvate to acetyl-CoA and some of the major pathways that compete with PHB synthesis for acetyl-CoA.

Equation 2-2: Embden-Meyerhof-Parnas pathway (glycolysis)

$$Glucose + 2NAD^+ + 2H^+ + 2ADP \rightarrow 2Pyruvate + 2NADH + 2ATP$$

Equation 2-3: Pentose phosphate pathway

$$Glucose + \frac{5}{3}NAD^+ + 2NADP^+ + \frac{11}{3}H^+ + \frac{5}{3}ADP \rightarrow \frac{5}{3}Pyruvate + CO_2 + \frac{5}{3}NADH + 2NADPH + \frac{5}{3}ATP$$

Equation 2-4 Entner-Doudoroff pathway

$$Glucose + NAD^+ + NADP^+ + 2H^+ + ADP \rightarrow 2Pyruvate + NADH + NADPH + ATP$$

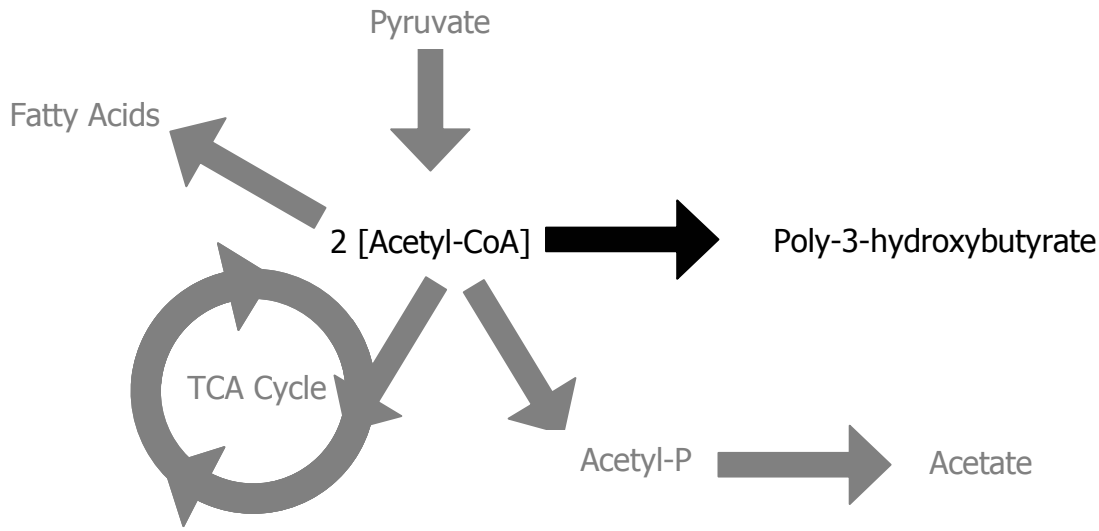


Figure 2-4: Metabolic pathways at acetyl-CoA node.
Acetyl-CoA is a two carbon precursor for biomass creation and energy pathways. It is primarily fed from the conversion of pyruvate, a glycolysis terminal product. Fatty acids and various TCA cycle intermediates are used as building blocks for biomass, and acetate production and TCA cycle are used to create ATP under anaerobic and aerobic production, respectively. Diversion of acetyl-CoA to PHB must be high, but carefully balanced to allow other necessary cellular functions to proceed. Arrows represent metabolic pathways.

There are differing views in the literature as to which molecules limit production of PHB in *E. coli* (discussed below). Acetyl-CoA and NADPH are required for the biosynthetic pathway, and adequate amounts of CoA and NADP+ are needed to allow for turnover in the pathway. Aside from the metabolites in the pathway, the three enzymes must have enough activity to provide high flux to PHB and K_m values that allow the enzymes to compete with other pathways for acetyl-CoA and NADPH.

Additional considerations must be made for the effects of PHB granules in the cytoplasm. Physical limitations must exist for the amount of PHB the cytoplasm can hold. By way of being solid phase, PHB avoids many of the toxic effects that can be experienced by soluble products, allowing for high titers, but eventually the bulky granules will affect cellular processes.

Genetic background choice has improved the physiology of *E. coli* and the genetic stability to improve PHB production. By testing many backgrounds of *E. coli* for their accumulation of PHB, it was observed that cloning/expression strains such as XL-1 Blue and *E. coli* wild type **B** accumulated more PHB than other cloning and wild type strains (Lee, Lee et al. 1994). Biomass growth and expression of the PHB pathway was better in XL-1 and **B** compared to other strains examined. The authors did note large amounts of filamentous *E. coli* in both the XL-1 and **B** strains. *FtsZ*, a cell division protein, mitigates filamentation that can lead to biofilm growth when overexpressed (Wang and Lee 1997). The biofilms limited growth properties of the cells and nutrient uptake. By overexpressing *ftsZ*, PHB titer was increased 6 fold.

An additional problem occurs with plasmids, that limits PHB formation. High copy number plasmids are required for substantial PHB formation, but under growth conditions, only a few copies of the plasmid are necessary to provide antibiotic

resistance or prevent cell death by *parB* (Lee, Chang et al. 1994). One group attempted to correct for this by designing a plasmid that suppresses copy number at low temperature (30°C) and induces plasmid replication at high temperature (38°C) (Kidwell, Valentin et al. 1995). Using this "temperature switch," the copy can be kept low during the beginning of production, and switch to high copy (and therefore high expression of the PHB pathway) once the desired biomass has been established.

In silico models of PHB metabolism

PHB synthesis has been modeled from both a kinetic-based model of the pathway from acetyl-CoA to PHB as well as large scale stoichiometric model that predict the effect on central carbon metabolism. A kinetic model of the pathway for *E. coli* was derived from a model for *A. eutrophus* (van Wegen, Lee et al. 2001). This model predicted that the flux to PHB is most sensitive to the acetyl-CoA supply, and not the NADPH availability. It also predicted that the activity of the NADPH-dependent reductase may control the flux of the pathway. A larger stoichiometric model was used to identify the importance of the Entner-Doudoroff (ED) pathway in PHB production and was validated experimentally by observing a decrease in PHB production after deactivating the 3-deoxy-6-phosphogluconate aldolase (*eda*) enzyme (Hong, Park et al. 2003). The ED pathway provides one NAPDH that is required for PHB synthesis, unlike

glycolysis, which only creates NADH, and the pentose phosphate pathway, which creates two NADPH.

NADPH Limitation vs Acetyl CoA limitation

The literature offers conflicting viewpoints as to what limits PHB productivity. Some have implicated NADPH as a limiting reagent under growth conditions, while others have ascerted acetyl-CoA limits productivity. Because these studies have focused on broad changes in the cell, although PHB accumulation has increased, it becomes difficult to determine what mechanisms the perturbations are acting. Additionally, some studies suffer from extremely poor growth rates, implying that PHB accumulation may still be in highly suboptimal conditions.

NADPH studies have focused on altering the redox balance in the cell through (a) phosphoglucose isomerase (*pgi*) deletion, (b) overexpression of pentose phosphate (PP) pathway enzymes (*zwf, gnd, talA,* and *tktA*), (c) gluconate feeding, or (d) transhydrogenase overexpression (Shi, Nikawa et al. 1999; Lim, Jung et al. 2002; Jung, Lee et al. 2004; Sanchez, Andrews et al. 2006; Song, Kim et al. 2006). These act through [A] preventing glucose from being metabolized through the Embden-Meyerhof-Parnas (EMP) pathway, and instead metabolized through the PP pathway or the Entner-Doudoroff (ED) pathway, making NADPH instead of NADH, [B] increasing activity of PP pathway causing

the same redirection as *[A]*, [C] feeding substrates naturally metabolized by the PP or ED pathways, and [D] catalyzing the exchange of electrons from NADH to NADPH. While these studies did change the availability of NADPH, it is unclear whether the observed increases in flux to PHB were a result of increased NADPH to drive the acetoacetyl-CoA reductase (AAR) reaction or altered regulation of the acetyl-CoA consuming reactions, making more acetyl-CoA available for PHB biosynthesis. TCA cycle consumption of acetyl-CoA is linked to redox in the cell through product inhibition by NADPH of isocitrate dehydrogenase (Dean and Koshland Jr 1993), and allosteric inhibition by high NADPH/NADP+ ratios that have been associated with lowered citrate synthase activity for growth on glucose (Lee, Kim et al. 1996; Lim, Jung et al. 2002). For PHB production on glucose, measurements of NADPH/NADP+ ratio was 3 and NADPH concentrations were 225 µM (van Wegen, Lee et al. 2001). This is higher than the NADPH K_m of AAR, 19 µM (Steinbuchel and Schlegel 1991), which should be sufficient to supply NADPH. Hong et. al. has shown that the ED pathway is used to provide NADPH for AAR in PHB production (unlike the EMP pathway which produces NADH) (Hong, Park et al. 2003).

Acetyl-CoA limitations may also be more complicated than expected. Acetate secretion is observed in PHB producing *E. coli* under aerobic conditions (Sanchez, Andrews et al. 2006),

implying an excess of acetyl-CoA is available. Phosphate acetyltransferase (*pta*) and acetate kinase (*ack*) deletions, which should decrease acetate production, have decreased PHB accumulation, implying that enzyme competition for the acetyl-CoA pool may not be the determining factor (Shi, Nikawa et al. 1999). Instead, the presence of acetyl phosphate may be required in order to activate the PHB synthase (Miyake, Schnackenberg et al. 2000) and without acetyl phosphate (the result of *pta* and *ack* knockouts) the PHB pathway may be down regulated. The Shi study also showed that the addition of α-methyl-glucoside, a glucose analog that retards glucose uptake through non-toxic competitive inhibition, decreases acid secretion, but does not change PHB accumulation. PHB productivity appears to be insensitive to glucose uptake, implying the pathway may exert metabolic control over the flux from acetyl-CoA to PHB.

A difficulty arises from comparing PHB studies that were based on a rich media, such as LB where a lot of carbon and organic nitrogen are available to the cell and do not need to be synthesized, compared to a minimal media where all biomass constituents must be derived from glucose and inorganic nitrogen. This will most certainly affect the NADPH required for catabolism of cellular components, and the carbon flux around and through acetyl-CoA. These fundamental differences in

metabolism between rich and minimal media may have different limitations in PHB formation.

The studies thus far have failed to look at the systemic changes that are induced by the perturbations to the system and have instead focused on perturbations that will have pleitropic changes. The redox state of the cell has far reaching effects on the metabolic network, and increases observed in PHB production may be due to regulatory effects that divert acetyl-CoA toward PHB rather than TCA cycle. It is important to observe the metabolic network as a whole, and make perturbations that are localized to determine the actual mechanisms of increased PHB accumulation.

Synechocystis PCC 6803

Synechocystis is a model organism for studying photosynthesis because it contains both photosystem 1 and 2 found in higher plants, but has the experimental convenience of being a fast growing, unicellular cyanobacteria. The *Synechocystis* genome sequence was published in 1996 (Kaneko, Sato et al. 1996) and genetic tools have been used to characterize the photosynthetic processes in the cell (Vermaas, Williams et al. 1987). This allows modern functional genomics approaches to be available for the organism.

PHB production in *Synechocystis* has been explored (Hein, Tran et al. 1998; Taroncher-Oldenberg, Nishina et al. 2000; Taroncher-Oldenburg and Stephanopoulos 2000; Sudesh, Taguchi et al. 2002; Wu, Shen et al. 2002). *Synechocystis* can use carbon dioxide and light as raw materials and can both carry out photosynthesis and convert sequestered carbon dioxide to PHB, eliminating the need for a second organism to do photosynthesis to provide sugars or fatty acids, which is required for other PHB fermentation processes. This *Synechocystis* one-step process could be achieved in both plants and microbial photosynthetic organisms (Madison and Huisman 1999). *Synechocystis* is thus an amenable organism for studying PHB production in the context of photosynthetic organisms, both as itself and as a model for PHB production in higher plants.

Physiological and biochemical aspects of PHB have been determined. The three enzymes for synthesizing PHB from acetyl-CoA have been identified in *Synechocystis* (Taroncher-Oldenberg, Nishina et al. 2000; Taroncher-Oldenburg and Stephanopoulos 2000). The pathway is primarily active when there are nutrient limitations (nitrogen or phosphate), but available carbon sources or reducing equivalents. The biochemical and genetic regulation of the PHB pathway has not been elucidated in the open literature.

PHB production has been altered both by nutrient feeding strategies and genetic alterations. While *Synechocystis* produces PHB at 1% dry cell weight (DCW) in BG11, a balanced nutrient broth for photosynthetic organisms, under low phosphate and low nitrogen conditions it has been shown to accumulate up to 7% DCW in wild-type *Synechocystis* (Sudesh, Taguchi et al. 2002; Wu, Shen et al. 2002). Supplementation with acetate has been observed to increase PHB accumulation, although it has not been determined whether it acts as an extra substrate for PHB or in a regulatory mode (Sudesh, Taguchi et al. 2002). Overexpression of the Type III PHB synthase was carried out, but did not increase PHB yield (Sudesh, Taguchi et al. 2002), implying that enzyme-level regulation and/or another step in the pathway limits flux to PHB. Wu et. al. deleted the ADP-glucose pyrophosphorylase to prevent *Synechocystis* from accumulating glycogen (Wu, Shen et al. 2002). In glycogen deficient strains of *Synechocystis*, 17% DCW PHB has been reported; the highest reported accumulation of PHB. *Synechococcus*, while a natural PHB producer, was engineered with the heterologous PHB pathway from *R. eutropha*, and accumulated PHB up to 25% DCW, the highest PHB accumulation observed for cyanobacteria (Takahashi, Miyake et al. 1998). It is likely that the recombinant expression avoided regulatory mechanisms that limit PHB production by the native pathway.

It is not understood what genetic/biochemical mechanisms are used to turn on or off PHB production in the different nutrient starvation conditions. Direct genetic control of PHB production may be advantageous to achieve high levels of PHB under optimal growth conditions, eliminating the need for nutrient starvation for commercial scale production. IME strategies will be useful in identifying how genetic control of PHB production is achieved in the cell.

3 Development of a high-throughput screen for PHB

3.1. Nile red fluorescence screen for PHB

3.1.1. Introduction

Combinatorial approaches do not require much prior knowledge, but instead rely on methods for the rapid assessment of the phenotypes of a diverse mutant library to identify clones with improved properties. Inverse metabolic engineering (IME) (Bailey, Shurlati et al. 1996; Jin, Alper et al. 2005) embodies the essence of the combinatorial approach, comprised of generating genomic perturbations, such as gene knockouts and over-expressions, and then screening them for the desired phenotype. High scoring clones are then isolated, and the genomic perturbation is identified. Many genomic perturbations have been developed based on transposons, plasmid overexpression, gene shuffling and other random mutagenesis methods (Alper, Jin et al. 2005; Alper, Miyaoku et al. 2005). An integral component of combinatorial approaches is a high throughput screen for efficiently probing the library diversity. Towards this end, a

high throughput screen for measuring intracellular PHB was designed for the two organisms, *Synechocystis* and *Escherichia coli*. The screen must be have enough capacity to measure 10^5 mutants with the precision and accuracy to discern differences in PHB on the order of 5% PHB (DCW).

Prior PHB screening methods used nile red a dye that stains PHB and other neutral lipids in bacteria (Ostle and Holt 1982). These staining methods can be classified into two categories: (1) non-lethal and qualitative, and (2) lethal and quantitative. Nile red staining protocols that keep cells viable have only been able to differentiate non-PHB producing from PHB producing cells (Spiekermann, Rehm et al. 1999). Presumably, this is due to inefficient membrane permeability of nile red in living cells. While this is helpful for identifying organisms that make PHB, it lacks the sensitivity necessary to detect incremental improvements one might expect in an engineered strain over a parental strain. Prior quantitative assessments of PHB based on nile red fluorescence have involved various fixing steps (Degelau, Scheper et al. 1995; Müller, Lösche et al. 1995; Gorenflo, Steinbuchel et al. 1999; James, Mauchline et al. 1999; Vidal-Mas, Resina-Pelfort et al. 2001). The fixing step, typically executed with an alcohol or acetone treatment, facilitates the permeation of the dye through the membrane. While this allows for accurate detection of PHB levels, the

lethal nature of the protocol prevents its use in a screen for detecting mutants with improved PHB accumulation.

A staining protocol that can accurately measure PHB content in a combinatorial screen must allow the stain to enter the cell, specifically stain granules, and maintain high viability. Dye binding specificity (and the resulting fluorescence intensity) is a thermodynamic property and a function of the staining conditions (temperature, ionic strength, etc). Cell permeabilization methods have been developed for other applications, such as extracting proteins from the outer membrane (Vazquez-Laslop, Lee et al. 2001) or transforming bacteria with plasmids (Sambrook, Fritsch et al. 1989). Adapting such methods to transport nile red across the membrane could allow efficient staining of the PHB granules. In this chapter, we describe the development of two novel methods for detecting PHB based on nile red fluorescence that is both quantitative and maintains cell viability. Each protocol can distinguish incremental differences in PHB appropriate for library screening in *E. coli* or *Synechocystis*.

3.1.2. Materials and methods

Bacterial strains and growth medium

Synechocystis PCC 6803 was grown at 30°C in BG_{11} medium (Rippka, Deruelles et al. 1979). Cultures were grown in a

light-tight incubator (E-36, Percival Scientific, Boone, Iowa) at 100 µmol of photons m^{-2} s^{-1} provided by cool white fluorescent lightbulbs. Cultures containing different amounts of PHB were prepared by growing them in $BG_{11(P)}$, a modified BG_{11} medium with 0.018 mM K_2HPO_4 (10% of the concentration in BG_{11}), and/or supplementing with 10 mM acetate for 8-12 days. Cells were cultured in Erlenmeyer flasks on a rotary shaker. A 0% PHB DCW control strain was made using a PHB synthase knockout, $phaEC_{syn}^{-}$ (Taroncher-Oldenburg and Stephanopoulos 2000).

E. coli (XL-1 Blue, Stratagene, La Jolla, Calif.) transformed with a modified pJOE7 (Lawrence, Choi et al. 2005) plasmid was cultured at 37°C in Luria-Bertani (LB) medium containing 20 g/L glucose and 25 µg/mL kanamycin. The modified pJOE7, called pAGL20, was kindly provided by Dr. Anthony Sinskey and contains the gene *phaAB* from *R. eutropha* encoding the β-ketothiolase and the acetoacetyl CoA reductase, *phEC* from *Allochromatium vinosum* encoding the two subunit PHB polymerase, and encodes kanamycin resistance. As a no PHB control, the same plasmid without the *pha* genes was also cultured. Optical density was used to track cell growth using an Ulstrapec 2100pro (Amersham Biosciences, Uppsala, Sweden). *Synechocystis* and *E. coli* were tracked by absorbance at 730nm and 600 nm respectively.

Staining and fluorescence activated cell sorting (FACS)

A nile red (Sigma-Aldrich, St. Louis, MO) stock solution was made by dissolving to 1 mg/mL in dimethyl sulfoxide (DMSO) unless otherwise noted. 3 µL of stock solution was added to 1 mL of staining buffer as indicated in the staining optimization. FACS was carried out on a FACScan (Becton Dickinson, Mountain View, CA) using the following settings; *Synechocystis* FSC=E00, SSC=411, FL-1=582, FL-2=551 and *E. coli* FSC=E00, SSC=411, FL-1=582, FL-2=535. Cells were excited with an air-cooled argon ion laser (488 nm), and FL-2 (585nm) was used to detect nile red fluorescence. Flow cytometry analysis was done on 50,000 cells using WinMDI 2.8.

Staining effectiveness was characterized by resolution, R_S (Equation 3-1), where M_n is the

Equation 3-1: Resolution equation

$$R_S = \frac{2(M_1 - M_2)}{\delta_1 + \delta_2}$$

geometric mean of the fluorescence distribution of n (n=1 is the PHB producing cell, n=2 is the no PHB control). δ_n is the standard deviation of the fluorescence distribution. R_S is a quantitative measure of the ability to differentiate two populations.

Cell viability was accessed by the ratio of the colony forming units (cfu) in the final stained preparation to cells from the media.

Chemical PHB analysis

PHB was analyzed as shown previously (Taroncher-Oldenburg and Stephanopoulos 2000). >10 mg of cells was collected from culture by centrifugation (10 min, 3,200×g). The resulting cell pellet was washed once with cold deionized water and dried overnight at 80°C. The dry cells were boiled in 1 ml of concentrated sulfuric acid for 60 min, then diluted with 4 ml of 14 mM H_2SO_4. Samples were centrifuged (15 min, 18,000×g) to remove cell debris, and liquid was analyzed by high pressure liquid chromatography using an Aminex HPX-87H ion-exclusion column (300 x 7.8 mm; Bio-Rad, Hercules, Calif.) (Karr, Waters et al. 1983). Commercially available PHB (Sigma-Aldrich, St. Louis, Mo.), processed in parallel with the samples, was used as a standard.

Synechocystis staining optimization

Synechocystis WT and $phaE_{syn}^-$ cultures were grown for 7 days to early stationary phase. *Dye concentration optimization*: Cells were centrifuged (5 min, 3000×g), and resuspended to A_{730}=0.4 in 0.9%(w/v) sodium chloride solution. 3 uL of

different concentration nile red solution was added to 1 mL of resuspended cells to final concentrations between 30 to 30,000 ng/mL. The mixture was incubated in the dark for 30 min and analyzed on the FACScan. *Staining conditions optimization*: Deionized water and 0.9%(w/v) sodium chloride were used to resuspend the cells for staining to an A_{730}= 0.4. 3 µL of 10 mg/mL nile red in DMSO was added to 1 mL of resuspended cells. The mixture was incubated in the dark for 30 min and analyzed on the FACScan. *Destain buffer optimization:* Staining was performed as above. After staining, cells were centrifuged (5 min, 3000×g), and resuspended to the same volume in:1%(w/v) sucrose, 1%(w/v) DMSO, phosphate-buffered saline (PBS), PBS + 1%(w/v) DMSO, 0.9%(w/v) sodium chloride + 1%(w/v) DMSO, or deionized water. Cells were incubated in the dark for 30 min and analyzed on the flow cytometer.

E. coli staining optimization

E. coli XL1-Blue harboring the modified pJOE and the no PHB control were cultured as described. *Shock optimization*: Cultures were grown to stationary phase. Sucrose shock, isopropanol treatment, dimethyl sulfoxide treatment, and heat shock permeabilization methods were tested for resolution and viability after the shock as follows. Sucrose shock was carried out as shown previously (Vazquez-Laslop, Lee et al. 2001). 1 mL

of culture was cooled to 4°C for 10 min. The cells were then centrifuged (3 min, 3000×g, 4°C) and resuspended in 1 mL ice-cold TSE buffer (10 mM Tris-Cl [pH=7.5], 20%(w/v) sucrose, 2.5 mM Na-EDTA). The TSE mixture was incubated on ice for 10 min then resuspended (3 min, 3000×g, 4°C) in 1 mL deionized water with 3µL 10 mg/mL nile red in DMSO. The solution was incubated in the dark for 30 min and analyzed on the FACScan. Isopropanol shocked cells were prepared by centrifuging (3 min, 3000×g) and resuspending in 70%(w/v) isopropanol for 15 min. Cells were then centrifuged (3 min, 3000×g) and resuspended in deionized water with 3µL of 10mg/mL nile red in DMSO and incubated for 30 min in the dark and analyzed on the FACScan. DMSO shock was performed by centrifuging (3 min, 3000×g) 1mL of cell culture. 50 µL of 10 mg/mL nile red in DMSO was added directly to the pellet. The pellet was quickly vortexed and diluted to 1 mL in water after incubating for 30 s. Cells were incubated for 30 min in the dark and analyzed on the FACScan. Heat shock was performed as in competent cell preparation (Sambrook, Fritsch et al. 1989). 1 mL of cells was cooled on ice for 10 min. Cells were then centrifuged (3 min, 3000×g, 4°C), and resuspended in 1 mL cold 80mM $MgCl_2$ / 20 mM $CaCl_2$. The sample was then centrifuged (3 min, 3000×g, 4°C) and resuspended in 1 mL 0.1 M $CaCl_2$ with 3µL of 10 mg/mL nile red in DMSO. Cells were heat shocked at 42°C for 90 s and incubated for 30 min in dark then

analyzed on the FACScan. *Concentration optimization*: Cells were prepared by sucrose shock using 3µL of different nile red solutions to a final concentration between 30 – 30,000 ng/mL. *Sucrose concentration optimization*: Cells were prepared by sucrose shock using TSE buffer with varying sucrose concentrations (0,5,10,15,20%(w/v)). Nile red of 10 mg/mL in DMSO was used for staining.

3.1.3. Results

Synechocystis protocol development

To maximize the resolution between PHB producing strains and controls in *Synechocystis*, nile red concentration and staining buffers were optimized to increase specific staining while minimizing non-specific staining. A destain step was introduced and evaluated presuming it would remove nonspecific stain from the cells. Figure 3-1(A) shows the effect of nile red concentration on resolution for *Synechocystis*. 3.3 µg/mL nile red was found to yield the best resolving power. Nile red's fluorescence intensity is a function of the ionic character of the solution. Water and 0.9%(w/v) sodium chloride were examined as staining buffers to test the effect of solution ionic strength on nile red staining and fluorescence. Water showed a much higher difference in cell staining compared to the sodium chloride solution (Figure 3-1(B)). Several destaining

solutions were examined to attempt to reduce non-specific staining. While the sodium chloride-based buffers performed worse than the non-ionic buffers, none of the destaining buffers significantly improved the resolution of the assay (data not shown).

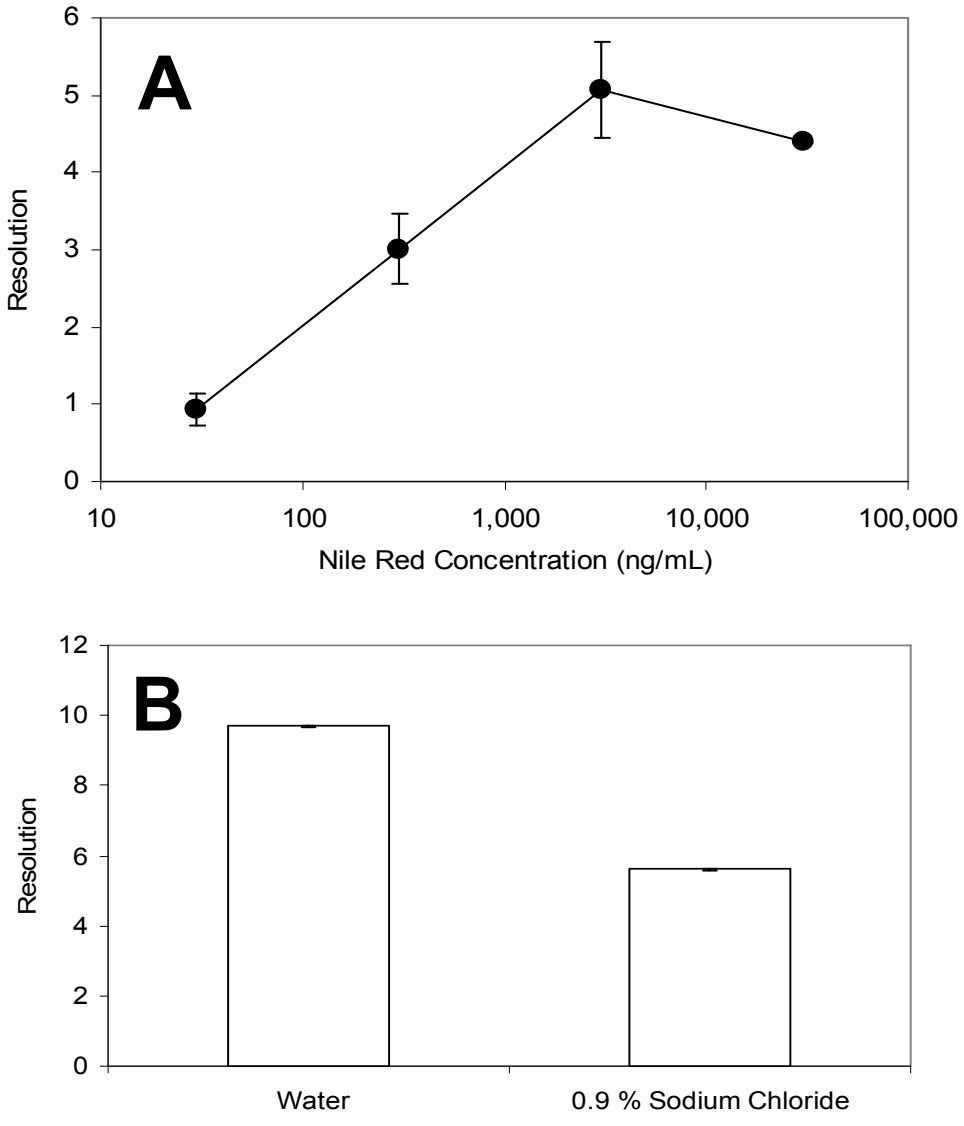


Figure 3-1: *Synechocystis* protocol optimization.
(A) Effect of nile red concentration on staining resolution; (B) Effect of buffer used for staining on resolution; RS is the resolution parameter (see text for definition) (all error bars represent standard deviation of duplicate experiments).

The optimized protocol is as follows: The cells were harvested by centrifugation (5 min, 3000×g) and resuspended to A_{730}=0.4 in deionized water. Next, 3 µL of a 1 mg/mL nile red in DMSO stock solution was added to 1 mL of the cell suspension and incubated in the dark for 30 min. Finally, the cells were immediately analyzed by flow cytometry.

Following protocol optimization, cultures grown under different conditions were assayed by the nile red staining protocol and chemical PHB analysis. As shown in Figure 3-2(A), the geometric mean of the flow cytometer measurement correlated very well with the analytical PHB measurements over a large dynamic range of PHB concentrations in different growth media. Comparison of the nile red fluorescence histogram for cells stained with 30 µg/mL nile red in 0.9 %(w/v) sodium chloride (Figure 3-2(B)) (protocol from (Gorenflo, Steinbüchel et al. 1999)) and the optimized protocol (Figure 3-2(C)) demonstrated a significant improvement in the overall staining. Non-specific staining in the overall population has been reduced by a shifting of the primary 0% PHB DCW peak to lower fluorescence, as well as a reduction of a secondary peak that was a result of strong non-specific staining.

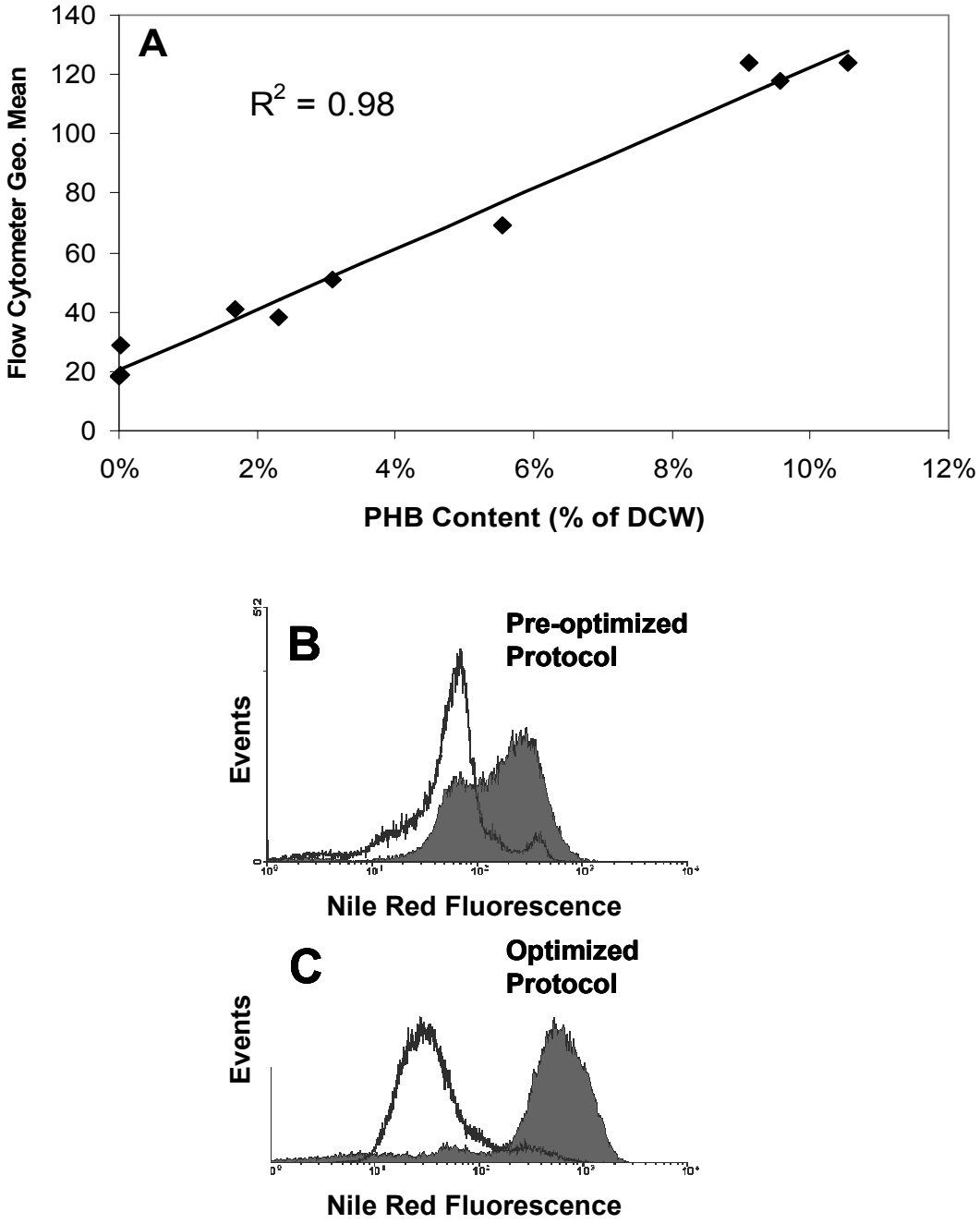


Figure 3-2: PHB staining effectiveness for *Synechocystis.*
(A) Correlation of fluorescence measurement from PHB staining protocol with chemical PHB measurement; (B) Histogram of 10% PHB DCW *Synechocystis* (gray) and 0% PHB DCW *Synechocystis* (black outline) using initial (low resolution) staining protocol; (C) Histogram of 10% PHB DCW *Synechocystis* (gray) and 0% PHB DCW *Synechocystis* (black outline) using final optimized staining protocol (see text for details).

E. coli protocol development

The optimized *Synechocystis* staining protocol was used as a starting point in investigating *E. coli* staining. Very poor staining of PHB was observed with this protocol, even in stationary phase where PHB accumulates to a level of approximately 40% DCW. Assuming that the observed staining differences are due to greater dye permeability through *Synechocystis* cell membranes, methods to improve the dye permeability of *E. coli* were examined. Figure 3-3(A) shows the resolution achieved by different permeabilization methods on stationary phase *E. coli* cells. Stationary phase *E. coli* was used to assure large differences in PHB. Of these, sucrose shock permeabilization gave the best resolution between high and no PHB strains, albeit with a very low viability of 0.022%.

To improve cell viability, the effect of varying the sucrose concentration used in the sucrose shock was studied. Exponential phase *E. coli*, which did not have as large a difference in PHB accumulation, was found to be more sensitive to the sucrose shock and was examined as a worst-case scenario for viability. Figure 3-3(B) shows the trade off between higher viability and better staining. Overall, a 10%(w/v) sucrose was found to yield similar staining characteristics to those obtained with higher sucrose concentration, while maintaining a

viability of 21%. Nile red concentration was varied to optimize resolution. A nile red concentration of 33 µg/mL gave the best separation of the peaks (Figure 3-3(C)). This increase in optimal nile red concentration compared to *Synechocystis* is most likely due to increased amounts of PHB in *E. coli*. Concentrations above this level were not soluble in the stock solution and gave considerable noise in the measurement (data not shown). Cell viability was not affected by nile red concentration (data not shown).

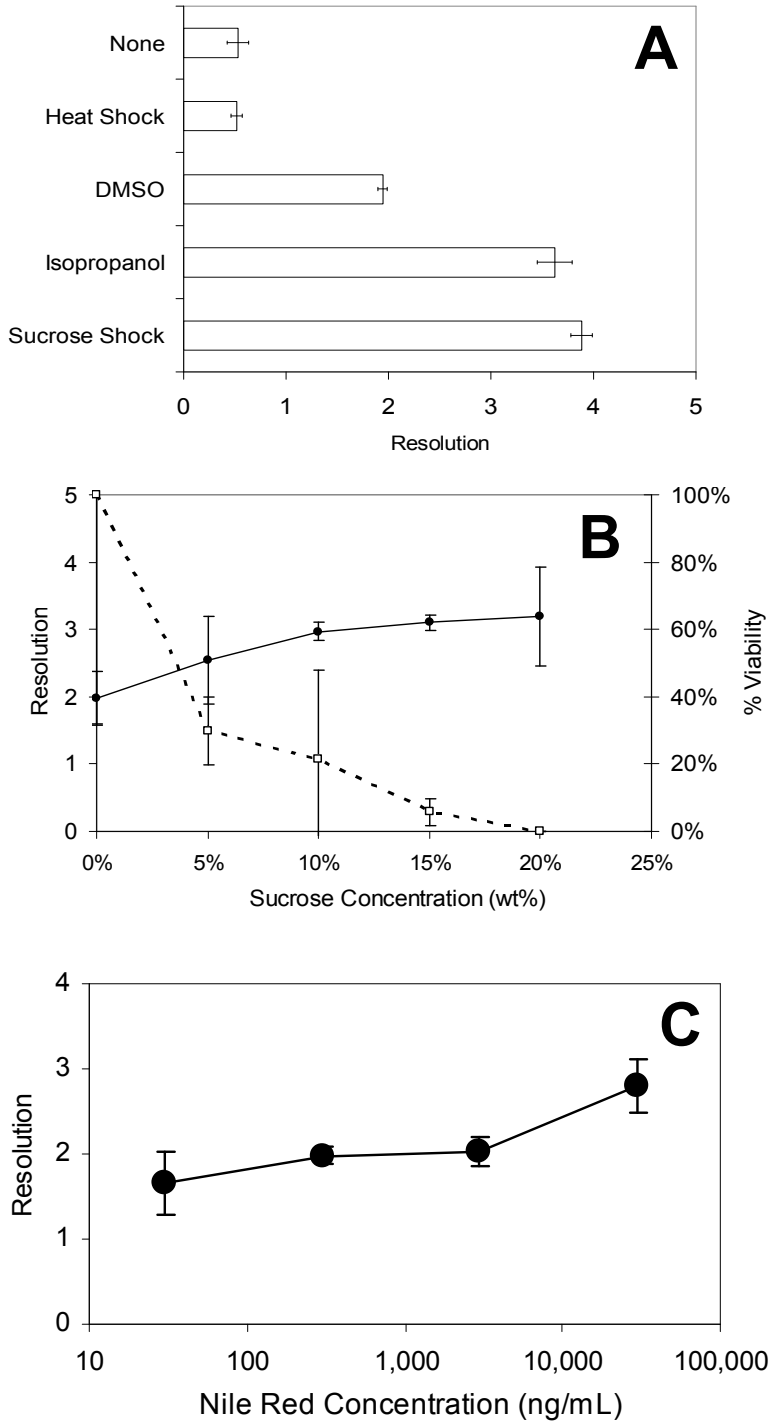


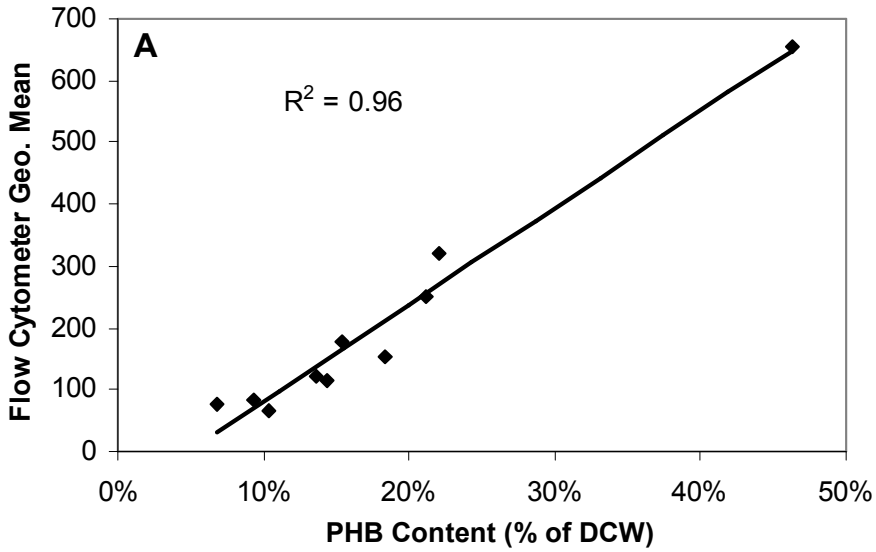
Figure 3-3: *E. coli* protocol optimization.
(A) Effect of different osmotic shock protocols on staining [see Materials and Methods for details]; (B) Effect of sucrose concentration on staining resolution and cell viability [●] Resolution of PHB and no-PHB controls [□] Viability of PHB control; (C) Effect of nile red concentration on staining resolution.

To further improve the viability of the cells after the staining procedure, different resuspension buffers used after the TSE incubation step were evaluated. Water, LB, and 1 mM magnesium chloride broth had a viability of 21%, 18%, and, 48% respectively. While even with the magnesium chloride only half the population survived, this only doubled the number of cells that would needed to be sorted in order to screen a library.

The optimized *E. coli* staining protocol is as follows: Cells were cooled to 4°C, and then harvested by centrifuging (5 min, 1000×g, 4°C). Next, cells were resuspended to A_{600}=0.4 in 10%(w/v) sucrose TSE buffer, and incubated on ice for 10 min. After this, the sample was centrifuged (5 min, 3000×g, 4°C) and resuspended to the same volume in 4°C 1 mM $MgCl_2$. This was followed by the addition of 3 µL of a 1 mg/mL nile red in DMSO to 1 mL of the cell suspension and incubated in the dark for 30 min. Finally, the cell population was immediately analyzed by flow cytometry.

Figure 3-4(A) shows the correlation between the PHB measurement obtained by the optimized *E. coli* nile red staining protocol and by chemical PHB analysis. The data shown in Figure 3-4(A) correspond to cells harvested at different time points along a growth curve over a broad range of PHB amounts from exponential to stationary phase. Figure 3-4(B) and (C) show the dramatic improvement in staining achieved in the final protocol.

As indicated in Figure 3-4(B), a large portion of the population was not stained in the original protocol but was stained after introducing the sucrose shock step.



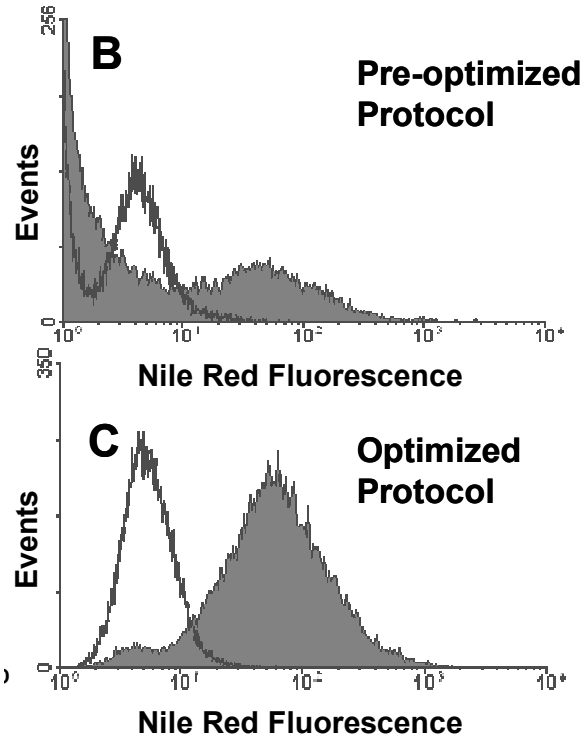


Figure 3-4: PHB staining effectiveness for *E. coli.*
(A) Correlation of fluorescence measurement from PHB staining protocol with chemical PHB measurement; (B) Histogram of ~40% PHB DCW *E. coli* (gray) and 0% PHB DCW *E. coli* (black outline) using *Synechocystis* staining protocol; (C) Histogram of ~40% PHB DCW *E. coli* (gray) and 0% PHB DCW *E. coli* (black outline) using final optimized staining protocol (see text for details).

3.1.4. Discussion

High resolution FACS experiments can sort cells from a population only to the extent that the fluorescence level of the cell correlates with the phenotype that is being examined. When trying to develop a staining protocol that yields good correlation between fluorescence and, in this case, PHB levels, important factors are, (a) promoting specific fluorescent molecule binding to the target (PHB), and, (b) allowing adequate access of the fluorescent molecule to the target.

In the nile red optimization for PHB fluorescence, molecule specificity was addressed by varying the environment for staining. This was especially important in *Synechocystis*. *Synechocystis* has multiple layers of thylakoid membrane which is used in its photosynthetic apparatus. These membranes provide a large area of lipid-like interfaces for non-specific binding of nile red. To minimize binding of the nile red to the thylakoid membrane, ionic strength of the medium and dye concentration were examined. Changing the ionic strength of the staining environment from a sodium chloride solution to deionized water, the nile red stained the PHB more specifically. Additionally, the dye concentration also strongly affected the resolution of the assay. In *E. coli*, this was not as important. The effect of dye concentration increased, unlike in *Synechocystis* where a maximum was observed. The ionic strength did not affect the

resolution in *E. coli* (data not shown). These observations can be attributed to the lack of large membrane structures in *E. coli*.

Providing consistent access of the fluorescent molecule to the target is also necessary for a quantitative fluorescence level. In *Synechocystis*, the stain readily permeated the cell and stained the PHB granules. Figure 3-2(B) shows that even in the prior staining protocol, all *Synechocystis* cells were being stained. This should be contrasted with Figure 3-4(B) which shows that a large portion of the *E. coli* did not stain at all for PHB in the non-optimized protocol. *Synechocystis* is a naturally competent cell, and as such is able to take up DNA molecules readily. This may imply the morphology of *Synechocystis's* membrane may allow it to take up nile red more readily than *E. coli*, which is not naturally competent. To improve the dye transport across the *E. coli* cell membrane, competent cell protocols and other permeabilization methods were attempted. Of these, sucrose shock permeabilized the cells in such a way that the nile red could enter the cytoplasm and stain the granule.

While the *E. coli* cells could now take up the nile red, most of the cells were killed in the process. Further optimization was required to increase the cell viability while retaining good staining properties. Adjusting sucrose

concentration and the buffers used improved the viability to 48%. This will allow an adequate efficiency for screening mutant libraries by FACS.

To validate the use of resolution (Eq. 1) as a metric for optimizing the protocol and to estimate the accuracy of the nile red fluorescence, the geometric mean of the fluorescence distribution was compared to a chemical PHB measurement of the culture. As there is presently no validated method for measuring PHB levels at the individual cell level, population average measurements, geometric mean of fluorescence and the whole culture chemical PHB measurement, were required to assess the quantitative accuracy of the staining protocols. The correlation between fluorescence and PHB content were greatly improved from initial staining experiments (data not shown) due to the improved staining of PHB granules and reduction in non-specific staining. The estimated error of prediction of PHB content from the geometric mean of fluorescence was ± 1.2% PHB DCW and ± 4.5% PHB DCW for *Synechocystis* and *E. coli* respectively (95% confidence of interval). From this, it can be inferred that the PHB levels on the single cell level can be estimated accurately based on the fluorescence measurement.

These protocols will allow single cell measurements of PHB levels in *Synechocystis* and *E. coli* to a precision that mutants with incrementally increased PHB accumulation can be sorted from

the library and characterized. Using FACS, 10 million cells can easily be assayed in less than 1 hr. While there will be a loss due to nonviable cells in the *E. coli* system, this loss does not prohibit the assay from screening genome scale libraries. As well, multiple cells of the same genotype will be present due to growth, increasing the likelihood of each library variant being screened.

Biological noise will most likely contribute false positives to the screen. Inherent in all single cell measurements is the cell-to-cell variation even in a clonal population. This is evident in Figure 3-2(C) and Figure 3-4(C). For the positive controls, a clonal population has a 10 fold difference in fluorescence within the population. This variation in PHB content will result in false positives being sorted as high PHB clones, while their average PHB content may be less.

The application of sucrose shock to allow *E. coli* to take up nile red is generalizeable to other bacteria and other small molecule dyes which do not permeate the membrane. By using such permeabilzation methods to allow impermeable fluorescent dyes to enter the cytoplasm, the number of phenotypes that can be screened in a high throughput fashion can be significantly increased. This will enable new fluorescence-based

combinatorial screens for other phenotypes where high throughput screens do not currently exist.

3.2. Dielectrophoresis screen for PHB

Besides the fluorescence of nile red, other approaches were initially investigated as means to sort cells based on PHB content, including dielectrophoresis, density gradient centrifugation, and momentum-based microfluidics. Density gradient centrifugation and momentum-based microfluidics were an attempt to separate cells based on their density or weight/size. While PHB content did alter these parameters of a cell, many other factors also changed this, and these methods proved incapable of reliably isolating mutants with high PHB content. Dielectrophoresis was pursued extensively but in the end was not as effective as the nile red fluorescence screen.

Dielectrophoresis field flow fractionation (DEP-FFF) is a chromatographic method for separating micron size particles by their electrical properties (Markx, Rousselet et al. 1997; Rousselet, Markx et al. 1998). DEP-FFF uses an electric force, which is dependent on electric properties and geometry of the cell, to levitate cells to different heights in a parabolic flow field. This microfluidic technique was explored as a possible screen for high PHB cells. DEP-FFF sorts based on electrical

properties of the cell and could be used for screening other phenotypes that would change the cell's electrical properties.

The DEP force is a function of intracellular content and the extracellular medium. The magnitude and direction of the DEP force is dependent on a variety of factors including the media conductivity, cytoplasmic conductivity, and the voltage, frequency, and geometry of the AC electric field (Gascoyne and Vykoukal 2002). We hypothesized that the cytoplasmic conductivity of a cell is a strong function of the amount of polymer accumulated. This is because the volume of PHB in the cell is nonconductive, and as this volume changes, so does the average conductivity of the cytoplasm. In a given set of conditions, cells with high PHB might have a different DEP force than cells with little PHB.

To test this hypothesis, we placed *E. coli* that were PHB+ or PHB-, as described above, consecutively on a quadrapole microelectronic microscope slide, overlaid with a solution containing the cells and 0.4 X phosphate-buffered saline. A 2.5V AC electric field was induced on the quadrapole. Figure 3-5 shows that PHB+ *E. coli* are attracted to electrodes when the AC field is at 1 MHz, while PHB- *E. coli* are not. At higher frequencies, both cells are repelled from the electrodes. The difference in forces was confirmation that DEP was affected by

PHB content, and could potentially be used for a separation of the two types of cells in a DEP-FFF device.

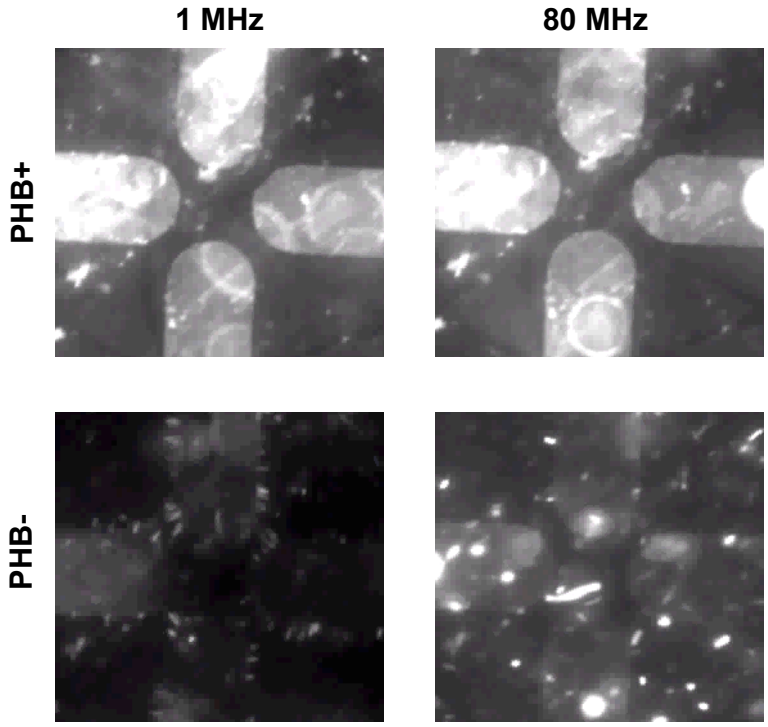


Figure 3-5: Dielectrophoretic force of *E. coli*.
Negative DEP forces pull PHB- cells to the electrodes at 1 MHz, but not PHB+ cells. Neither cell is attracted at 80 MHz. XL-1 Blue (pAGL20) [PHB+] and XL-1 Blue (pAGL19) [PHB-] were suspended in a 2 parts phosphate buffered saline (PBS) and 3 parts water. A 2.5 V AC electric field was induced at 1 MHz and 80 MHz. pAGL19 is the same backbone plasmid as pAGL20, without the PHB operon.

Field flow fractionation takes advantage of these PHB-dependent DEP forces. Figure 3-6 gives a schematic representation of the device. A sample loop is used to load cells into the front of the DEP-FFF chamber. Flow is stopped, and the cells are allowed to settle to the bottom of the chamber. At this time, the signal generator is turned on, inducing the electric field. The DEP force pushes the cells vertically away

from the electrodes (see Figure 3-6 inset). The strength of the DEP force decays with distance from the electrodes. When the field is turned on, the cells equilibrate at a height where the DEP force and gravitational force are equal. Since the magnitude of the DEP force is a function of the PHB content of the cells, the equilibrium height will be different for cells with different amounts of PHB.

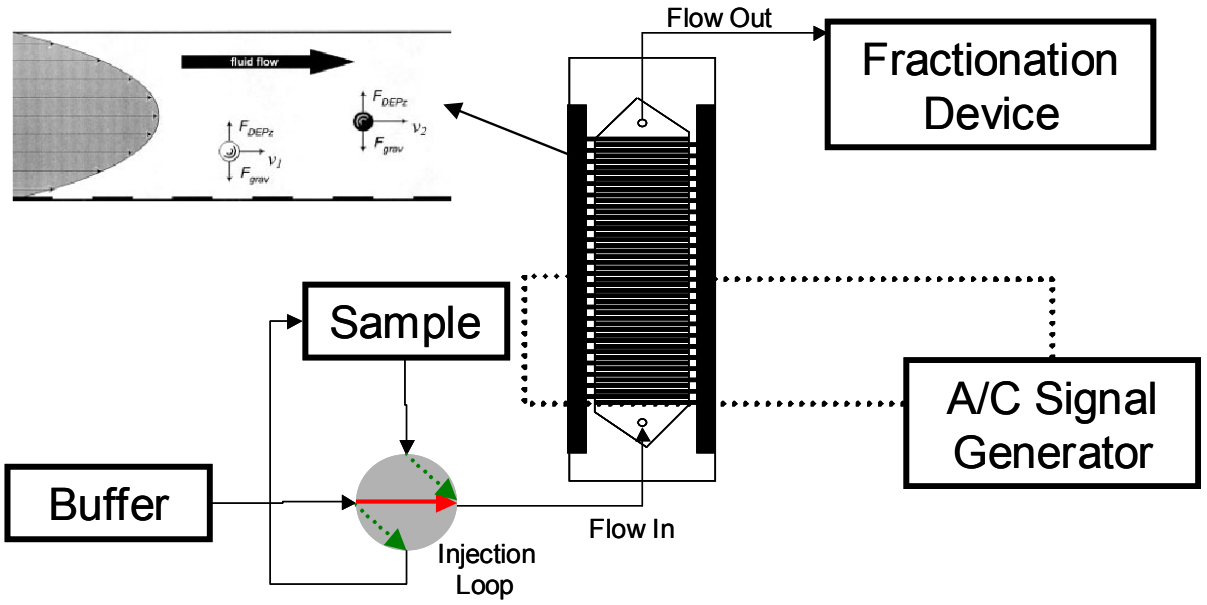


Figure 3-6: Schematic representation of DEP-FFF.
See text for description. Upper left inset from Gascoyne and Vykoukal (2002).

After the cells have reached their equilibrium height, buffer is pushed through the device. The parabolic flow profile causes cells that are near the centerline to travel quickly through the device and cells near the bottom to travel slowly (all cells are well below the half height to avoid convolution do to the upper half of the parabola). This causes high PHB

cells to have a short residence time in the device and are collected in early fractions.

Several parameters had to be defined in order to correctly separate cells based on PHB. Geometric (flow chamber dimensions, electrode width and spacing) and operational (solution conductivity, voltage and frequency of AC field, fluid flow rate) parameters affect the ability to discriminate on PHB. While several of these parameters were adjusted to improve separation conditions in the DEP-FFF device, satisfactory separations were never achieved. Although DEP force was a function of PHB accumulation, other cellular properties, such as viability, cell size, and culture age, affected DEP and convoluted the separation. This was in contrast to nile red fluorescence, which was primarily a function of cellular PHB content. DEP-FFF was at not as mature as FACS and would required significant further investigation to have a useful screen. While currently this technology is not ready for mass use, further development promises significantly cheaper approaches to sorting, compared to the large capital costs of FACS machines.

4 Identification of gene disruptions for increased poly-3-hydroxybutyrate accumulation in *Synechocystis* PCC 6803

4.1. Introduction

Given that a high throughput screen has now been established, we turned our attention to identifying genetic nodes that increase PHB accumulation in *Synechocystis*. Although all polyhydroxyalkanoates ultimately use carbon dioxide and light as a feedstock, *Synechocystis* offers a direct route, unlike heterotrophic bacteria, which would require cultivation of starches in plants, purification, and finally fermentation. A one step PHB synthesis process, whereby a single organism is employed to both carry out photosynthesis and convert sequestered carbon dioxide directly to PHB, would eliminate the current need for two organisms: (1) to carry out photosynthesis to create sugars and/or fatty acids, and (2) to convert the sugars/fatty acids to PHB. This process can be achieved in both plants and microbial photosynthetic organisms (Suriyamongkol, Weselake et al. 2007). Library mutagenesis approaches are not

scaleable to plant systems, making cyanobacteria an attractive platform for IME approaches to improve PHB production. Identifying gene targets in *Synechocystis* is interesting to both engineer the organism and guide rational approaches for engineering plants for improved PHB production.

In this study, we use an IME approach to generate a library of transposon-based gene disruptions in *Synechocystis.* This library is screened for improved PHB accumulation using a high throughput fluorescence activated cell sorting screen and gene disruptions are identified that increase PHB accumulation. We have thus obtained improved strains as well as identified genetic markers for further improvements in other photosynthetic systems.

4.2. Materials and methods

4.2.1. Strains, culture conditions, reagents

Synechocystis PCC 6803 was grown at 30°C in BG11 medium (Rippka, Deruelles et al. 1979). Cultures were grown in a light-tight incubator (E-36, Percival Scientific, Boone, Iowa) at continuous 100 µmol of photons m^{-2} s^{-1} provided by cool white fluorescent lightbulbs. Nutrient modifications to the growth medium were: BG11(P), a modified BG11 medium with 0.018 mM K_2HPO_4 (10% of the concentration in BG11); BG11(A), BG11 supplemented with 10 mM sodium acetate; or BG11(P)(A), BG11(P) supplemented

with 10 mM sodium acetate. Cells were cultured in 250 mL Erlenmeyer flasks with a foam stopper to allow gas transfer on a rotary shaker at 100 rpm.

DH5α (Invitrogen, Carlsbad, CA) was used for cloning steps. pUC19 (Invitrogen, Carlsbad, CA) was used for various cloning steps and as a suicide vector for recombination in *Synechocystis*. Kanamycin (kan) was used at 50 µg/mL for DH5α and 100 µg/mL for *Synechocystis*.

All restriction enzymes, ligases, Taq DNA polymerase, PCR and cloning reagents were purchased from New England Biolabs (Beverly, MA) and cloning was performed using standard protocols (Sambrook, Fritsch et al. 1989).

4.2.2. Development of transposon library in *Synechocystis*

A library of *Synechocystis* genomic DNA fragments cloned into pUC19 was kindly provided by DuPont Co. *In vitro* transposon mutagenesis was performed on the purified plasmid library using EZ::TN<Kan-2> Insertion Kit (Epicentre Biotechnologies, Madison, WI). This kit includes a transposon DNA fragment containing a kanamycin resistance marker held in the active site of a transposase that will randomly insert the transposon into the plasmids in the reaction mixture. Plasmids were subsequently transformed into DH5α, and transformants were plated on kanamycin Luria-Bertani (LB) agar plates, so only

plasmids with successful insertions were selected and amplified. About 20,000 kan^R colonies were scraped from the LB plates and the plasmids were isolated using Qiagen Miniprep kit (Qiagen, Hilden, Germany). The plasmids bearing segments of *Synechocystis* genomic DNA with transposon insertions were then transformed into *Synechocystis* by electroporation (Chiaramonte, Giacometti et al. 1999). Wild type (WT) *Synechocystis* was grown to exponential phase (A_{730}~0.5), washed twice with 0.3 volume deionized water, and finally resuspended in 0.02 volume (A_{730}~14) deionized water. 40 µL cells were mixed with 2 µL mutagenized plasmid library (2.5 µg) and electroporated at 2.5 kV, 25 µF, 200 Ω, using a Gene Pulser (Bio-rad, Hercules, CA) with 0.2 cm gap distance electroporation cuvette. The transformed *Synechocystis* was plated on BG11 agar plates with kanamycin and colonies that have the transposon recombined into the genome were selected and scraped into a single freezer stock after 10 days.

4.2.3. Flourescence-based PHB measurement and screening

The *Synechocystis* library was inoculated into BG11, BG11(P), BG11(A), BG11(P)(A) to an A_{730}=0.022 and cultured for 14 days. Cells were quantitatively stained with nile red to give a fluorescence signal proportional to PHB content (Tyo, Zhou et al. 2006). Briefly, cells were resuspended in deionized water to an

A_{730}=0.4. 3µL of a 1mg/mL nile red (Sigma-Aldrich, St. Louis, MO), in dimethyl sulfoxide solution was used to stain intracellular PHB granules to fluoresce at 585 nm, and 1 uL of a 1 mM bis-(1,3-dibutylbarbituric acid)trimethine oxonol (oxonol) (Invitrogen, Carlsbad, CA) was used as a live/dead stain which fluoresces at 530 nm. Cells were incubated in the dark for 30 min at room temperature then sorted on a MoFlo fluorescence activated cell sorter (FACS) (Dako, Carpinteria, CA). Cells were sorted based on the PE filter set (for nile red fluorescence) and FITC filter set (for oxonol) Cells were collected that did not fluoresce for oxonol (living) and were in the top 0.1% of the nile red fluorescence distribution (high PHB). Collected cells were plated on BG11 agar, and individual colonies were isolated. The location of the transposon in high PHB producing isolates was identified using thermal asymmetric interlaced (TAIL) PCR (Liu and Whittier 1995) using primers homologous to the transposon insert (Table 4-1 TAIL primers).

Table 4-1: Oligonucleotides used in *Synechocystis* study.

Name	*Description*	*Sequence (5'-3')*
Identifying Transposon Location		
TAIL AD1		NTCGASTWTSGWGTT
Kan-2 TAIL #1		ACACTGGCAGAGCATTACGCTGACTTG
Kan-2 TAIL #2		CGGCGGCTTTGTTGAATAAATCGAACT
Kan-2 TAIL #3		CAGACCGTTCCGTGGCAAAGCAA
Integrating Transposons to WT		
slr1670(sense)	NdeI	GGAATTCCATATGAATTTACCTAGAGGTGTTTACGCCCCTGGG
slr1670(anti)	SacI	ACGCGAGCTCAGTAATTCCAGATCACAGGCCGCCG
sll0377(sense)	NdeI	GGAATTCCATATGTTGTTGATATTTTCCCAGTGTCGGCGGAGCT
sll0377(anti)	SacI	ACGCGAGCTCTCCCAATTCCTCAATACCTTCAATGCGA
sll0683(sense)	NdeI	GGAATTCCATATGTTGCTGTACCCATCGGCGTT
sll0683(anti)	SacI	ACGCGAGCTCGCATTAACCCTTCAATCTTCAAG
sll8035(sense)	NdeI	GGAATTCCATATGAACTGGGCTGAGTCCACAATGCGA
sll8035(anti)	SacI	ACGCGAGCTCCGTAGTACCCGCATTAACTGCCCCG
Validating Transposon Integrations		
slr1670(s) VAL		TGGGCAAAATCCCGTTGGGACTC
sll0461(s) VAL		CGTTAGGCGACTCTGGATGGGATCA
sll0565 (a) VAL		GCATTGTGGAACTGGCCCAACG
sll0377(a) VAL		AGCTGGGGCCAACACTCCTAAACCC
sll0683 (s) VAL		TGGATTCCCCTCTTTAGCACCCCG

4.2.4. Transferring transposon insertions to parental strain

To verify the identified gene disruptions were sufficient for the observed changes in PHB production, the same transposon insertion was introduced into a WT *Synechocystis*. This was accomplished by one of two methods, electroporation of PCR products (Chiaramonte, Giacometti et al. 1999), or recombination from a suicide vector (Vermaas, Williams et al. 1987). For electroporation, a fragment of DNA including the transposon and 750 bp flanking *Synechocystis* DNA on either side was amplified by PCR from the genomic DNA of the isolated mutant and gel purified. *Synechocystis* was prepared for electroporation as above. Suicide vector recombination relied on amplifying a similar PCR fragment that was cloned into pUC19 using NdeI and

SacI sites that were introduced in the primers. The pUC vector with cloned *Synechocystis* flanked transposon was used to recombine the transposon to the same genetic location in WT *Synechocystis* by incubating the plasmid with exponentially growing cells and relying on the natural competency of *Synechocystis* (Vermaas, Williams et al. 1987). Transformants were selected on kanamycin BG11 agar plates in both methods. Correctly transferred transposons were verified by PCR.

4.2.5. PHB analysis

PHB was analyzed as in Chapter 3.2.3 Chemical PHB Analysis

4.3. Results

4.3.1. Library development

While transposon mutagenesis has been used routinely in other organisms to create gene disruption libraries that elicit different phenotypes, traditional approaches have proved unsuccessful in *Synechocystis*. Instead a two step approach was used (Bhaya, Takahashi et al. 2001). *In vitro* transposon mutagenesis was used first to insert kanamycin markers into plasmids that bear random fragments of *Synechocystis* genomic DNA (Figure 4-1(I)). After *in vitro* transposon mutagenesis of the *Synechocystis* genomic library, 20,000 Kan^R colonies harboring a genomic library with insertional mutagenesis were rescued for

amplification and isolation of mutagenized plasmids. In a second step, these mutagenized plasmids were transformed into wild type *Synechocystis* (Figure 4-1(II)), and the transposon integrated into the genome by homologous recombination determined by the flanking DNA on the plasmid (Figure 4-1(III)). The plasmid bearing the *Synechocystis* genomic DNA can not replicate in *Synechocystis* and was subsequently lost. Because the flanking DNA on each plasmid was from a different region of the genome, the transposons inserted and disrupted individual genes in the genome. 10,000 mutants with transposons in random locations were generated by this method. Given 3,168 ORFs, this gives about 3x coverage, or 95% confidence that each gene has been disrupted. The constructed library can be screened for many different phenotypes and represents a useful library for photosynthetic organisms.

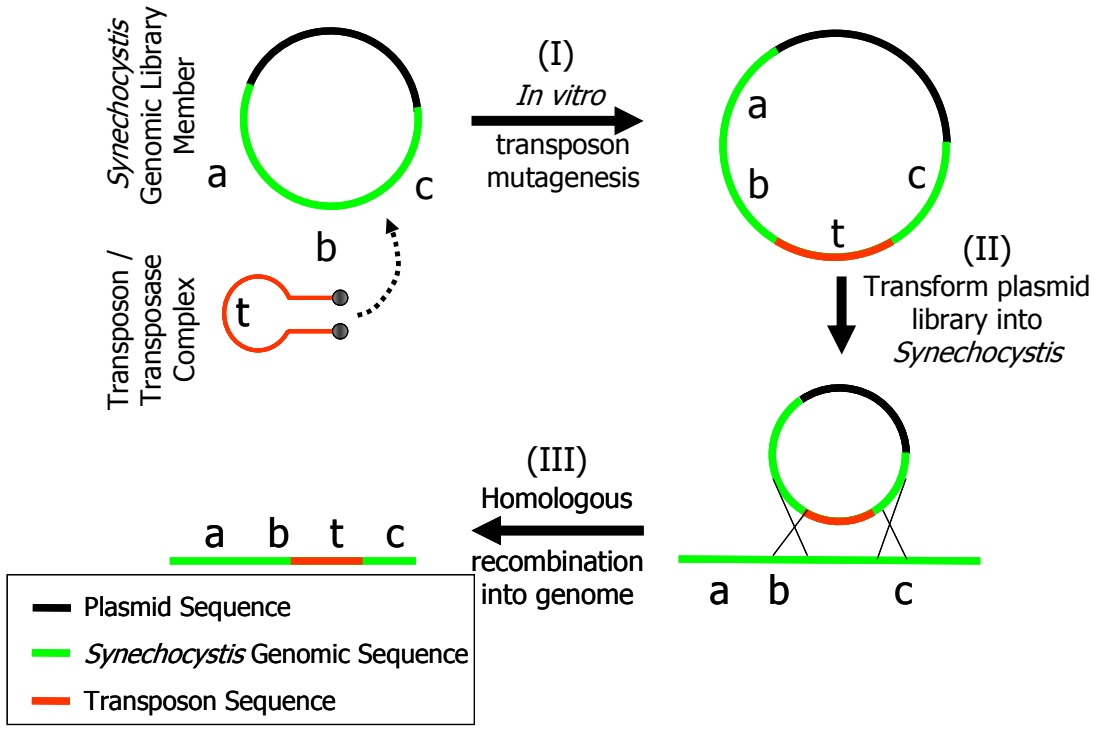


Figure 4-1: Library development strategy.
(I) A commercial in vitro transposon / transposase kit is used to randomly insert transposon DNA into members of a Synechocystis genomic library cloned into pUC19. [a,b,c are consecutive arbitrary sequences in the Synechocystis genome. t is the transposon sequence including a kanamycin resistance marker.] (II) The mutagenized genomic library is transformed into WT Synechocystis. (III) DNA sequences flanking the transposon are recombined with homologous fragments in the genome. This integrates the transposon and causes a disruption in the gene. The location of the integration will be determined by the region of DNA that was on the plasmid.

4.3.2. Library screening

The transposon mutagenesis library was grown in four different media supplementations known to elicit varied PHB accumulation, BG11, BG11(P), BG11(A), and BG11(P)(A). After 14 days, cell populations were sorted by FACS and individual mutants with high fluorescence from nile red were isolated.

Figure 4-2(a) shows the sorting configuration of the FACS for the library grown in BG11. FITC fluorescence of oxonol was used to identify dead cells so that only viable cells were sorted. Over the four nutrient conditions, 34 clones were isolated from the screen and further characterized.

Each isolated mutant was regrown under conditions identical to those under which it was sorted. PHB accumulation was estimated using the nile red fluorescence protocol after 7 and 14 days (Figure 4-2(b)). Fifteen clones that displayed significantly higher nile red fluorescence at one or both of these time points were further assayed using the more accurate HPLC-based method. Of the 15 clones that were isolated from the FACS screening, improved PHB accumulation was confirmed by HPLC for only 8 (8 of 31 true positives isolated from the initial screen). Figure 4-3 shows the measured PHB in these isolated clones. The transposon insertion site for clones that had increased PHB was identified using TAIL PCR to amplify DNA adjacent to the transposon. The genome location was identified by sequencing the adjacent DNA and BLASTing against the *Synechocystis* genomic sequence (http://www.kazusa.or.jp/cyanobase/).

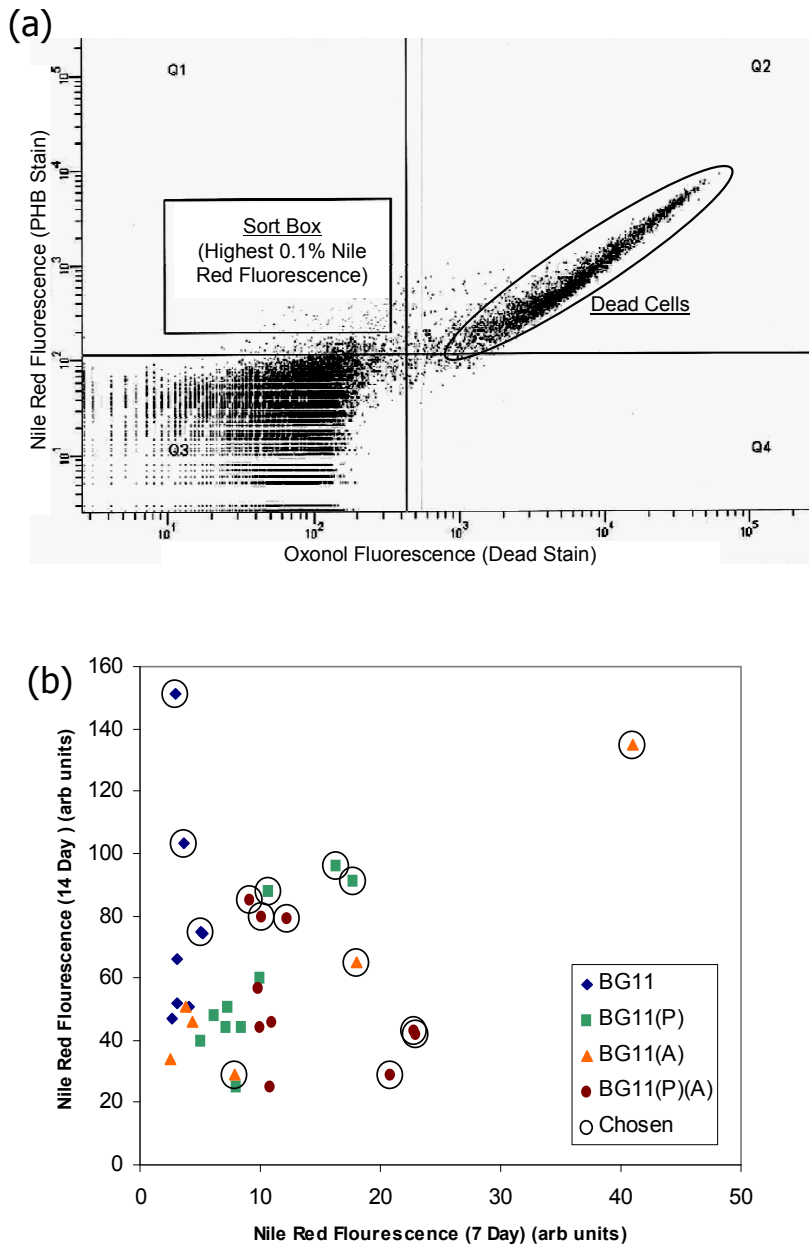


Figure 4-2 Library screening.

(a) Distribution of fluorescence measured by FACS in library grown in BG11. Cells from the top 0.1% of the nile red (PHB) fluorescence distribution that were below the threshold for the dead stain (oxonol) fluorescence were chosen (square box). Dead cells tend to emit high fluorescence in the nile red fluorescence range and would lead to sorting many non viable cells, but were avoided by detecting dead cells from the high oxonol fluorescence (oval) (b) Nile red (PHB) fluorescence of cultures grown from isolated clones after 7 and 14 days. Circled clones were chosen based on high nile red fluorescence at 7 or 14 days. ♦ - Clones isolated from BG11 ■ – Clones isolated from BG11(P) ▲ – Clones isolated from BG11(A) ● – Clones isolated from BG11(P)(A)

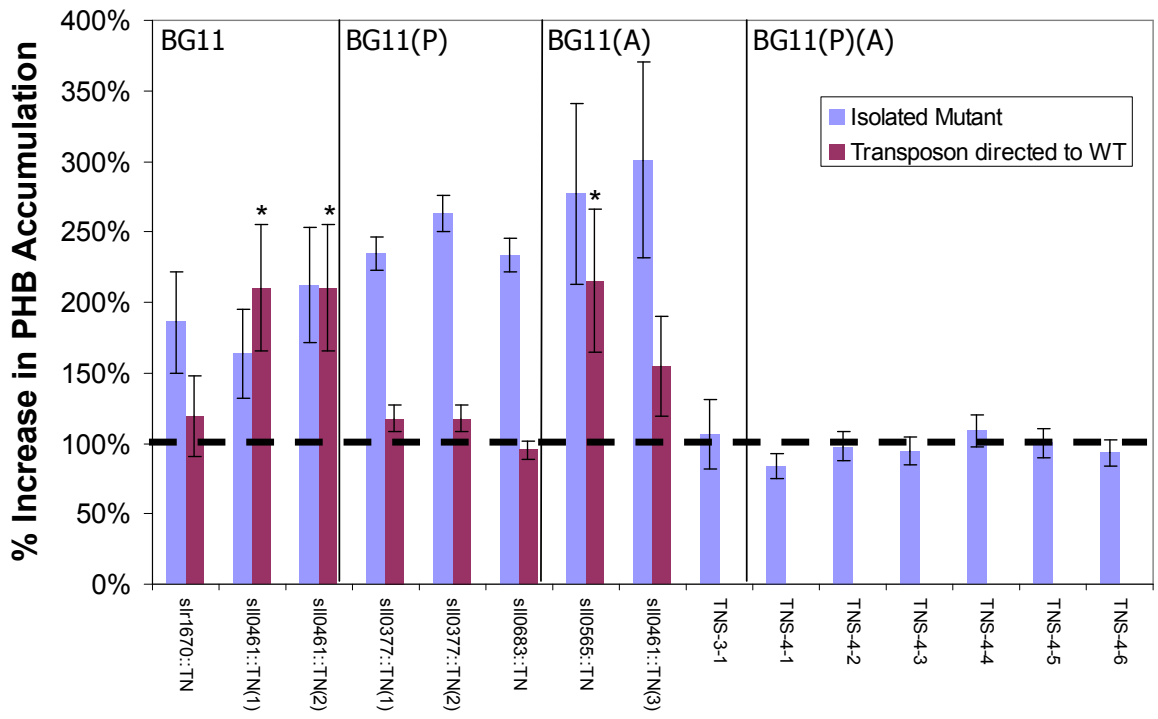


Figure 4-3 Quantitative analysis of isolated clones and transposons systematically introduced to WT by HPLC.

PHB from isolated clones that were circled in **Figure 4-2**(b) was analyzed (blue). For isolated clones that showed improvement, the transposon insertion was inserted into a WT strain at the same location to verify the transposon was sufficient for increasing PHB. The dotted line represents the WT accumulation of PHB in each condition (BG11/BG11(P)/BG11(A)/BG11(P)(A)). WT PHB accumulation: BG11 = 3.5% BG11(P) [10% Phosphate] = 6.6%; BG11(A) [10 mM Acetate supplement] = 2.6% BG11(P)(A) [10% Phosphate with 10 mM Acetate supplement] = 12.5% (All are %PHB (DCW)).
* - p-values <0.05

To verify that the transposon insertion was sufficient for the increase in PHB accumulation, the transposon was introduced at the same site in the WT strain and the latter was assayed for improved PHB accumulation. Figure 4-3 (Transposon directed to WT) shows the PHB accumulation of mutants identified from the screening step and the PHB accumulation after the transposon was

transferred to a new WT strain. Of the 8 initial isolates, only 4 transposon insertions were found to elicit increased PHB upon transferring the transposon to a WT strain, suggesting that additional chromosomal mutations may have contributed to the PHB improvements of the initial isolates.

4.3.3. Location of transposon insertions

Three insertions were found in sll0461 and one insertion in sll0565 (Figure 4-4). sll0461 is a 1,263 bp ORF encoding a gamma-glutamyl phosphate reductase (*proA*) used in the metabolic pathway from glutamate to proline. sll0461 is the first ORF in an operon with two other genes (ssl0900 and sll0462), encoding a very small hypothetical protein and a predicted permease, respectively. A total of three clones were found with insertions in sll0461, two from the BG11 culture, and one from the BG11(A) culture. The two mutants isolated from BG11 both had insertions in the same site, 770 bp from the start of the gene. Translation of this transcript would yield a 267 aa product including 257 aa from the original sll0461, but truncating the last 164 aa. The location of the transposon insertion isolated from the BG11(A) occurred at the end of the ORF at 1,112 bp, only changing the last 50 aa of the C-terminus. In both of these insertion events, the RNA polymerase is unlikely to transcribe through the transposon, preventing

transcription of ssl0900 and sll0462. It can not be determined by this study which of these three ORFs play a role in the altered PHB anabolism.

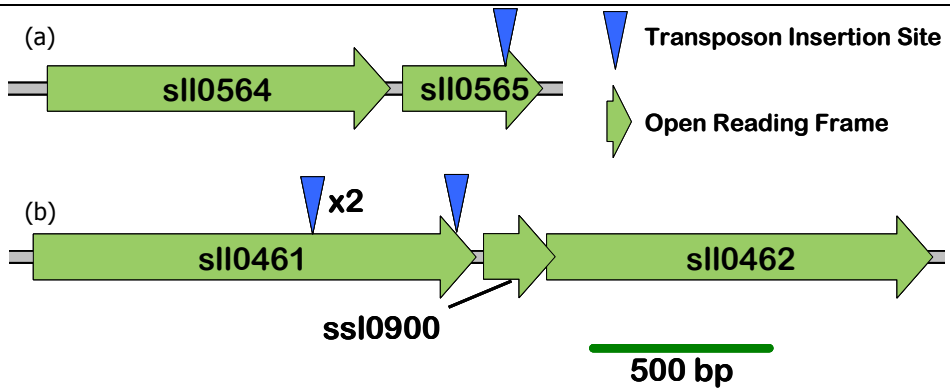


Figure 4-4 Context of transposon insertions.
(a) Locations of transposon insertions in sll0565. (b) Locations of transposon insertions in sll0461. Green arrows are annotated ORFs.

sll0565 is a 401 bp ORF of unknown function that is the second ORF in a 2 gene operon. The transposon in sll0565 was inserted at 307 bp, changing the last 31 aa, but keeping the length almost the same (133 aa for WT, 135 aa with transposon insertion). This insertion occurs at the end of the operon, and does not inhibit the transcription of any downsteam ORFs.

4.3.4. PHB accumulation in improved strains

Growth and PHB accumulation phenotypes of the improved mutants, which were identified in either BG11 or BG11(A), were characterized in all four media conditions (BG11,BG11(P),BG11(A), and BG11(P)(A)). Figure 4-5(a)-(d) show the growth profiles for the transposon insertions in the two ORFs compared to wild type.

Without phosphate limitations (Figure 4-5(a),(c)), all cultures grew up to A_{730}~12, while with lowered phosphate (Figure 4-5(b),(d)), biomass could only accumulate to A_{730}~3. Under all growth conditions, there was no difference in the growth profile of the two mutants. PHB accumulation at 7 days was not different from WT, but differences arose at 14 days (Figure 4-6). While both mutants had higher PHB accumulation in BG11, BG11(P), and BG11(A), the increases in BG11 were the largest. Interestingly, sll0565::kan was found to have improved yields of PHB in BG11, even though it was isolated from the BG11(A) condition, and both were found to have slight increases in PHB accumulation in BG11(P), though neither were found in this condition during the screen. This implies that the mode of PHB accumulation in both strains is independent of acetate presence, does not inhibit growth, and becomes more pronounced at high cell densities.

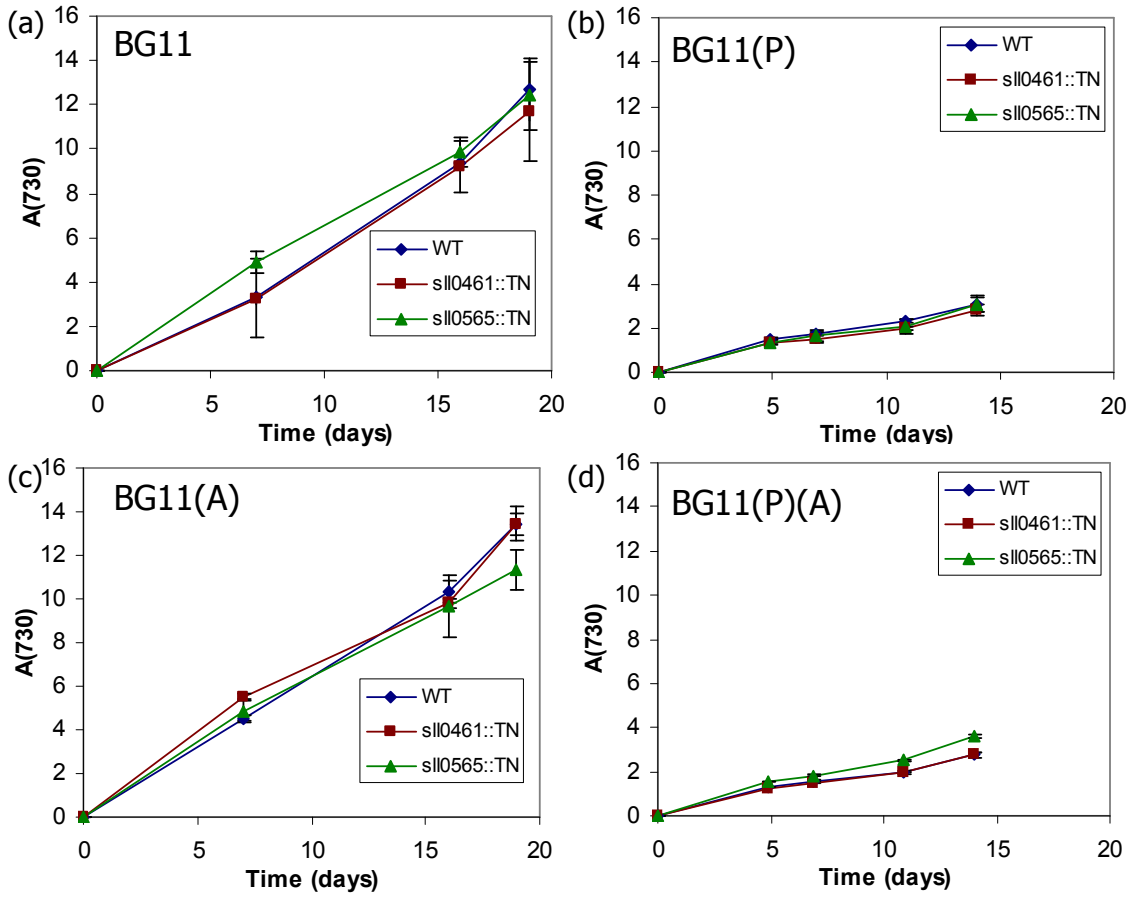


Figure 4-5 Growth time course data for improved mutants in different growth media. (a) Growth of improved mutants in BG11. (b) Growth in BG11(P) [10% Phosphate]. (c) Growth in BG11(A) [10 mM Acetate supplement]. (d) Growth in BG11(P)(A) [10% Phosphate with 10 mM Acetate supplement]. ♦ - WT Synechocystis; ■ – Transposon inserted in sll0461 Synechocystis; ▲ – Transposon inserted in sll0561 Synechocystis.

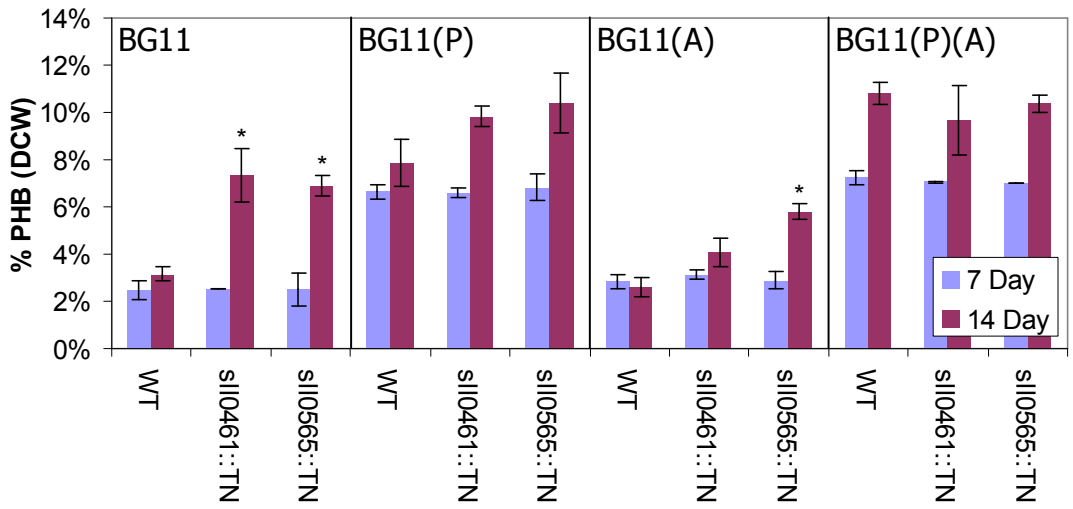


Figure 4-6 PHB accumulation of *Synechocystis* transposons.
7 and 14 days. Blue – 7 days; Red – 14 days. BG11, BG11(P), BG11(A), BG11(P)(A) and strains as in Figure 4-5. * - p-value < 0.05.

4.4. Discussion

Inverse metabolic engineering strategy is dependent on genetic diversity in the library, and thereby the phenotypes that can be produced. Transposon mutagenesis is an attractive technique for introducing genetic diversity because it only delivers one change per cell and the genetic change can be easily located and transferred to other host cells. Genetic methods for direct transposon delivery are not available for *Synechocystis* as for more traditional model organisms like *E. coli* and *S. cerevisiae*. This forced a two step approach to transposon mutagenesis. Both steps had to provide library sizes much greater than the number of *Synechocystis* genome ORFs in order to provide good overall coverage. The first step, *in*

vitro transposon mutagenesis of the plasmid library, readily yielded ~20,000 transpositions when transformed into *E. coli*. The second step, transforming the mutagenized plasmids to *Synechocystis*, yielded ~1,500 transformations per standard reaction (~2×10^6 cells), consistent with previous observations (Kufryk et al. 2002). Seven parallel transformations were used to scale up to the target size of 10,000 clones.

There are inherent difficulties with screening for a phenotype by only measuring one cell compared to measuring a large clonal population. This difficulty arises from the effect of sample size on the ability to estimate the mean of a population. In this study, only 25% of the clones isolated from the FACS with the highest single cell PHB levels showed increased PHB accumulation upon measuring the population grown from that cell. Unlike colony-based and 96 well-type screening, which rely on the averaged measurement of many cells from a clonal population, thereby precisely estimating the mean, FACS allows the measurement of only one cell, reducing the confidence that the measurement represents the mean. This is compounded by a very large variance caused by biological noise. In a clonal population accumulating PHB, the coefficient of variation in PHB fluorescence is ~100% (data not shown). The ability to separate cells whose means are not significantly larger than the biological variance is inherently difficult, because the

biological noise will be larger than the expected increases in PHB. Protein evolution, which often uses single cell measurements, can obtain 10 fold increases in phenotype in a round of screening (Crameri, Raillard et al. 1998). However, as was the case in this study, increases of 20%-50% are typical for metabolic fluxes, making the improved mutants difficult to isolate due to the biological noise.

It was also observed that for many of the mutants that did exhibit PHB increases, the transposon insertion alone was not sufficient to elicit the improved PHB phenotype (Figure 4-3). This was the case in five of the eight mutants that had improved PHB, including the highest PHB accumulating mutants. This implies that there are other genetic changes or epigenetic effects contributing to PHB accumulation independently or in cooperation with transposon insertions. This is particularly true in the case of phosphate starvation, where significant increases in PHB were observed (up to 16% PHB (DCW)), but not linked to the transposon insertion (Figure 4-3, BG11(P)). Accessing the nature of these changes was not possible in this study.

Two insertions were found to be sufficient for improved PHB production in *Synechocystis* under various culture conditions. These insertions were in *proA*, an enzyme in the proline biosynthetic pathway, and an unknown protein. sll0461 encodes

one of two copies of *proA* found in *Synechocystis*, sll0373 being the other. It is interesting to note that while three mutants were isolated with insertions in sll0461, none were found in sll0373. Proline acts as an important protective osmolyte in higher plants and reduced cytoplasmic proline may signal other stress responses that favor PHB production. In this context we note that Δ^1-pyrroline-5-carboxylate synthase (*P5CS1*), a sll0461 homologue in higher plants, is expressed in response to dehydration and high salt (Ábrahám, Rigó et al. 2003). sll0462, a gene most likely silenced by the insertion in sll0461, is homologous to predicted permeases involved in the purR regulon in *Nostoc punctiforme* PCC 73102. The insertions in the sll0461/sll0462 gene cluster affected PHB accumulation more dramatically at higher cell densities (14 day vs 7 day) and appeared to be suppressed by the presence of acetate in the media. The transposon does not affect growth of the mutant, and the increased diversion of carbon to PHB suggests that there is excess carbon dioxide fixing capacity that could be harnessed for product formation. sll0565 has significant homology to the myosin heavy chain (*AT3G53350*) in *Arabidopsis thaliana*, which is involved in locomotion along filamentous actin in plants (Kinkema, Wang et al. 1994). As with sll0461, the insertion did not affect growth, and the differences in PHB accumulation were observed later in culture growth. However, unlike the

insertions found in sll0461, the sll0565 mutant performed about the same regardless of acetate addition.

While a biochemical mechanism could not be inferred by the location of the insertions, promising genetic loci have been identified for future studies. Recently, there has been an emphasis on heterologous expression of the PHB operon in plants (Bohmert et al. 2000; Menzel et al. 2003; Poirier 2002). As *Synechocystis* fixes carbon dioxide through the Calvin cycle, it is likely that targets found in *Synechocystis* could be relevant in future efforts to engineer higher plants for PHB production.

In this work, we have generated genetic diversity by delivering gene insertions via transposon mutagenesis to *Synechocystis*. The library was grown under different conditions known to affect PHB accumulation. Using a fluorescence-based assay, the amount of PHB was estimated on a single cell level, and cells with high fluorescence for PHB were isolated from the library and characterized. Insertions in two ORFs, sll0461 and sll0565, were found to significantly increase PHB in standard BG11 media, as well as, smaller improvements in acetate supplementation and phosphate limitation.

5 FME perturbations for PHB productivity in *E. coli*

5.1. Introduction

This chapter represents a shift from the combinatorial approaches of inverse metabolic engineering (IME) described in Chapter 3 & 4 to a more rationally-driven forward metabolic engineering. IME approaches were pursued at great length in *E. coli*, but were not successful. Although mutants could be found that increase PHB, these mutants could not successfully be related to the genetic perturbations that were introduced in the libraries. As will follow, a case was built that the PHB productivity is controlled by the 3 step PHB biosynthetic pathway. The PHB pathway was not genetically perturbed in the previous IME screenings, meaning changes outside the pathway could not increase PHB.

In particular, we studied PHB formation in the setting of growth phase product formation. Extensive literature, that was review in Chapter 2, has focused on batch phase operations, where the PHB is made primarily in stationary phase. Because PHB is an intracellular product, there must always be enough

cellular biomass to surround the PHB granules. It follows then, that PHB production could be achieved in a continuous manner, where biomass formation and PHB production occur simultaneously.

Growth phase product formation requires a delicate balance between product formation and biomass formation (growth). Given that specific substrate uptake rates are finite, the resources diverted to biomass formation and product formation can not exceed a defined rate forcing a trade off between the two. While product formation is the objective of a biochemical process, increasing product formation will eventually inhibit growth by starving the cell for resources necessary for growth or by the toxicity of accumulated products or intermediates in the product forming pathway, thus reducing the rate of biocatalyst production. From an overall productivity, a precisely defined product formation rate is essential to maximize yield and productivity.

This is especially true in the case of poly-3-hydroxybutyrate, where the product is expected to use a large portion of the glucose uptake and is stored intracellularly (Anderson and Dawes 1990), thereby potentially hindering growth rate significantly. PHB production in recombinant *E. coli* is attractive for a number of reasons: high yields are possible, purification of the polymer from *E. coli* is much easier than from the native organism, and there are no known PHB

depolymerases in *E. coli* (Madison and Huisman 1999). Chemostat processes may be attractive for PHB production in *E. coli* because continuous process enjoy greater utilization of installed capital, lower equipment sizes, and increased volumetric productivities (when calculated including down-times inherent in batch processes), which may be helpful for this low value product. Furthermore, only a limited amount of PHB can be stored intracellularly, requiring simultaneous generation of both biomass and PHB for continuous production. To date little work has explored continuous production for PHB. Previous studies have shown that PHB accumulation primarily occurs in stationary phase and very little accumulation happens in log phase growth (Wang and Lee 1997). These characteristics would not be amenable for growth phase accumulation, and thus new engineering strategies need to be employed to improve growth phase accumulation.

Pathway activities have been implicated as a possible flux limitation. *In vitro* enzyme activity measurements of the PHB pathway and kinetic models have implicated limitation in the pathway activity (Sim, Snell et al. 1997; van Wegen, Lee et al. 2001). In particular, AAR activity has been measured as low, although no direct experiments have shown whether this limits flux to PHB. In this work, we study the effects of stepwise overexpression of the PHB pathway on pathway flux. Separately,

the growth rate is varied through dilution rate in a nitrogen limited chemostat to vary biomass demand for acetyl-CoA. These experiments show that the activity of the PHB pathway primarily controls PHB flux, and careful tuning of that pathway activity will be required for optimal product/biomass formation ratio.

5.2. Materials and methods

5.2.1. Strains, plasmids, primers and media

Strains and plasmids used in this study are listed in Table 5-1. pAGL20, a modified pJOE7 kindly provided by Anthony Sinskey, contains the genes *phaAB* from *R. eutropha*, encoding the β-ketothiolase and the acetoacetyl coenzyme-A reductase, *phaEC* from *Allochromatium vinosum*, encoding the two-subunit PHB polymerase on a kanamycin resistant backbone (Lawrence, Choi et al. 2005). pZE21 is a ColE1 plasmid with kanamycin resistance and green fluorescent protein (*gfp*) driven by a P_L-tetO promoter (Lutz and Bujard 1997).

Table 5-1: Strains and plasmids used in *E. coli* study

Name	*Description*	*Reference*
Strains		
XL1-Blue	Cloning/Expression Strain of *E. coli*	Stratagene (La Jolla, Calif.)
K12 *recA*::*kan* TGD (cat+PHB)	30 tandem copies of PHB biosynthetic operon from pAGL20 on *E. coli* genome	(Tyo and Stephanopoulos, Submitted)
Plasmids		
pAGL20	PHB biosynthetic pathway on modified pJOE7	(Lawrence, Choi et al. 2005)
pZE21	Medium copy plasmid (ColE1 origin, kan^R)	(Lutz and Bujard 1997)
pZE-Cm	pZE21 with kan^R replaced with Cm^R	
pZE-Cm-phaA	pZE derivative with *cat* and R. eutrophus *phaA*	
pZE-Cm-phaB	pZE derivative with *cat* and R. eutrophus *phaB*	
pZE-Cm-phaEC	pZE derivative with *cat* and R. eutrophus *phaA*	
pZE-Cm-gfp	pZE derivative with *cat* and *gfp*	
pZE-Cm-phaECAB	pZE derivative with *cat* and pAGL20 PHB operon	
pZE-kan-tacpha	pZE derivative with pAGL20 PHB operon driven by strong promoter (p_{tac})	

Cloning was performed using standard techniques and materials from New England Biosciences (Beverly, MA), and all cloning steps were performed in DH5α (Invitrogen). Table 5-2 lists all primers used for PCR. pZE21 was digested SacI/AatII and ligated to a chloremphenical acetyl transferase PCR product from pAC184 bearing the same sites to create pZE-Cm. Promoterless PCR products of *phaA*, *phaB*, *phaEC*, and *phaECAB*, were generated from pAGL20 and cloned into the KpnI/MluI sites of pZE-Cm for the systematic overexpression study. These plasmids were co-transformed with pAGL20 into XL-1 Blue and characterized. Whole operon promoter replacement was accomplished by synthesizing the tac promoter from

oligonucleotides and cloning into the AatII/EcoRI site of pZE21. A promoterless *phaECAB* PCR product from pAGL20 was then cloned into the KpnI/MluI sites. This plasmid was transformed into XL-1 Blue and characterized.

Table 5-2: Oligonucleotides used in *E. coli* study.

Name	*Restriction Site*	*Sequence (5'-3')*
Cm (s)	AatII	AAA***GACGTC***CGTTGATCGGCACGTAAGAGGTTCC
Cm (a)	SacI	AA***GAGCTC***CCTTAAAAAAATTACGCCCCGCCC
phaA (s)	KpnI	GG***GGTACC****GC*ATGACTGACGTTGTCATCGTATCCGC
phaA (a)	MluI	CG***ACGCGT***CGGAAAACCCCTTCCTTATTTGCG
phaB (s)	KpnI	GG***GGTACC****GC*ATGACTCAGCGCATTGCGTATGTGAC
phaB (a)	MluI	CG***ACGCGT***CCGACTGGTTGAACCAGGCCG
phaEC (s)	KpnI	GG**GGTACC**GACGGCAGAGAGACAATCAAATCATG
phaEC (a)	MluI	CG**ACGCGT**ATGGAAACGGGAGGGAACCTGC
tac promoter (s)		GAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGG
tac promoter (a)	AatII / EcoRI	***AATT***CCACACATTATACGAGCCGATGATTAATTGTCAACAGCTC***ACGT***
qPCR phaA (s)		CGTTGTCATCGTATCCGCCG
qPCR phaA (a)		GACTTCGCTCACCTGCTCCG
qPCR phaB (s)		GTGGTGTTCCGCAAGATGAC
qPCR phaB (a)		CGTTCACCGACGAGATGTTG
qPCR phaE (s)		GGAGCAGAGCCAGTATCAGG
qPCR phaE (a)		CACCCTGGATGTAGGAGCCC

K12 *recA::kan* TGD(cat+PHB) was created using tandem gene duplication, and contains 25 copies of the *phaECAB* operon in tandem on the *E. coli* genome at the λ phage integration site(Tyo and Stephanopoulos 2008). (Details of this strain will be presented in Chapter 6.) This strain was used for the chemostat studies because of the improved genetic stability over plasmid-based expression systems.

Strains were cultured in the minimal media, MR with 20 g/L glucose at 37°C (called MR from here forward). MR medium was prepared as defined previously (Wang and Lee 1997). Luria-Bertani (LB) broth was used for growth on solid media and standard preparations of cells.

5.2.2. Systematic overexpression studies

XL-1 Blue was transformed with the prescribed plasmid(s). Colonies were picked and grown in a 14 mL culture tube with 5 mL of LB for 12 hrs. 0.5 mL of the culture was inoculated into a 14 mL culture tube with 5 mL MR and cultured for 12 hrs. Cells were then inoculated into a 250 mL Erlenymer flask with a working volume of 50 mL MR + 10 mg/L thiamine to a starting A_{600}=0.015. Cultures were grown to steady state growth and harvested at A_{600} = 2.0 - 2.5. Prior experiments showed cells were at steady state, exponential growth at this point in the cell culture(data not shown). Growth rate, % PHB (DCW), and transcriptional measurements were taken as described below. Specific PHB productivity was estimated as the product of growth rate and percent PHB (DCW), which is exact at the limit of steady state growth. 34 µg/mL chloramphenicol and/or 25 µg/mL kanamycin was used to maintain plasmids in each experiment.

5.2.3. Chemostats

Nitrogen-limited chemostat experiments were performed in a 3 L stirred glass vessel using the BioFlo 110 modular fermentation system (New Brunswick Scientific, Edison, NJ) with a 1 L working volume. Bioreactor controllers were set to pH = 6.9, adjusted by 6 N NaOH through controller, 30% dissolved oxygen, controlled by adjusting feed oxygen concentration, and temperature at 37°C, controlled by a thermal blanket and cooling coil. Gas flow was set at 3 L/min and agitation at 400 rpm. Antifoam SE-15 (Sigma-Aldrich, St. Louis, MO) was diluted in water and added by peristaltic pump to control foaming.

Sterile MR media without antibiotics was fed by peristaltic pump at flowrates of 50,100, and 300 mL/h and culture broth was removed from the reactor using a level-stat. 10 mL samples were taken for characterization after four residence times at a given flowrate. Approach to steady state was verified by monitoring glucose concentration in reactor. Three measurements, each spaced one residence time apart, were taken at each steady state. Ammonium, the only nitrogen source, was monitored semi-quantitatively by NH_4^+ test strips (Merck KGaA, Darmstadt, Germany), and was always below 10 mg/L (the lowest measurable value on the test strip). Chemostats were run in replicate, and glucose was in excess at all points except $D=0.1h^{-1}$, where a dual nutrient limitation existed. Contamination was checked by

streaking broth on nile red LB plates and checking that all colonies were PHB^+ by nile red fluorescence.

5.2.4. Analytical methods

Cell densities were monitored at 600 nm using an Ulstrapec 2100pro (Amersham Biosciences, Uppsala, Sweden). Cell and PHB concentrations were determined as previously described in Chapter 3.2.3. Glucose concentrations were measured by a glucose analyzer (Yellow Springs Instruments, Yellow Springs, OH). Acetate was measured by high-pressure liquid chromatography using an Aminex HPX-87H ion-exclusion column (300 7.8 mm; Bio-Rad, Hercules, Calif.) and 14 mM H_2SO_4 mobile phase at 50°C and 0.7 mL/min. Cell culture supernatant was kept at -20°C before analysis.

5.2.5. Transcriptional analysis

RNA was isolated from growing cells at A_{600} = 2.0 using the RNEasy Protect Bacteria mini kit (Qiagen USA, Valencia, CA), according to the manufacturer's protocol, and diluted to 20 ng/µL total RNA. 60 ng isolated RNA was converted to cDNA using the ImpromII reverse transcriptase (Promega, Madison, WI) with 1 µg random hexamers according to the manufacturer's protocol. cDNA concentrations of *phaA*, *phaB*, and *phaE* were determined by qPCR using a Biorad iCycler and the iQ SYBR Green Supermix (Biorad) using 5 µL of a 100-fold dilution of the cDNA reaction,

according to the manufacturer's protocol. qPCR primers are in Table 2. A standard curve was made using dilutions of pAGL20, cut once to relax supercoiling. Controls without reverse transcriptase were used to verify that contaminating DNA was negligible.

5.2.6. Flux balance analysis

Flux balance analysis was done with a previously existing stoichiometric model (Edwards and Palsson 2000), supplemented with three metabolic reactions and one exchange flux which together allow for PHB synthesis from acetyl-CoA: i) 2 AcCoA → AcAcCoA + CoASH; ii) AcAcCoA + NADPH → 3HBCoA + NADP; and iii) 3HBCoA → PHB + CoASH. The exchange flux was PHB → ∅, representing the net formation and export of PHB from the metabolically active portion of the cytoplasm.

The stoichiometric trade-off between growth yield and PHB yield (on glucose) was calculated as described by Bell and Palsson (see the introduction of (Bell and Palsson 2005)). Briefly, the PHB yield was constrained to a certain value *s*, biomass yield was maximized subject to the usual FBA constraints as well as the PHB constraint. By repeating the optimization at varying values of *s*, we obtained the full projection, in the two dimensions of PHB yield and biomass yield, of the feasible region of flux space.

Calculations were done with MATLAB software (The Mathworks, Natick MA), and the TOMLAB interface (Tomlab Optimization, Inc., Trefasgatan, Sweden) to the linear program solver CPLEX 9.0 (ILOG, Gentilly, France).

Biomass and PHB yields calculated from the model were converted to productivities by multiplying by the specific glucose uptake rate. Glucose consumption was measured at 1.1 g glucose $gRCW^{-1}$ h^{-1}, yielding a maximal growth rate (from stoichiometric model) of 0.57 h^{-1} and maximal PHB production rate of 530 mg PHB / $gRCW^{-1}$ h^{-1}.

5.3. Results

5.3.1. Validating conditions for balanced growth

Measuring PHB specific productivity was a key parameter in the analysis of PHB pathway regulation. The direct measurements, growth rate and PHB content, must be taken while the cells are growing at steady state. While chemostats were used in the second part of the study to measure productivities, chemostats were not conducive to measuring many different strains in parallel. Instead, a protocol was developed to measure PHB at pseudo-steady state, or balanced, growth in shake flasks. To this end, we measured the PHB accumulation from early log phase through early stationary phase in a shake flask (Figure 5-1). Because it was difficult to have a repeatable lag phase, data

was analyzed vs A_{600} rather than vs time. As can be seen, PHB content is high as the cell begins to grow, due to accumulation from the stationary phase inoculum. As the initial effects are diluted away by growth, the PHB content falls. Finally, as the cells enter stationary phase, growth slows, and the PHB content rises again. Between A_{600} = 2-3, the PHB is constant, and represent a regime where PHB accumulation is at steady state. To verify the PHB accumulation was at steady state, balanced growth, the strain were subcultured into pre-warmed shake flasks as the culture approached A_{600}=2, and the PHB content was measured over several subculturings (data not shown). PHB content did not change, confirming that the PHB accumulation was at steady state. The region between A_{600} = 2-3 was chosen for further assays.

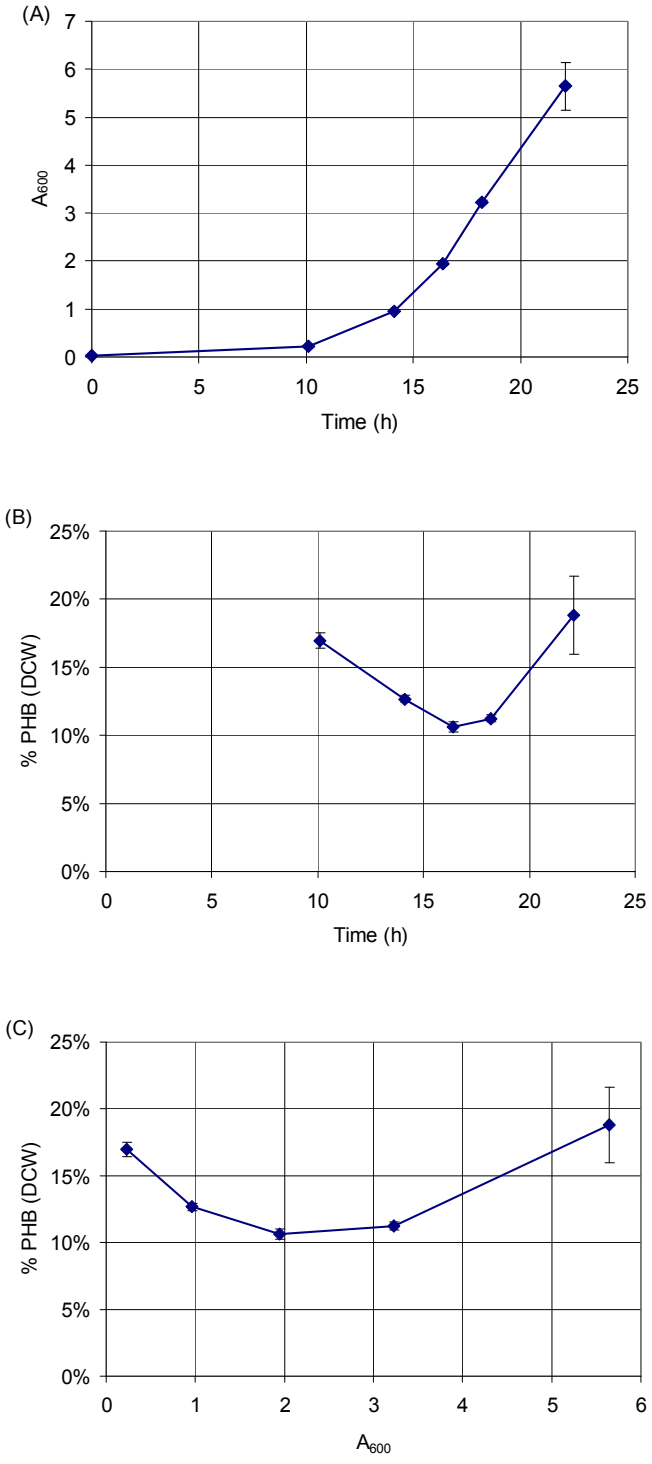


Figure 5-1: PHB accumulation during batch growth.
XL-1 Blue (pAGL20) was grown in a shake flask in MR. PHB content was measured as the culture reached different cell densities. (A) Cell density vs time. (B) PHB vs time. (C) PHB vs Cell density. Direct comparisons between experiments were best done based on cell density rather than time, due to highly variable lag phases.

5.3.2. Model estimates of PHB pathway flux control

To better understand how the PHB flux might respond to changes in activity of the PHB pathway enzymes, a sensitivity analysis was performed using the model by van Wegen, et. al. (2001). Matlab code used for this analysis is in Appendix I. Figure 5-2 shows the variation of *phaA*, β-ketothiolase (β-KT), and *phaB* (AAR), while *phaEC* (PHB synthase) is held constant. Other model parameters, such as precursor and co-factor concentrations, were set at published values for *E. coli.* The model shows that at the native expression levels, the PHB flux is most sensitive to changes in AAR. After AAR activity is increased 3-fold however, β-KT becomes limiting. This matched well with our results which showed *phaB* limited flux to PHB, but after the *phaB* limitation is relieved, whole operon expression was necessary to achieve higher levels.

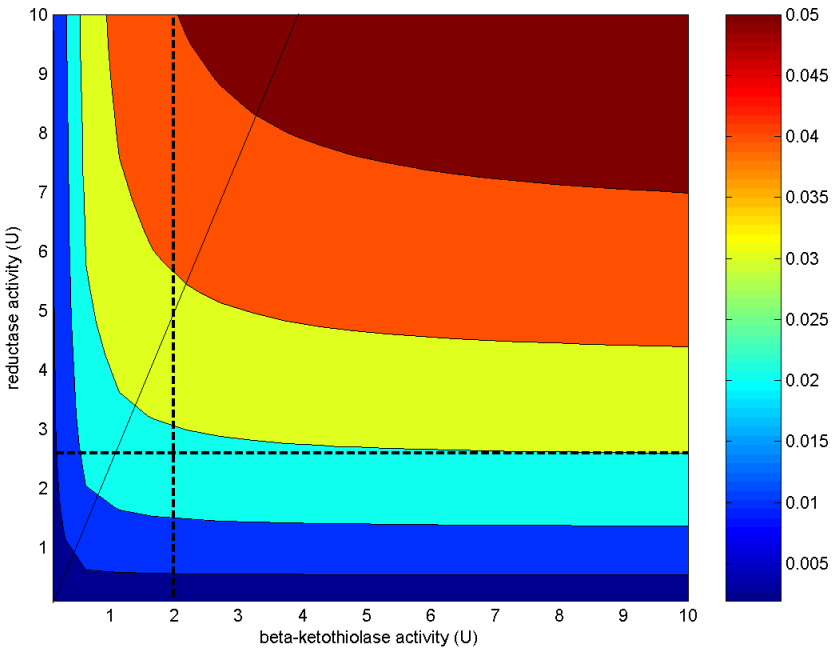


Figure 5-2: Predicted enzyme activity effects on PHB productivity.
Intersection of dotted lines are the measured activities of β-ketothiolase and acetoacetyl-CoA reductase. Solid line is the ratio of the two enzymes that has the steepest ascent (flux control equally distributed). Heat plot is increasing flux to PHB [μM PHB monomer min^{-1}].

5.3.3. Systematic overexpression of PHB pathway

The sensitivity of PHB specific productivity to perturbations in PHB pathway expression levels was analyzed by sequential overexpression of the *phaA, phaB, and phaEC* genes in the background of a second plasmid (pAGL20) that expressed the entire PHB operon from the native *R. eutrophus* promoter. We hypothesized that the flux control would lie in the activity of the product forming pathway and not in the precursor availability (acetyl-CoA or NADPH) during growth on glucose. Figure 5-3 shows the increase in transcript levels and the

corresponding phenotypic change in PHB production for each of the three overexpressions compared to the controls. GFP overexpression was included to control for additional burden to the cell by a plasmid/gene product unrelated to PHB. Figure 5-3(a) shows the transcript levels were increased by two- to five-fold for each of the three genes. *phaB* exerts the most flux control over the system and increases the growth-phase specific PHB productivity by 37% (Figure 5-3(d)). The other two steps in the pathway, *phaA* and *phaEC*, do not increase the flux to PHB upon overexpression, and therefore do not exhibit flux control over the pathway. While GFP did increase the PHB accumulation (Figure 5-3(b)), this is primarily due to a decrease in growth rate (Figure 5-3(c)), while productivity remained unchanged (Figure 5-3(d)). To our knowledge, this is the first experimental determination that *phaB* is the limiting step in the PHB pathway. The maximal specific growth rate was inversely related to the PHB productivity (Figure 5-3(c)), as might be expected for either (1) an increased diversion of carbon from biomass to PHB or (2) more intracellular PHB causing toxicity in the cytoplasm, as seen in the PHB accumulation (Figure 5-3(b)).

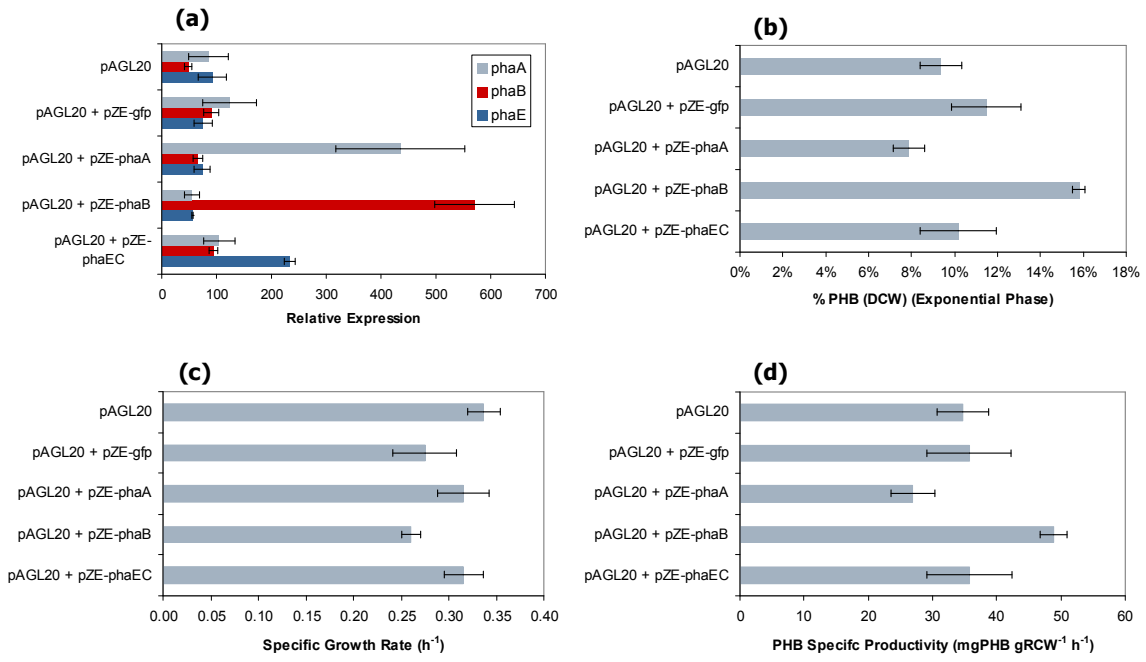


Figure 5-3: Systematic overexpression of steps in PHB pathway.
phaA, *phaB*, *phaEC*, and *gfp* were overexpressed in the presence of a second plasmid containing the complete PHB operon. Strains were characterized in exponential phase. (a) RNA levels (b) Exponential phase PHB accumulation (c) Specific growth rates (d) Specific PHB productivity. RCW – Residual cell weight, non-PHB fraction of biomass.

Overexpression of the entire pathway by complementing the pAGL20 plasmid with a second plasmid driving the whole PHB operon or by only one plasmid with a strong P_{tac} promoter revealed further increases in flux were possible. Figure 5-4 shows the transcriptional and phenotypic effects of overexpressing the entire operon. While *phaB* overexpression removed the AAR limitation, by overexpressing all three genes, additional limitations were overcome, resulting in productivity increases of 170% (270% above pAGL20). While the highest transcriptional levels were measured when both plasmids were

expressing the operon (pAGL20 + pZE-phaECAB), the highest PHB productivity was with only one plasmid using a strong P_{tac} promoter. This is presumably because additional carbon and energy requirements involved in replicating a second plasmid limit PHB productivity.

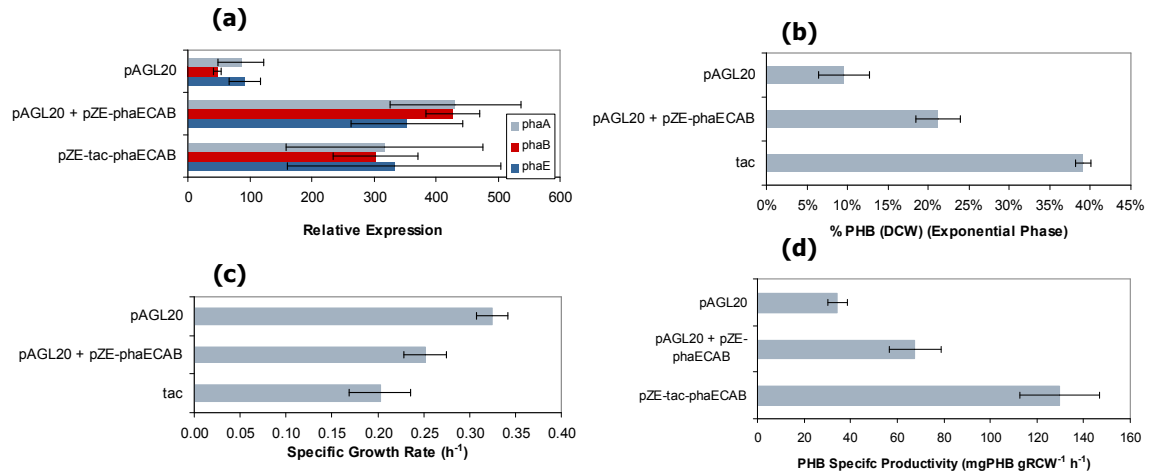


Figure 5-4: Overexpression of entire PHB pathway.
The PHB operon was expressed by one or two plasmids with different promoters. Strains were characterized in exponential phase. (a) RNA levels (b) Exponential phase PHB accumulation (c) Specific growth rates (d) Specific PHB productivity. RCW – Residual cell weight, non-PHB fraction of biomass.

As before, increases in PHB productivity were correlated with decreases in growth rate. This was specifically true for other promoters (P_{T5}, P_{rrnB}) stronger than P_{tac} (data not shown). When inoculated in MR medium, strains with these promoters grew extremely slow (growth rates of ~0.03 h^{-1}), and the strains were genetically unstable: cells defective in PHB production quickly overtook the culture (data not shown). While genetic instabilities made these strains impossible to characterize, it

is suspected that PHB productivities exceeded a threshold beyond which resource consumption or toxicity due to large PHB granules reduces the growth to almost nothing. This represents a regime of PHB pathway expression where the PHB/biomass ratio has become too high and is prohibitively toxic to the cell.

5.3.4. Nitrogen-limited chemostat PHB production

PHB productivity measurements in chemostats allowed us to vary growth rates by controlling the dilution rate under aerobic, nitrogen-limiting conditions. In contrast to the overexpression experiments, where the PHB productivity was changed and the cells were allowed to grow at their maximal growth rate, here the growth rate is constrained, but the relative expression of the PHB pathway is held constant. Table 5-3 shows the measured parameters of the strain at three dilution rates. Figure 5-5 shows the specific uptake and diversion of glucose to various metabolites and biomass.

Table 5-3: PHB production in chemostat

Dilution Rate		*0.05 h^{-1}*	*0.10 h^{-1}*	*0.30 h^{-1}*
% PHB (DCW)		49%	37%	22%
Yield (g PHB/ g GLU)		0.10	0.09	0.08
% of Theoretical Max Yield		20%	20%	17%
Concentrations [g/L]	X	3.1	3.3	3.5
	PHB	2.9	2.0	1.0
	ACE	2.5	2.5	1.1
	GLU	0.3	9.1	17.8
Volumetric Productivities [g/(L h)]	X	0.15	0.33	1.04
	PHB	0.15	0.20	0.29
	ACE	0.13	0.25	0.32
	GLU	1.48	2.09	3.66
Specific Productivities [g/(g biomass h)]	X	0.05	0.10	0.30
	PHB	0.05	0.06	0.08
	ACE	0.04	0.08	0.09
	GLU	0.48	0.63	1.05

X – non-PHB cell weight; PHB – poly-3-hydroxybutyrate; ACE – acetate; GLU - glucose

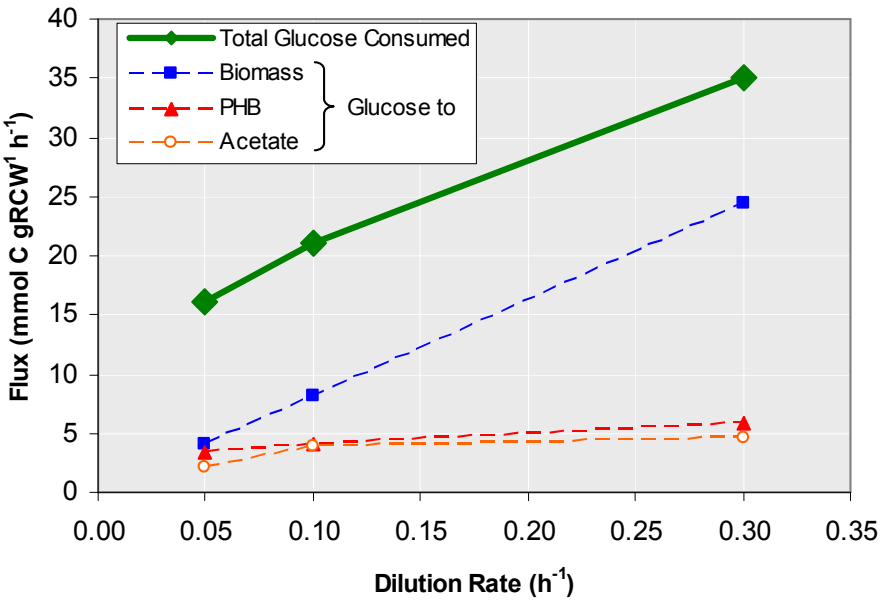


Figure 5-5: C-mol fluxes in PHB nitrogen-limited chemostat.
K12 recA::kan TGD(cat+PHB) was grown in a nitrogen-limited chemostat. Steady state rates of metabolite formation/consumption were measured and converted to C-mols. Glucose consumed for biomass (non-PHB), PHB, and acetate was based on assumed yields. Assumed yields on glucose; Biomass (residual cell weight) - 0.41 g/g, PHB - 0.48 g/g, Acetate – 0.66 g/g.

An excess of glucose (under nitrogen limiting conditions) provided an abundant substrate supply for PHB accumulation, although at the lowest dilution rate, the glucose was also completely consumed. This is presumably due to increased maintenance requirements in the cell. Non-PHB biomass concentration was constant at all dilution rates and was determined by the concentration of nitrogen in the feed (and the biomass yield of *E. coli* on ammonium). Changes in specific glucose uptake were accounted for by the change in glucose demand for biomass as seen by the parallel lines in Figure 4. Although the amount of PHB increased in the reactor with decreasing dilution rate, the yield and specific productivity remained relatively constant. PHB specific productivity only changed by a factor of 1.6 for the six fold change in growth rate, which was the least change for all metabolites and biomass measured. Acetate production was present, consuming as much glucose as PHB, and was also insensitive to the dilution rate, even though the dissolved oxygen was kept above 30% throughout the experiment.

5.3.5. Comparison of experimental observations to a stoichiometric maximal model

The observed trend between PHB productivity and growth rate is well inside the stoichiometrically optimal envelope of PHB productivity as function of growth rate. Figure 5-6 shows the data from the overexpression studies and chemostat study overlaid with the stoichiometric envelope constraining the relationship between the flux to PHB and the flux to growth. The envelope was based on a genome-scale metabolic network model of *E. coli* metabolism (Edwards and Palsson 2000) modified to include the PHB reactions.

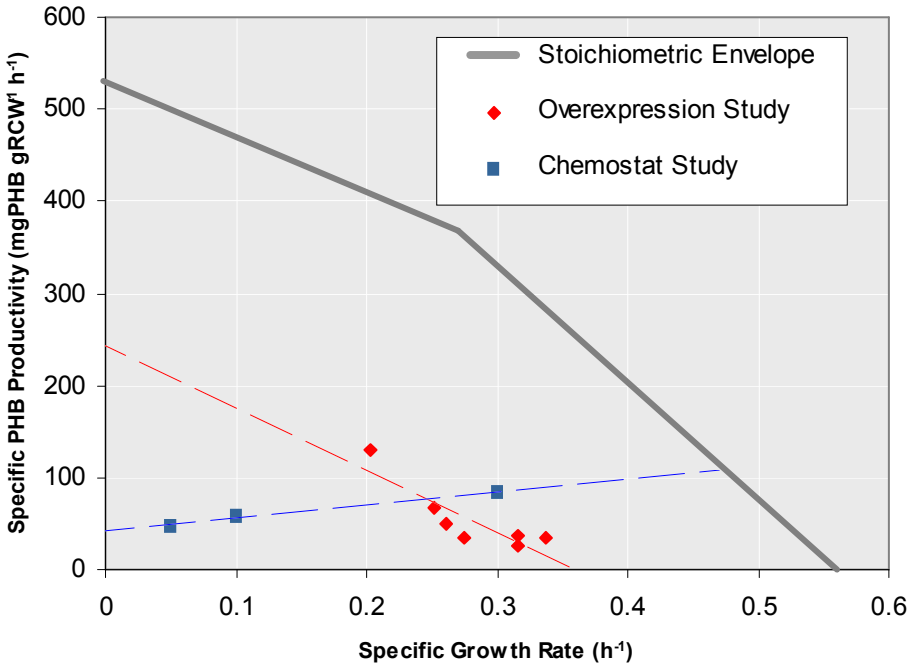


Figure 5-6: Comparison of overexpression and chemostat study to stoichiometric model for PHB in *E. coli*.
PHB productivities and specific growth rates from overexpression and chemostat studies are compared to the maximal stoichiometric trade off. Line represents an *E. coli* stoichiometric model (see text). Stoichiometric model was used to predict the biomass yield at different PHB yields. Data points for overexpression and chemostat studies are for all perturbations attempted. Red line – Linear regression of overexpression study; Blue line – Linear regression of chemostat study

In the overexpression study, pathway expression was changed but cell growth was unconstrained. That is, each strain grew in batch culture at its strain-specific maximum. The trend in the overexpression data shows a tradeoff between product formation and growth, which is broadly similar to the predictions of the stoichiometric model. The overexpression data falls well to the left of stoichiometrically optimal growth rate envelope. Part of the reason for the discrepancy is *E. coli* K12's well known

acetate overflow metabolism, which allows faster growth at high glucose concentrations, at the penalty of stoichiometrically submaximal biomass yields. Another reason may be the use of an XL-1 Blue background strain in this study, as opposed to the wild-type K12 for which the stoichiometric model was developed.

However, the observed tradeoff between PHB productivity and cell growth was close to stoichiometrically optimal, as parameterized by the slopes of the lines in Figure 5-6. Under the low-growth segment of the stoichiometric envelope, the stoichiometrically optimal tradeoff between PHB and biomass was 610 mg PHB / gDCW. The slope of the high-growth segment was ~1280 mgPHB/gDCW. The best-fit line through the overexpression data showed that the observed trade-off was approximately 680 mg PHB / gDCW. This analysis suggests that PHB productivity competes with cell growth for access to carbon, and is limited by the flow of carbon into the PHB biosynthesis pathway. Granule toxicity and other non-stoichiometric affects appear less important.

In contrast, in the nitrogen-limited chemostat at low dilution rates, we hypothesized that specific glucose uptake rates would remain relatively unchanged, allowing glucose previously required for biomass formation to be diverted to PHB, resulting in an increased PHB productivity. Surprisingly, the data did not support our hypothesis and instead revealed a flat

relationship between PHB productivity and growth rate. At lower growth rates, *E. coli* was unable to direct surplus glucose available in the chemostat environment toward PHB production, although it should be feasible in the stoichiometric space calculated. This fact suggests that glucose uptake was regulated, even under nitrogen limitation, such that acetyl-CoA pool sizes were approximately constant regardless of growth rate or glucose availability.

5.4. Discussion

This study focuses on the effect of pathway enzyme expression and growth rate on PHB productivity. PHB screening studies, an inverse metabolic engineering approach, was carried out for *E. coli* in an analogous manner to that for *Synechocystis* as described in Chapter 4. However, increasing PHB productivity by introducing random genetic diversity and isolating improved PHB mutants, was unsuccessful in finding improved mutants that could be linked to the genetic changes in the library. The genetic diversity introduced in the *E. coli* libraries, unlike the *Synechocystis* libraries, would be expected to perturb the global metabolic network and would most likely produce improved mutants by increasing the supply of precursors (acetyl-CoA or NADPH). Because no mutants could be found that improved PHB productivity, it was hypothesized that limitations in the

product-forming pathway (which would not be expected to be perturbed in the libraries) may control the rate of PHB formation. This lead us to a forward metabolic engineering approach for improving PHB productivity in *E. coli*.

PHB productivity was very sensitive to pathway expression, and in particular to AAR. While other studies have implied AAR is limiting, to our knowledge, this is the first direct experimental determination *in vivo.* The systematic overexpression approach used here was able to unambiguously identify *phaB* as exerting the most flux over PHB production. Overexpression of the entire operon led to much higher PHB productivities. At the highest expression levels using P_{T5} and P_{rrnB}, the diversion of acetyl-CoA to PHB almost stopped cell growth on glucose (data not shown). This would indicate that when PHB flux is high enough to bring the acetyl-CoA and/or NADPH concentrations to below the K_m for the PHB pathway, the levels are already to low for cellular functions to behave normally, or else intracellular PHB granules have began to strongly interfere with cellular processes. These results show that PHB productivity can be manipulated over all possible fluxes by only manipulating the expression of the product-forming pathway, and that precursor shortages may occur, not in a fashion where the product-forming pathway is underutilized,

but perhaps where cellular function is inhibited by loss of an essential intermediates.

These limitations in the PHB pathway are consistent with the low *in vitro* AAR activities observed in previous studies (Sim, Snell et al. 1997; van Wegen, Lee et al. 2001). In our experiments, as in most studies, PHB is expressed in medium/high copy (+25) from plasmids or tandem genes. AAR activities in *R. eutrophus* have been reported at 0.2-0.3 U/mg total protein (Park and Lee 1996), while AAR activities in *E. coli* are much less at 0.05 U/mg total protein (Lee, Kim et al. 1996). This five fold change in activity may be the result of problems in the expression or activity of AAR. Additionally, PHB synthase has been shown to be regulated by acetyl phosphate levels, and may be repressed under conditions in *E. coli* (Kessler and Witholt 2001). This agrees with the acetate production observed in the chemostat. Diverting carbon from acetate to PHB would increase PHB yield but may require protein engineering on the PHB synthase to deregulate PHB synthase from allosteric effects that inhibit its activity, such as acetyl phosphate regulation.

PHB productivity reveals the metabolic changes more clearly than PHB accumulation. While our strategy of perturbing the metabolic network through changes in growth rate under nitrogen-limited conditions is not more sophisticated than other techniques described previously, those experiments focused on

PHB accumulation. PHB accumulation measurements are subject to both changes in PHB productivity and growth rate. This was seen in the GFP control in Figure 5-3 and PHB accumulation and productivity in Table 5-3. In both cases the PHB specific productivity was constant, although percent PHB (DCW) varied.

It is important to note that altering the levels of the PHB synthase has been show to have an effect on the quality of the PHB, specifically the molecular weight (Sim, Snell et al. 1997). In this study we have focused on the flux to the PHB pathway and the quantity of PHB. For commercial production, appropriate expression of the PHB synthase will be determined by the desired MW.

Understanding that the flux control is primarily in the product-forming pathway allows the engineer to have a great deal of control over PHB productivity by controlling the expression levels of the PHB pathway. For batch production, expression systems that suppress PHB production during growth phase and turn it on to a high level in stationary phase would be the most effective, by completely separating biocatalyst formation from product formation. Continuous production could either take place as two chemostats in series with the first generating the biocatalyst and the second producing the product (analogous to the batch process), or as a single chemostat where the biocatalyst and product are produced simultaneously. In the

latter case, a precisely tuned expression level would be required to maintain a PHB/biomass ratio that maximizes total productivity. As observed, an upper limit of expression exists where too much carbon is diverted to PHB resulting in a PHB/biomass ratio which has very low cell viability and growth.

Thus, through simple and systematic studies perturbation of gene expression levels, we resolved the long-standing debate in the literature about the rate-controlling step of PHB production: AAR is rate-limiting. More than two decades of prior study did not reach this result because most experimental perturbations in prior approaches suffered from diverse, and often unknown, pleiotropic effects. Our study was the first to carefully manipulate relevant variables in isolation of more global systemic changes. The increases in specific productivity that are now possible should improve economics of both existing batch processes, as well as open up opportunities for continuous processes.

6 Tandem Gene Duplication

6.1. Introduction

As mentioned in Chapter 5, as PHB expression level was increased, mutations quickly became a problem. It is unlikely that the increase in PHB expression initiated mutations but did significantly increase the growth advantage caused by cells that did not make PHB (an increase in growth rate from 0.03 h^{-1} to 0.32 h^{-1}). This overtake of the culture by mutants that loss productivity happened in the course of seven doublings in the shake flask (Materials and Methods, Chapter 5). While PHB productivities had reached very attractive levels as the result of high PHB expression levels, the genetic stability, and resultant loss of productivity, made even batch culture highly problematic.

Plasmids are an important tool for the recombinant production of proteins and heterologous enzymes in many organisms and are responsible for the production of many specialty chemicals, biologics, small molecule drugs, and other chemicals. However, this is particularly true in systems where

large gene doses are required, as we saw with PHB expression. Plasmids have problems with genetic instability that reduce the number of active recombinant alleles in a cell that arise from (1) segregational instability, the result of unequal distribution of plasmids to daughter cells, and (2) random mutations to alleles carried on these plasmids. Because recombinant alleles typically reduce growth rate, both mechanisms can result in subpopulations that grow faster and have fewer intact recombinant alleles than the parent cell, resulting in decreased productivity. Selectable markers and partitioning elements mitigate the effects of segregational instability. Such methods have allowed advances by avoiding plasmid-less population; however, they are suboptimal because it only imposes a minimal copy number and can require expensive antibiotics. These genetic instabilities inherent in plasmid propagation have negatively impacted the development of continuous processes and have contributed to the majority of recombinant processes to batch.

Selectable markers and partitioning elements are not able to prevent DNA mutations that eliminate activity or prevent plasmid containing DNA mutations from propagating. For a plasmid containing a selectable marker and an expression allele in a cell under the appropriate selection conditions, the system will evolve to a population of plasmids that have intact

selectable markers, but have lossed productivity of the desired recombinant protein. We put forward that this activity loss is not determined by the sequential inactivation of different copies due to DNA replication errors, but by 'allele segregation,' that is the accumulation of plasmids with inactive production alleles due to the random distribution of the original and mutated alleles from mother to daughter cell, biased by a growth advantage of the daughter cell containing fewer copies of the original allele.

The loss of productivity due to allele segregation is expected to be much faster than by sequential DNA replication errors alone. For a typical 40 copy plasmid recombinant expression system (see Appendix I for details), a basal DNA mutation rate of 5.4×10^{-10} errors/bp copied (Drake, Charlesworth et al. 1998) will result in a DNA mutation in ~13 generations. If random replication errors prevailed, it would take ~520 generations to lose activity of all copies of the allele. Allele segregation is expected to happen much faster due to the probability that after a mutated plasmid is copied, both copies of a mutated allele can be inherited by the same daughter cell. A simple statistical calculation using binomial distribution of mutant plasmids from mother to daughter cells, and weighted by an expected growth advantage for cells containing more mutant plasmids predicts that a subpopulation containing only mutant

plasmids could form in ~20 generations and overtake the culture in an additional ~12. Allele segregation is therefore expected to dramatically decrease the time that all copies of an allele are inactivated in a culture from ~520 generations to ~45. A method of expressing many copies of a recombinant allele that avoids allele segregation significantly increase the lifetime that a population would be productive in a biotech process.

Tandem gene duplication (TGD) is a naturally occurring phenomenon that generates many head-to-tail repeats of a DNA sequence in the genome (Zhang 2003), giving rise to many copies of an allele that are not subject to allele segregation. TGD has been studied extensively in the context of evolutionary biology, but to date has not been shown to be genetically stable for recombinant expression (Xiaohai Wang 1996; Olson, Zhang et al. 1998). By avoiding allele segregation, TGD can play a very important role in the introduction of new pathways and expression of single genes impacting product formation in microorganisms and other application of metabolic engineering and synthetic biology.

The TGD mechanism causes gene duplication events to occur between two regions of high sequence homology driven and selected by a growth advantage conferred for each additional copy (Figure 6-1(a)). An unequal crossover between the two homology regions on opposite strands (Figure 6-1(b)) occurs

after the replication fork. This crossover event is *recA* dependent in prokaryotes and yields two strands of DNA, one containing two copies of the region between the homologous sites and another with that region deleted (Figure 6-1(c)). Each of these strands is passed down to one daughter cell (Figure 6-1(d)). If a growth advantage is conferred by multiple copies of the region, the cell containing the tandem genes will begin to overtake the population and be available for further rounds of duplication (Figure 6-1(e)). This process can continue, and has been observed to make up to 200 copies of the inter-homologous region (Andersson, Slechta et al. 1998). If conditions change such that a growth advantage is given for fewer copies, the copy number can be reduced by the same process. To prevent this, *recA* is deleted (Figure 6-1(f)), thereby preventing the crossover step and fixing the copy number.

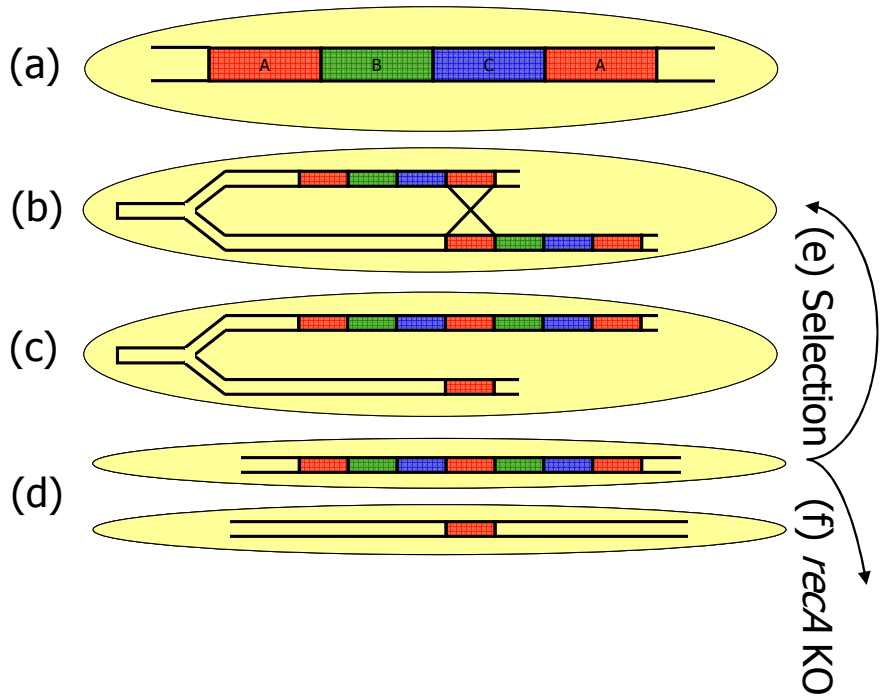


Figure 6-1: Tandem gene duplication: construction, amplification, and stabilization. (a) A construct is delivered to the genome that contains an antibiotic marker and the gene(s) of interest flanked by homologous regions (b) *recA* mediates an uneven homologous crossover between regions flanking the antibiotic marker and gene of interest. (c) This generates a strand with two copies of the cassette and another with a deletion. (d) One daughter cell inherits the two repeats while the other daughter cell looses the insert. (e) Antibiotics are used to provide a growth advantage for the cell with increased repeats. The process can then repeat itself to increase the number of duplications. (f) Finally, *recA* is deleted to prevent further change in copy number. [A] Homologous region (1 kb of *chlB* from *Synechocystis* PCC6803) [B] Antibiotic Resistance Gene (chloramphenicol acetyl transferase) [C] Gene of Interest (Polyhydroxybutyrate operon)

The engineering strategy was to construct a DNA cassette containing gene(s) of interest and chloramphenicol acetyl transferase (*cat*), flanked on both sides by identical, non-coding 1 kb regions of foreign DNA that has low homology to any other region of the *E. coli* genome. The large identical regions served as homology regions for the crossover event, and

increased tolerance to chloramphenicol was used to provide a growth advantage for cells containing duplicated genes. The construct was delivered to a wild type *E. coli* genome and subcultured in increasing concentrations of chloramphenicol. Once the cells have developed a resistance to the desired concentration of chloramphenicol and therefore have reached a desirable number of duplications, *recA* was deleted to prevent any further increase or decrease in copy number. At this point, the strain expressed the gene(s) according to the gene dosage and did not require the chloramphenicol to maintain the copy number. In this work we describe the methods to achieve the engineering strategy described above and present experimental evidence for the improved genetic stability that is achieved by tandem gene duplication and its application to metabolic engineering of the PHB pathway.

6.2. Materials and Methods

6.2.1. Strains and media

E. coli K12 was used for the host strain of all tandem gene duplications. DH5α (Invitrogen) was used for cloning steps. XL1 Blue was used for long term plasmid comparison experiment. Lysogens and transfer strains used in the λInCh protocol were kindly provided by Dr. Dana Boyd and Dr. Jon Beckwith (Boyd, Weiss et al. 2000). BW 26,547 *recA::kan* Lambda *recA*+, generated

by Barry Wanner, used for P1 phage transduction of the *recA* deletion allele to the TGD strains, was received from Bob Sauer. Genomic DNA was isolated from *Synechocystis* PCC6803 using the Wizard Genomic Purification kit (Promega). pZE-kan-tacpha is as described in Chapter 5. Luria-Bertani broth was used for routine culturing and for tandem gene amplification. PHB production was measured when grown in minimal MR medium (Wang and Lee 1997) with 20 g/L glucose.

6.2.2. Construction of TGD strain

All oligomers used in cloning are in Table 6-1. pTrcHis2B (Invitrogen) was digested by SphI & ClaI, blunted by Mung Bean Nuclease and ligated to yield only the pBR322 origin and β-lactamase. Overlap PCR (Pont-Kingdon 2003) was used to connect a 1 kb region of *chlB* from *Synechocystis* PCC6803 to the chloramphenicol acetyl transferase (*cat*) from pAC-LYC. The resulting DNA, followed by another copy of the *chlB* were sequentially cloned into the multicloning site of the pTrcHis2B-derived plasmid. This plasmid, pTGD, was used as a general integration platform (Figure 6-2). The *phaECAB* operon with *tac* promoter and optimal RBS from a modified pAGL20 were cloned into the SphI/MluI sites. The resulting DNA constructs were transferred from the plasmids to the genome of K12 *E. coli* using the λInCh protocol (Boyd, Weiss et al. 2000). This protocol

delivers the plasmid construct to the genome, followed by a recombination step that removes the phage DNA from the genome.

Table 6-1: Oligonucleotides used in TGD study.

Name	*Description*	*Sequence (5'-3')*
Int P1	chlb (s)	GCACCCTACGCATCGCCAGTTCTT
Int P2	chlB (a)	CGCTCTCGGGCAAACTTTTCTGTGTT
Int P23	Overlap primer for Anneal-Extend PCR	GGAACCTCTTACGTGCCGATCAACGGC CCCGTTGTCTTCACTGATCAACACT
Int P3	cat (s)	CGTTGATCGGCACGTAAGAGGTTCC
Int P4	cat (a)	CCTTAAAAAAATTACGCCCCGCCC
Lam P1(s)	BamHI used for pTGD construction	TATCGGATCCCCAGTTCTTTCAAAAAC GTCCACGCC
Lam P4(a)	SacI SphI used for pTGD construction	GGGAGCTCAGCATGCCCTTAAAAAAAT TACGCCCCGCCC
Lam P7(s)	EcoRI MluI used for pTGD construction	GGAATTCAACGCGTCCAGTTCTTTCAA AAACGTCCACGCC
Lam P8(a)	XbaI used for pTGD construction	CGTCTAGAGCCCCGTTGTCTTCACTGAT CAACACT
PHB (s)	SphI used for pTGD cloning	GCAAGCATGCAGCTTCCCAACCTTACC AGAGGGCG
PHB (a)	MluI used for pTGD cloning	GCACGCGTCGGCAGGTCAGCCCATATG CAG
Cm qPCR (s)	qPCR to measure copy number of *cat* in genome or in plasmids	CGCCTGATGAATGCTCATCC
Cm qPCR (a)	qPCR to measure copy number of *cat* in genome or in plasmids	AGGTTTTCACCGTAACACGC
bioA qPCR (s)	qPCR to normalize *cat* copy number to number of copies of genome	GTGATGCCGAAATGGTTGCC
bioA qPCR (a)	qPCR to normalize *cat* copy number to number of copies of genome	GCGGTCAGACGCTGCAACTG

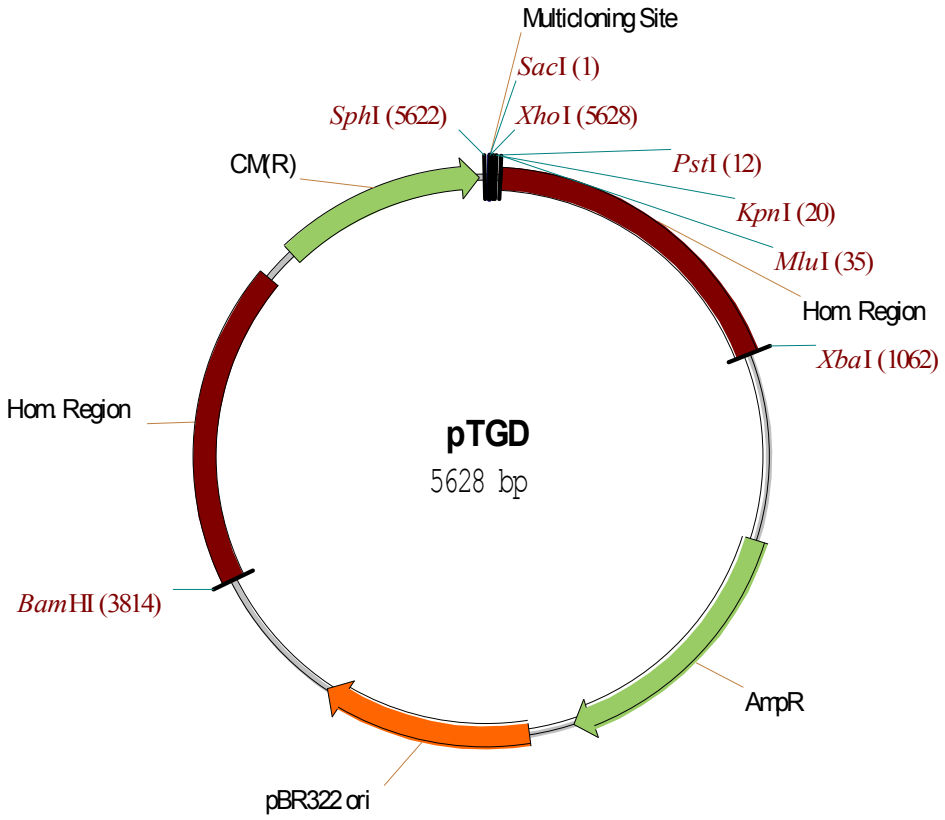


Figure 6-2: Plasmid map for pTGD.
Hom Region- 1 kb DNA from *Synechocystis* used for *recA*-based DNA crossover. CM(R)-chloramphenicol acetyl transferase (chloramphenicol resistance), AmpR – betalactamase (ampicillin resistance), ori – origin of replication.

Amplification of the construct was accomplished by subculturing the resulting strains in LB (1:100), increasing the chloramphenicol concentration in doublings from 13.6 µg/mL to the desired concentration (as high as 1,360 µg/mL). *recA* deletion was accomplished by P1 phage transduction of the *recA*::*kan* allele to the TGD strain. *recA*- was routinely validated by UV sensitivity to 3,000 µJ energy using a Stratalinker UV crosslinker (Stratagene, La Jolla, CA).

6.2.3. Subculturing genetic stability assay

Strains were grown in 5 mL LB in 14 mL culture tubes at 37°C without antibiotics. After cultures reached stationary phase (typically ~12 hrs), 50 µL were diluted into a fresh 5 mL of LB (100x dilution). This was repeated a total of seven times. At which point, genomic DNA was purified and copy numbers measured by qPCR. A sample of the cells was grown in MR medium for 3 days and PHB analyzed.

6.2.4. Exponential phase genetic stability assay

Cells were cultured by continous subculture. Cells were grown in 250 mL erlyenmeyer shake flasks with 50 mL MR media at 37°C and 225 rpm. Cells were innocculated at A_{600}=0.015 and were allowed to grow to late exponential phase (typically A_{600} = 2.0). Cells were subcultured by innocculating the culture to A_{600}=0.015 in a prewarmed shake flask as above. By this, cells were continuously growing at maximal growth rate and did not enter stationary phase during the course of the experiment. Specific growth rate was estimated in each subculture based on the A_{600} data points taken. PHB accumulation was measured as below and used to calculate PHB productivity as the product of specific growth rate and PHB accumulation, an approximation that holds at steady state.

6.2.5. qPCR measurement of copy number

Copy numbers were detected by qPCR on genomic DNA isolated from the appropriate strains using the Wizard Genomic Purification kit (Promega). qPCR was carried out on a Biorad iCycler using the iQ SYBR Green Supermix (Biorad). *cat* copy numbers was detected and compared to the copy number of *bioA*, a nearby native gene in the genome. Table 6-1 has primers used for qPCR. Genomic DNA containing only one copy of the construct was diluted and used as a standard curve.

6.2.6. Poly-3-hydroxybutyrate Assay

End point PHB accumulation of the strains was determined by inoculating an overnight culture of LB (1% v/v) into a 250 mL shake flask with 50 mL of MR media. Cells were cultured at 37°C for 72 hrs at 225 rpm. Cells were harvested by centrifugation, and PHB content was analyzed by hydrolysis to crotonic acid followed by HPLC analysis as described in Chapter 3.2.3.

6.2.7. FACS Assay

Percentage of cells producing PHB was determined by fluorescence activated cell sorting (FACS) using nile red to stain the PHB granules. Samples were collected from the exponential phase subculturing experiment, and allowed to grow to late stationary phase (to maximize the PHB in a cell). 200

µL of cell culture was mixed with 800 µL isopropanol to fix the cells. This was incubated for 10 min and then resuspended in 10 mg/mL MgCl with 3 µg/mL nile red. Cells were stained with nile red for 30 min in the dark and analyzed on a Becton Dickinson FACScan. Cells were scored as containing PHB if the fluorescence was above the 95th percentile of a distribution of cells containing no PHB.

6.3. Results

A DNA construct that would promote gene duplication via TGD on the *E. coli* genome was constructed. A shuttle plasmid, pTGD, was constructed that would deliver the required DNA elements to the genome with the λInCh genomic integration system (Boyd, Weiss et al. 2000) (details in Materials and Methods). pTGD contains two identical, non-coding 1 kb regions of *Synechocystis* DNA flanking *cat* and a multicloning site. Upon delivery to the genome of *recA*+ *E. coli*, the construct was amplified by serial subculturing in increasing concentrations of chloramphenicol resulting in higher chloramphenicol tolerance due to more copies (called TGD(*cat*)) (Figure 6-3). In equivalent *recA*- strains, increased resistance to chloramphenicol could not be induced by serial subculture (Figure 6-3). A second strain was created with the PHB operon cloned into the multicloning site of pTGD (called TGD(*cat* + *PHB*)). It behaved the same as TGD(*cat*) in

response to chloramphenicol depending on the *recA* genotype (Figure 6-3).

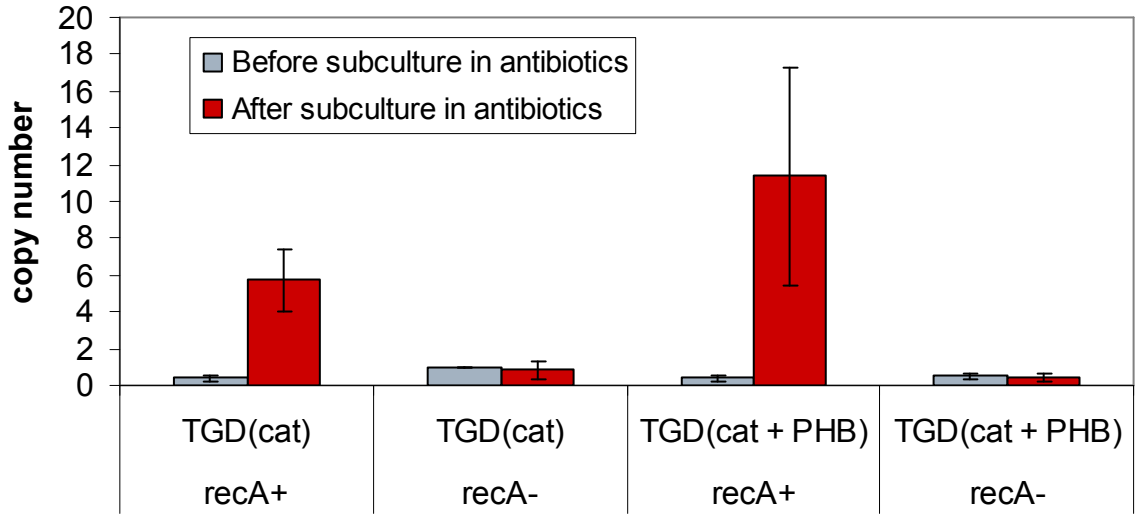
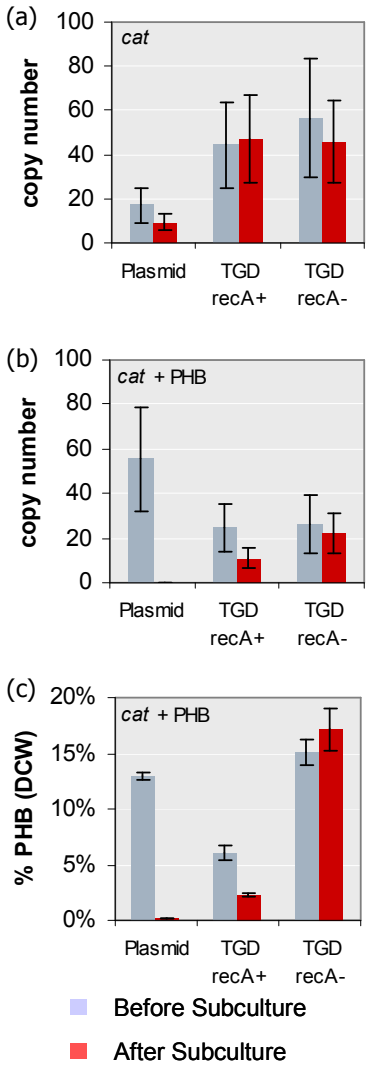


Figure 6-3: *recA* is required for copy number amplification.
TGD strains with and without *recA* were subcultured in antibiotics. For strains that were *recA*+, amplification occurred, as additional copies could be made through *recA*-mediated recombination. *recA*- strains could not do the recombination, and therefore not increase their copy number.

The stability of copy number was determined by subculturing the strain seven times (~45 generations) in the absence of antibiotics for plasmids and TGD(*cat*) constructs (Figure 6-4(a)). As expected, the plasmid copy number was significantly reduced. Tandem gene constructs, both *recA*+ and *recA*-, faired much better and did not have reduced copy number. To impose conditions that place a metabolic burden on the cell due to increased copy number, and increase the growth disadvantage of recombinant expression, the PHB-containing strain was subjected to the same

subculturing as above (Figure 6-4(b)). With the PHB operon, the plasmid copy number had been reduced to less than 0.1 %. The TGD (*cat* + *PHB*) *recA*+ strain copy number was reduced, due to the possibility for the DNA crossover event to reduce the copy number. However, the reduction was not as dramatic because DNA crossover, which occurs in ~10^{-2} cells, or ~7 generations, for high homology regions (Anderson and Roth 1981), is not expected to be as frequent as segregation. The TGD(*cat* + *PHB*) *recA*- strain was not able to reduce its copy number, even with the selection pressure imposed by the PHB operon. The ability of *recA* to act as an 'on/off' switch for changing copy number represents an important control to determine and set the copy number. The PHB accumulation in late stationary phase tracked with the copy number measurements of the strains (Figure 2(c)). The plasmid made almost no PHB after subculture, while the TGD(*cat* + *PHB*) *recA*- construct produced the same amount of PHB as before subculture. This confirms that the gene copies are both present and active after many rounds of subculture in TGD *recA*-strains.

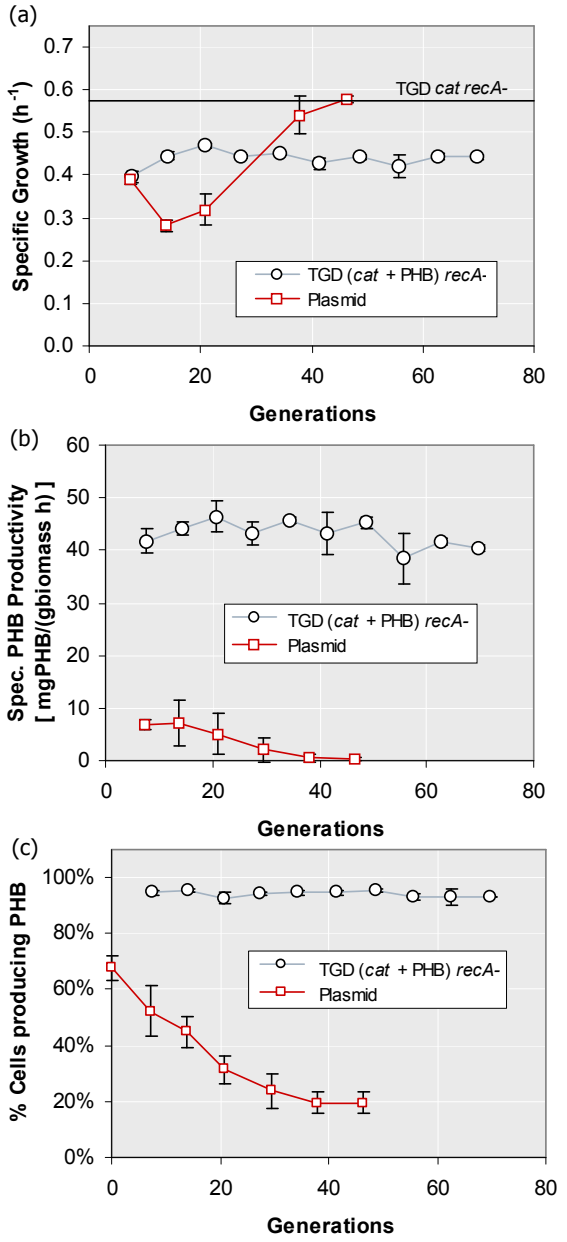


(a) Copy numbers before and after seven rounds of subculture for identical constructs of *cat* that were expressed as multiple copies on (1) a plasmid, (2) multiple tandem genes in a *recA*+ strain or (3) multiple tandem genes in a *recA*- cell without antibiotics. (b) As above, except all strains contain *cat* and the PHB operon, which significantly reduced the growth rate associated with the multiple copies of each gene, allowing cells with loss of PHB activity to have a growth advantage. (c) PHB accumulation before and after seven rounds of subculturing. Plasmids were lost even without the burden of PHB, while a *recA*+ tandem gene construct lost copies at a slower rate than plasmids. TGD *recA*- constructs were stable for the length of the experiment in all cases. Copy numbers were measured by qPCR on DNA extractions before and after the subculturing. Copy number was determined by the ratio of copies of *cat* (on plasmid or in TGD) compared to the genome (as measured by a gene near the integration site, *bioA*).

Figure 6-4: TGD *recA*- is more stable than plasmids or TGD *recA*+ without antibiotics.

To test the long term PHB productivity, TGD(*cat* + *PHB*) *recA*- was maintained in exponential phase by subculturing for 70 generations without antibiotic maintenance and compared to plasmid-based PHB strain *with* antibiotic maintenance. Unlike the prior experiment, the plasmid-based strain (but not the TGD strain) will be grown in the presence of antibiotics, requiring the resistance cassette to be active, but not necessarily the

PHB operon. Figure 6-5 shows the growth rates, specific PHB productivity, and percentage of cells producing PHB over time for the two strains. Without antibiotics, PHB productivity of the plasmid system went to zero in five doublings (data not shown). The average specific growth rate of the TGD strain was 0.44 h^{-1} compared to 0.58 h^{-1} for the same strain without PHB production. This difference in growth rate would provide a driving force for a mutant population to overtake the culture had it occurred. The plasmid system evolved to the maximal growth rate in 40 generations, which was accompanied by a loss of PHB productivity. The TGD system maintained >90% of its PHB productivity up to 70 generations. PHB plasmid maintained resistance to the antibiotics throughout the subculture, however, PHB only 20% of the cells were producing any PHB. An estimated 27 generations had passed from the cell transformed with the plasmid until the beginning of the subculturing experiment, making the total of 67 generations in line with the estimates of Appendix I. This shows that TGD constructs without selection pressure maintain recombinant expression for much longer than plasmids, even when antibiotic selection is present.



TGD (*cat* + PHB) *recA*- and K12 *recA*- (pZE-tac-pha), referred to as Plasmid, were subcultured to maintain exponential phase for 70 generations. (a) Specific growth rate change with subculturing. Line is TGD *cat recA*- strain (with no PHB). Significant differences in growth show a selection pressure to evolve to a growth rate faster than was allowed under that PHB production conditions. However, even with a possible 25% improvement in growth rate, no faster growing mutants emerged for the TGD strain. This is compared to 40 generation for the plasmid-based PHB system. (b) PHB specific productivity in exponential phase. TGD(*cat* + PHB) maintains 90% of the PHB productivity throughout and at a significantly higher level than the plasmid tested. Plasmid-based PHB system lost all measurable productivity by 40 generations. (c) Percentage of cells producing PHB. Flow cytometry used to quantify nile red-stained PHB content in cells. Percentage is fraction of the population that fluoresced higher than the 95th percentile of a fluorescence distribution for the negative control. TGD (*cat* + PHB) *recA*- – strain as above. Plasmid – K12 *recA*- (pZE-tac-PHB). Generations based on the doubling time of the parent strain.

Figure 6-5: Long term performance of tandem PHB operon.

The ability to control gene dose is a useful variable for metabolic engineering and synthetic biology, because it allows precise control over expression. By varying the concentration

of antibiotic used during the amplification, a range of copy numbers could be produced. Figure 6-6 shows the effect of chloramphenicol on copy number in the constructs. Without the PHB operon, the copy number reaches a maximum of 45 copies with a linear range from 0 – 500 µg/mL chloramphenical. With the PHB operon, there is an upward trend in copy number throughout the range of antibiotics used, did not increase beyond ~25 copies. It is possible that the PHB operon limits the copy number that is achievable. This is most likely due to the competition between the growth *advantage* due to the increase copies of the *cat* gene working against the growth *disadvantage* caused by an increase in the copies of the PHB operon. At low copy number, the increased growth advantage by an additional copy number of *cat* outweighs the disadvantage of more PHB operons. However, at higher copy number, the reverse becomes true, and the copy number reaches an equilibrium that is lower than if the PHB operon is not present.

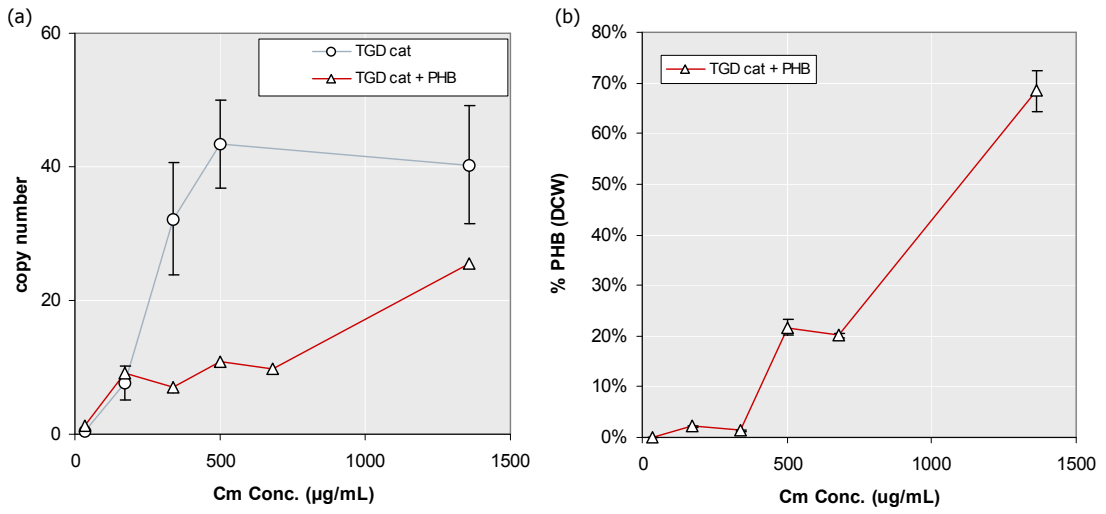


Figure 6-6: Copy number of TGD can be varied fined-tuned expression. Strains were amplified to different copy numbers by growing them at different concentrations of chloramphenicol before deleting *recA*. (a) Copy number of *cat* in genome. (b) PHB accumulation.

6.4. Discussion

Tandem genes offer a superior method for high copy expression in many respects. Most importantly, TGD avoids allele segregation, thereby circumventing the shortest step to the propagation of mutant alleles. This is achieved by linking all of the copies on a single strand of DNA in the genome. By linking them in the genome, each daughter cell is guaranteed to receive one copy of each of the tandem genes. Therefore, the mechanism for losing recombinant expression is by DNA copy errors, which has been predicted to be much slower than allele segregation.

As well, some properties of TGD are uniquely useful for synthetic biology and metabolic engineering. The copy number of the tandem genes is constant, due to it being in the genome, compared to plasmids of which a growing population will have a distribution of copy numbers (Bentley and Quiroga 1993). Additionally, if plasmids with relaxed origin of replications are used for expression, the copy number can vary greatly between growth and stationary phase (Paulsson and Ehrenberg 2001). In synthetic control circuits, a transient variation in copy number could be problematic in the control scheme, leading to unintended responses and artifacts in the circuit. Because of the constant copy number, TGD should be helpful in the design of gene circuits employed in synthetic biology studies. Furthermore, the copy number can be tuned by the strength of antibiotic used for the gene amplification step as shown above. This is equivalent to tuning promoter strength and allows for a continuous range of expression for a given gene by controlling the copy number. This control over expression is also attractive for finely tuned genetic circuits.

Tandem gene constructs meet many of the other properties of plasmids. Unlike other genomic integration approaches that are limited to one copy, the copy numbers of TGD can be varied over much of the same range as plasmids. This is shown both in the measurement of copy number and in PHB productivity, verifying

that high expression is possible. As with the *par* locus or auxotrophic markers, the tandem genes can be propagated without any external antibiotic.

Unlike prior attempts to engineer TGD, this method uses *recA* as an 'on/off' switch to control the change in copy number of the tandem genes (Olson, Zhang et al. 1998). Without the deletion of *recA*, the utility of the TGD constructs would be extremely hindered as copy number would decrease over time.

Allele segregation must be avoided for long term expression. While minimal genome approaches promise increased genetic stability by removal of transposons, prophage, and other genetic loci that would otherwise increase the mutation frequencies (Kolisnychenko, Plunkett et al. 2002), the fidelity of DNA polymerase, and associated DNA repair mechanisms, which are not expected to be improved by minimal genomes, ensure that DNA mutations will eventually arise. If the limiting time scale of mutation propagation is allele segregation, these improvements in DNA mutation frequency will have little effect on the longevity of plasmid expression in these synthetic strains. However, multiple copies of the desired pathway in the genome, as achieved by TGD, will be much more stable.

TGD could be deployed in many different biotech applications. TGD constructs could be particularly useful in the production of low-value added commodity chemicals.

Continuous processes have inherent economic advantages over batch processes, but have not traditionally been used in the biotech industry because of issues with sterility and genetic mutation. TGD significantly extends the time that recombinant genes can be expressed in continuous systems and may prove useful in low value products such as biofuels. Tandem gene duplication events are ubiquitous in nature, and could therefore be applied in many prokaryotic and eukaryotic host organisms, by delivering analogous DNA constructs and inactivating *recA* homologues. We expect this method to have wide industrial application for recombinant expression because of the stability, high expression, and lack of need for plasmid maintenance. In the case of PHB production, where high expression levels will inhibit PHB production, due to large diversion of carbon to PHB, as seen by the stoichiometrically maximal calculations (Figure 5-4), TGD stands to make a significant difference in how long PHB accumulation can be sustained.

7 Conclusions and Recommendations

7.1. Conclusions

This study identified genetic changes that would increase poly-3-hydroxybutyrate accumulation in two systems using two methodologies. Specifically, forward metabolic engineering (FME) and inverse metabolic engineering (IME) methodologies were used to set up a framework for choosing or identifying these genetic changes in (A) a photosynthetic system (*Synechocystis)*, using light and carbon dioxide and a native PHB pathway, or (B) a heterotrophic system (*E. coli*), using sugar and a recombinant PHB pathway. This latitude of biological system and approach shows the necessity of multifaceted approaches to strain improvement, and this study has shown that synergies can be exploited between the different approaches.

For IME approaches, methods were developed to introduce transposons into *Synechocystis*. These transposon libraries can serve as a resource for screening other phenotypes in a photosynthetic system. For the purposes of this study, a screen for PHB was developed. Careful optimization of salt

concentrations and permeabilization methods made it possible to have a quantitative relationship between nile red fluorescence and PHB accumulation, while keeping the cells viable. This method, in conjunction with fluorescence activated cell sorting (FACS), allowed large transposon libraries of *Synechocystis* to be sorted, and mutants with increased PHB to be isolated. Through this, it was determined that two genes, *proA* and *sll0565*, would increase PHB accumulation in a variety of conditions, and led to the hypothesis that proline and osmolyte balancing may play an important role in the regulation of PHB accumulation in *Synechocystis*. However, IME strategies failed to increase PHB production in *E. coli*.

FME strategies in *E. coli* were employed and focused on the effect of changing expression levels in the product-forming pathway and engineering ways of improving genetic stability. As mentioned, IME strategies, analogous to those carried out in *Synechocystis*, did not reveal changes in the genome that would increase PHB production in *E. coli*. We hypothesized that expression levels in the product-forming pathway controlled the flux from acetyl-CoA to PHB (these expression levels would not likely be perturbed in the mutagenesis methods used for IME in this study). Systematic overexpression of the three steps in the product-forming pathway revealed that *phaB*, the acetoacetyl-CoA reductase, exerts the most flux control over the pathway.

Beyond this, increased expression of the entire pathway allowed even higher flux levels. This implies that, after *phaB*, other expression limitations in the pathway exist. High PHB pathway expression increased PHB flux, but the high PHB/biomass ratio severely reduced growth rate, allowing PHB-deficient mutants to quickly overtake the culture.

To reduce mutation problems, tandem gene duplication was developed. It was hypothesized that the propagation of mutated alleles in multicopy plasmids from (a) the first mutant plasmid among many native plasmids to (b) all plasmids being a copy of the mutant was much faster than DNA copy errors. Probabalistic calculations predicted this was primarily a result of the random distribution of plasmids from mother to daughter cells coupled with an increased growth advantage for more copies of the mutant allele, termed 'allele segregation.' Tandem genes avoid allele segregation by having all copies in tandem on the genome, forcing ordered inheritance of each copy of the allele. Calculations predicted a 10 fold increase in genetic stability, and experiments found at least a 3 fold increase. TGD does not require antibiotics, have a stable copy number, and expression can be tuned over a range of copy numbers. The increased genetic stability will make chemostats more attractive, as fermentations can be extended beyond what would be possible with plasmids.

Aside from commenting on the specific conclusions of PHB production in microorganisms, it is helpful to comment on the use of forward and inverse metabolic engineering strategies in a strain improvement program. FME will always be more effective when the amount of information in a system is high. Forward approaches lend themselves to hypothesis driven experiments, where a negative result can be as definitive as a positive result (which is not the case in IME screening experiments). Forward approaches also focus on a few measurements guided by relevant hypotheses. Typically, experimental methods for measuring few samples are more available than measuring many, as is required by screening approaches (This can require significant effort to develop an effective screen for metabolic engineering applications). But forward approaches can only be engineer systems to the extent of what is known.

Inverse strategies, by converse, do not need the amount of information that forward strategies need. These approaches rely on being able to test many hypotheses rapidly (such as the hypothesis that each gene in the genome increases PHB), uncovering relationships that were not previously known. IME adds distal knowledge, by asking many global hypothesis that can be answered simultaneously.

FME and IME work best in parallel, allowing information gathered from IME to direct FME experiments. This was

particularly true for *E. coli*. Although literature implied precursor shortages limited PHB, by performing the global IME studies, it was hypothesized that other constraints were limiting the system, leading to the first experimental determination that flux control in the PHB pathway is limiting. It is our hope that likewise, the information from the *Synechocystis* study will be useful to researchers studying PHB metabolic engineering in plants.

7.2. Recommendations

With regard to further engineering of *Synechocystis*, the most interesting pursuits would be to apply the knowledge gathered in this thesis to PHB production in higher plants. Commercial production of PHB in *Synechocystis* is unlikely to be economical, due to high capital costs for light bioreactors. However, plants, which can be cultivated at lower costs, may be attractive, although yields are currently prohibitively low. Of particular interest is the investigation of the effects of proline biosynthesis and/or osmolyte stress on PHB accumulation in higher plants. This is outside the expertise of our lab, but we hope appropriate researchers may find our results useful.

PHB production in *E. coli* shows much promise, and indeed has already been implemented industrially (Kennedy 2007). Questions have been left unanswered concerning why the activity

of the PHB pathway is so low, given the high copy number of the operon. While high flux was achievable with increased expression, enzyme activities need to be carefully studied for allosteric regulation, and when possible deregulate the enzymes by amino acid level mutagenesis. Understanding and engineering these enzymes may increase PHB flux without the need for very high levels of expression and may minimize other secretion products, such as acetate.

The observation that PHB productivity can be controlled by expression of the PHB operon allows for many possible modes of operations (Chapter 6 Discussion). As continuous processes may be attractive, further work on one-step or two-step chemostats will likely answer those questions. Studying glucose-limited chemostats, and running chemostats with strains that have different PHB operon expression levels will reveal if yields and productivities can be achieved that would make these processes commercially attractive. Glucose-limited chemostats may also have reduced acetate production.

Tandem gene duplication should find application in several metabolic engineering systems. Tandem genes will be most advantageous in applications that require many doublings of the cells and where growth is significantly reduced by the recombinant expression. The copy number attainable by TGD appears to be limited by the growth disadvantage associated with

high copies of PHB. One way to overcome this would be to repress the PHB genes (or other genes of interest) during the TGD amplification step. This would allow the copy number to increase without the growth disadvantage imparted by each new copy of the gene of interest. After *recA* deletion has been completed, the repression can be removed, and all copies will have full expression. This should increase the attainable copy numbers, and thereby the expression levels.

Tandem gene duplication should be very useful as metabolic engineering shifts from a paradigm of high value products, such as biologics and specialty chemicals, to low value products, like commodity polymers and biofuels. TGD does not require antibiotics in production and more importantly the ability to stay productive for long periods in chemostats should significantly reduce production costs.

In general when planning a strain improvement program, we would propose a strategy of (A) doing forward metabolic engineering to the extent possible on the wild type strain using available data. In parallel with this, develop methods for mutagenesis and screening of the desired phenotype using the wild type strain. After FME options have been exhausted, (B) do inverse metabolic engineering (using the IME methods developed in *(A)*) on the current-best engineered strain. New genetic loci,

identified in (B), should be investigated in detail, using FME approaches.

References

Abe, C., Y. Taima, et al. (1990). "New bacterial copolyesters of 3-hydroxyalkanoates and 3-hydroxy-w-fluoroalkanoates produced by *Pseudomonas oleovorans*." Polym Commun **31**: 404-406.

Ábrahám, E., G. Rigó, et al. (2003). " Light-dependent induction of proline biosynthesis by abscisic acid and salt stress is inhibited by brassinosteroid in *Arabidopsis*." Plant Mol Biol **51**(3): 363-372.

Alper, H., Y. S. Jin, et al. (2005). "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli." Metab Eng **7**(3): 155-164.

Alper, H., K. Miyaoku, et al. (2005). "Construction of lycopene-overproducing E-coli strains by combining systematic and combinatorial gene knockout targets." Nat Biotechnol **23**(5): 612-616.

Alper, H., J. Moxley, et al. (2006). "Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production." Science **314**(5805): 1565-1568.

Anderson, A. J. and E. A. Dawes (1990). "Occurrence, Metabolism, Metabolic Role, and Industrial Uses of Bacterial Polyhydroxyalkanoates." Microbiological Reviews **54**(4): 450-472.

Anderson, P. and J. Roth (1981). "Spontaneous Tandem Genetic Duplications in Salmonella typhimurium Arise by Unequal Recombination between rRNA (rrn) Cistrons." Proc Natl Acad Sci U S A **78**(5): 3113-3117.

Andersson, D. I., E. S. Slechta, et al. (1998). "Evidence That Gene Amplification Underlies Adaptive Mutability of the Bacterial lac Operon." Science **282**(5391): 1133-1135.

Bailey, J. E. (1991). "Toward a Science of Metabolic Engineering." Science **252**(5013): 1668-1675.

Bailey, J. E., A. Sburlati, et al. (1996). "Inverse Metabolic Engineering: A Strategy for Directed Genetic Engineering of Useful Phenotypes." Biotechnol Bioeng **52**: 109-121.

Bailey, J. E., A. Shurlati, et al. (1996). "Inverse metabolic engineering: A strategy for directed genetic engineering of useful phenotypes." Biotechnol Bioeng **52**(1): 109-121.

Baptist (1963). U.S. patent 3,072,538. USA.

Baptist (1963). U.S. patent 3,107,172. USA.

Bell, S. L. and B. R. Palsson (2005). "Phenotype phase plane analysis using interior point methods." Comput Chem Eng **29**(3): 481-486.

Bentley, W. E. and O. E. Quiroga (1993). "Investigation of subpopulation heterogeneity and plasmid stability in recombinant *Escherichia coli* via a simple segregated model." Biotechnol Bioeng **42**(2): 222-234.

Bhaya, D., A. Takahashi, et al. (2001). "Light regulation of type IV pilus-dependent motility by chemosensor-like elements in Synechocystis PCC6803." Proc Natl Acad Sci U S A **98**(13): 7540-7545.

Boyd, D., D. S. Weiss, et al. (2000). "Towards Single-Copy Gene Expression Systems Making Gene Cloning Physiologically Relevant: Lambda InCh, a Simple Escherichia coli Plasmid-Chromosome Shuttle System." J Bacteriol **182**(3): 842-847.

Bro, C., B. Regenberg, et al. (2003). "Transcriptional, Proteomic, and Metabolic Responses to Lithium in Galactose-grown Yeast Cells." J Biol Chem **278**(34): 32141-32149.

Burgard, A. P., P. Pharkya, et al. (2003). "OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization." Biotechnol Bioeng **84**(6): 647-657.

Burger, R., M. Willensdorfer, et al. (2006). "Why Are Phenotypic Mutation Rates Much Higher Than Genotypic Mutation Rates?" Genetics **172**(1): 197-206.

Chiaramonte, S., G. M. Giacometti, et al. (1999). "Construction and characterization of a functional mutant of Synechocystis 6803 harbouring a eukaryotic PSII-H subunit." Eur J Biochem **260**(3): 833-843.

Crameri, A., S.-A. Raillard, et al. (1998). "DNA shuffling of a family of genes from diverse species accelerates directed evolution." Nature **391**(6664): 288-291.

Curley, J. M., B. Hazer, et al. (1996). "Production of poly(3-hydroxyalkanoates) containing aromatic substituents by *Pseudomonas oleovorans*." Macromolecules **29**: 1762-1766.

De Koning, G. J. M., H. M. M. V. Bilsen, et al. (1994). "A biodegradable rubber by crosslinking poly(hydroxyalkanoate) from *Pseudomonas oleovorans*." Polymer **35**: 2090-2097.

De Koning, G. J. M., M. Kellerhals, et al. (1997). "A process for the recovery of poly(3-hydroxyalkanoates) from pseudomonads.2. Process development and economic evaluation." Bioprocess Eng **17**: 15-21.

Dean, A. M. and D. E. Koshland Jr (1993). "Kinetic mechanism of Escherichia coli isocitrate dehydrogenase." Biochemistry **32**(36): 9302-9309.

Degelau, A., T. Scheper, et al. (1995). "Fluorometric Measurement of Poly-Beta Hydroxybutyrate in Alcaligenes-Eutrophus by Flow-Cytometry and Spectrofluorometry." Appl Microbiol Biotechnol **42**(5): 653-657.

Doi, Y. (1990). Microbial polyesters. Yokohama, Japan., VCH Publishers, Inc.,.

Drake, J. W., B. Charlesworth, et al. (1998). "Rates of Spontaneous Mutation." Genetics **148**(4): 1667-1686.

Duckworth, H. W., D. H. Anderson, et al. (1987). "Structural basis for regulation in gram-negative bacterial citrate synthases." Biochem Soc Symp **54**: 83-92.

Edwards, J. S. and B. O. Palsson (2000). "The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities." Proc Natl Acad Sci U S A **97**(10): 5528-5533.

Farmer, W. R. and J. C. Liao (2000). "Improving lycopene production in Escherichia coli by engineering metabolic control." Nat Biotechnol **18**(5): 533-537.

Fell, D. A. and S. Thomas (1995). "Physiological control of metabolic flux: the requirement for multisite modulation." Biochem J **311**(1): 35.

Fischer, E., N. Zamboni, et al. (2004). "High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived 13C constraints." Anal Biochem **325**(2): 308-316.

Gascoyne, P. R. C., Y. Huang, et al. (1992). "Dielectrophoretic separation of mammalian cells studied by computerized image analysis." Meas Sci Technol **3**(5): 439.

Gascoyne, P. R. C. and J. Vykoukal (2002). "Particle separation by dielectrophoresis." Electrophoresis **23**(13): 1973.

Gill, R. T. (2003). "Enabling inverse metabolic engineering through genomics." Curr Opin Biotechnol **14**(5): 484-490.

Gorenflo, V., A. Steinbuchel, et al. (1999). "Quantification of bacterial polyhydroxyalkanoic acids by Nile red staining." Appl Microbiol Biotechnol **51**(6): 765-772.

Hein, S., H. Tran, et al. (1998). "Synechocystis sp. PCC6803 possesses a two-component polyhydroxyalkanoic acid synthase similar to that of anoxygenic purple sulfur bacteria." Arch Microbiol **170**: 162-170.

Holmes, P. A. (1986). U.S. patent 4,620,999. USA.

Hong, S. H., S. J. Park, et al. (2003). "In silico prediction and validation of the importance of the Entner-Doudoroff pathway in poly(3-hydroxybutyrate) production by metabolically engineered Escherichia coli." Biotechnol Bioeng **83**(7): 854-863.

Huisman, G. W., E. Wonink, et al. (1991). "Metabolism of poly(3-hydroxyalkanoates) (PHAs) by Pseudomonas oleovorans. Identification and sequences of genes and function of the encoded proteins in the synthesis and degradation of PHA." J Biol Chem **266**(4): 2191-2198.

James, B. W., W. S. Mauchline, et al. (1999). "Poly-3-hydroxybutyrate in Legionella pneumophila, an energy source for survival in low-nutrient environments." Appl Environ Microbiol **65**(2): 822-827.
Jendrossek, D., A. Schirmer, et al. (1996). "Biodegradation of polyhydroxyalkanoic acids." Appl Microbiol Biotechnol **46**(5): 451-463.
Jin, Y.-S. and G. Stephanopoulos (2007). "Multi-dimensional gene target search for improving lycopene biosynthesis in Escherichia coli." Metab Eng **9**(4): 337-347.
Jin, Y., H. Alper, et al. (2005). "Improvement of Xylose Uptake and Ethanol Production in Recombinant Saccharomyces cerevisiae through an Inverse Metabolic Engineering Approach." Appl Environ Microbiol **71**(12): 8249-8256.
Jung, Y. M., J. N. Lee, et al. (2004). "Role of tktA4 gene in pentose phosphate pathway on odd-ball biosynthesis of poly-beta-hydroxybutyrate in transformant Escherichia coli harboring phbCAB operon." J Biosci Bioeng **98**(3): 224-227.
Kaneko, T., S. Sato, et al. (1996). "Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. 2. Sequence determination of the entire genome and assignment of potential protein-coding regions." DNA Res **3**: 109-136.
Karr, D. B., J. K. Waters, et al. (1983). "Analysis of Poly-Beta-Hydroxybutyrate in Rhizobium-Japonicum Bacteroids By Ion-Exclusion High-Pressure Liquid-Chromatography and Uv Detection." Appl Environ Microbiol **46**(6): 1339-1344.
Kennedy, V. B. (2007). Plastics that are green in more ways than one. MarketWatch. New York City.
Kessler, B. and B. Witholt (2001). "Factors involved in the regulatory network of polyhydroxyalkanoate metabolism." J Biotechnol **86**(2): 97-104.
Kidwell, J., H. E. Valentin, et al. (1995). "Regulated expression of the Alcaligenes eutrophus pha biosynthesis genes in Escherichia coli." Appl Environ Microbiol **61**(4): 1391-8.
Kinkema, M., H. Wang, et al. (1994). " Molecular analysis of the myosin gene family in Arabidopsis thaliana." Plant Mol Biol **24**(4): 1139-1153.
Kolisnychenko, V., G. Plunkett, Iii, et al. (2002). "Engineering a Reduced Escherichia coli Genome." Genome Res **12**(4): 640-647.
Lawrence, A. G., J. Choi, et al. (2005). "In Vitro Analysis of the Chain Termination Reaction in the Synthesis of Poly-(R)-beta-hydroxybutyrate by the Class III Synthase from Allochromatium vinosum." Biomacromolecules **6**(4): 2113 -2119.

Lee, I. Y., M. K. Kim, et al. (1996). "Regulatory effects of cellular nicotinamide nucleotides and enzyme activities on poly(3-hydroxybutyrate) synthesis in recombinant Escherichia coli." Biotechnol Bioeng **52**(6): 707-712.

Lee, S. Y. and H. N. Chang (1995). "Production of poly(3-hydroxybutyric acid) by recombinant Escherichia coli strains: genetic and fermentation studies." Can J Microbiol **41 Suppl 1**: 207-15.

Lee, S. Y., H. N. Chang, et al. (1994). "Production of poly(beta-hydroxybutyric acid) by recombinant Escherichia coli." Ann N Y Acad Sci **721**: 43-53.

Lee, S. Y., K. M. Lee, et al. (1994). "Comparison of recombinant *Escherichia coli* strains for synthesis and accumulation of poly-(3-hydroxybutyric acid) and morphological changes." Biotechnol Bioeng **44**(11): 1337-1347.

Lee, S. Y., Y. K. Lee, et al. (1995). "Stimulatory effects of amino acids and oleic acid on poly (3-hydroxybutyric acid) synthesis by recombinant Escherichia coli." J Ferment Bioeng **79**(2): 177-180.

Lim, S. J., Y. M. Jung, et al. (2002). "Amplification of the NADPH-related genes zwf and gnd for the oddball biosynthesis of PHB in an E-coli transformant harboring a cloned phbCAB operon." J Biosci Bioeng **93**(6): 543-549.

Liu, Y. and R. Whittier (1995). "Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking." Genomics **25**(3): 674-81.

Lu, J. L. and J. C. Liao (1997). "Metabolic engineering and control analysis for production of aromatics: Role of transaldolase." Biotechnol Bioeng **53**(2): 132-138.

Lutz, R. and H. Bujard (1997). "Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements." Nucl Acids Res **25**(6): 1203-1210.

Madison, L. L. and G. W. Huisman (1999). "Metabolic Engineering of Poly(3-Hydroxyalkanoates):From DNA to Plastic." Microbiol Mol Biol Rev **63**(1): 21-53.

Markx, G. H., J. Rousselet, et al. (1997). "DEP-FFF: Field-flow fractionation using non-uniform electric fields." J Liq Chromatogr Rel Technol **20**(16-17): 2857-2872.

Masamune, S., C. T. Walsh, et al. (1989). "Poly-(R)-3-hydroxybutyrate (PHB) biosynthesis: mechanistic studies on the biological Claisen condensation catalyzed by ß-ketoacyl thiolase." Pure Appl Chem **61**: 303–312.

Miyake, M., J. Schnackenberg, et al. (2000). "Phosphotransacetylase as a key factor in biological

production of polyhydroxybutyrate." Appl Biochem Biotechnol **84**(1): 1039-1044.
Müller, S., A. Lösche, et al. (1995). "A flow cytometric approach for characterization and differentiation of bacteria during microbial processes." Appl Microbiol Biotechnol **43**(1): 93-101.
Niederberger, P., Prasad R, et al. (1992). "A strategy for increasing an in vivo flux by genetic manipulations. The tryptophan system of yeast." Biochem J **287**(2): 473-479.
Olson, P., Y. Zhang, et al. (1998). "High-Level Expression of Eukaryotic Polypeptides from Bacterial Chromosomes." Protein Expression Purif **14**(2): 160-166.
Ostergaard, S., L. Olsson, et al. (2000). "Increasing galactose consumption by Saccharomyces cerevisiae through metabolic engineering of the GAL gene regulatory network." Nat Biotechnol **18**(12): 1283-1286.
Ostle, A. G. and J. G. Holt (1982). "Nile Blue-a As a Fluorescent Stain For Poly-Beta- Hydroxybutyrate." Appl Environ Microbiol **44**(1): 238-241.
Parekh, S., V. A. Vinci, et al. (2000). "Improvement of microbial strains and fermentation processes." Appl Microbiol Biotechnol **54**(3): 287-301.
Park, J. S. and Y. H. Lee (1996). "Metabolic characteristics of isocitrate dehydrogenase leaky mutant of Alcaligenes eutrophus and its utilization for poly-ß-hydroxybutyrate production." J Ferment Bioeng **81**(3): 197-205.
Paulsson, J. and M. Ehrenberg (2001). "Noise in a minimal regulatory network: plasmid copy number control." Q Rev Biophys **34**(1): 1-59.
Peoples, O. P. and A. J. Sinskey (1989). "Poly-Beta-Hydroxybutyrate (Phb) Biosynthesis in Alcaligenes-Eutrophus H16 - Identification and Characterization of the Phb Polymerase Gene (Phbc)." J Biol Chem **264**(26): 15298-15303.
Peoples, O. P. and A. J. Sinskey (1989). "Poly-Beta-Hydroxybutyrate Biosynthesis in Alcaligenes-Eutrophus H16 - Characterization of the Genes Encoding Beta-Ketothiolase and Acetoacetyl-Coa Reductase." J Biol Chem **264**(26): 15293-15297.
Pont-Kingdon, G. (2003). Creation of Chimeric Junctions, Deletions, and Insertions by PCR. Methods in Molecular Biology: PCR Protocols. J. M. S. Bartlett and D. Strirling. Totowa, NJ, Humana Press Inc. **226**: 511-515.
Rippka, R., J. Deruelles, et al. (1979). "Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria." J Gen Microbiol **111**(MAR): 1-61.
Rousselet, J., G. H. Markx, et al. (1998). "Separation of erythrocytes and latex beads by dielectrophoretic

levitation and hyperlayer field-flow fractionation." Colloids and Surfaces a-Physicochemical and Engineering Aspects **140**(1-3): 209-216.
Sambrook, J., E. F. Fritsch, et al. (1989). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press.
Sanchez, A. M., J. Andrews, et al. (2006). "Effect of overexpression of a soluble pyridine nucleotide transhydrogenase (UdhA) on the production of poly(3-hydroxybutyrate) in Escherichia coli." Biotechnol Progr **22**(2): 420-425.
Santos, C. N. S. and G. Stephanopoulos (2008). "Combinatorial engineering of microbes for optimizing cellular phenotype." Curr Opin Chem Biol **12**(2): 168-176.
Sauer, U. and U. Schlattner (2004). "Inverse metabolic engineering with phosphagen kinase systems improves the cellular energy state." Metab Eng **6**(3): 220-228.
Schubert, P., A. Steinbuchel, et al. (1988). "Cloning of the Alcaligenes-Eutrophus Genes for Synthesis of Poly-Beta-Hydroxybutyric Acid (Phb) and Synthesis of Phb in Escherichia-Coli." J Bacteriol **170**(12): 5837-5847.
Schubert, P., A. Steinbüchel, et al. (1988). "Cloning of the Alcaligenes-Eutrophus Genes for Synthesis of Poly-Beta-Hydroxybutyric Acid (Phb) and Synthesis of Phb in Escherichia-Coli." J Bacteriol **170**(12): 5837-5847.
Shi, H., J. Nikawa, et al. (1999). "Effect of modifying metabolic network on poly-3-hydroxybutyrate biosynthesis in recombinant Escherichia coli." J Biosci Bioeng **87**(5): 666-677.
Shi, H. D., J. Nikawa, et al. (1999). "Effect of modifying metabolic network on poly-3-hydroxybutyrate biosynthesis in recombinant Escherichia coli." J Biosci Bioeng **87**(5): 666-677.
Sim, S. J., K. D. Snell, et al. (1997). "PHA synthase activity controls the molecular weight and polydispersity of polyhydroxybutyrate in vivo." Nat Biotechnol **15**(1): 63-67.
Song, B. G., T. K. Kim, et al. (2006). "Modulation of talA gene in pentose phosphate pathway for overproduction of poly-beta-hydroxybutyrate in transformant Escherichia coli harboring phbCAB operon." J Biosci Bioeng **102**(3): 237-240.
Spiekermann, P., B. H. A. Rehm, et al. (1999). "A sensitive, viable-colony staining method using Nile red for direct screening of bacteria that accumulate polyhydroxyalkanoic acids and other lipid storage compounds." Arch Microbiol **171**(2): 73-80.
Steinbüchel, A. and B. Füchtenbusch (1998). "Bacterial and other biological systems for polyester production." Trends Biotechnol **16**: 419-427.

Steinbuchel, A. and H. Schlegel (1991). "Physiology and molecular genetics of poly(beta-hydroxy-alkanoic acid) synthesis in *Alcaligenes eutrophus*." Mol Microbiol **5**(3): 535-542.

Stephanopoulos, G. (1999). "Metabolic Fluxes and Metabolic Engineering." Metab Eng **1**(1): 1-11.

Stephanopoulos, G. and J. J. Vallino (1991). "Network Rigidity and Metabolic Engineering in Metabolite Overproduction." Science **252**(5013): 1675-1681.

Styczynski, M. P., J. F. Moxley, et al. (2007). "Systematic identification of conserved metabolites in GC/MS data for metabolomics and biomarker discovery." Anal Chem **79**(3): 966-973.

Sudesh, K., K. Taguchi, et al. (2002). "Effect of Increased PHA synthase activity on polyhydroxyalkanoates biosynthesis in Synechocystis sp. PCC 6803." Int J of Bio Macro **30**(97).

Suriyamongkol, P., R. Weselake, et al. (2007). "Biotechnological approaches for the production of polyhydroxyalkanoates in microorganisms and plants -- A review." Biotechnol Adv **25**(2): 148-175.

Taft-Benz, S. A. and R. M. Schaaper (1998). "Mutational analysis of the 3'-->5' proofreading exonuclease of Escherichia coli DNA polymerase III." Nucl Acids Res **26**(17): 4005-4011.

Takahashi, H., M. Miyake, et al. (1998). "Improved accumulation of poly-3-hydroxybutyrate by a recombinant cyanobacterium." Biotechnol Lett **20**(2): 183-186.

Taroncher-Oldenberg, G., K. Nishina, et al. (2000). "Identification and analysis of the polyhydroxyalkanoate-specific beta-ketothiolase and acetoacetyl coenzyme A reductase genes in the cyanobacterium Synechocystis sp strain PCC6803." Appl Environ Microbiol **66**(10): 4440-4448.

Taroncher-Oldenburg, G. and G. Stephanopoulos (2000). "Targeted, PCR-based gene disruption in cyanobacteria: inactivation of the polyhydroxyalkanoic acid synthase genes in Synechocystis sp PCC6803." Appl Microbiol Biotechnol **54**(5): 677-680.

Tyo, K. and G. Stephanopoulos (2008). "Tandem Gene Duplication: an alternative to plasmids for high copy, stable expression." (submitted).

Tyo, K., H. Zhou, et al. (2006). "High-Throughput Screen for Poly-3-Hydroxybutyrate in Escherichia coli and Synechocystis sp. Strain PCC 6803." Appl Environ Microbiol **72**(5): 3412.

Tyo, K. E., Y.-S. Jin, et al. (submitted). "Identification of gene disruptions for increased poly-3-hydroxybutyrate accumulation in Synechocystis PCC 6803." Biotechnology & Bioengineering.

Tyo, K. E. and G. N. Stephanopoulos (submitted). "Tandem Gene Duplication: an alternative to plasmids for high copy, stable expression." Nat Biotechnol.

Tyo, K. E. J., C. R. Fischer, et al. (submitted). "Effects of Genetic and Metabolic Perturbations on PHB Productivity in Escherichia Coli." Metab Eng.

Van Wegen, R. J., S. Y. Lee, et al. (2001). "Metabolic and kinetic analysis of poly(3-hydroxybutyrate) production by recombinant Escherichia coli." Biotechnol Bioeng **74**(1): 70-80.

Vazquez-Laslop, N., H. Lee, et al. (2001). "Molecular Sieve Mechanism of Selective Release of Cytoplasmic Proteins by Osmotically Shocked Escherichia coli." J Bacteriol **183**(8): 2399-2404.

Vermaas, W. F. J., J. G. K. Williams, et al. (1987). "Sequencing and Modification of Psbb, the Gene Encoding the Cp-47 Protein of Photosystem-Ii, in the Cyanobacterium Synechocystis-6803." Plant Mol Biol **8**(4): 317-326.

Vidal-Mas, J., O. Resina-Pelfort, et al. (2001). "Rapid flow cytometry–Nile red assessment of PHA cellular content and heterogeneity in cultures of Pseudomonas aeruginosa 47T2 (NCIB 40044) grown in waste frying oil." Antonie Van Leeuwenhoek **80**(1): 57-63.

Wang, F. and S. Y. Lee (1997). "Production of poly(3-hydroxybutyrate) by fed-batch culture of filamentation-suppressed recombinant Escherichia coli." Appl Environ Microbiol **63**(12): 4765-4769.

Wang, F. L. and S. Y. Lee (1997). "Production of poly(3-hydroxybutyrate) by fed-batch culture of filamentation-suppressed recombinant Escherichia coli." Appl Environ Microbiol **63**(12): 4765-4769.

Webb, A. (1990). U.S. patent 4900299. USA.

Weikert, C., F. Canonaco, et al. (2000). "Co-overexpression of RspAB Improves Recombinant Protein Production in Escherichia coli." Metab Eng **2**(4): 293-299.

Wieczorek, R., A. Pries, et al. (1995). "Analysis of a 24-kilodalton protein associated with the polyhydroxyalkanoic acid granules in Alcaligenes eutrophus." J Bacteriol **177**(9): 2425-2435.

Williams Sf, Martin Dp, et al. (1999). "PHA applications: addressing the price performance issue. I. Tissue engineering." Intern J Biological Macromol **25**: 111-121.

Wong, H. H., R. J. Van Wegen, et al. (1999). "Metabolic analysis of poly(3-hydroxybutyrate) production by recombinant Escherichia coli." J Microbiol Biotechnol **9**(5): 593-603.

Wu, G. F., Z. Y. Shen, et al. (2002). "Modification of carbon partitioning to enhance PHB production in Synechocystis sp PCC6803." Enzyme Microb Technol **30**(6): 710-715.
Xiaohai Wang, Z. W. N. A. D. S. (1996). "G418 Selection and stability of cloned genes integrated at chromosomal." Biotech Bioeng **49**: 45-51.
York, G. M., J. Stubbe, et al. (2002). "The Ralstonia eutropha PhaR Protein Couples Synthesis of the PhaP Phasin to the Presence of Polyhydroxybutyrate in Cells and Promotes Polyhydroxybutyrate Production." J Bacteriol **184**(1): 59-66.
Zhang, J. (2003). "Evolution by gene duplication: an update." Trends in Ecology & Evolution **18**(6): 292-298.

Appendix I – Kinetic model of PHB pathway

The PHB pathway was analyzed using the model by van Wegen, et. al. (van Wegen, Lee et al. 2001). Details of model are in paper.

Matlab code for kinetic model

```
function x = PHB(rmax_A, rmax_B, A);

%Keith Tyo
%10/23/06

%Parameters and equations taken from
%van Wegen, et. al.  Biotech. & Bioeng.  74(1) 70-80

% Will be solving 3 equations for
% (1) flux [r] (2) acetyl-acetyl-CoA conc [AA] (3) 3-HB conc [HB]

% Will be optimizing flux by varying rmax_A and r_max_B activity.


%Solved for concentrations
%Initial Guess
AA = 0.14;
HB = 27.5;

%Activities
%rmax_A = 2;
%rmax_B = 2.6;
rmax_EC = 0.7653; % This one will be set

rmax = [rmax_A
      rmax_B
      rmax_EC];

%Metabolite concentrations
%A = 475; %uM
C = 47.5; % based on a AcCoA/CoA = 10
P = 60;
PH = 240;
pH = 7.6;
```

```
metab = [   A
            C
            P
            PH
            pH];


%Optimization variable
r = 0.5;

%Find zeroes of function
xo = [   r
         AA
         HB];

options = optimset;

x = fsolve(@kin_eqn, xo, options, rmax, metab)

return;

function X = kin_eqn(xo, rmax, metab)

rmax_A = rmax(1);
rmax_B = rmax(2);
rmax_EC = rmax(3);

A = metab(1);
C = metab(2);
P = metab(3);
PH = metab(4);
pH = metab(5);

%Constants
Keq = 500*10^(7.2-pH);

r = xo(1);
AA = xo(2);
HB = xo(3);

X = zeros(1,3);

X(1) = rmax_A*(A^2-25000*AA*C)/(840*A+A^2+67.7*A*C)-r;

X(2) = rmax_B*(AA*PH-
1/Keq*P*HB)/(95+19*AA+5*PH+AA*PH+2897/Keq*HB+1532/Keq*P+93/Keq*HB*P)-r;
```

```
X(3) = rmax_EC*HB/(720+HB)-r;

return;
```

Appendix II – Plasmid propagation model

Plasmid System

We'll assume the propagation of a mutation that overtakes a population can be characterized by three steps.

1) DNA replication error
2) Propagation of mutated plasmid to form subpopulation with only mutated plasmid
3) Overtake of parent population by mutant population

Assumptions

40 copies of a plasmid

Each plasmid contains a 5 kb region of expression

DNA replication error

Estimate of DNA Mutation Rate

Although apparent DNA mutation rates can vary widely (10^{-11} to 10^{-6} errors/bp copied) given the state of the cell or location of the DNA, 5.4×10^{-10} errors / bp copied [or 1 error every 1.85×10^{9} bases copied] is a basal DNA mutation rate consistent with many observations(Drake, Charlesworth et al. 1998; Taft-Benz and Schaaper 1998; Burger, Willensdorfer et al. 2006).

However, not every mutation can be expected to result in a loss of function for the enzyme.

Expectation of number of doublings for one mutation

$$Total\ bp\ copied = (Total\ replications) \times (Length\ of\ DNA) \times (\#\ of\ copies)$$

$$1.85 \times 10^9 = (\sum_{i=0}^{n} 2^i) \times (5000\ bp) \times (40)$$

$$n = 13\ doublings$$

It is expected that one copy will be mutated in ~13 doublings. If grown from a single cell, that means 1 cell in 8,192 will be mutated.

Propagation of mutant plasmid

A simple segregated subpopulation balance model was used to estimate the relative abundance of subpopulation that differ in the number of copies of the mutant plasmid for a given number of generations and an initial distribution of mutant plasmid copy numbers. The model assumes a binomial distribution of plasmids from mother to daughter cells. This distribution will be weighted at each division by a linear growth advantage that favors cells with more of the mutant allele plasmids. This

model is a theoretical simplification of a model by Bentley & Quiroga that assumes plasmids replicate at the same rate as the cells, as would be expected for steady state(Bentley and Quiroga 1993). The unmutated plasmids are not explicitly accounted for in this model.

The binomial distribution was taken as

$$\delta_{ij} = \left(\frac{(2j)!}{(2j-i)!\,i!}\right)\left(\frac{1}{2}\right)^{2j}$$

Where δ_{ij} is the probability that a mother cell with *j* plasmids generates two daughter cells, one with *i* plasmids one another with *j-i* plasmids. The 2j term is used because it is assumed the plasmids double at the same rate as the cells do, as would be expected at steady state growth.

The model was used to account for the number of mutated plasmids in a given cell. The dependence of growth rate on plasmid copy number was assumed to be linear and increase with increasing number of mutant plasmids.

$$\mu = \frac{\mu_{fast} - \mu_{slow}}{C_{\max}} C + \mu_{slow}$$

C = Number of mutant copies of plasmid

μ = Growth rate of strain with C mutant copies

μ_{fast} = Growth rate of strain containing only mutant plasmids (0.32 h^{-1})

μ_{slow} = Growth rate of strain containing only original plasmids (0.15 h^{-1})

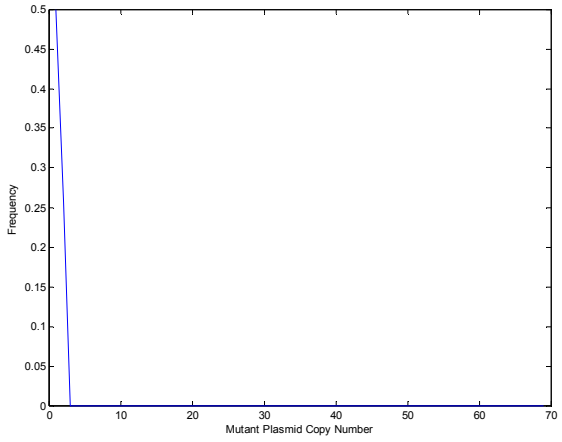
C_{max} = Copy number of plasmid (40)

Values for difference in μ_{fast}, μ_{slow} , and C_{max} were based on experimental observations of a PHB strain, before and after loss of PHB productivity.

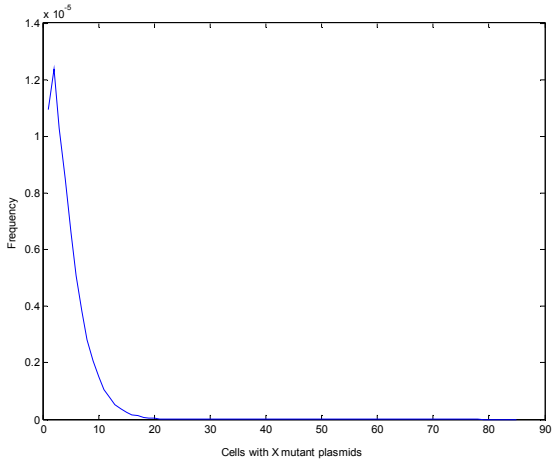
Matlab software package was used to iterate the number of generations (see below for code). Briefly, a vector, *N*, whose elements, *i*, are the number of cells with (*i*-1) mutant plasmids. This vector is expanded to *N_doub*, where each element is moved two twice its index, except the zero plasmid (i=1). *N_doub* is multiplied by the matrix of the binomial distribution to calculate the new distribution of plasmids. This distribution is weighted by the linear growth advantage before iterating for *n* generations.

These results are for a starting distribution of 1 cell with 1 mutant plasmid (the 0^{th} population is not plotted to

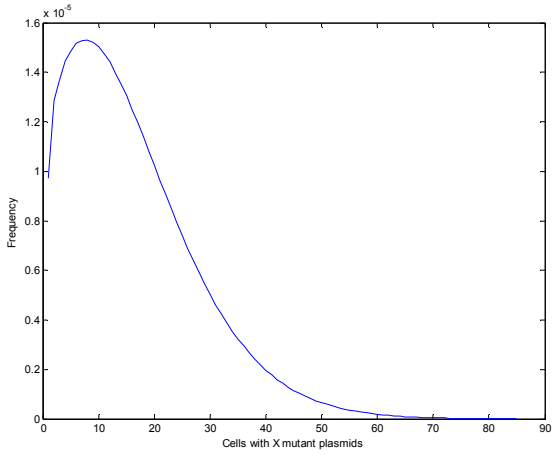
maintain scale features). By the 20th generation a mutant plasmid subpopulation has formed and broken away from the parental population.



1st Generation



10th Generation



20^{th} Generation

Depending on the expression used for plasmid replication rate, the segregation to a cell only containing mutant plasmids would take ~20 doublings.

Overtake of parent population by mutant population

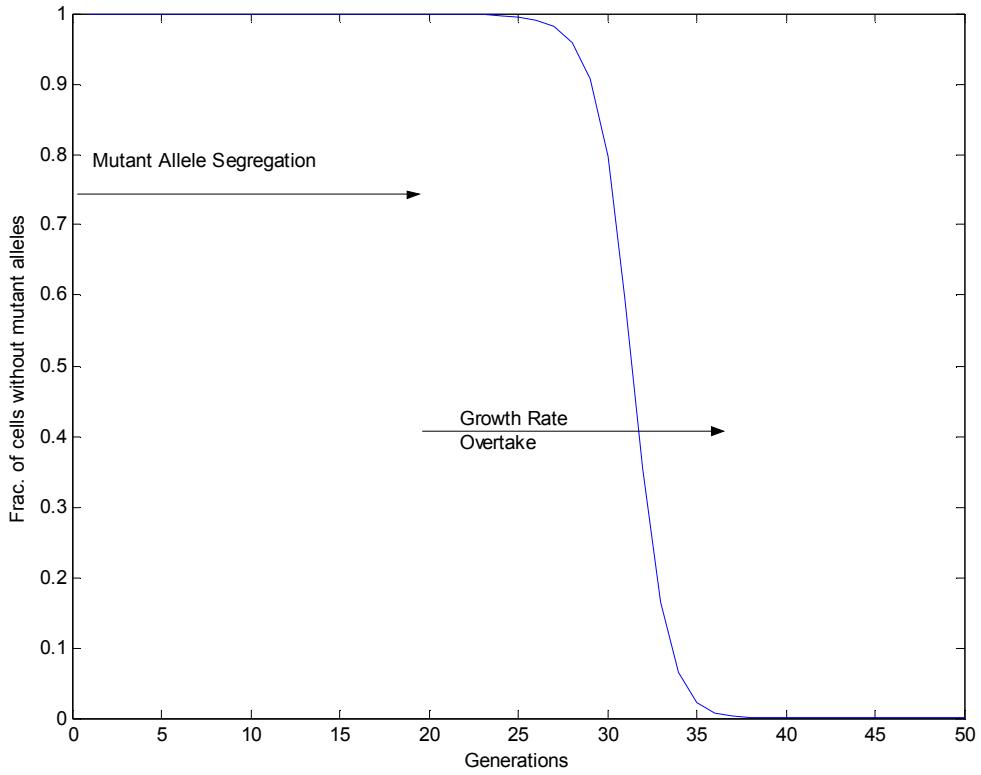
Once a second population of mutant plasmids has segregated, the ratio of mutant to parental populations follows this expression

$$Ratio = \frac{M}{P} = \frac{M_o}{P_o} e^{(\mu_m - \mu_P)t}$$

An initial ratio of 1 mutant and 1.3×10^5 parental cells (based on 17 doublings of the parental strain) and the specific growth rates used above, it was estimated to take 69 hrs or 15

doublings of the parental strain for the mutant to overtake the population.

This is also predicted by the distribution model.



Matlab code for propagation model

```
function [N_running] = binomial(n)

% This population balance model consist of a vector, N, with i indices, where each element has copy
% number (i-1) of the mutant plasmid.  At each generation, the probability
% of daughter cells inheriting a given copy number is calculated for each
% mother cell subpopulation and summed.  The distribution of copy numbers
% is then weighted by a linear function that gives a growth advantage for
% increased copies of the mutant plasmid.

% The unmutated plasmid is not specifically accounted for in this model,
% and the total number of plasmids is assumed to be at steady state (as in
% the plasmids replicate at the same rate as the cells).

% The starting distribution is based on the DNA mutation rate calculations
% that it will take 13 generations for a mutant to arise, leading to 1
% mutant in 2^13 ~8192.

% This function iterates a binomial distribution of plasmids from mother to
% daughter cells.  It is weighted at each generation by a growth advantage
% observed for XL1 Blue growing with
%   (a) pZE-tac-pha [Specific growth rate = 0.15 h-1]
%   (b) null PHB mutant of pZE-tac-pha [Specific growth rate = 0.32 h-1]
%  The copy number for this plasmid is ~40 in exponential phase.

% The input n, is the number of generations to calculate.



% Max concentration of plasmids / cell (ie max calculated copy number)
Npmax   = 85;  %Arbitrary, but must be high

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Plasmid segregation - binomial distribution

for j=1:Npmax+1
   i = 1;
   while i-1 <= 2*(j-1)
      deltaij(i,j) = (factorial(2*(j-1))/(factorial((2*(j-1))-(i-1))*factorial(i-1)))*0.5^(2*(j-1));
      i = i + 1;
   end
end

% Truncate probabilities for copy numbers above maximum copy number
```

```
trunc_prob = sum(deltaij(Npmax+1:2*Npmax+1,:));
deltaij(Npmax+1,:) = trunc_prob;
deltaij(Npmax+2:2*Npmax+1,:) = [];


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Set growth advantage matrix
slope = (.32 - .15)/(Npmax - 0); % Based on experimental observation
int = .15; % Based on experimental observation

% Make linear growth advantage from slope and intercept given.
y = zeros(Npmax+1,1);

for x=1:Npmax+1
   y(x) = x*slope + int;
end

for i=1:Npmax+1
   G(i,i) = y(i);
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Vector of starting distribution
No = zeros(Npmax+1,1);

% Distribution based on DNA mutation rate calculation
No(2) = 1;
No(1) = 8192;

% Normalize distribution
No = No/sum(No);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Propogate n times

disp('Number of Generations');
disp(n);

i = 1;

while i <= n

   N_running(:,i) = No;

   N = G*deltaij*No;
```

```
    % Normalize N
    N = N./sum(N);

    No = N;
    i = i+1;
end

plot(N)

return
```